LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

The Journey toward the Patient-Centered Medical Home: A Grounded, Dynamic Theory of Primary Care Transformation

ANDRADA TOMOAIA-COTISEL, MPH, MHA

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the
University of London

MARCH 2018

Department of Health Services Research & Policy

Faculty of Public Health Policy

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded in part by the United States Agency for Healthcare Research & Quality, the World Health Organization, the London School of Hygiene and Tropical Medicine and the University of Utah.

Research group affiliation(s): University of Utah/Care By Design™ Research Team

AHRQ/Hershey Transforming Primary Care Research Group

LSHTM/European Observatory on Health Systems and Policies

WHO/Alliance for Health Policy and Systems Research

ABSTRACT

Introduction: This dissertation develops a grounded and dynamic theory of primary care transformation (PCT) in a health service delivery organization (HSDO) implementing the Patient-Centered Medical Home model, in the United States of America. The focus of this theory is on the structural facilitators and challenges to achieving and sustaining high quality primary care.

Methodology: Fieldwork performed included semi-structured interviews conducted across the HSDO (n=82), direct observations (n=10 clinics) and archival review. This dissertation utilizes newly-developed methods for theory development and validation, in concert with existing system dynamics methods; with an improved potential to integrate findings across quantitative and qualitative research directions.

Results: My theory illuminates how the actions of various stakeholders (medical assistants, managers, clinicians and patients) interact with each other and with the fundamental characteristics of primary care service delivery to create diverse transformation trajectories.

Two types of leverage points are identified: policies and preferences. The latter are more difficult to modify as they require changing stakeholders' mental models. It is the combination of policies and preferences *interacting within the system structure* that produces hoped-for and feared transformation trajectories. There is no policy that induces success regardless of stakeholder preferences. There are some preferences that induce success or failure regardless of the policies being implemented.

Conclusion: Sustaining success requires understanding the system structure within which policies and preferences operate – how decisions are made, their consequences, and the delays involved. Otherwise, transformation risks being overwhelmed by unintended consequences, misunderstood system behavior or impatience. This work presents an improved understanding of what PCT involves, and of how operational and cognitive aspects intersect.

Overall, this work is more than a study of transformation. It presents theory, methods and a case for the development of an integrative methodology and paradigm.

KEY WORDS

primary care, tenets, tensions, task-shifting, task-sharing, transformation, Patient-Centered Medical Home (PCMH), mixed methods, health services research, system dynamics, systems thinking, causal loop diagrams, validation, mental models, cognitive limitations, paradigm, paradigm crossing, transition zones, integrative paradigm, interplay, stories with numbers, emotional engagement

FUNDING SUPPORT

While I did not receive a dissertations-specific grant or contract, I did benefit from funded-work that supported my dissertation research. This is the funding that I wish to acknowledge in this section.

Specifically, this work was funded in part by grants from the United States Agency for Healthcare Research & Quality (AHRQ) and conference support from the World Health Organization, the London School of Hygiene and Tropical Medicine and the University of Utah.

The primary funding source other than my own funds was the AHRQ funded "Transformed Primary Care – Care By Design" grant R18 HS19136-01 (PI: Michael K. Magill). This grant was part of the AHRQ Transforming Primary Care Practice grants: https://www.ahrq.gov/professionals/systems/primary-care/tpc/index.html.

Additional grants and contracts that funded work which informed my general understanding during this time included: an AHRQ conference grant (R13 HS0219933-01; PI: Robert A. Gabbay), the AHRQ "Successful Strategies—Care Management" contract R18 HS020106-01 (PI: Michael K. Magill) as well as a subcontract with Econometrica, Inc. (2235-000-UTAH; PI: Magill), which was operating under AHRQ contract HHSA2902007 TO No. 5.

ACKNOWLEDGEMENTS

First and foremost, I wish to thank Sam, John, Costica and Maria, as well as my extended family, for their love and support along this journey.

Thank you also to Dr. Karl Blanchet and Dr. Zaid Chalabi, my supervisors, who provided much needed guidance and opportunities to grow not only within my dissertation but also in other work. Thank you to my advisory committee who provided guidance in preparation for the initial proposal (the upgrading document) as well as at critical points since; they are Dr. Bernd Rechel, Professor Martin McKee, Professor Sir Andy Haines and Dr. Jenny Neuburger. Also, at LSHTM more broadly, thank you to my friends and advisors Dr. Nicki Thorogood and Professor John Cairns.

This study would not have been possible without the encouragement of Professor Michael K. Magill, Professor Debra L. Scammon, Dr. Julie Day, Rachel Day, and Professor Timothy W. Farrell. It also benefited from the feedback of others on the University of Utah Care By Design™ Research Team, namely Professor Norman Waitzman and Professor Jaewhan Kim. One could say this work started in the summer of 2008 in Ghana when I first met Mike and we talked about our mutual desire to strengthen primary care. I joined his team and, as a multidisciplinary research team embedded within a health services delivery organization, we applied for grants to study the transformation the organization was undergoing. Creative, mixed methods approaches were not only welcomed but also encouraged. Thank you for investing in me, for giving me the opportunity to organically craft this work and for mentoring me throughout. I would also like to thank the managers, clinicians and medical assistants at the Community Clinics and University Hospital for their contributions.

Likewise, I thank those outside the organization who provided guidance and insight. These include: from the Hershey Transforming Primary Care Research Group, Dr. Bob Gabbay, Dr. Rob Reid, Dr. Leif Solberg and Dr. Neil Calman as well as invited guest Dr. Mark Freidberg; from the World Health Organization, Dr. Taghreed Adam; and many from the System Dynamics Society, most notably Professors Deborah & David Andersen, Professor Rod MacDonald, Gary Hirsch, Professor Yaman Barlas, Professor Hazhir Rahmandad, Professor John Barton, Professor Andy Ford, Professor Khaled Saeed and Professor Jay W. Forrester.

The first time I attended the annual conference, I had the opportunity to have lunch with Jay, Yaman and others and hear Jay's thoughts regarding mental data. I have often thought back to that day as I developed my methods. John was first to talk through my ideas with me at the conference. Gary first oriented me to the grammar of Causal Loop Diagrams, David pointed me to purposive text analysis, Andy helped me to see my reference mode, Khaled introduced the idea of slicing complex problems and Hazhir encouraged me to think about methods for *saturation* and *shared understanding* in qualitative model validation. Last but certainly not least, Rod mentored me through the simulation model development process.

Thank you to AHRQ for funding the initial project and for approving the work proposed in this dissertation as well as for the Hershey conference grant that spawned the Hershey Transforming Primary Care Research Group that was helpful as I developed this work and my broader thinking on methods and primary care transformation.

Thank you to AHRQ, LSHTM and WHO/Alliance for Health Policy and Systems Research for supporting me in attending conferences to develop and disseminate this work. It is in such venues that I gained useful feedback as the work progressed and now as I think about what future research might be useful.

Thank you my friends at the MIT/Albany/WPI PhD Colloquium, as well as the MIT and WPI SD student groups who provided general and project-specific feedback and insights as my work progressed. Thank you also to my research degree friends at LSHTM, especially Nilla Backman, Svetlana Anker, Emma Garnett, Jorn Jacobsen and Maureen Seguin.

Thank you to Unity Health Care, Inc. and Mary's Center, two Federally Qualified Health Centers in the Washington DC area, where I worked and gained a passion for providing high quality primary care to everyone and saw that it can be done! Thank you to the University of Utah, University Healthcare as it allowed me to study how this can be done in an integrated health system. Thank you to the RAND Corporation, as it allowed me to work part-time so that I could finish this dissertation while at the same time providing a stimulating work environment that allowed me to keep up to date with the research and policy developments in this area.

Thank you to Stanford University's Human Biology program and the Haas Center for Public Service that started me on my journey toward studying health systems from an interdisciplinary perspective.

I would like to end these acknowledgements by thanking Sydne & Scott S. Parker, Dr. Phil R. Lee, Dr. Geoff Heller, Professor Kurt Stange, Dr. John A. Peterson, Lynette & Phil Selander, Loretta & Doug Allred and Kathy & Dr. John W. Gardner, Lorene & Jerry Jump, and Jane & Ronald L. Sellers for their long-standing encouragement, example and mentorship as they have worked *to improve the human condition*.

DECLARATION OF OWN WORK

I, Andrada Tomoaia-Cotisel, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

TABLE OF BOXES

TABLE OF MOST COMMON ACRONYMS

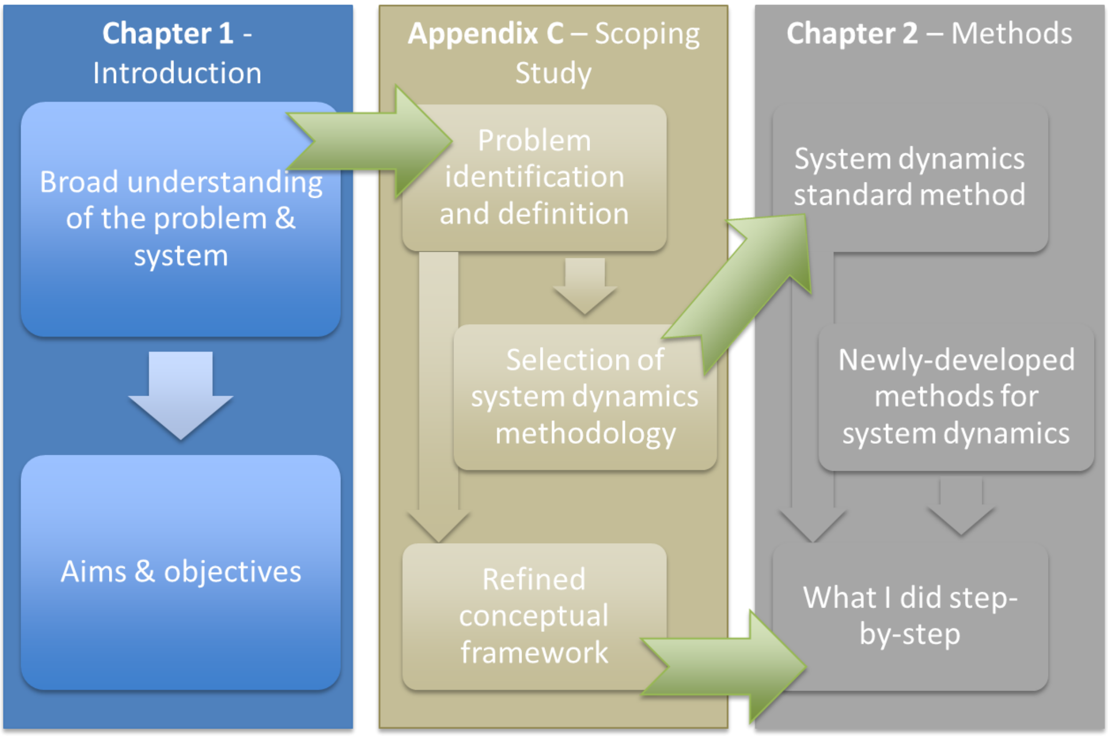
Acronym	Term
CLD	Causal Loop Diagram
CLD Combination	Causal Loop Diagram Combination
CM-S	Conceptual Model Saturation
HSDO	Health service delivery organization
MD(s)/Clinician(s)/CL	Clinicians (licensed practitioners able to manage their own patients). May include doctors, nurse practitioners and physician assistants. MD is the acronym used in simulation model variables and CL references to specific study participants. They are interchangeable. Participants may also refer to these individuals as "providers". Otherwise, the term "clinician" is used.
MA(s)Clinical Staff Member(s)	Primary care team members (staff members working in teams with clinicians). Initially, these personnel are medical assistants. As task shifted become more complicated, care teams utilize personnel of other professions such as nurses, social workers and pharmacists (See Chapter 1 for more detail). MA is the acronym used in model variables and references to specific quotations.
MMHSR	Mixed methods in health services research
PCMH	Patient-Centered Medical Home
PCT	Primary care transformation
RIQ (Causal or Cognitive)	Rigorously-interpreted quotations. There are two types: causal and cognitive
SD-S	System Dynamics Saturation
SDM	System Dynamics Modeling
SIM-S	Simulation Model Saturation
SMM-S	Shared Mental Model Saturation

TABLE OF ACRONYMS

Acronym	Term
AHRQ	United States Agency for Healthcare Research and Quality
ANOVA	Analysis of variance
CLD	Causal Loop Diagram
CLD Combination	Causal Loop Diagram Combination
CM	Center Manager
CM-S	Conceptual Model Saturation
CPT	Current Procedural Terminology Code - a code characterizing the patient visit for insurance billing (payment) purposes.
CptV1	Conceptual Validity Subtype #1
CptV2	Conceptual Validity Subtype #2
CptV3	Conceptual Validity Subtype #3
CptV4	Conceptual Validity Subtype #4
DV1	Data Validity Subtype #1
DV2	Data Validity Subtype #2
DV3	Data Validity Subtype #3
EV1	Experimental Validity Subtype #1
EV2	Experimental Validity Subtype #2
FV1	Formulational Validity Subtype #1
FV2	Formulational Validity Subtype #2
FV3	Formulational Validity Subtype #3
HSDO	Health service delivery organization
ID	Identification number
MD(s)/Clinician(s)/CL	Clinicians (licensed practitioners able to manage their own patients). May include doctors, nurse practitioners and physician assistants. MD is the acronym used in simulation model variables and CL references to specific study participants. They are interchangeable. Participants may also refer to these individuals as “providers”. Otherwise, the term “clinician” is used.
MA(s)Clinical Staff Member(s)	Primary care team members (staff members working in teams with clinicians). Initially, these personnel are medical assistants. As task shifted become more complicated, care teams utilize personnel of other professions such as nurses, social workers and pharmacists

Acronym	Term
	(See Chapter 1 for more detail). MA is the acronym used in model variables and references to specific quotations.
MMHSR	Mixed methods in health services research
NM	Nurse Manager
PC	Primary care
PCMH	Patient-Centered Medical Home
PCT	Primary care transformation
RIQ	Rigorously-interpreted quotations. There are two types: causal and cognitive
SD-S	System Dynamics Saturation
SDM	System Dynamics Modeling
SIM-S	Simulation Model Saturation
SMM-S	Shared Mental Model Saturation
TPC	The Transforming Primary Care grant
TS	Task-shifting
US	United States of America

CHAPTER 1 INTRODUCTION



1.1 INTRODUCTION

> "Theory without empirical research is empty, empirical research without theory is blind." Bourdieu (p. 774-775) [1]

For primary care to reach its theoretical potential, the *theory* of how primary care improves over time must be better understood in its complexity, using *methods* up to the task. Two aims emerged as I pursued this dissertation: a theoretical aim and a methodological aim.

***Theoretical Aim:* to develop a better *theory* of primary care transformation (PCT)**. On the one hand, there is a broad consensus about primary care, namely: it plays a central role in a health system and works best when its four main tenets (access, continuity, comprehensiveness and coordination) are in place. On the other hand, interventions putting this theory into practice, such as the Patient-Centered Medical Home (PCMH) in the United States (US), have proven to be challenging to implement and to evaluate. Therefore, while the value of and desire for a health system based on primary care is clear, it is sometimes less apparent how it can be attained.

***Methodological Aim:* to select adequate *methods,* and create new ones when needed, to understand the dynamics of PCT**. On the one hand, primary care is not constrained in that it does not have a clear research paradigm like other medical professions[2]. On the other hand, research in primary care involves many intangible variables, making empirical research and policy formulation challenging *(ibid.)*. Therefore, while there is paradigmatic flexibility, researchers have been slow to use, adapt and develop the methods which are suitable in these conditions.

This dissertation contributes to meeting these aims by developing a grounded and dynamic theory of PCT in a PCMH-implementing health service delivery organization (HSDO), in the US (objective #1). The focus of this theory is on the structural facilitators and barriers to achieving and sustaining high quality primary care. Contributing to outstanding research questions in mixed methods in health services research (MMHSR) involves three more objectives: addressing paradigm issues (objective #2), presentation of results that engage the subconscious, emotional level on which decision making is based (objective #3) and addressing issues of empirical validity (objective #4).

This dissertation has grown organically from the mixed methods work of a multi-disciplinary, organizationally-embedded, research team studying PCT at the University of Utah; herein referred to as the Mixed Methods Project. In 2010, the US Agency for Healthcare Research & Quality awarded 14 Transforming Primary Care grants to retrospectively describe the process and content of transformation toward PCMH that had occurred in various settings across the US, including one to the Mixed Methods Project at the University of Utah[3].

My primary tasks as a member of this team were: to design and collect portions of the qualitative data, to design and perform mixed methods analyses, and to disseminate findings (see Appendix A for full list of dissemination products). Both University of Utah Institutional Review Board approval and London School of Hygiene and Tropical Medicine Ethics Committee approval were received for the analysis performed in this dissertation (see Appendix B).

In the next two sections, I will present my research aims in greater depth. Section 1.2 focuses on primary care and Section 1.3 focuses on methods. Each section identifies the *relevant* problems (shown in **boldface underlined font**) and the corresponding objectives. Section 1.4 presents an overview of the dissertation.

1.2 PRIMARY CARE

1.2.1 DEFINITION OF PRIMARY CARE

In this section, I first describe primary care, then I analyze formal definitions and discuss how primary care is changing. *Primary care* (as referred to in this dissertation and as broadly practiced in the US) is described considering the following five questions: What is primary care? What is delivered? Who delivers primary care services? Who receives primary care, and when? Where and how are services delivered? How are these services paid for?

What is primary care?

Primary care is the patient's first point of contact with the health system. With the exception of needing emergency services, the patient begins seeking care in primary care and is then referred to specialty care as needed. Thus, primary care is the first step in a continuing health care process, where the patient develops a long-term relationship with the clinician (i.e., the person delivering the primary care services)[4].

At times, the term *primary health care* is conflated with the term *primary care;* however, *primary health care* denotes a more holistic perspective as it also includes coordination with other sectors in addition to the health sector[1][5]. This dissertation uses the term *primary care* as it focuses specifically on primary care services delivered within the health sector and is the term commonly used in the US (the context of this work).

What is delivered?

Primary care is the part of the health system that treats and/or coordinates care for the patient as a whole to promote health, prevent disease, treat acute symptoms and manage chronic conditions. Services offered in primary care in the US include preventive services[4, 6] (e.g., immunizations, colonoscopy), acute services[4] (e.g., laceration repair, foreign body removal), chronic disease management services[7] (e.g., patient education, monitoring blood glucose level), as well as some simple in-office procedures depending on provider preference[8] (e.g., Papanicolaou smear, biopsy, new-born circumcision)[8]. For getting access to services outside primary care,

[1] "[Primary health care] involves, in addition to the health sector, all related sectors and aspects of national and community development, in particular agriculture, animal husbandry, food, industry, education, housing, public works, communications and other sectors; and demands the coordinated efforts of all those sectors" (p.2).

the clinician can coordinate[2] care, for example, by sending referrals to specialty, receiving reports from specialty and hospital visits and conferring with other providers regarding the patient's evolving care plan[4].

Who delivers primary care services?

Providers practicing in primary care may have been trained in a range of medical specialties, including: family medicine, general practice, general internal medicine, geriatrics, general pediatrics, obstetrics and gynecology, and sports medicine[9]. Primary care clinicians (herein, clinicians) are professionals with postgraduate training; including physicians, nurse practitioners, and physician assistants[10]. All are licensed practitioners able to manage their own patients (although the level of physician supervision required differs by US states)[11, 12].

Often, clinicians work in teams, known as care teams. Teams may be directed by one or two clinicians or more diffusely by a group of clinicians. Care team members (referred to in this dissertation as clinical staff members) may include unlicensed personnel such as medical assistants, licensed personnel such as nurses and social workers, as well as postgraduate-trained personnel such as nutritionists and pharmacists[13]. These additional team members may be referred to by their training/background (as above) or by their role (e.g., care manager, transitions navigator, health educator)[14-16]. While team make-up varies, it is most common for clinical staff members who are medical assistants to work full time on the care team with the clinician, while licensed personnel work on an on-call basis. Clinical staff members perform various tasks, depending on their training, license, and prescribed role within the care team. For example, Tomoaia-Cotisel *et al.*[16-18] reported on the range of policies for organizing and delivering care management[19][3] services across the US. Sometimes, clinicians deliver these services. In other practices, particular staff members are dedicated to providing these services. Yet, other practices divide these services across clinicians and staff members. The training of clinical staff members providing these services also varies, as do the services provided.

[2] This can be by playing the role of a gatekeeper, although, in the US, the insurance and legal professions (through malpractice litigation) also determine the services offered, and which services a patient can receive even without consent of their primary care doctor, to some extent.
[3] Defined as "a set of activities designed to assist patients and their support systems in managing medical conditions and related psychosocial problems more effectively" (p.2).

Who receives primary care, and when?

Patients of all ages receive primary care for the large majority of their health care needs. Access to this care is mediated by the cost of services and one's ability to pay. To receive adequate primary care, healthy individuals are to have periodic visits for screening, counseling and comprehensive physical examination. How often these are *needed* depends on a patient's age, gender and other circumstances[20], although an annual visit is *preferred* for patients to maintain a strong relationship with their clinician[21, 22]. Otherwise, primary care services are received as needed; for example, patients with chronic conditions make recurring visits and other patients seek care when a need (e.g., feeling sick) is identified.

Clinicians are referred to as having a *patient panel* – that is, that they have a group of patients for whom they take stewardship. Some practices use a proactive model where their goal is to take responsibility for the health of their panel as a population (e.g., by following up with patients after a discharge from the hospital, or mailing educational materials), not just those seeking visits. In other cases, a reactive model is used where patients are not assigned to a clinician or care team, but see whoever is available, and where most services are provided on an as-needed basis.

Where and how are primary care services delivered?

Most often, the patient comes to the office of the clinician. The patient is escorted to the exam room where concerns, which often involve social and behavioral aspects in addition to disease, are discussed, progress is reviewed, tests are ordered, referrals are made, and care plans are updated. Primary care is also sometimes, although less frequently, delivered in the patient's place of residence (i.e., the home, or a nursing home).

How are these services paid for?

The US has an insurance-based[4] health care system. When services are delivered to a patient, clinicians generate a *claim* which characterizes the visit overall using a Current Procedural Terminology Code (i.e., CPT Code)[23]. This claim is sent to the patient's insurance company and/or to the patient for payment.

[4] There are both private insurance providers and public ones – commonly referred to as *payers*.

If the patient chooses not to have insurance, then he/she is required to pay the full cost of care out of pocket. For insured patients, the insurance company reviews the claim as well as justification for the CPT code and either approves or denies the claim. If it is approved, then the insurance company pays a portion or all of the cost, according to the contract. If the claim is denied, then the patient must pay for it in its entirety. If the patient cannot/does not pay the amount owed, then the cost is absorbed by the primary care practice.

Because primary care encompasses many different types of services and can often be taken for granted, much effort has been placed in setting out aspirational, formal, research and operational definitions of primary care. The Alma Ata Declaration provides an aspirational definition[5]. At the same time, the US Institute of Medicine provides a formal definition of what primary care constitutes in the US[24, 25]. A quarter-century later, Starfield and colleagues[26] provide a formal definition for researchers to use, observing that research on the effectiveness of primary care needed to use a more specific, formal definition in the study design in order to have meaningful findings for health policy. Put forth by the American Academy of Family Physicians, the American Academy of Pediatrics, the American College of Physicians, and the American Osteopathic Association, the PCMH Joint Principles presents an even more specific definition, which aims to operationalize the aspirational principles laid out in Alma Ata[27].

There are four tenets that are common to each of these definitions: access, coordination comprehensiveness and continuity. These are the key features that distinguish primary care from secondary and tertiary care, and it is described in general terms as follows (based on Kozakowski[4]):

- **Access** refers to the responsibility of primary care to be the first step in patients' health care process;
- **Continuity** refers to the responsibility of primary care to help patients through health problems, which cannot be solved in one visit implying a continuous process of care, and the importance of patients' developing a long-term relationship with the clinician;
- **Coordination** refers to the responsibility of clinicians to work together with other professionals who are responsible for addressing patients' health problems; and
- **Comprehensiveness** addresses the responsibility of clinicians to provide services from a holistic perspective and according to the relevant clinical guidelines. These tenets are the structural pillars of the edifice of primary care.

Table 1.1 below shows excerpts from these formal definitions focusing on the four tenets.

Table 1.1 Primary Care - from Global Aspiration to The Patient-Centered Medical Home

	Global Aspiration Alma-Ata 1978[5] (p. 1-2)	**Formal Definition** US Institute of Medicine 1978[25] (p. 1)	**Research Definition** Starfield *et al.* 2005 [26] (p. 458)	**Operational Definition** PCMH Joint Principles 2007[27] (p. 1-2)
Access	(1) "first level of contact … with the … health system" (2) "universally accessible"	"accessible health care services"	"first-contact access for each new need"	"Enhanced access to care is available through systems such as open scheduling, expanded hours and new options for communication between patients, their personal physician, and practice staff."
Continuity	"first element of a continuing health care process"	"developing a sustained partnership with patients"	"long-term person- (not disease) focused care"	"Personal physician - each patient has an ongoing relationship with a personal physician trained to provide first contact, continuous and comprehensive care."
Coordination	(1) "sustained by integrated, functional and mutually supportive referral systems, leading to the progressive improvement of comprehensive health care" (2) "[includes coordination with sectors in addition to the health sector; for example,] agriculture, animal husbandry, food, industry, education, housing, public works, communications"	"the provision of integrated [health care services]"	"coordinated care when it must be sought elsewhere"	"Care is coordinated and/or integrated across all elements of the complex health care system (e.g., subspecialty care, hospitals, home health agencies, nursing homes) and the patient's community (e.g., family, public and private community-based services). Care is facilitated by registries, information technology, health information exchange and other means to assure that patients get the indicated care when and where they need and want it in a culturally and linguistically appropriate manner."
Comprehensiveness	"addresses the main health problems in the community, providing promotive, preventive, curative and rehabilitative services accordingly"	"by clinicians who are accountable for addressing a large majority of personal health care needs"	"comprehensive care for most health needs"	"Whole person orientation - the personal physician is responsible for providing for all the patient's health care needs or taking responsibility for appropriately arranging care with other qualified professionals. This includes care for all stages of life; acute care; chronic care; preventive services; and end of life care."

Table 1.1 shows that PCMH is an organized, specific concept for putting the four primary care tenets into practice. Nevertheless, the theory supporting PCMH recognizes that the process of successfully putting these tenets into practice (i.e., PCMH implementation) also requires changes beyond delivery of care, which comprise the *transformation* of primary care (PCT):

> "The patient-centered medical home is four things: 1) the fundamental tenets of primary care: first contact access, comprehensiveness, integration/coordination, and relationships involving sustained partnership; 2) new ways of organizing practice; 3) development of practices' internal capabilities, and 4) related health care system and reimbursement changes. All of these are focused on improving the health of whole people, families, communities and populations, and on increasing the value of healthcare."[28] (p. 601)

Important differences between PCMH and traditional primary care in the US include:

- New roles for and/or new types of staff members
- A team-based care approach
- Shifting tasks from the clinician to other team members (e.g., the pharmacist, the social worker, the care manager, the medical assistant, the health coach), and
- Innovative payment models (e.g., based on value of work performed).

1.2.2 PROBLEM: FULL VALUE OF PRIMARY CARE HAS YET TO BE REALIZED

The well-established value[29-31] of primary care involves better quality[32], better health, lower cost, as well as lower inequity[25, 26, 33-43]. There is, though, a misalignment between these benefits and the cost of improving primary care. The benefits "accrue [to society] at the level of the patient's lived experience outside of health care, and at the levels of the healthcare system, community, workforce and population" (p.604); thus, the value of primary care can be found largely by looking outside the health care system[28]. However, in the US, the cost of investing in improvements to primary care, like PCMH, is primarily borne by independent primary care clinics as well as by integrated HSDOs[28].

Along with the misalignment of benefits and costs, there is also a paucity of experience with "the process and intended and unintended consequences of transforming current practices into [PCMHs]"[28] (p. 601). This situation places primary care HSDOs and clinics in a difficult position, where they must choose between undertaking a major transformation, at their own expense, with a limited understanding of how to succeed, and maintaining the status quo (with its own problems, discussed later).

In order to learn how the hundreds of pilots and demonstrations across the US worked to overcome these challenges, the US Agency for Healthcare Research and Quality commissioned research (in 2010) "to better understand challenges faced by primary care practices as they transform into PCMHs"[3] (p. 1). In so doing, the hope was to narrow the gap between theory and practice. Each site selected has contextual as well as implementation differences[8, 44-57].

The effort at the University of Utah Community Clinics occurred semi-autonomously from the parent organization, University of Utah Health (herein the HSDO). It began in 2003, following a business turnaround, with innovation allowed and encouraged[58].

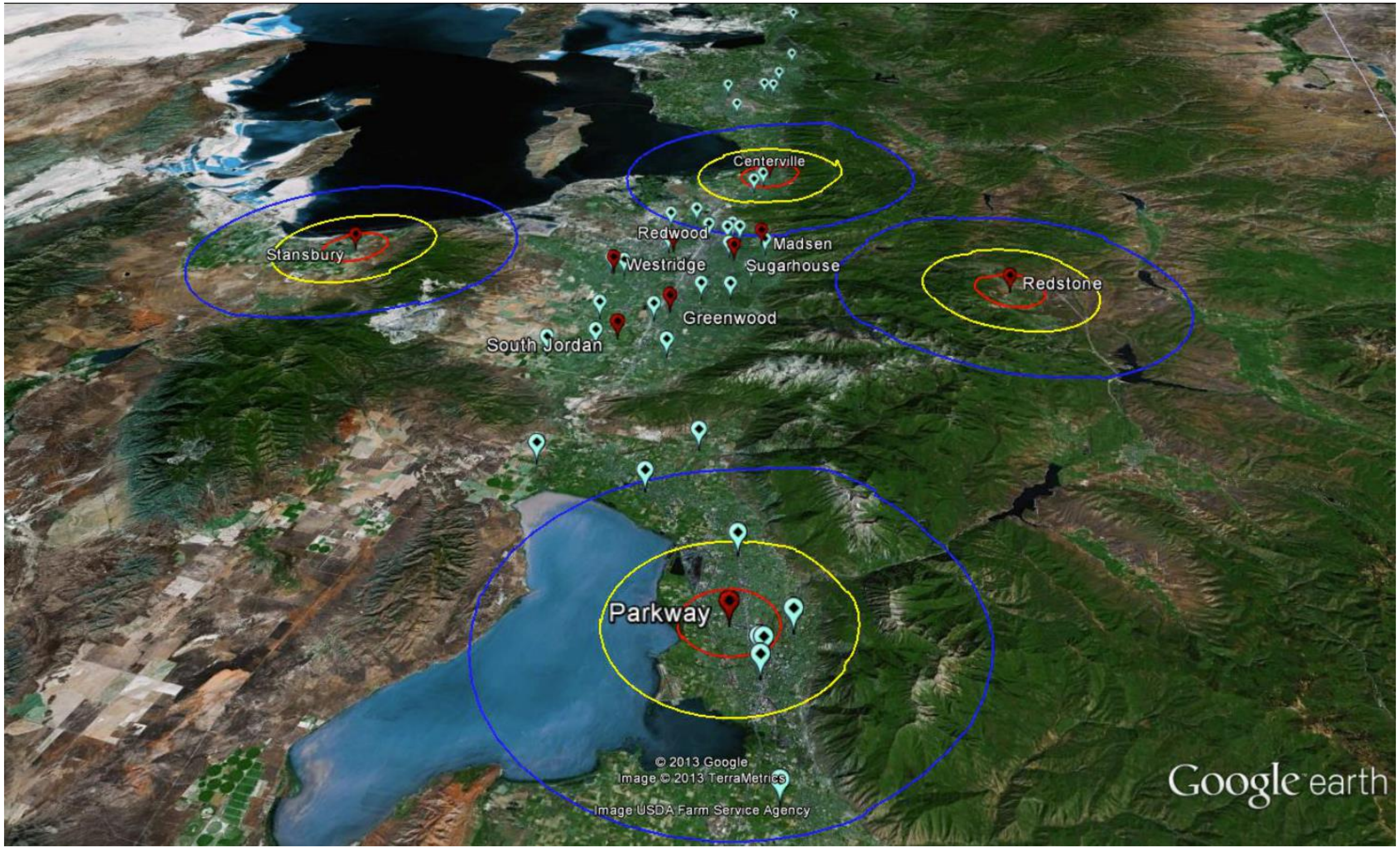
University of Utah Health, the Mountain West's only academic health care system, is one of the health systems providing care for Utahans and residents of five of the surrounding states. Its three-part mission involves excellence in: patient care, education and research. University of Utah Health is an integrated HSDO, including four hospitals, which provides primary care in 12 community clinics. It also has its own health insurance plan: University Health Plan[59].

At the time data were collected (2011), there were 10 community clinics, with 70 clinicians, which served 100,000 active patients (more than 200,000 primary care visits) per year. Clinics ranged in size, with the smallest clinic having four clinicians and approximately 9,000 visits and the largest clinic having 14 clinicians and approximately 17,500 visits[56].

The HSDO's community clinics employ management staff consisting of healthcare administrators, medical directors, and nurse managers. Among other duties, mangers are tasked with staying profitable, with any profits being absorbed by the HSDO.

Figure 1.1 displays the 10 community clinics (red points) along with similar clinics (blue points). There is no internal competition for 4 of the 10 clinics (as shown by the red two-mile, yellow five-mile and blue ten-mile radius circles around those clinics). They are all located in the Greater Salt Lake City metropolitan region[60].

Figure 1.1 **Map of the 10 University of Utah Community Clinics, Greater Salt Lake City Metropolitan Region, Utah, US**



NOTE: This Figure displays the Greater Salt Lake City metropolitan region. The 10 community clinics (red points) and similar clinics (blue points) are shown. No internal competition exists for 4 of the 10 clinics (as shown by the red two-mile, yellow five-mile and blue ten-mile radius circles).

Now, in 2017, it has been almost 10 years since the earliest PCMH pilots were launched. While no meta-analysis or systematic literature review exists that would point to the ability of PCMH to deliver the promised value, peer-reviewed publications do provide some insights: (1) transforming primary care is challenging[61-63], (2) clinics vary in how they put the concepts of PCMH into practice[16, 54, 63, 64] and (3) clinics vary in the extent to which they improve outcomes (e.g., cost, quality, patient satisfaction)[15, 65-70].

Despite this variation, the consensus on the merits of primary care (and thus of PCMH implementation as the current blueprint for PCT) is strong enough to ensure that PCMH adoption is likely to continue growing – whether through public policy or organically. This means clinics adopting PCMH will not be able to begin their transformation with certainty of succeeding. Improvements to theory, such as a better understanding of the complex interactions between the primary care tenets in transformation, have the potential both to improve clinics' chances of success and to refine researchers' questions.

The US Institute of Medicine reaffirms the importance of primary care as the "logical foundation of an effective health care system" (p. v) and as being "essential to achieving the objectives that together constitute value in health care" (p. 2)[71]. **The past decades have seen a growing recognition of the value of primary care and a growing sense of urgency to narrow the gap between theory and the reality of primary care in practice.**

The transformation of primary care sought by PCMH matters because it is designed to address important problems arising in health systems where the primary care tenets are deficient, including:

> **There is a need for more comprehensive care:** Half of the US population suffers from chronic conditions[72]; these conditions are *uncontrolled* for: half of those with hypertension[73], more than 80% of those with hyperlipidemia[74] and 43% of those with diabetes[75]. McGlynn *et al.* report that clinicians are only able to provide 55% of chronic and preventive services[76]. Estimates indicate that it would take 21.7 hours per day for a clinician to deliver comprehensive services to a panel of 2,500 patients[77-79].

> **There is a need for greater access**: At times, the US population also uses hospitals for conditions that are considered primary care treatable and/or primary care preventable, whether entering the Emergency Department, hospital admissions or readmissions[80]. Research has shown that greater access to clinicians addresses primary care treatable concerns and reduces preventable hospital visits[81-83].

There is a need for more coordination among providers: The US population also experiences care fragmentation, which is when there is insufficient infrastructure to have specialty providers coordinate patient care with the clinician. This can result in adverse consequences, in particular for patients with chronic and mental health conditions[15-18, 84, 85].

There is a need for an added measure of continuity between provider and patient: Patients' longitudinal continuity with a provider is reinforced by the interpersonal relationship that forms between them, and vice versa[86-88]. This relationship enhances the provider's ability to be person-focused and to be aware of the person's context such that "care [is] integrated and prioritized across acute and chronic illness, preventive, psychosocial, and family care"[89] (p. 294). Low continuity exacerbates the deficiencies in the other three tenets described above[90].

These societal problems are felt deeply in the US as its citizens recognize that the US has high health care spending and only moderate outcomes[76, 91, 92]. These trends persist when the US is compared to other countries[25, 26, 33-36, 93, 94].

The US is not unlike other World Health Organization Member States in this respect. In the 2008 World Health Organization report *Primary Health Care - Now More Than Ever*[95], Director General Margaret Chan laments: "despite enormous progress in health globally, our collective failures to deliver in line with [Alma Ata] values are painfully obvious and deserve our greatest attention" (p. viii). As part of "a shift towards… more comprehensive thinking about the performance of the health system as a whole" *(ibid.)*, Member States have demanded knowledge regarding how they can achieve more "equitable, inclusive and fair" health systems and meet the growing demand for primary care *(ibid.)*. The World Health Organization calls for "re-organiz[ing] health services around people's needs and expectations, so as to make them more socially relevant and more responsive to the changing world", as a crucial step toward resolving the "intolerable gaps between aspiration and implementation" (p. ix)[95].

1.2.3THEORETICAL AIM & OBJECTIVE #1

The dissertation relies on the following problem statement for this research:

> *Primary care transformation has been and continues to be an elusive target. In the short term, implementation is hard and failure abounds. In the long term, some practices reach successful implementation. We lack sufficient understanding of the structure of primary care, and of the policies that can impact this structure.*

This problem statement is the cumulative result of all phases of this dissertation. As presented in Appendix C, the Scoping Study, tensions were found to exist within the structure of primary care such that the four tenets of primary care and the context of transformation at the HSDO influenced each other, where the hoped-for levels of implementation could not be reached in all tenets at the same time. These tensions involve complex interactions within the underlying causal structure of primary care which contribute to the observed failure and success modes.

While the value of the primary care tenets is well understood, current theory lacks an understanding of the complex interactions between them as well as their interaction within the system of care already in place. This understanding is necessary in order to realize the aspirations of health care systems worldwide.

My theoretical aim is to develop a better *theory* of primary care transformation. This aim has one theoretical objective (Objective #1): to develop a grounded, dynamic theory of PCT in order to build understanding of the key structures generating the primary care health service delivery system behaviors of difficulty, failure and success that HSDOs experience when implementing system improvements such as PCMH. In meeting this aim, this research will develop useful theory for anyone engaged in PCT (including clinicians, managers and other influential policy-makers in health service delivery systems), facilitating improved transformation.

1.3HEALTH SERVICES RESEARCH

In this section, I describe how, in order to reach the above *theoretical* aim/objective #1, an additional *methodological* aim would also need to be met. Section 1.3.1 introduces mixed methods in health services research (MMHSR) and describes a problem in MMHSR methodology: methods for studies seeking to inform policy have not yet reached their potential (according to Miller *et al.*[96]). In Section 1.3.2, the methodological research aim and objectives #2-4 are described.

1.3.1PROBLEM: FULL POTENTIAL OF MIXED METHODS HAS YET TO BE REALIZED

Miller at al explain that, whereas health services research in general aims to "provid[e] valid characterizations of the complex interactions among components of … care delivery systems", the purpose of mixed methods studies is to "*fully* capture the *complex* interactions among components, including interactions among multiple levels of analysis and *over time*"[96] (p. 2125, emphasis added). Therefore, when studying change in health services delivery systems, researchers often employ a mixed methods study design. MMHSR studies are expected to build understanding of both the *structure* and the *dynamics* of health care delivery systems and to do so in a *grounded*, empirically valid way.

The theoretical objective of this dissertation is to build understanding of a complex structure, of the structure-behavior link, and of policy options, fitting within the scope of MMHSR.

The expectations placed on MMHSR studies are high, but they are even greater when considering how to influence policy. Miller *et al.* explain that by reporting "perspectives and experiences" as "stories to accompany numbers", MMHSR studies have the *potential* to engage "policy makers, system leaders, and practitioners in dialogue about the nature of the research and the implications of the findings"[96] (p. 2129). **In order for MMHSR research to affect change, it must reach these stakeholders at a deep level, motivating them to change policies and behaviors. The authors said that methods for this have not yet been developed.**

There are three areas where prevailing MMHSR approaches could improve to more effectively reach these policy-making stakeholders: (1) addressing paradigm issues, (2) engaging the subconscious, emotional level on which decision making is based, and (3) addressing issues of empirical validity. Developing methods which address these areas will contribute to MMHSR's attaining of what Miller *et al.* envision to be the future of MMHSR findings: "a mosaic from which an emotionally engaging and empirically valid research story is created"[96] (p. 2129).

1.3.2 METHODOLOGICAL AIM OF THE STUDY

In order to succeed at the theoretical objective (objective #1 above), I needed to use *methods* up to the task. Thus, a methodological aim emerged: to select *adequate* methods, and create new ones when needed, to understand the dynamics of PCT. My standard for adequacy is that the methods used meet the three needs for improvement identified by Miller *et al.*[96] above, namely: addressing paradigm issues (objective #2), presentation of results that engage the subconscious, emotional level on which decision making is based (objective #3) and addressing issues of empirical validity (objective #4).

This study uses System Dynamics[5] Modeling (SDM) methodology (as justified in Appendix C and explained in Chapter 2). In the process of seeking to address the three areas raised above, this dissertation *also* contributes to new developments in SDM and its application. The following sections define the problem that each objective seeks to address.

1.3.2.1 OBJECTIVE #2: ADDRESSING PARADIGM ISSUES

In their introductory text to the social sciences, Hoover & Donovan[97] define *science* as "the reduction of uncertainty … (using) what observation can accomplish" (p. viii). *Social science* is distinct in its responsibility to answer "the theorist's most basic question**:** *'what can be done to improve the human condition – and what matters are beyond our ability to change?'*" (*ibid.*, emphasis added). In fulfilling this responsibility, the limitations of methodology must be taken into account because, "(no) approach holds all the answers … every approach has pitfalls and openings to prejudice. Choosing the appropriate methodology, or combination of methodologies, is the critical consideration" (p. ix). This choice is critical because getting the methods wrong can mean research runs the risk of incorrectly placing improvable issues outside the boundary of "our ability to change" (p. viii) when they are indeed within reach[97].

MMHSR studies are commissioned in efforts to expand this boundary; implicitly assuming that either qualitative or quantitative methods used in isolation would not be appropriate. For the complex problems that MMHSR seeks to understand, there are important philosophical issues with choosing either the interpretivist (subjective) or the positivist (objective) paradigm, or placing either one in a position above the other[96].

Philosophers of social science, such as Trigg, have argued that philosophical issues arise because of the tendency of each paradigm to focus on a certain type of problem. Approaches

[5] System dynamics is the proper term for what is discussed in this dissertation. It is different from systems dynamics, systems dynamic, and dynamical systems. At times these terms are used improperly in the literature, so I make this clarification.

from the positivist paradigm focus on individual-level problems with measurable variables, often involving agents. Approaches from the interpretivist paradigm focus on context-level problems with intangible variables, often involving societal structures. Such individual/society (or agent/structure) issues are "the major problem facing social science" (p. 205) because they lead to research that ignores the "intertwined relationship of individual and social setting" (p. 207), which plays out over time through "the fact of unintended consequences" (p. 207)[98].

Public health scholars Sale *et al.* argue that, while holistic research is certainly needed in the domain of public health, mixing methods raises important philosophical issues because using qualitative and quantitative methods requires the use of multiple paradigms – depending on how they are mixed, research can enter the territory of *paradigm incommensurability.* When this paradigm problem occurs, information is lost and data can be misrepresented. To be empirically valid, according to Sale *et al.*, MMHSR studies should work to achieve complementarity until a paradigm shift occurs and social science develops an *integrative* methodology[99].

Therefore, for Trigg, the paradigm issue is the individual/society problem and for Sale *et al.* it is paradigm incommensurability. One approach to dealing with these paradigm issues is SDM, which has its own limitations (described below). Therefore, in this work, (objective #2) I seek to address paradigm issues raised above by addressing outstanding questions in SDM regarding its potential to act as an integrative paradigm and methodology.

According to the research of the system dynamics theorist David Lane, SDM use in practice follows an *integrative methodology*: "the social theoretic assumptions inferred from [SDM] practice are seen to stretch from objectivism and across social system theory into social action theory and, arguably, are consistent with *social theories that aim to <u>integrate</u> objective and subjective positions*"[100] (p. 455, emphasis added).

While all work in SDM integrates numbers and stories, prior to Lane's theoretical work[101-106], its position as a potentially-integrative methodology was not clear. Lane points out that the interpretivist tools which consider *social realities* within organizations are under-developed (e.g., solutions that are implementable and acceptable, dealing with norms and values, creating culturally feasible changes)[105]. Such tools would "attach importance to the conceptualization of situations [and] the creation of shared understanding of the different perspectives [of stakeholders]"[105] (p. 113).

The interpretivist roots of SDM are found in its use of *decision functions*[6] – equations that represent the way *information* is brought together in the mind of a decision-maker. These decisions are understood to follow policies which may or may not have ever been documented descriptively and to involve processing information from informal sources and perceptions[107] (p. 118). Therefore, Jay Forrester, the system dynamics founder, felt that selecting variables and linking them in decision functions, was not only the second most important step in the SDM model development process (after setting model boundaries), but that it was also the most difficult one[107] (p. 118).

Morecroft pointed out that "there is very little guidance for the earliest and sometimes most challenging step of initial conceptualization"[108] (p.14) (see also Morecroft[109, 110]). A decade later, Peterson & Eberlein[111] bemoaned the state of the SDM literature for (still) failing to articulate the procedures for how equations should be written, explaining that it required researchers to be able to make "unconscious or intuitive leaps" (p. 172). Contemporaneously, Richmond argued that such leaps cross a gap which appears for novices to be an "abyss"[112] (p. 145). Much SDM scholarship has worked to improve conceptualization methods[113, 114] (p. 110, 442)[115]. That said, the issues of social realities and shared understanding raised by Lane[105] and the setting of boundaries and selection and linking of variables raised by Forrester[107] (p. 118) have received some but relatively less, attention(see Sterman[113] (p. 138-139) and others[116, 117]). For SDM to function as an integrative methodology, the methods for addressing these questions need to be improved.

1.3.2.2 OBJECTIVE #3: PRESENTATION OF RESULTS THAT ENGAGE THE SUBCONSCIOUS, EMOTIONAL LEVEL ON WHICH DECISION-MAKING IS BASED

Policy-makers, whether elected officials, managers or physicians, make policies using reasoning that exists at a subconscious, emotional level, based partly on beliefs, intuition and tradition. Miller *et al.* argue that in order to persuade policy makers to change these policies, MMHSR needs *a mosaic that tells stories with numbers* in order to better engage them at this level[96].

In this dissertation, the term *mental model* is reserved for what is known in SDM as a *mental model of a dynamic system*; defined as:

> "A relatively enduring and accessible, but limited, internal conceptual representation of an external system (historical, existing or projected) whose structure is analogous to the perceived structure of that system"[118] (p. 414, parentheses in original).

[6] The term **decision function** is used interchangeably with the terms **policy function** and **operating policies** in system dynamics and in this dissertation.

The SDM research concept of mental models, therefore, considers the subconscious level of decision-making posited by Miller. This aspect is explained in more detail by Dolye and Ford:

> "Part of mental models may be altered, deleted or added on a time scale of minutes or seconds. Yet, a mental model considered as a whole, while continually changing in detail, may endure in memory in some form for years or decades. The phrase 'relatively enduring' means that the term 'mental model' should be reserved for cognitive structures that are stored in a potentially permanent state in long-term memory rather than for structures that are stored only temporarily (on the order of seconds or minutes) in short term or working memory."[119] (p. 17)

While SDM methods are designed for emotional engagement[120-122], several aspects are rarely mentioned, specifically, how individual interview transcripts can be used to:

- Explore the relationship between individual-level and shared understanding
- Generate and test the validity of qualitative and quantitative models
- Produce visualizations

Therefore, in this work, I seek to develop ways to present results that engage the subconscious, emotional level on which decision making is based (objective #3).

1.3.2.3 OBJECTIVE #4: ADDRESSING ISSUES OF EMPIRICAL VALIDITY

Formal validation in social science includes tools such as *Adjusted R-square* in multivariate regression modeling and *Saturation* in grounded theory. Researchers use these tools to determine when sufficient confidence in findings has been attained.

Under SDM, "validation is a gradual process of building confidence in the usefulness of a model—inherently a social, judgmental, qualitative process: models cannot be proved valid but can be judged to be so"[100] (p. 454). System Dynamics founder Jay W. Forrester considered two broad categories of validity: ultimate validity and interim validity. Ultimate validity can only be assessed when a model is *used* over time to improve a real system[107] (p. 115,117). Hence, because the goal is for models to be *used*, the interim goal for validity is *usefulness*[100, 123][7] and the maxim for validation in SDM is *all models are wrong, some are useful* (see Sterman[113] (p. 890) and[10]). The validation tools typically used in SDM are described in the Methods Chapter.

In MMHSR, empirical validation of qualitative and quantitative findings is expected, but this is complicated by the paradigm issues discussed above. While SDM validation methods are designed for validating structure, behavior and culture from both qualitative and quantitative

[7] Lane shows that this criteria, specifically "usefulness with respect to some purpose" (p. 184 in Barlas (1996)), is not just an aim of validation in SDM, but is also considered the end goal in most similar disciplines.

viewpoints, the quantitative methods are better developed than the qualitative ones. Therefore, in this work, I seek to develop a novel set of validation methods that use qualitative information for validating both qualitative and quantitative system dynamics models (objective #4).

1.4 DISSERTATION OVERVIEW

This section presents an overview of this dissertation. A description of each chapter is provided, including for the Scoping Study - Appendix C.

Chapter 1 Introduction – This chapter introduced a broad understanding of the problem and system within which it was found. It also presented the aims and objectives of this research (see Table 1.2).

Table 1.2 Aims & Objectives of this Dissertation

My theoretical aim is **to develop a better *theory* of primary care transformation (PCT); specifically,** to develop a grounded, dynamic theory of PCT in order to build understanding of the key structures generating success and failure (**objective #1**). In meeting this aim, this research will develop useful theory for anyone engaged in PCT, facilitating improved transformation. The dissertation relies on the following problem statement: *Primary care transformation has been and continues to be an elusive target. In the short term, transformation is hard and failure abounds. In the long term, some practices reach successful transformation. We lack sufficient understanding of the structure of primary care, and of the policies that can impact this structure*. While the value of the primary care tenets is well understood, current theory lacks an understanding of *the complex interactions* between them as well as their interaction within the system of care already in place. This understanding is necessary in order to realize the aspirations of health care systems worldwide.
My methodological aim is **to select adequate *methods,* and create new ones when needed, to understand the dynamics of PCT**. In the course of striving to meet this aim, this work seeks to address outstanding research questions in MMHSR by developing contributions to SDM methods. This effort involves three more objectives: addressing paradigm issues (**objective #2**), presentation of results that engage the subconscious, emotional level on which decision making is based (**objective #3**) and addressing issues of empirical validity (**objective #4**).

Appendix C Scoping Study – This appendix identifies and defines the problem in more detail. It also presents the analysis leading to the selection of system dynamics modeling and leading to the clarified problem statement and conceptual framework. It is mentioned here (after Chapter 1) to clarify the sequence of events but it is an appendix (rather than a chapter of the thesis) because it was preliminary to the SDM work that is the focus of this dissertation.

Chapter 2 Methods – This chapter describes the system dynamics standard methods as well as the system dynamics methods that I developed. Both are then described in context of this research (what I did step-by-step).

Chapter 3 Results & Validation – This chapter presents the models developed and the validation results. The simulation results and the Theoretical Model are also presented. Finally, meta-level validation results (reflecting on the research process overall) are presented.

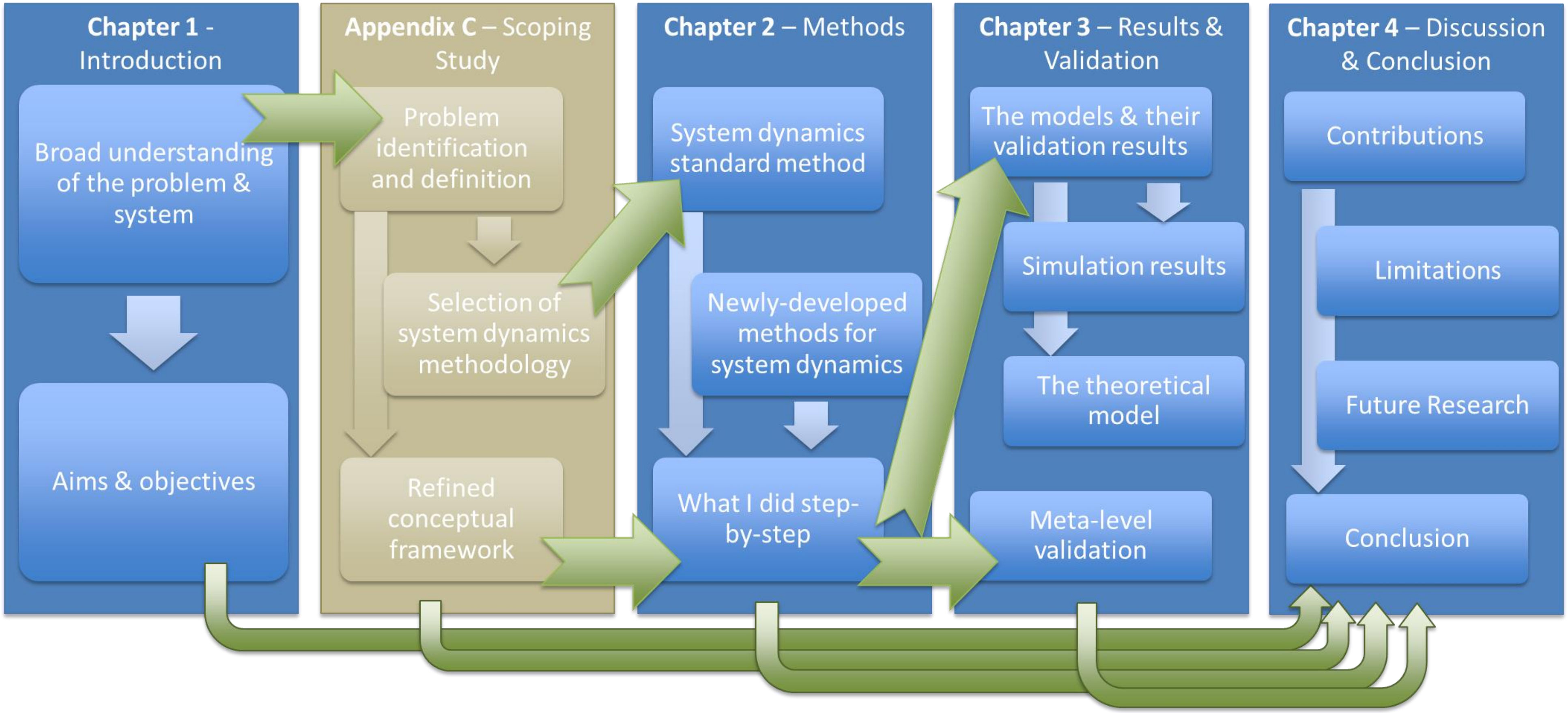
Chapter 4 Discussion & Conclusions – This chapter ties together the previous chapters. I discuss the theoretical and methodological contributions of this work as well as its limitations and areas for future research.

Figure 1.2 visualizes the flow of this dissertation. Dark-blue rectangles identify chapters. Light-blue objects identify chapter contents. Arrows show how contents are linked (blue for links inside the chapter and green for links across chapters). Not shown in this figure are the other appendices included in this dissertation, namely:

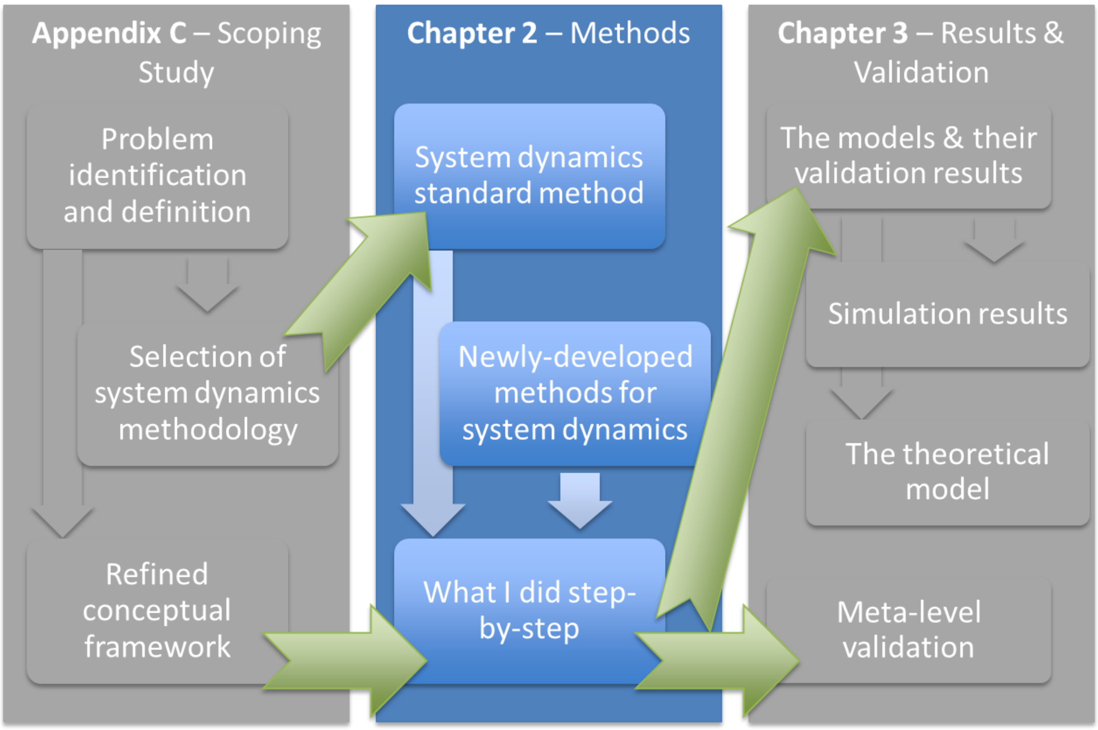
- Appendix A Publications & Presentations Relevant to this Dissertation
- Appendix B Ethical Approvals
- (Appendix C Scoping Study – mentioned above)
- Appendix D Validation Purpose and Types in System Dynamics
- Appendix E Newly-Developed Validation Methods (developed during this research)
- Appendix F Model Revision & Documentation (both quotations and equations)
- Appendix G Behavior & Policy Sensitivity Analysis Summary

This figure is used on each chapter's title page to orient the reader to the contents of the chapter as well as how the chapter builds on previous work and leads to the next chapter.

Figure 1.2 Overview of this Dissertation



CHAPTER 2 METHODOLOGY AND METHODS



2.1 INTRODUCTION

Miller *et al.*[124] see the purpose of health services research as explicating the structure of the healthcare system. Similarly, Stange *et al.* view the field as "examining the effect of systems on the structure, process, and outcomes of health care."[2] (p. 289) This dissertation explores the structure of the primary care system by developing a white-box, dynamic model – a model where the purpose is distinct from forecasting (e.g., time-series, regression models) as the model describes "how the real system actually operates in some aspects"[123] (p. 185).

One of the fundamental features of primary care is its concern for patients' health over time[2]. The SDM methods used in this dissertation focus on developing theory around the question of how the causal relationships in this system interact over time. A simulation model is developed to enable research into what these interactions mean for policy and decision-making. In other words, SDM is used to uncover the system structure underlying these dynamics, and converting what were muddled descriptions into an effectively-specified theory of the system dynamics of primary care transformation (PCT).

This chapter presents the modeling and validation methods employed. All of the items referenced in this introduction are defined and described in greater detail in the sections that follow.

> Section 2.1 orients the reader to the **overall approach**
>
> Section 2.2 defines SDM and describes the **standard SDM methods**
>
> Section 2.3 describes **newly-developed methods** created as part of this dissertation and integrates them with standard SDM methods (for both model development and validation)
>
> Section 2.4 presents the (standard and newly-developed) **methods as they were used** in this dissertation
>
> Section 2.5 summarizes the result of the **methods development aim**

The methods used in this dissertation are grouped into four phases as presented in Table 2.1 below. For each phase, I list its purpose, as well as the corresponding input and output. This table begins with the problem definition identified in the scoping study (Appendix C). Phase 1 elicits participants' mental models[8] using the semi-structured interview transcripts. CLDs are produced for each individual participant. In Phase 2, these CLDs are iteratively merged to

---

[8] See Section 1.3.2.2 for the definition of mental model used in SDM and in this dissertation.

produce a shared mental model which finally becomes the Conceptual Model. Phase 3 uses the Conceptual Model to build the Simulation Model. Phase 4 uses the Simulation Model to develop the Theoretical Model and policy recommendations. The Theoretical Model summarizes my contribution to the theory of primary care – it is my most comprehensive *dynamic hypothesis* and the result of my grounded dynamic analysis. There are two cross-phase steps shown on the right of this table. Throughout this research, the problem statement (see Section 1.2.3) is improved. Also, at appropriate points along the way, validation methods are employed to verify results (and make revisions as needed).

Table 2.1 Methods Summary Table

Study Phase	Purpose	Input	Output		
Scoping Study	Problem Definition	Mixed Methods Project Data	Problem Definition & Choice of SDM	**Cross-Phase Step:** Iterative Improvement of Problem Statement & Dynamic Hypothesis	**Cross-Phase Step:** Validation
Phase 1	Mental Model Elicitation	Semi-Structured Interviews	Participant CLDs		
Phase 2	Develop Conceptual Model	Participant CLDs	Conceptual Model		
Phase 3	Develop Simulation Model	Conceptual Model	Simulation Model		
Phase 4	Policy Analysis	Simulation Model	Policy & Theory Results		

These methods iteratively develop and validate a problem statement and a dynamic hypothesis. This hypothesis is expressed in the form of models, beginning with CLDs and culminating in the development of a theoretical model. Model validation can also contribute to model development when improvements are made. Table 2.2 below presents the models I developed[9] and a brief summary of the role each one plays in overall model development.

[9] Implementing validation tests may results in periodic model revisions, these are labeled accordingly (e.g., I marked models made during the SMM-S Test as SMM2.1, SMM2.2 … SMM2.x; but these are referred to as SMM2 in this dissertation) – not shown in Table 2.2 above.

Table 2.2 Model Summary Table

Model name	Role in model development
Participant CLDs	The CLDs that were generated; one specific to each individual interview.
Team CLDs, Clinic CLDs	The CLDs that were generated in implementing CLD Combination; one specific to each team and clinic in the *model development set* of interviews.
SMM1	Shared mental model 1 is the result of CLD Combination. It is the first model that is assumed to be a shared mental model (i.e., it is the first draft of the dynamic hypothesis).
SMM2	Shared mental model 2 is the version of the model after modifications made during the **Shared Mental Model - Saturation (SMM-S) Test:** Do the different clinics agree on the structure of the system?
SMM3 / Conceptual Model	Shared mental model 3 is the version of the model after modifications made during the **Conceptual Model - Saturation (CM-S) Test:** Does an additional clinic agree with SMM2 on the structure of the system? (SMM3 is also referred to as the Conceptual Model).
SIM1	Simulation model 1 is the first quantitative version of the model. It is produced using SMM3 as the blueprint.
SIM2	Simulation model 2 is the version of the model after modifications made during the **Simulation Model - Saturation (SIM-S) Test:** Does an additional clinic agree with SIM1 on the structure *and behavior* of the system?
SIM3	Simulation model 3 is the version of the model after policy analysis structures were added to SIM2.
Theoretical Model	It is a visualization of the Simulation Model, bringing together the policy structure diagrams that describe the Simulation Model.

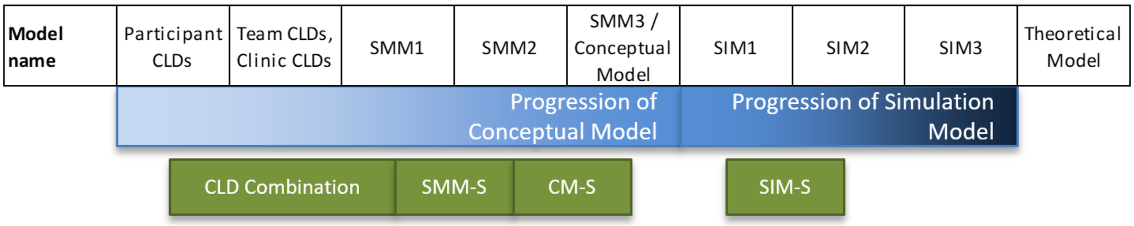
Table 2.3 below places each of these models inside its respective phase of the research. Model development methods and model validation methods performed within each phase are also referenced. Newly-developed methods (proposed in this dissertation) are marked with an asterisk.

Table 2.3 Model Iterations & Validation Methods

Study Phase	Purpose	Model Development Method	Model Iterations	Validation Methods: Model Tests	Validation Methods: Meta-level Tests
Scoping Study	Problem Definition	N/A	N/A	N/A	SD-S* Data Suitability* Methods Suitability* Stakeholder Suitability* Cessation
Phase 1	Mental Model Elicitation	Purposive Text Analysis Mild Pruning	Participant CLDs	None	
Phase 2	Develop Conceptual Model	CLD Combination* Pruning	Team CLDs, Clinic CLDs SMM1 SMM2 SMM3/ Conceptual Model	SMM-S Test* CM-S Test* Stakeholder Dialogue	
Phase 3	Develop Simulation Model	Simulation Modeling	SIM1 SIM2	SDM Standard Methods SIM-S Test* Stakeholder Dialogue	
Phase 4	Policy Analysis	N/A	SIM3 Theoretical Model	Stakeholder Dialogue	
* = Newly-developed methods for SDM, proposed in this dissertation.					

Figure 2.1 below presents the sequence of model iterations, progressing from the participant CLDs to the Theoretical Model. It also indicates where the newly-developed model development method (CLD Combination) and model validation tests are performed (SMM-S, CM-S and SIM-S Tests). Other validation methods listed in Table 2.3 above are not shown here. This figure is used throughout this dissertation to orient the reader to where the model and/or method being described is located in the sequence (using gray background).

Figure 2.1 Model Progression Sequence



2.2SYSTEM DYNAMICS

This section defines SDM and presents the standard SDM methods for model development and model validation. These methods are given an in-depth treatment because, in this chapter, I also describe and justify my newly-developed methods for SDM (Section 2.3). These descriptions give the reader a necessary foundation to understand my research process (Section 2.4), and therefore for considering my contributions not only to Aim 1 (theoretical aim - Section 4.2.1) but also to Aim 2 (methodological aim - Section 4.2.2 through Section 4.2.4).

2.2.1AN ORIENTATION TO SYSTEM DYNAMICS

Dynamic complexity acknowledges that complex systems have component parts, which interact in ways that produce counterintuitive behaviors[125]. Complex systems are "dynamic, evolving, and interconnected" (p. 506), resulting in "feedback [loops], time delays and stocks and flows" *(ibid.)* that are not understood by typical "static, narrow, and reductionist" *(ibid.)* mental models[126].

Humans are unable to mentally infer the dynamic behavior of complex systems[127-132], which leads to failures in performance in dealing with such systems[130, 133-137]. Thus, a mental model that accounts for feedback loops, time delays, as well as stocks and flows is required to understand the behavior of such systems.

SDM attempts to capture the causal structure of complex phenomena, which is understood best by the stakeholders who experience these phenomena most directly. This is because these stakeholders' experiences have impacted their mental models in long-term memory. These experiences are assumed to imprint elements (i.e., variables, links, delays and feedback loops) of the causal structure of complex phenomena on mental models[119].

The *theory of system dynamics* proposes that social systems present evolving behaviors, which can be explained by "endogenous processes represented by feedback loops, rates, and stock variables"(i.e., the components of a system dynamics model)[138] (p. 474). It also provides principles on how these components should be used in constructing models[138]. The purpose of a system dynamics model is not necessarily to forecast (*black-box* or *correlational* model) but rather to explore how certain aspects of the real system actually operate (*white box* or *causal-descriptive* model)[123]. These models can be used to develop theories, gain insights, and assist in developing solutions.

According to the original work on SDM – System Dynamics founder Jay W. Forrester's book, Industrial Dynamics, – mathematical models can be useful for building understanding and can "impart precision to our thinking"[107] (p. 57). Modeling helps to convert subjective mental data into a more precise but not necessarily more accurate form. Such a model is a bridge between the subjective realm of "what we *believe* to be the nature of the system under study" (ibid., emphasis in original) and the objective realm of numbers. Forrester envisioned the goal of such SDM studies to be improving mental models by correcting errors in assumptions, eliminating prejudices, and inconsistencies "with the *qualitative* nature of the real world"[107] (p. 58, emphasis in original). As accuracy becomes a goal, the correspondence of model behavior to numbers in the real system takes on more importance.

Commenting on the use of SDM in public health, Sterman points to three challenges stemming from complexity, namely that it hinders: *generation of, learning from, and implementation of polices on the basis of* evidence[10][126]. He argues that there are only *effects*; that "'side effects' are not a feature of reality, but a sign that the boundaries of our mental models are too narrow, our time horizon too short."[126] (p. 505)

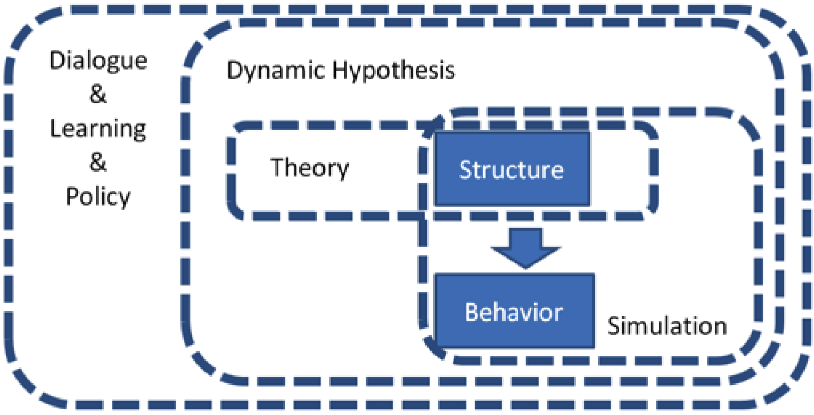
In brief, SDM is "an iterative process in which the modeler will test a *dynamic hypothesis* that represents a feedback *theory* or causal *structure* generating a series of *behaviors* over time [seen via *simulation*], allowing the problem *actors to learn* about the situation, and to design or redesign their guidance policies"[139] (p. 275, emphasis added). In so doing, a model can *build understanding* among participants of how their current mental models contribute to an important problem[140].

SDM sees complex systems as having a structure that causes the observed behavior (Figure 2.2, solid blue boxes). A problem is identified. A theory is posited for the mechanisms involved in generating the problematic behavior (dashed box encompassing theory for the structure producing the observed problem). This theory can be represented via a simulation model that includes all of the structure needed to endogenously generate the problematic behavior (dashed box labeled simulation). The simulation model can be used to test the dynamic hypothesis that the theory posits, providing the needed environment to improve the theory (dashed box labeled dynamic hypothesis).

[10] *Generation* **–** for some important public health problems a randomized controlled trial would be unethical; in others there are ripple effects elsewhere in the system; in yet others there are long time delays that make it difficult to explore intergenerational effects. *Learning* – diffusion of innovation occurs slowly and unevenly. *Implementation* – Policies often fail when all stakeholders are not included in their design.

The SDM research process is designed to be participatory- individuals improve their mental models (learning) as their ideas are represented in the model (dialogue) and the full range of dynamic implications of the policies operating in the system is explored (dashed box labeled *dialogue and learning and policy*). This description is visualized in Figure 2.2 below. Throughout this figure, dashed lines mean that these represented aspects may change as people learn more about the system structure causing the problematic behavior.

Figure 2.2 System Dynamics Modeling



This figure is adapted from multiple sources[113, 138, 139].

In so doing, *SDM* adds a level of empowerment, fighting against the belief that "we are helpless victims of forces that we neither influence nor comprehend"[126]. Simulations provide a safe place to experiment with different conditions, time delays, interventions, etc. Extreme condition tests allow for learning more about a system's structure and dynamics. People using models to simulate these extreme conditions attain a better understanding of the system and potential emergencies that may arise without endangering the real system as the entire experience is simulated. This experience of interacting with the simulation results in improving one's mental model.

Health systems are characterized by policy resistance, that is, situations where a policy or administrative decision, like a particular aspect of PCT designed to solve a problem, actually causes the problem to get worse and/or makes it more intractable[126, 141]. By including these *unintended consequences,* system dynamics is well-suited to finding and designing sustainable policy solutions. Policy resistance does not refer to individuals having a resistant attitude toward a change in policy, but rather it refers to a characteristic of complex systems where a change in policy changes system behavior, but not as intended by policy-makers[141, 142]. This may be due to the fact that policies intended to improve one aspect of the complex system will tend to result in sub-optimal levels for other aspects of the system[143].

2.2.2 FROM MODEL TO THEORY – RESEARCH IN SYSTEM DYNAMICS MODELING

A model is "a tangible aid to imagination and learning, a transitional object to help people make better sense of a partly understood world"[114] (p. 404). It is something, with which a stakeholder can interact, to improve their mental model. As expert practitioners in modeling broadly (and specifically in SDM) use varying terms for describing the purpose of modeling as well as for the different types of models, Table 2.4 presents the terms used in this dissertation for the purpose of modeling and for the types of models used, alongside some commonly used terms.

Table 2.4 Comparison Between Terms in this Thesis and Commonly Used Terms

<table>
<tr><th></th><th>This Dissertation</th><th>Morecroft[114] (p. 110,442)</th><th>Lane[105]</th><th>Oral & Kettani[144]</th></tr>
<tr><td>Purpose</td><td>Problem Statement</td><td>Problem Articulation</td><td>Appreciation of the Situation</td><td>Managerial Situation</td></tr>
<tr><td rowspan="3">Types of Models</td><td>Conceptual Model</td><td rowspan="3">Formal Models
• Conceptual
• Algebraic</td><td>Communicated Conceptual Model</td><td>Conceptual Model</td></tr>
<tr><td>Simulation Model</td><td rowspan="2">Formal Model</td><td>Formal Model</td></tr>
<tr><td>Theoretical Model</td><td>Theoretical Model</td></tr>
</table>

This dissertation presents three types of models: *Conceptual*, *Simulation* and *Theoretical* models. Both the Conceptual Model and the Simulation Model are formal models. The Conceptual Model is a visualization of the tensions in the structure of primary care that induce the diverse transformation trajectories identified in the problem statement. The Simulation Model is the algebraic representation of those tensions. The Theoretical Model presents a summary of the understanding gained from this process – it is a simplified visualization of the Simulation Model. In other words, the chronology of models is: *Conceptual Model* to *Simulation Model* to *Theoretical Model*.

SDM assumes that the relationship between individuals and context can be disentangled conceptually such that stakeholders can address important problems. This is done by developing a dynamic hypothesis based on participants' stated understanding of the causal structure of the system behind a given problem. This hypothesis is codified explicitly in both the conceptual and the simulation models. Simulation identifies the impact of unintended consequences and the leverage for addressing them. Lane refers to the result as a *minor content theory* that is made up of causal relationships between variables complete with units of measure and measurements[103, 145]. This theory also "makes the concepts and causal processes identified by problem owners concrete and explicit"[138] (p. 474). These claims are concepts that can be tested empirically; thus helping model users to understand what can be done about a problem and what is beyond control.

In a review of SDM literature, Kopainsky and Luna-Reyes[138] pointed out that, in conceptual terms, SDM model building and theory building are similar processes: both work on the basis of iterative improvements. Indeed, in the applied sciences, modeling and theory serve essentially the same purpose, that is, both seek to better understand phenomena or systems in order to improve decision-making or action in the area under study[146, 147]. Model building in SDM can include mathematical approaches where researchers seek to discover patterns in data, or it can rely entirely on qualitative approaches where stakeholders' stories are converted into *mental models* that are then built into computer simulations[139]. In both cases, SDM results in simulations to "close the loop in the theory building process"[138] (p. 483).

In the first column, Table 2.5 below identifies the sequence of research steps for applied theory building as outlined by Lynham[147]. In the second column, this table identifies steps for SDM[139]. These steps parallel those outlined by Lynham. In the following subsections, I will describe each step.

Table 2.5 Theory Building & The System Dynamics Methodology

The Research Method for Applied Theory Building	System Dynamics Methodology
Conceptual Development	**Conceptualization** Problem articulation
	Dynamic hypothesis
Operationalization	**Simulation Phase** Model formulation
(Dis-)confirmation	**Testing** Model behavior
	Model evaluation
Application	**Implementation** Policy analysis
	Use
***Continuous* refinement and development of the theory**	**Iteration***
Note: While system dynamics modelers may differ in the exact terms used and their segmentation of the steps, the content areas and their progression are the same. *- Iteration happens within steps as well as within a project and across projects. The point is models are continuously developed and then brought back to stakeholders or raw data for improvement.	

2.2.2.1 CONCEPTUALIZATION

Conceptualization consists of generating a problem statement and dynamic hypothesis.

2.2.2.1.1 PROBLEM DEFINITION

A problem statement is drafted, which focuses subsequent work such that a system boundary is made clear—including the necessary elements (variables, links, delays and feedbacks), the appropriate time scale, and sufficient details so that the problem under investigation is endogenously produced. Without this, a model can quickly become a messy diagram that can overwhelm policy makers without adding value.

Sweeney *et al.*[148] recommend thinking purposefully when choosing the time horizon for the analysis. They also emphasize the importance of defining the problem context and audience. Defining these elements differently will likely change the variables to include or exclude (i.e., *model boundary*) and thus produce a different result. Sterman[113] (p. 86) recommends the following questions to consider:

- What is the key problem?
- What are the key concepts and system variables?
- What is the time horizon?
- What happened to these concepts/variables in the past? What is likely to happen to these concepts/variables in the future? (these are known as Reference Modes)

For example, we may be interested in the functioning of a primary care clinic. If we are interested in patient flow within the clinic, we may focus on a workday. If we are interested in how that clinic's care management program impacts readmissions, we may focus on a week to a month. If we are interested in patient behavior change efforts, we may focus on three months to one year. The audience may be the clinic staff, but it could also be the clinic management or the regional policy-maker.

To develop the problem statement, a review of the relevant literature should first be conducted to better understand the problem under investigation. Also, discussions with stakeholders can provide added direction and clarity. In order to represent stakeholder perspectives in a balanced manner, a systematic method can be used to select key participants[149, 150].

2.2.2.1.2 DYNAMIC HYPOTHESIS

The articulated problem guides the modeler in formulating a dynamic hypothesis that explains the problem behavior as being endogenously produced via feedback structures among key system variables[113] (p. 86). Methods for formulating a *dynamic hypothesis* include:

- Interacting with the client to build visualizations of the dynamic hypothesis based on the client's mental model[113] (p. 80-81)
- Holding group model building sessions where stakeholders are engaged in the development of the visualizations[114] (p. 112)[151]
- Analyzing existing qualitative data (e.g., meeting transcripts)[117]

These methods are also used for eliciting simulation model insights (Section 2.2.2.2).

The specific method chosen depends upon the data and resources available and skills/comfort level of the researcher. The key to identifying dynamically important information is not in how the original source was generated, but in the skill of the researcher to recognize when a system structure exhibits the signs of feedbacks, time delays and nonlinearities.

For example, purposive text analysis utilizes existing text (or it can be used to analyze newly-generated text)[117] while one-on-one and group model building require interaction with stakeholders[152]. Purposive text analysis relies heavily upon qualitative analysis skills whereas one-on-one and group model building relies heavily upon interviewing and group facilitation skills. One-on-one model building may better capture minority thoughts whereas group model building relies upon skillfully crafted groups and real-time facilitation to protect the minority's comfort level with sharing thoughts. If there is a need to build a shared understanding of the system, then group model building process will lead to a better result.

One way to map a dynamic hypothesis is with the use of a Causal Loop Diagram (CLD)[11][113] (p.28-29)[118, 153, 154]. A CLD is one way of visualizing the causal linkages in a system[155]. They function as an early step toward developing a dynamic hypothesis, or a summary of one's dynamic hypothesis from a simulation model[121]. . The dynamic hypothesis could represent a theory, an individual's mental model, a group's (shared or not shared) understanding.

[11] There has been disagreement among US-based (e.g., Doyle and Sterman) and UK-based SDM researchers (e.g., Lane) regarding whether or not system dynamics models used to represent mental models in general (and CLDs by implication) are equivalent to, or a form of, *cognitive map*. I use the term *CLD* for CLD representations of peoples' and groups' mental models.

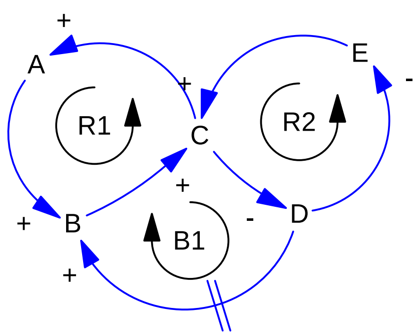
CLDs are one of many tools used for mapping causal structure[156]. Others include: model boundary diagrams, sector maps[12], stock and flow diagrams, and policy structure diagrams[113] (p. 86,97-102).

As CLDs are used in this dissertation, below, I present the way to read them. For additional information, please see, for example, Tomoaia-Cotisel *et al.*[154].

> ***Variables*** identify system components[113] (p. 152).
>
> ***Arrows*** link variables to indicate a cause-effect relationship. Each arrow has a *polarity* to show the nature of the causal relationship (denoted by a plus or minus sign at the arrow head). Polarity between two variables describes the relationship between changes in those variables where all other variables in the system are held constant. When ***time delays*** exist between cause and effect variables, they are marked by a double line crossing the stem of the arrow[113, 115] (p.138-139).
>
> Causal relationships among variables can create two types of ***feedback loops***. ***Reinforcing (R) loops*** describe a positive feedback process where the loop generates a self-reinforcing behavior—as in a vicious cycle or a virtuous cycle. ***Balancing (B) loops*** describe a negative feedback process where the loop generates a self-correcting, goal-seeking, or stabilizing behavior. The rate of change decreases over time, and the loop brings the system to an equilibrium[113] (p.138-139).

Figure 2.3 presents a hypothetical CLD with two reinforcing loops and one balancing loop with a delay.

Figure 2.3 A Causal Loop Diagram



[12] Also referred to as subsystem diagrams.

2.2.2.2 SIMULATION MODEL

A simulation model[13] is "an *inference engine* to diagnose performance problems; a virtual world *to experience* dynamic complexity and stimulate imagination; and a laboratory *to design and test* new policies and strategies"[114] (p. 89, emphasis added). Simulation model formulation involves "positing a detailed structure and selecting the parameter values"[139] (p. 276) for the problem of interest. This is done using the dynamic hypothesis and operational thinking – focusing on "how things really work… [rather than on] how things would theoretically work, or how one might fashion a bit of algebra capable of generating realistic-looking output"[157] (p. 127) – as well as additional insights from those experiencing the problem.

Methods for eliciting these insights are similar to those presented for formulating a dynamic hypothesis (Section 2.2.2.1.2 above). The difference is that this time, the methods are focused on the simulation model rather than on visualizations of the dynamic hypothesis (corresponding emphasis added below):

- Interacting with the client *to acquire data, ask clarifying questions* (e.g., to build table functions, clarify structural flaws uncovered in the model building process)[113] (p. 585-595)
- Holding group model building session where stakeholders are engaged in the *development of the simulation model*[114] (p. 112)[151]
- Analyzing existing *qualitative data* (e.g., meeting transcripts)[117]

Stock and flow diagrams communicate the detailed structure of SDM simulation models. Whereas CLDs emphasize feedback structures, stock and flow diagrams emphasize the physical structure: tracking accumulations (termed 'stocks', 'levels' or 'states') and how they move through the system (termed 'flows' or 'rates')[113] (p. 102). Two additional types of variables are used in stock and flow diagrams: auxiliary variables represent calculations and constants bring in information from outside the system. Finally, stock and flow diagrams are converted into stock and flow *models* using "friendly algebra"[114] (p. 84, 115). There are two types of equations: *stock equations* calculating the change in the stocks, and *decision functions*[14] calculating what changes to make to those stocks based on information which is taken in from across the current values of the stocks. These range in complexity from "a simple equation"

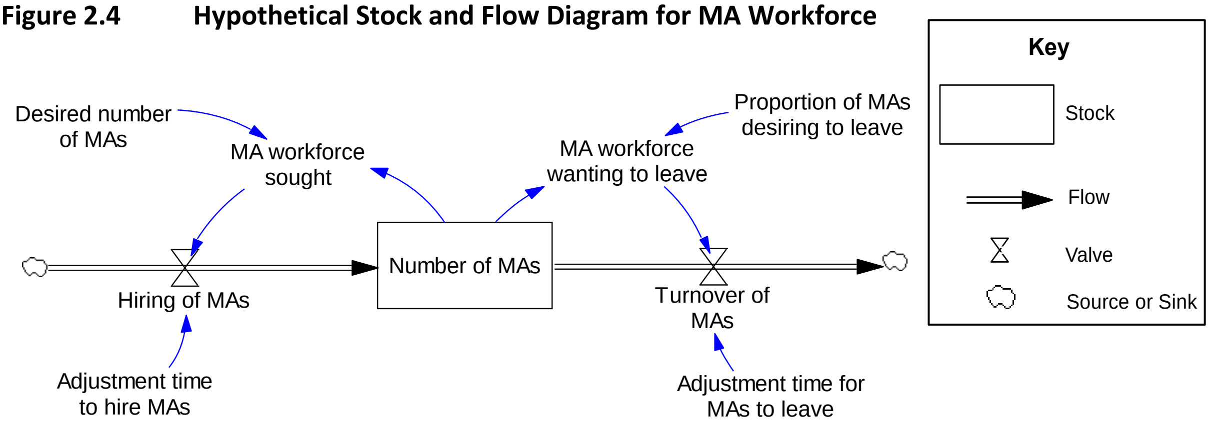
[13] This term is lower case when referring to simulation models generally. It is capitalized when specifically referring to the final simulation model (SIM 2 or SIM 3) developed in this dissertation.
[14] The term **decision function** is used interchangeably with the terms **policy function** and **operating policies** in system dynamics and in this dissertation.

(p. 69) to “a long and elaborate sequence of computations that progresses through the evaluation of a number of intermediate concepts”[107] *(ibid.)*.

As stock and flow diagrams are used heavily in this dissertation, below, I present the way to read them.

Figure 2.4 below presents a hypothetical stock and flow structure for the number of medical assistants employed in a clinic. The stock *Number of MAs* tracks the accumulation of medical assistants (MAs) over time. At any one point in time, it holds the number of MAs currently employed. Flowing in are newly hired MAs. Flowing out are MAs that are leaving the organization (i.e., turnover). On the left, the organization desires to have 10 MAs. It monitors the number of MAs and when it is below the desired level, it initiates the hiring process. On the right, a measure is found for how much the MAs desire to leave their employment. Both flows have an adjustment time; that is the amount of time it takes the organization to hire MAs (for the inflow) and the amount of time it takes for MAs to depart (for the outflow). Where MAs come from before they are hired (source) and where MAs go after they leave the clinic (sink) are outside the model boundary; therefore, they are represented by clouds. The equations for this structure are presented in Box 2.1 with units in red, inside parentheses (e.g., (MAs)). In the equations, *dt* represents the time step.

Figure 2.4 Hypothetical Stock and Flow Diagram for MA Workforce



Box 2.1 Equations Corresponding to Figure 2.4

Stock

Number of MAs (MAs) = $\int$ [hiring of MA - turnover of MA] *dt*

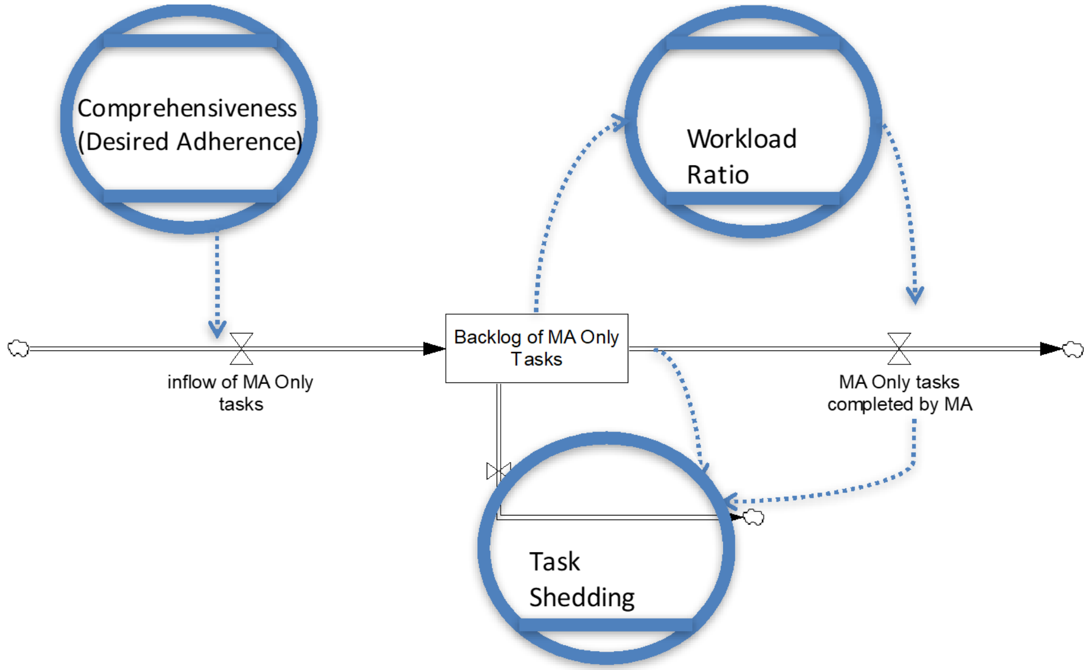
Initial value = 10

Flows

Hiring of MAs (MAs/Month) = MA workforce sought / adjustment time to hire MAs

SDM uses diagrams to visually communicate the essential parts of a simulation model. Policy structure diagrams "focus on operating policies and strip away all the underlying formulational detail"[114] (p.162). Operating policies[15] are visually represented as "a large circle with information inputs and outputs" and include two types: one "that interprets information and hands it on to other parts of the [system]" and another that "subsumes a flow rate where a decision-making process … leads to action"[114] (p. 162-163). As Chapter 3 uses these diagrams to walk through the simulation model as well as to present the Theoretical Model, I present an example in this section. Figure 2.5 below has both types of operating policies: (1) those transforming and passing information ("*Comprehensiveness (Desired Adherence)*" and "*Workload Ratio*") and (2) those leading to action ("*Task Shedding*").

Figure 2.5 Hypothetical Policy Structure Diagram for MA-only Tasks



[15] The term **decision function** is used interchangeably with the terms **policy function** and **operating policies** in system dynamics and in this dissertation.

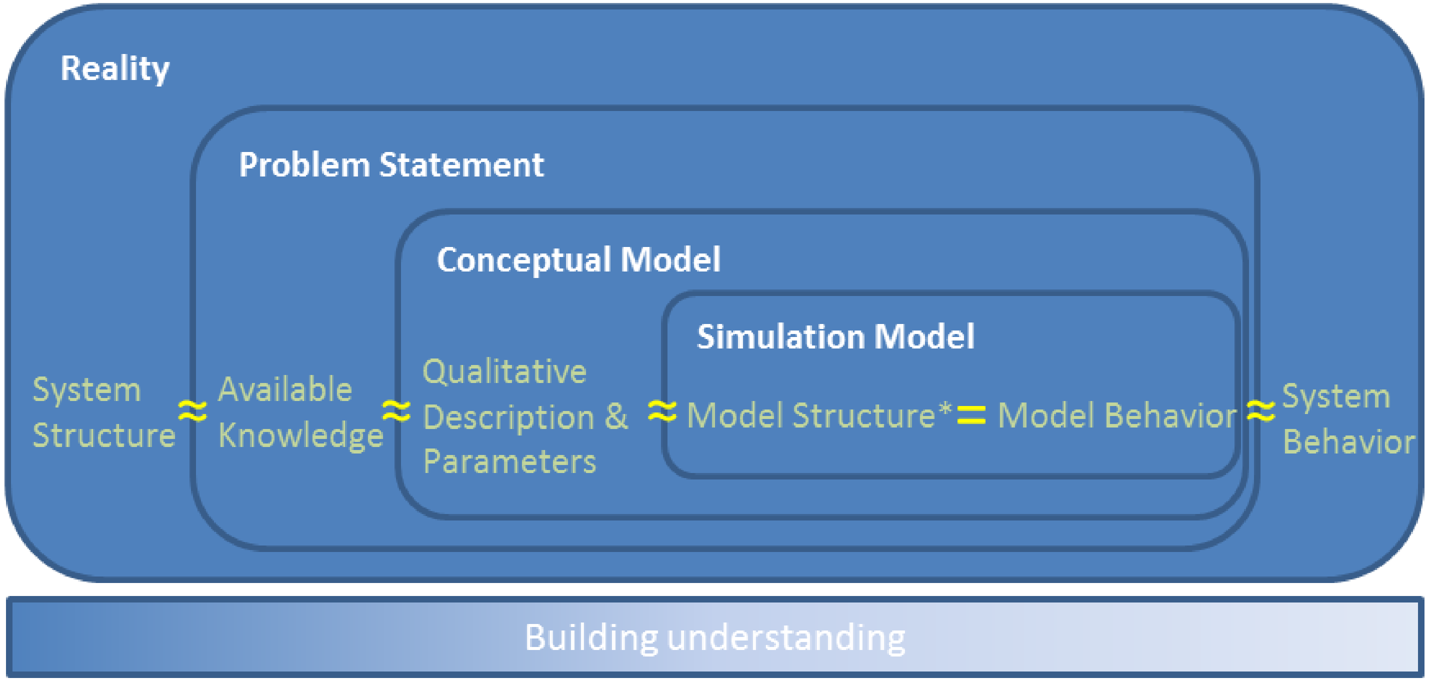
> "The *presumption* of model significance rests on two foundations. Primarily, confidence depends on how *acceptable the model is as a representation* of the separate organizational and decision-making details of the actual system. Secondarily, confidence is confirmed by the correspondence of total model *behavior to that of the actual system*." Industrial Dynamics[107] (p. 117, emphasis added)

Under SDM, "validation is a *gradual process of building confidence in the usefulness* of a model—inherently a social, judgmental, qualitative process: models cannot be proved valid but can be judged to be so"[100] (p. 454). The social aspect involves developing a model that stakeholders as a group find acceptable because they agree that it is useful. Stakeholder agreement on usefulness is the goal, not absolute truth[158]. As Richardson & Pugh explain, validation in SDM is also a process: "an ongoing mix of activities embedded throughout the iterative model building process"[159] (p. 311).

Figure 2.6 below illustrates the different aspects of an SDM study requiring validation (the yellow text), using a basic view of the increasingly narrow problem definition as one moves (from left to right) from the real world to the quantitative simulation model and of the increasing understanding along the way (the *building understanding* box becomes lighter). There is a model building process from left to right, where on the right, one reaches the maximal understanding via simulation approximating system behavior. In that model building process, it is imperative to validate along the way[16] so as not to carry forward an early mistake and so as to build confidence in the model (and thus the understanding that it generates). The only equal sign in the figure is that from model structure to model behavior. The other links between yellow phrases are shown via an approximation symbol (the squiggly equal sign). Each time this symbol is used, there is a need for validation that the approximation made between the two sides of the sign is sufficiently good.

[16] See Table 2.6 and Table 2.10 for validation tests that can be implemented. See Table 2.9 and Table 2.11 for how these tests contribute to different types of validity.

Figure 2.6 Validation is Embedded Throughout the SDM Research Process



* System boundaries, interacting variables, values of parameters

This figure was developed based upon my reading of works on SDM's theory of validation[100, 105, 107, 123, 160] (p.115-129).

White-box models like SDM models are *causal-descriptive* rather than *correlational*, meaning that they "are statements as to how real systems actually operate in some aspects… [each is] a 'theory' about the real system [and] must not only reproduce/predict its behavior but also *explain* how the behavior is generated…"[123] (p. 185). Therefore, both structural and behavioral tests are necessary to validate the model.

Table 2.6 below presents a list of tests commonly used in SDM model validation. These tests are not necessarily mutually exclusive. Furthermore, not all of them have to be performed. There is not an order of importance for the validation tests. Questions such as how *do* research teams choose to stop changing the model[161], how many tests are performed, how are the tests chosen, and when and how the results are reported are still unanswered in SDM[162] and in modeling broadly[163]. Nevertheless, typically, modelers select a battery of validation tests from the list in Table 2.6 below, presumably until they are sufficiently confident in the model and its usefulness that they choose to stop changing the model structure (and/or when time or money run out)[123, 161, 164]. The most common tests for studies with numerical data are the *Structure Confirmation - Theoretical, Mode Reproduction Ability,* and *Behavior Sensitivity* tests. For studies, like this one, where over-time numerical data are not available, the *Extreme Policy, Structure Boundary Adequacy* and *Extreme Conditions in Equations* tests are used most [162].

For each test, a definition is also provided. This table is adapted from Table 1 provided in Lane[105]. Test names were modified to reflect both Lane's[105] and Barlas's[123] list of tests. These lists compile tests developed and described in numerous other sources[146, 165, 166]. Newer texts refer to these same tests[114] (p. 441)[113] (p. 858-861). Table 2.6, Table 2.7, Table 2.9 and Section 2.3.2 Model Validation below provide explanations of what these tests mean.

Table 2.6 Model Validation Tests as used in the SDM Standard Method

Name of test	**Definition**[123] (p. 189-195)} and[105] (p. 123)
Parameter Confirmation - Theoretical	"Evaluating the constant parameters ... [against generalized knowledge about the system that exists in the literature] both conceptually and numerically."[123]
Parameter Confirmation - Empirical	"Evaluating the constant parameters against knowledge [obtained directly from] the real system, both conceptually and numerically."[123]
Structure Confirmation - Empirical	"Comparing the model structure with information (quantitative or qualitative) obtained directly from the real system being modeled."[123]
Structure Confirmation - Theoretical	"Comparing the model structure with … [generalized knowledge about the system that exists in the literature]."[123]
Behavior Reproduction	"Does the [simulation model's] behavior match any historical data and/or the reference mode?"[105]
Extreme Policy	"When policies are pushed to extremes, are the [simulation model's] behaviors reasonable?"[105]
Mode Reproduction Ability	"With different past policies, does the [simulation model] yield behaviors consistent with other [examples] of the system"[105]
Behavior Prediction	"Does [simulation model] reproduce the anticipated behavior in future/hypothetical situations?"[105]
Anomalous/Surprise Behavior	"Have odd behaviors been studied to show that either: 1) they are anomalous, needing [simulation model] corrections to remove them? [or] 2) the [simulation model] yields insights into a previously unrecognized mode?"[105]
Perspectives Boundary Adequacy	"Do the models support debate on different perspectives in the [Problem Statement] concerning: 1) choice of modelling approach used? 2) System dynamics issue addressed? 3) Goals to be achieved? [and] 4) Policies for doing so?"[105]
Structure Boundary Adequacy	"Do the models contain sufficient and appropriate variables, policies and feedback loops to address the issue that they are being built to study?"[105]
Dimensional Consistency	"Checking the right-hand side and left-hand side of each equation for dimensional consistency"[123]
Extreme conditions in equations	"Are the outputs of policies reasonable if [simulation model] inputs take extreme values?"[105]
Behavior Sensitivity	"Are the previous behavior tests compromised by the plausible changes in parameter values?"[105]
Behavior Boundary Adequacy/Structure Sensitivity	"Does the [simulation model] contain sufficient and appropriate variables, policies and feedback loops to address the issue when this is tested by adding new pieces of relevant structure and examining the resulting behavior?"[105]
Policy Sensitivity	"Are the suggested [policies and recommendations] robust to plausible parameter changes?"[105]
Policy Boundary Adequacy	"Does the addition of more possibly relevant structure change the [policies and recommendations]?"[105]
Norms/Values Boundary Adequacy	"Do the models support debate concerning, and represent the behavior of the relevant actors': 1) goals (are the desired states acceptable?) 2) Policies (are the actions based on discrepancies between goal and actual conditions acceptable within the culture?)"[105]
Roles Boundary Adequacy	"Are the feedback links in the models consistent with the abilities of current actors in the system to access, interpret and employ information?"[105]

Forrester considered two potential sources of validity for SDM research: ultimate validity and interim validity. Ultimate validity can only be assessed when a model is *used* over time to improve a real system[107] (p. 115,117). Hence, because the goal is for models to be *used*, or, to have "usefulness with respect to some purpose"[123] (p. 184), the interim goal for validity in SDM is *usefulness.* The maxim for validation in SDM is *all models are wrong, some are useful*[113] (p. 890)[10].

Interim validity is assessed in an iterative process: first, validate structure, then validate behavior. Validation both follows *and* initiates further model development. If a model passes structure tests, but fails behavior tests, then the structure needs to be changed, initiating additional structure testing.

I acknowledge that the term validation brings with it a non-trivial degree of controversy within (as well as outside) the field of system dynamics. There is an emerging preference on the part of some system dynamicists to use the term *verification* instead[167, 168]. This dissertation uses the term validation as it is accepted in system dynamics and also more broadly understood by the health services research audience.

2.2.2.4 IMPLEMENTATION

The goal of the SDM modeling-validation process is to develop a model for use. Goals for models are that they be ones which stakeholders can have confidence in and which can be used and evaluated/improved over time. The use of SDM models is to identify policies which will plausibly address the problem or situation in the most important ways.

Now that model validation tests are completed and one is comfortable moving forward by using the model to gain understanding, research activities shift to identifying, creating the structure for, and running the model for various policy changes and environmental (external) conditions. Specifically, after the researcher identifies "consistent stories about alternative futures" that might arise as well as policies that could be implemented, the researcher creates model structure to incorporate these system changes into the model[114] (p. 272). Next, *what-if* scenarios are run to see the effect of each policy in isolation. Sensitivity analysis is run to test the robustness of each policy recommendation to different scenarios and uncertainties. Then, policies are run in tandem to see the impact of policy interactions (e.g., synergy)[113] (p. 86).

2.3 NEWLY-DEVELOPED METHODS FOR SYSTEM DYNAMICS MODELING IN THIS DISSERTATION

Kopainsky and Luna-Reyes[138] explored the relationship between system dynamics and theory building. They began with a review of papers published in the System Dynamics Review between 2003 and 2006 (20% of all issues of that journal up to that time). They found that three fourths of model-based papers report only on the mathematical formulations in their simulation models, without documenting data collection and analysis methods on which those formulations are based. The authors concluded from this review that there is room for improvement in adequately grounding model structure in data. Next, they compared the theory-building process of SDM with those of case study research and grounded theory. Their recommendation for future work in SDM was to use a "rigorous process for the identification of concepts and relationships among them, that is, for the formulation of the dynamic hypothesis" and to develop methods for "integrating findings from different cases into a more general theory of an important and recurring problem"[138] (p. 482).

In the course of developing a grounded, dynamic theory of primary care transformation (PCT), I also developed new SDM methods for both integrating findings across cases (see Section 2.3.1) and a rigorous process for formulating a dynamic hypothesis (see Section 2.3.2). They were developed to address needs of this research, and are designed to be used in concert with the SDM Standard Method.

Materials on the conceptual basis of formal methods of systems thinking and SDM[107, 113, 169] were useful for designing model development methods. For research requiring the design or improvement of formal validation tools in SDM, Barlas[123] recommends comparing the formal validation tools and typical process presented in that work to the work presented in Lane[105] for SDM practitioners involved in many different types of SDM activities. Both articles compile previous work in simulation modeling and validation, broadly across modeling disciplines as well as in SDM. Both were useful in designing and assessing the model validation methods that are presented here.

This research utilizes interview data that was gathered using the *purposive sampling* and *oversampling* techniques[17] in order to ensure that the interviews would provide sufficient information on the complex problem of PCT (see Section 2.4.2).

[17] Note: I did not develop these methods.

Purposive sampling involves using a sample that selects participants based on important characteristics[170]. This approach requires the involvement of "those closest to the process", in this case HSDO management staff, to guide sample selection[171] (p. 171). The goal is to select "useful samples for learning about process performance and the impact of changes over time" (p. 175). These samples aim to maximize the diversity of perspectives on "the fragments of experience we are most interested in" (p. 172)[171].

I apply the term *oversampling* to indicate that, while typically qualitative studies use a sample of less than 50 interviews[172]), I conducted 82 interviews. The aim in selecting this sample size was to complement purposive sampling by *more fully* capturing the diversity of perspectives on PCT, in the hope of *ensuring* saturation: "the point of diminishing return where increasing the sample size no longer contributes to new evidence"[172] (p. 83, 107) by sampling well beyond that point (albeit in a non-probabilistic way).

Each of these interviews was then transcribed providing what is known in SDM as a mental database. This mental database included interviews from participants working across a range of clinic environments, expressing divergent attitudes toward the process and concept of PCT and representing a range of clinician practice types. These participants included managerial staff and clinical staff: the clinicians, and MAs. The scoping study (Appendix C) identified these two stakeholder types as being the best source of data[18] on the problem of interest. This finding served to reduce the scope of the mental database to clinical professions.

I subsequently segmented the interview database into those set aside for *model development* (consisting of the interviews from 80% of the clinics in the database) and those set aside for *model validation* (consisting of the interviews from the remaining 20% of clinics). Within each group, there is also a *saturation reserve* (3 clinics for model development and 1 for validation). Should more information be needed to complete model development or model validation, the *saturation reserve* interviews could be consulted.

The interviews provided a rich mental database to draw from for eliciting mental models from text, with sufficient sample size to enable segmenting of clinical staff interviews for model development and validation purposes. This section first presents a new method for model development called *CLD Combination*, with accompanying rationale. Second, it presents the methods for model validation, with accompanying rationale.

[18] This decision is verified in Phase 2, by including interviews from the next hierarchical level of the HSDO (managers) in validation.

2.3.1 MODEL DEVELOPMENT

2.3.1.1 RATIONALE FOR NEW MODEL DEVELOPMENT METHODS

In his book Industrial Dynamics[107], Forrester admonishes the reader that the work of selecting and linking variables in equations is the most important part of SDM, after setting model boundaries. The reader is cautioned that this task is not only important, but it is also the most difficult aspect of the SDM methodology. While simple equations may be used to calculate "the inherent results of the physical state of the system", formulating equations becomes more difficult when they represent *decision functions*[19] – where information is brought together in the mind of a decision-maker and important decisions are made[107] (p. 118). Forrester informs the reader that the information for decision functions is found in descriptive, informal sources and includes policies that may not have ever been documented before.

The difficulty in formulating these equations comes from two sources: the difficulty of formulating mathematical equations to describe complex processes and the difficulty of eliciting the necessary information to represent decisions. Even so, Peterson and Eberlein[111] observe that the SDM literature fails to articulate the procedures for how equations should be formulated, including decision functions. The process is described as one of trial and error where a researcher *guesses* at equations relying on "unconscious or intuitive leaps to achieve a trial equation" (p. 172) after which a draft simulation model is improved by trial and error "until time runs out or satisfactory behavior is achieved" (p. 161)[111]. As writing equations in SDM involves a leap from stories to numbers, it is one of the key ways that SDM proposes to integrate qualitative and quantitative methods; yet it is still unclear how it is done.

Efforts to address this issue have resulted in textbooks with clear, formal guidance for naming variables, visualizing relationships and drawing table functions[113] (p. 585-595)[114] (p. 228-231) as well as methodological works which provide guidance on the associated group facilitation methods[151]and on formal techniques for documenting model development using text sources[116, 117, 173]. How should the different perspectives on causality be treated?

According to Lane, the process of resolving diverse perspectives into a model of shared cognition remains unclear in the formal SDM methodology[105, 106] (and in related fields as well[174]).

[19] The term decision function is used interchangeably with the terms policy function and operating policies in system dynamics and in this dissertation.

In order to develop and validate the models used in this dissertation, I needed to clarify the steps that are used in SDM for developing decision functions from raw descriptive data (in my case, from interview transcripts). This dissertation is attempting to access the understanding of complex phenomena in PCT, which exists in mental models of stakeholders who experience them most directly: clinicians and MAs involved in PCT. It is assumed that, in their interviews, they have provided the information necessary for accessing this understanding. This is because their experience with PCT has impacted their mental models in long-term memory, which they accessed during the semi-structured interview process. The newly-developed method presented in the next section provides one rational way of integrating these perspectives.

2.3.1.2 CAUSAL LOOP DIAGRAM COMBINATION

Model name	Participant CLDs	Team CLDs, Clinic CLDs	SMM1	SMM2	SMM3 / Conceptual Model	SIM1	SIM2	SIM3	Theoretical Model

This dissertation utilized a new method for model development. Because it is a method for merging causal loop diagrams, it is called *CLD Combination*. CLD Combination begins with a distinct CLD for each interview, including only the variables and links that were mentioned by that participant. CLDs are combined, two at a time, until one CLD is produced.

In this dissertation, the term *shared mental model* (SMM) is reserved for this level of abstraction – the aggregated understanding obtained from combining individual mental models assuming that the resulting cognitive structure could be obtained from any individual when specifically and appropriately probed. In other words, the term *shared mental model* is used for the CLD at highest level of aggregation – not because all participants explicitly identify the same structural elements but because it is assumed that, as practitioners with deep experience in PCT, they have access to the entire understanding in the shared mental model even though they only drew on parts of it during the semi-structured interview.

Once validated (procedures described later in this chapter), the shared mental model becomes the *Conceptual Mode*[20]*l*. This model serves as the blueprint for developing the simulation model, representing the system structures at play as described by those practicing within them. It is assumed that system participants have the best available understanding of the structure of the system. What is lacking is not new or better information but combining these stakeholder understandings. This approach is in accordance with the spirit of the original SDM methodology as laid out in Forrester's book, Industrial Dynamics:

> "Industrial dynamics models are built on the same information and evidence used for the manager's usual mental model of the management process. The power of industrial dynamics models does not come from access to better information than the manager has. Their power lies in their ability to use more of the same information and to portray more usefully its implications."[107] (p. 117)

Preparing[21] CLDs for CLD Combination involves: coding the interviews, creating corresponding CLDs, *mildly* pruning[22] each CLD, and, finally, revising the variable names so that the same variable has the same name across CLDs.

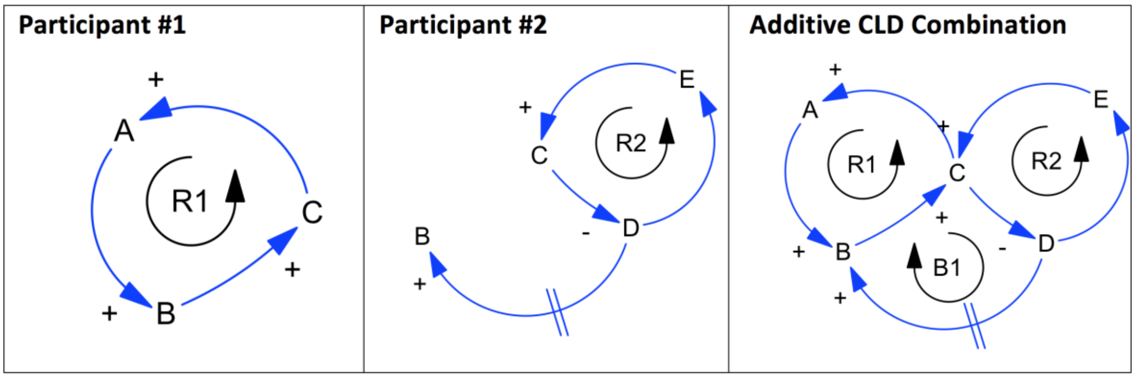
[20] This term is capitalized when referring to the specific conceptual model developed in this dissertation.
[21] See Section 2.4.3 for details on the preparatory process.
[22] Mild pruning: keep delays and any feedbacks. Full pruning: keep only feedbacks of at least 3 links.

In this preparatory process, the same concepts (i.e., the variables and the relationships between variables) are considered repeatedly. My learning in earlier CLDs influenced later interpretations. For example, the later CLDs used more consistent variable names than earlier ones. Prior to CLD Combination, all CLDs were reviewed and variable names were updated to reflect the same variable name when the same concept was represented across CLDs.

CLD Combination involves ordering the CLDs (for combining) using criteria that correspond to individual interviews in some logical way. In this research, participants interviewed were all part of one Health Service Delivery Organization (HSDO). Two or more clinicians were interviewed in each clinic. Those clinicians worked on separate teams. For each clinician, one MA from their team was also interviewed. These organizational affiliations (team, clinic, HSDO) were used as the logical criteria for CLD Combination: CLDs corresponding to individual interviews were combined into Team CLDs; then these Team CLDs were combined into Clinic CLDs; and, finally, Clinic CLDs were combined into one HSDO CLD. In so doing, the process of CLD Combination develops a shared mental model representing the causal mechanisms as described in the interviews.

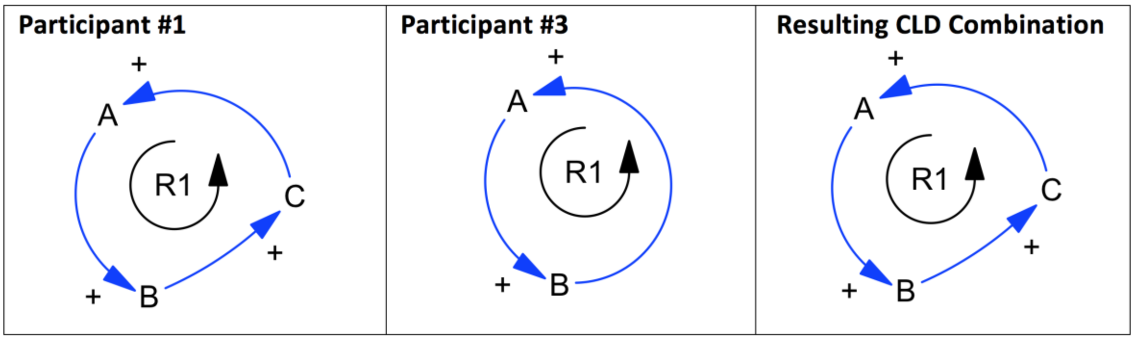
Comparison is an important task in all stages of CLD Combination. I started with the most detailed CLD and revised it by comparing it with the next most detailed CLD. When two CLDs are combined and all differences are complementary, a simple additive process of CLD Combination is possible as shown in Figure 2.7 below. For example, if Participant #1 identified feedback loop A→B→C→A and Participant #2 identified feedback loop C→D→E→C as well as a delay for D→B, the combined CLD would have all of these structures.

Figure 2.7 CLD Combination – Addition



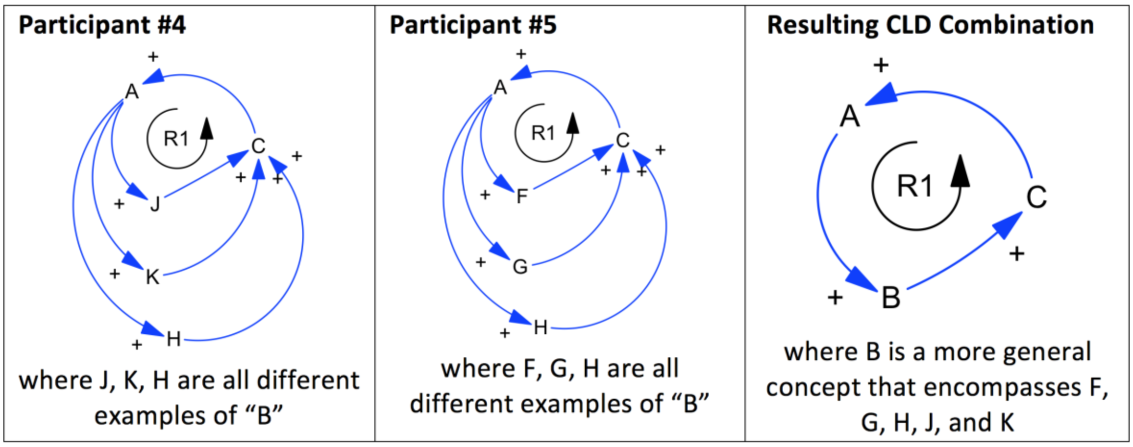
Sometimes the differences are not complementary and a decision must be made about which one is more accurate. In these cases, the result of comparing the different understandings from the CLDs was that I improved my understanding of the concepts in question and a more accurate description of system structure emerged in the combined CLD. One practice I employed was to select the more detailed understanding of a particular system structure, as shown in Figure 2.8 below. For example, if Participant #1 identified feedback loop A→B→C→A and Participant #2 identified feedback loop A→B→A, the combined CLD would have the structure A→B→C→A.

Figure 2.8 CLD Combination – Selection



Furthermore, there are times when a participant provides a lot of detail complexity that does not add to the dynamic complexity. While useful in understanding what that particular concept entails, it is not useful in generating the CLD or the simulation model. In these instances, the additional detail complexity was merged into a variable name that could encompass the various details given. For example: interviews saw each MA skill as a potentially interesting phenomenon on its own. This was part of the search for the *must-have* elements of Patient-Centered Medical Home. Representing each skill separately added little value to CLDs since their place on the CLD was all the same. None of them explained a new causal mechanism from the rest. Therefore, it was conceptually the same to aggregate them as *MA Capabilities*. A generic example is presented in Figure 2.9 below.

Figure 2.9 CLD Combination – Merging Variables Providing Detail Complexity



In the final level of CLD Combination, CLDs are merged into a shared mental model for all participants interviewed. Unlike previous combinations that relied upon *mild* pruning, this time full pruning is implemented (Yearworth's rule regarding only keeping feedback loops with three links was enforced[116]). In this dissertation, this final level involved combing the clinic CLDs into the HSDO CLD (i.e., the shared mental model).

By following these rules, all stakeholder perspectives are valued equally. Participant characteristics (e.g., being the clinician rather than the MA, having a longer tenure at the organization) did not influence which system structure was chosen. In the example just given, the more detailed causal mechanism is used in the resulting CLD, irrespective of which participant gave each description. Elements and boundaries which were exposed and defined in purposive text analysis of individual interviews represent findings from participants with different *perspectives* as well as different *behaviors*. In CLD Combination, the different perspectives on the structure of the underlying system are now compared, iteratively. With each comparison, more generic model structures are distilled, simplifying/aggregating structures when appropriate and choosing the more detailed structure when structures conflict. This method results in the definition of the elements and boundaries of the shared mental model.

2.3.2 MODEL VALIDATION

Newly-developed model validation methods in this dissertation are based on the SDM practice of *Reflective Modeling*, where the modeler/researcher subjects the model to testing in order to "uncover flaws and hidden assumptions, challenge preconceptions, and expose assumptions for critique and improvement"[113] (p. 858).

In order to better understand my methods, an in-depth treatment of validation is provided in Appendix D; here I present a summary. Then, the rationale for these new validation methods is presented. Finally, the newly-developed model validation methods are described. Barlas[123] recommends considering Lane[105] in designing new validation methods. Therefore, as I present the validation overview, rationale, and newly-developed validation methods, I draw heavily from this source.

2.3.2.1 VALIDITY TYPES SUMMARY

The validity types are the goals of model testing. The validity types considered in SDM literature[105] are presented in Table 2.7 below: conceptual, formulational, experimental, and data validity. For each type, the corresponding subtypes (column two) and proposed short names (column three) are also listed. In the last two columns, I mark whether the subtype is one of my newly-proposed subtypes (i.e., proposed as part of this dissertation) or whether it has a newly-extended definition in this dissertation.

Table 2.7 Summary of Validity Types and Subtypes

Validity Types	Subtype	Short name	Proposed Subtype	Extended Definition
Conceptual Validity	CptV1	Variables & boundaries	x	
	CptV2	Links	x	
	CptV3	Saturation	x	
	CptV4	Culture	x	
Formulational Validity	FV1	Language limitations		x
	FV2	Conceptual equals simulation		
	FV3	SDM guidelines		x
Experimental Validity	EV1	Structural design		
	EV2	Insights gained		
Data Validity	DV1	Mental data	x	
	DV2	Written data	x	
	DV3	Numerical data	x	

I developed brief definitions for each of the validity subtypes, based on relevant literature (see Appendix D for more detail). I present these in Table 2.8 below. Underlined text calls out the key concepts which inform the *short names* for each subtype.

Table 2.9 below presents the validity tests previously presented in Table 2.6 above, now also including their objectives, focus and type of validity gained. This table is adapted from Table 1 provided in Lane[105]. Test names were modified to reflect both Lane and Barlas's lists of tests[105, 123]. CptV and DV subtypes are collapsed[23], as they are conceptualized by Lane[105]. The *Structure Confirmation – Empirical* test has an "o" for FV1 and DV (specifically DV 1 and 2). While not in Lane[105], as described above, I believe that these additions are warranted.

[23] A version that disaggregates CptV and DV is provided in Table 2.11 below.

Table 2.8 Validity Types, Subtypes & Their Definitions

Validity Types	Subtype	Short name	Subtype Definitions
Conceptual Validity	**CptV1**	Variables & boundaries	Concerns the extent to which the variables and boundaries in the model are well-developed and validated.
	CptV2	Links	Concerns the extent to which the relationships (i.e., links) in the model are well-established and validated.
	CptV3	Saturation	Concerns the extent to which new and relevant data regarding the elements in the model have ceased to emerge.
	CptV4	Culture	Concerns the extent to which the relevant actors feel that the cultural aspects of the research are addressed (i.e., cultural acceptability, bounded rationality, and non-coercive approach).
Formulational Validity	**FV1**	Language limitations	Concerns the extent to which the limitations of language impacted the development of the simulation model (i.e., have concepts in the conceptual model been omitted or distorted in this translational process?).
	FV2	Conceptual equals simulation	Concerns the extent to which the simulation model is shown to be consistent with the structure and behavior described in the conceptual model and problem statement (i.e., is there a satisfactory representativeness between the simulation and the conceptual understanding?).
	FV3	SDM guidelines	Concerns the extent to which the simulation model conforms to SDM construction guidelines (e.g., not having artificial min/max functions to fix a bug in the model).
Experim ental Validity	**EV1**	Structural design	Concerns the design and carrying out of simulation runs to test the structural design of the model (i.e., the trial and error of equation writing referred to by Peterson[111]).
	EV2	Insights gained	Concerns the usefulness, accuracy, and robustness of the insight gained from simulation model runs (i.e., the model's ability to adequately reproduce real-world behavior).
Data Validity	**DV1**	Mental data	Concerns the extent to which the mental data is reliable/appropriate, accessible and sufficient.
	DV2	Written data	Concerns the extent to which the written data is reliable/appropriate, accessible, and sufficient.
	DV3	Numerical data	Concerns the extent to which the numeric data is reliable/appropriate, accessible, and sufficient.
NOTE: These definitions are based on the work of Lane as well as my reading of Oral & Kettani and the review by Tsopiapsis *et al.*, and the three rules for theoretical saturation from Grounded Theory. Please see Appendix D for more on validity types[105, 144, 163, 166].			

Table 2.9 Validation Tests Matched to their Objective, Focus and Validity Type

		Focus of test			**Types of validity**						
Objective	**Name of test**	Structure	Behavior	Culture	CptV	FV1	FV2	FV3	EV1	EV2	DV
Relevance to and consistency with Appreciation of the Situation	Parameter Confirmation - Theoretical	x			x		x				x
	Parameter Confirmation - Empirical	x			x		x				x
	Structure Confirmation - Empirical	x			x	o	x				o
	Structure Confirmation - Theoretical	x			x		x				
	Behavior Reproduction		x				x				x
	Extreme Policy		x				x				
	Mode Reproduction Ability		x				x				
	Behavior Prediction		x							x	
	Anomalous/Surprise Behavior		x							x	
	Perspectives Boundary Adequacy			x	x						
Suitability for purpose	Structure Boundary Adequacy	x			x	x					
	Dimensional Consistency	x						x			
	Extreme conditions in equations	x						x			
	Behavior Sensitivity		x						x	x	
	Behavior Boundary Adequacy/Structure Sensitivity		x		x				x		
	Policy Sensitivity		x							x	
	Policy Boundary Adequacy		x							x	
	Norms/Values Boundary Adequacy			x	x		x				
	Roles Boundary Adequacy			x	x		x				

(CptV) Conceptual Validity, (FV) Formulational Validity 1-3, (EV) Experimental Validity 1-2, (DV) Data Validity. See Lane[105] for more information

2.3.2.2 RATIONALE FOR NEW MODEL VALIDATION METHODS

SDM methodology requires iterative model building and validation[100, 159, 160]. As I progressed through model conceptualization (and beyond), I needed validation methods that would test that the model building was progressing appropriately (i.e., that flaws were being identified such that the model could be improved) as well as methods that would test for cessation (i.e., that I could move to the next stage of model development: from qualitative to quantitative model building) (see Section 2.2.2.3).

The original conceptualization of validation in SDM included a strong argument for the use of non-quantitative model validation tools (i.e., model validation tools relying on qualitative data). Forrester provided two reasons supporting the use of both a quantitative and a qualitative treatment of validity, 1) since many aspects of SDM model development are based on qualitative *mental data,* then the same type of data should be used to validate such a model, and 2) that it expands the number of problems which can be studied when "a preponderant amount of human knowledge is in nonquantitative form… model building and model validation do not [need to] stop at the boundary where numerical data fail"[107] (p. 129).

In a paper on modeling and validation broadly (not only SDM), operations researchers Oral & Kettani reaffirm Forrester's observation that the richest source of data for modeling is "mental data bases" consisting of mental models, and that such data have a place in model validation[144] (p. 226). Of the formal methods for qualitative validation, these authors describe SDM's inherent advantages in conceptualization and conclude that, when it is "using the cognitive capacities of the relevant actors fully [it] is perhaps the most promising [method] … for determining the validity of a given 'conceptual model'"[144] (p. 227).

Efforts to address this gap (between qualitative and quantitative model validation) have resulted in the development of methods for qualitative validation in Group Model Building (where the group could disconfirm during the workshop)[152], and structured stakeholder dialogues (proposed by Forrester[107] (p. 263,268) and implemented in a structured way in the *dis-confirmatory interview*[175]). However, stakeholders are not always available throughout the modeling process (i.e., because of financial constraints, changing access to stakeholders via management personnel or interest changes). In these cases, it would be useful to have methods that would be able to use contemporaneously collected data (i.e., textual *mental databases*, for example, a subset of interviews set aside for validation). However, qualitative validation methods using such data did not exist.

As I had used interviews to develop the shared mental model, I needed a method that would validate the structure of this model using interview data before I felt comfortable moving forward to using it as a blueprint for building the simulation model. The purpose of developing these new validation methods was for "confronting the model with data and expert opinion"[113] (p. 80-81). SDM considers mental data to be the most useful data for model development and assumes that it can be elicited by semi-structured interview. This dissertation presents newly-developed methods that can utilize this type of data for qualitative model validation – designed to confront the evolving versions of the model with mental data contained in interview transcripts.

2.3.2.3 NEWLY-DEVELOPED VALIDATION METHODS

This section presents the newly-developed validation methods. These methods are described in detail in Appendix E. In this section, I introduce these methods by providing a brief overview of the goals of each method. For ease of the reader, I use the full method name in this section; hereafter, each method's corresponding acronyms are used. I will explain later in the chapter (Table 2.13) how these methods map to those listed in Table 2.6. Additional information regarding this mapping is also provided in Appendix E.

Table 2.10 below lists the new methods, their acronyms and focus questions as well as the data set used and level of analysis. Model-specific validation tests are used in the various phases of model development. Meta-level validation tests enable periodic reflection on the *meta-level* validity – aspects that relate to the overall project.

Table 2.10 Formal Validation Methods Designed for Textual Mental Databases

			Data Set Used		Level	
Method name	**Acronym**	**Focus Question**	**Model Development Set**	**Model Validation Set**	**Model**	**Meta**
Shared Mental Model Saturation	SMM-S	Do the different clinics agree on the structure of the system?	x		x	
Conceptual Model Saturation	CM-S	Does an additional clinic agree with SMM2 on the structure of the system?		x	x	
Simulation Model Saturation	SIM-S	Does an additional clinic agree with SIM1 on the structure *and behavior* of the system?		x	x	
System Dynamics Saturation	SD-S	Does an additional clinic confirm the usefulness of SDM as the modeling approach?		x		x
Methods Suitability	N/A	What are the trade-offs of the specific modeling approach used in this research?				x
Data Suitability	N/A	Are the data upon which the model is based used within the scope of their limitations?				x
Stakeholder Dialogue Suitability	N/A	Have stakeholders been appropriately engaged throughout the research project?				x

Three of these methods (CM-S, SIM-S, and SD-S) use a set of interviews that was left unused throughout the model development process. This subset was selected based on participant characteristics (in this dissertation, having clinicians, clinical staff and management personnel practicing in one context; i.e., all participants from one clinic). Once selected, these interviews were set aside until model validation (herein, *the model validation set*).

These methods use tests which are not pass/fail per se. The first three tests are used to identify strengths and weaknesses of the model. The remaining four tests are used to identify strengths and weaknesses in the overall research approach. All these tests are performed as described in Appendix E and results of these tests are then presented along with model findings. It is then up to the audience to judge.

Shared Mental Model Saturation (SMM-S) is proposed here to have been reached when it has been demonstrated that (1) the shared mental model appears to describe the system structure underlying the problem statement and (2) the addition of one more data source from the saturation reserve is not likely to modify the existing shared mental model (in the case of this dissertation, this would be another interview or clinic set of interviews).

The process of CLD Combination draws on the bank of CLDs representing individuals' mental models as interpreted via Purposive Text Analysis. SMM-S goes one step back to check the shared mental model against the individual clinic CLDs. In so doing, it mitigates against the risk that essential elements were lost in the process of CLD Combination.

Conceptual Model Saturation (CM-S) is proposed here to have been reached when it has been demonstrated that the addition of one interview is not likely to modify the Shared Mental Model[24]. I start with the existing Shared Mental Model (that was just verified via SMM-S). That model is revised as needed during CM-S, until the addition of one more interview is not likely to modify it. The method involves coding a new set of interviews from scratch (the *model validation set* of interviews). This mental data is rigorously analyzed to validate the *structure* of the Shared Mental Model. At this point, once the CM-S Test is passed, the CLD is referred to as the *Conceptual Model*.

This method provides a formal way of reflecting on the conceptual model. In so doing, there is the recognition that even researchers' mental models of what they study are hard to change. This process helps to surface researcher assumptions and to allow participants' views to question existing assumptions including on important concepts such as model structure, system boundaries and parameters.

Simulation Model Saturation (SIM-S) is proposed here to have been reached when it has been demonstrated that the simulation model does not contain radical departures from the participants' interviews. Participant perspectives are accessed in the results of CM-S and used here to validate the structure *and behavior* of the simulation model.

This method provides a formal way of reflecting on the simulation model (e.g., its purpose, structure, parameters and behavior) by way of comparison with the problem definition, dynamic hypothesis and participant perspectives. For structure, this reflection involves exploring the limitations and boundaries of the model; the limitations being concepts in quotations which were simplified in the model and the boundaries being concepts which were left out of the model, treated as outside the system, as constants, or as aggregated concepts. For behavior, this reflection includes exploring how the model is able to generate the scenarios described in participants' interviews. In so doing, this method checks that the large number of elements *created* in translating the conceptual model into the simulation model is consistent with the way participants perceive the workings of the system.

---

[24] This term is now capitalized since the shared mental model has been validated via the SMM-S test, producing SMM2. SMM2 is being validated in this test.

System Dynamics Saturation (SD-S) is proposed here to have been reached when the extent to which the validation set interviews verify that the target group sees SDM as a useful way of addressing the issue has been demonstrated (CptV4 - culture). This is demonstrated by rigorously analyzing relevant segments of the coded transcripts generated in CM-S, segments which, demonstrate participant perceptions of issues which SDM is used to study (e.g., mental models, systems with causal structures which include significant time delays, feedback loops and non-linearity).

This test verifies that participants' mental models contain causality and delays and that system structure is perceived to be causing their problem. Furthermore, this test verifies that these participants understand that they have mental models as characterized in SDM, that these models are important, that these models can (and indeed must) change, and that doing so involves emotional engagement. In verifying these things, this test demonstrates the extent to which participants are likely to feel that, when presented with it, SDM research is appropriate and acceptable even though they were not directly exposed to it prior to their interview.

Methods Suitability is proposed here to be determined by reflecting on the methods used in the research. This test relates to both the modeling process (i.e., how diverse understandings are brought together) and the modeling language(s) (what sorts of models are used to represent those understandings) used in the research project. In so doing, the researcher provides a documentation of the trade-offs considered in the project's overall methodological approach and justifies the researcher's reasoning, exposing the methodological validity claims to evaluation by others.

Data Suitability is proposed here to be determined by reflecting on the three types of data that are possible, the data availability, and their limitations. This test relates both to model documentation and model validation. It should be started prior to model development and updated during model development. In so doing, the researcher provides a documentation of the data which were used in model development and validation, exposing the data validity claims to evaluation by others.

Stakeholder Dialogue Suitability is proposed here to be determined by reflecting on the research project and the extent to which stakeholders have been engaged. In so doing, the researcher provides a documentation of the project's approach to stakeholder dialogue and justifies the researcher's reasoning, exposing the validation claims using evidence from stakeholder dialogue to evaluation by others.

Table 2.11 identifies the type(s) of validity that I have found each method to assess.

Table 2.11 Validity Types Ascribed to Each Newly-developed Validation Method

	Types of validity											
	CptV1	CptV2	CptV3	CptV4	FV1	FV2	FV3	EV1	EV2	DV1	DV2	DV3
Shared Mental Model Saturation (SMM-S)	x	x	x	x	x							
System Dynamics Saturation (SD-S)				x								
Conceptual Model Saturation (CM-S)	x	x	x	x	x							
Simulation Model Saturation (SIM-S)	x	x		x	x	x		x				x
Stakeholder Dialogue Suitability	x	x		x	x	x	x	x	x			
Methods Suitability				x	x							
Data Suitability										x	x	x
(CptV) Conceptual Validity, (FV) Formulational Validity 1-3, (EV) Experimental Validity 1-2, (DV) Data Validity. See Lane[105] for more information												

Each cell marked with an “x” in Table 2.11 above corresponds to one row in Table 2.10 below. In Table 2.12, the first column lists the validity type, defined in Appendix D. The second column identifies the method. The third column provides the question(s) that each newly-developed method contributes to addressing for each particular type of validity. The final column identifies the data that are compared in addressing the question(s) including the version of the model and the data source(s).

Table 2.12 Questions that Each Newly-developed Validation Method Addresses, Sorted by Validity Type

Validity Type	Method	Question	Data Compared
CptV1 Variables & boundaries	SMM-S	Are the variables and boundaries of the shared mental model well-developed and validated?	SMM1 compared to the problem statement and CLDs at one level of aggregation lower than SMM1
	CM-S	Are the variables and boundaries of the Shared Mental Model well-developed and validated?	SMM2 compared to the problem statement and model validation set interviews
	SIM-S	Are the variables and boundaries of the simulation model well-developed and validated?	SIM1 compared to the model validation interviews
	Stakeholder Dialogue	Are the variables and boundaries of the models well-developed and validated?	Models compared to stakeholder perspectives
CptV2 Links	SMM-S	Are the relationships in the shared mental model well-established and validated?	SMM1 compared to the CLDs at one level of aggregation lower than SMM1
	CM-S	Are the relationships in the Shared Mental Model well-established and validated?	SMM2 compared to the model validation set interviews
	SIM-S	Are the relationships in the simulation model well-established and validated?	SIM1 compared to the model validation set interviews
	Stakeholder Dialogue	Are the relationships in the models well-established and validated?	Models compared to stakeholder perspectives
CptV3 Saturation	SMM-S	Have new and relevant data regarding the elements in the shared mental model ceased to emerge?	SMM1 compared to the CLDs at one level of aggregation lower than SMM1
	CM-S	Have new and relevant data regarding the elements in the Shared Mental Model ceased to emerge?	SMM2 compared to the model validation set interviews
CptV4 Culture	SMM-S	Are bounded rationality and cultural acceptability evident in the relationships?	SMM1 compared to the CLDs at one level of aggregation lower than SMM1
	CM-S	Are bounded rationality and cultural acceptability evident in the relationships?	SMM2 compared to the model validation set interviews
	SIM-S	Are bounded rationality and cultural acceptability evident in the simulation model?	SIM1 compared to the model validation set interviews
	SD-S	Does the target group see SDM research as a useful way of addressing the issue?	The research project undertaken thus far considered through the lens of the model validation set interviews
	Stakeholder Dialogue	Do stakeholders see SDM research as a useful way of addressing the issue? Are bounded rationality and cultural acceptability evident in the models?	Models compared to stakeholder perspectives
	Methods Suitability	Does the modeling process address cultural acceptability and cognitive limitations and use a non-coercive approach?	Reflection

Validity Type	Method	Question	Data Compared
FV1 Language limitations	SMM-S	Do the relationships in the shared mental model distort aspects of the clinic CLDs?	SMM1 compared to the CLDs at one level of aggregation lower than SMM1
	CM-S	Do the relationships in the Shared Mental Model distort aspects of the model validation set participants' statements?	SMM2 compared to the model validation set interviews
	SIM-S	Have important elements been omitted or distorted because of the procedures used in the creation of the simulation model (i.e., "the discipline of the programming language"[105] (p. 120))?	SIM1 compared to the model validation set interviews
	Stakeholder Dialogue	Do the relationships in the model distort aspects as they are conceptualized by stakeholders?	Models compared to stakeholder perspectives
	Methods Suitability	Have important elements or relationships been omitted or distorted because of the procedures used in the model development process?	Reflection
FV2 Conceptual equals simulation	SIM-S	How consistent is the simulation model with the structure and behavior that it was meant to represent?	SIM1 compared to the problem statement and dynamic hypothesis
	Stakeholder Dialogue	How consistent is the model with the structure and behavior that it is meant to represent?	Models compared to stakeholder perspectives
FV3 SDM guidelines	Stakeholder Dialogue	How well do the models conform to SDM guidelines? (e.g., CLD grammar, stock and flow diagram and model grammar)	Models compared to stakeholder perspectives
EV1 Structural design	SIM-S	Do experiments in the simulation model challenge or support its structure?	SIM1 compared the model validation set interviews
	Stakeholder Dialogue	Do experiments in the simulation model challenge or support its structure?	Models compared to stakeholder perspectives
EV2 Insights gained	Stakeholder Dialogue	Are the model's analytical insights at an acceptable level of quality (i.e., usefulness, accuracy, robustness)? (discuss with stakeholders)	Models compared to stakeholder perspectives
DV1 Mental data	Data Suitability	Are the mental data upon which the model is based used within the scope of their limitations?	Reflection
DV2 Written data	Data Suitability	Are the written data upon which the model is based used within the scope of their limitations?	Reflection
DV3 Numerical data	SIM-S	Is the numerical data for use in the simulation model reliable/appropriate, available and sufficient?	SIM1 compared the model validation set interviews
	Data Suitability	Are the numerical data upon which the model is based used within the scope of their limitations?	Reflection
(CptV) Conceptual Validity, (FV) Formulational Validity 1-3, (EV) Experimental Validity 1-2, (DV) Data Validity. Please see Lane[105] for more information.			

The proposed model validation methods described here correspond to tests in the standard SDM method, as shown in Table 2.13 below. From right to left, the table presents: newly-developed validation methods, their foci, the standard method tests that they address, and the purposes to which the tests contribute. This table presents the new validation methods in terms of the tests to which they correspond. This presentation is based on combining two tables above (Table 2.9 *and* Table 2.11).

Table 2.13 New Formal Validation Methods in Their Context

		Focus of test			New Validation Method						
Purpose of validation test	**Name of test**	Structure	Behavior	Culture	**Shared Mental Model Saturation**	**System Dynamics Saturation**	**Conceptual Model Saturation**	**Simulation Model Saturation**	**Stakeholder Dialogue (Suitability)**	**Methods Suitability**	**Data Suitability**
Relevance to and consistency with Appreciation of the Situation	Parameter Confirmation - Empirical	x			x		x	x	x		x
	Structure Confirmation - Empirical	x			x		x	x	x		x
	Perspectives Boundary Adequacy	x		x		x			x	x	
Suitability for purpose	Structure Boundary Adequacy	x			x		x	x	x	x	
	Norms/Values Boundary Adequacy			x	x		x	x	x		
	Roles Boundary Adequacy			x	x		x	x	x		

For Stakeholder Dialogue Suitability, suitability is in parentheses because, while Stakeholder Dialogue is not a new validation method, Stakeholder Dialogue Suitability is a new method. It is the Stakeholder Dialogue that verifies each of these tests and Stakeholder Dialogue Suitability that verifies that appropriate stakeholder groups were engaged appropriately.

2.4 METHODS

This section summarizes the standard and newly-developed SDM methods as they were used in this dissertation. Table 2.14 below orients the reader to the sub-sections that follow: the four phases and two cross-cutting steps. As the research progressed through the phases, new model iterations were developed. This table also lists the model development and model validation methods as well as model iterations and how they correspond to the various phases of this work. This table is based on Table 2.1 and Table 2.3 presented earlier. The purpose of this table is to include the corresponding locations of the methods *and results* sections.

In the sections below, Section 2.4.1 presents a brief description of the iterative improvement of the problem definition and dynamic hypothesis. Sections 2.4.2 to 2.4.4 and 2.4.6 describe the data collection and analysis methods employed during phases one through four. Section 2.4.5 details the iterative process of model validation implemented across these research phases.

Table 2.14 Methods Summary, Model Iterations, Validation Methods & Locations in the Thesis

Study Phase	Purpose	Input	Output	Model Development Method	Methods Location	Model Iteration	Results Location	Cross-Phase Step: Model Validation Methods	Methods Location	Results Location	Cross-Phase Step: Meta Validation & Locations	**Cross-Phase Step:** Iterative Improvement of Problem Statement & Dynamic Hypothesis
Scoping Study	Problem Definition	Mixed Methods Project Data	Problem Definition & Choice of SDM	N/A	Appendix C	N/A	Appendix C	Stakeholder Dialogue	Appendix C	Appendix C	Cessation (Section 2.4.5.1 and 3.5.1) Data Suitability* (Section 2.4.5.2 and 3.5.2) Methods Suitability* (Section 2.4.5.3 and 3.5.3) SD-S* (Section 2.4.5.7 and 3.5.6) Stakeholder Suitability* (Section 2.4.5.8 and 3.5.7)	
Phase 1	Mental Model Elicitation	Semi-Structured Interviews	Participant CLDs	Purposive Text Analysis Mild Pruning	Section 2.4.2	Participant CLDs	Section 3.2	None	None	None		
Phase 2	Develop Conceptual Model	Participant CLDs	Conceptual Model	CLD Combination* Pruning	Section 2.4.3	Team CLDs Clinic CLDs SMM1 SMM2 SMM3	Section 3.3	SMM-S Test* CM-S Test* Stakeholder Dialogue	Section 2.4.5.4 & Appendix F	Section 3.5.4 & Appendix F		
Phase 3	Develop Simulation Model	Conceptual Model	Simulation Model	Simulation Modeling	Section 2.4.4	SIM1 SIM2	Section 3.4	SDM Standard Methods, SIM-S Test* Stakeholder Dialogue	Section 2.4.5.5 & Appendix F	Sections 3.5.5.1 & 3.5.5.2 & 3.5.5.3		
Phase 4	Policy Analysis	Simulation Model	Policy & Theory Results	N/A	Section 2.4.6	SIM3 Theoretical Model	Section 3.6	Stakeholder Dialogue	Section 2.4.5.6	Section 3.5.5.3		

* = Newly-developed methods for SDM, proposed in this dissertation.

2.4.1 CROSS-PHASE STEP - ITERATIVE IMPROVEMENT OF THE PROBLEM DEFINITION & DYNAMIC HYPOTHESIS

In the scoping study (Appendix C), a problem statement (i.e., purpose) was drafted based on preliminary qualitative and quantitative analyses. This problem statement was periodically updated throughout the model development and validation process. Questions considered along the way, to develop (and iteratively update) the problem definition, are listed in *problem definition* above. With each phase, additional results provided the opportunity for reflection and for the creation of a more explicit definition of the problem.

My dynamic hypothesis[25][113] (p. 86) was also updated along the way. All of the model development and model validation steps leading to the development of the Theoretical Model refined it.

2.4.2 PHASE 1 – MENTAL MODEL ELICITATION

Purposive sampling[170, 171][26] was used to select five primary care clinics deemed to be representative (in their context) of the larger organization (the 10 HSDO clinics) – interviews from these clinics are the *model development set*. Purposive Text Analysis was performed on all clinician and MA interviews at these five clinics (n=20 interviews). Interviews from a sixth clinic (n=10 interviews) were set aside for validation once the model was developed – these are the *model validation set*. The remaining HSDO interviews were set aside for saturation reserve: three clinics for saturation reserve associated with model development and one clinic for saturation reserve associated with validation. Thus, in total, interviews from 80% of clinics were designated for model development and interviews from 20% of clinics were designated for validation. These designations are presented in Table 2.15 below.

Table 2.15 Designation of Clinics for Model Development & Validation

	Analysis Set	**Saturation Reserve**	**Group Total**
Model Development	5	3	8
Model Validation	1	1	2
TOTAL	6	4	10

[25] A dynamic hypothesis explains the problem behavior as it is endogenously produced via feedback structures among key system variables. See Section 2.2.2.1.2 for more on dynamic hypotheses.

[26] To be clear, there is no methodological relationship between purposive sampling and Purposive Text Analysis. Purposive sampling was used two times: first, (by HSDO managers) selecting respondents and second, (by me) segmenting the mental database (as in Table 2.15).

Clinics were selected based on their context, with the goal of maximizing the variation in contextual factors. HSDO management personnel characterized each clinic's context. Descriptions of the five selected clinics are presented in Table 2.16 below.

Table 2.16 Clinic Characteristics

	Clinic A	Clinic B	Clinic C	Clinic D	Clinic E
Urbanization	Suburban	Urban	Suburban	Urban	Suburban
Distance from UHC	Mid-range	Close	Far	Mid-range	Far
Patient Diversity	High	Low	Moderate	Moderate	Moderate
Training Clinic (residency)	Rotation	Continuity	Rotation	Rotation	Rotation
MD Availability (most)	Full Time	Part Time	Full Time	Mix	Full Time
Ancillary Specialty	Many	None	Many	Many	A few
Market Competition	High	Moderate	Low	High	Moderate

Clinicians interviewed practicing in these clinics span all the practice clusters (described in the Appendix C Scoping Study). Therefore, it is assumed that using these clinics increases the likelihood that the analysis captures the range of perspectives held among front-line employees within the HSDO.

Purposive Text Analysis is used to develop CLDs from these interviews. Whereas interviews were used for thematic coding in the scoping study, they are now used for model development. To satisfy the assumptions of purposive text, the verbatim transcripts of the interviews must represent:

Participants' "sophisticated [or first-hand] knowledge of the system"[117]

Participants interviewed in the Mixed Methods Project were: one clinician and one medical assistant (herein, MA) from each care team in the HSDO's 10 community clinics as well as relevant managers. The community clinics Quality Director as well as clinic center managers performed the selection using purposive sampling. Selection criteria were: 1) to have one clinician and one MA from each team 2) to capture a range of approaches to implementation and 3) to capture a range in the length of involvement in the transformation process (e.g., initial implementation, introduction into a clinic mid-way through implementation).

Candid discussions, where participants "are not grand-standing or taking rhetorical positions that they do not believe in strongly"[117] (p. 314)

Interviews were performed in a private room at the respective clinic. Participants had the option to terminate the interview at any time. They signed a consent form to participate.

While the research team was embedded within the HSDO, I (the interviewer) was not. It was made clear to respondents that HSDO staff would not have access to respondent interviews and that their statements would not be presented in an identified manner at any time. The semi-structured interviews were gathered with ethical approval from the London School of Hygiene and Tropical Medicine as well as from the University of Utah.

By employing a systematic coding procedure that treats the data in a consistent manner, Purposive Text Analysis overcomes the temporal and spatial distance between data source and researcher which may introduce biases (i.e., the researcher's own assumptions about the system investigated; e.g., variables considered important for one researcher might be ignored by another)[117].

Purposive Text Analysis is a method that was first developed and applied by Kim in a study that generated a stock and flow diagram from a series of verbatim transcripts from the US Federal Reserve Board's Open Market Committee (an important U.S. policy-making group) meetings[117]. The stock and flow diagram represents the policy makers' mental models communicated and shared during the meeting leading to their collective decision. When the researcher cannot verify the diagram with the original stakeholder, systematic coding and documenting allows the researcher to leave a documentation of data-to-diagram linkage and, where feasible, creates an opportunity for the diagram to be examined by others[117].

This dissertation uses this same method to generate CLDs. The coding procedure can be summarized as follows:

1. Define the problem of focus.
2. Select data segments within the problem boundary. Each data segment consists of one argument and its supporting rationales.
3. From each data segment, identify the cause variable, effect variable, and the polarity of the relationship.
4. Represent each causal relationship in a simple words-and-arrow diagram.
5. Collect and merge the words-and-arrow diagrams into a collective CLD. In doing so, collapse similar variables using a common variable name.
6. Assign unique identifiers to data segments and CLD variables and causal links. As the coding progresses, document the data segments where each CLD variable or causal link is elicited.

Purposive Text Analysis proceeded as follows: clinician and MA interview transcripts were coded to identify portions where structural relationships were discussed, focusing on statements describing a cause and effect relationship. Simultaneous to coding in computer-aided qualitative data analysis software (NVivo[176]) respondent mental models were visualized in system dynamics modeling software (Vensim[177]) as the coded causal relationships were then translated into a CLD.

An example illustrating how purposive text analysis was used is presented below in Figure 2.10. One MA describes one aspect of task shifting within her team as she describes whether she is allowed to remain in the examination room during patient visits. This figure presents the statement and shows how it was subdivided into four causality arguments, and a corresponding words and arrow diagram, in the words of this MA.

Once individual CLDs were completed, variable names were standardized across all CLDs. This was achieved by entering variable names from one individual CLD into MS Excel and then updating that list (by adding and/or modifying the variable names) as additional CLDs were reviewed. The final set of variable names was then used to create a standardized CLD for each individual.

These CLDs were then pruned. Pruning elicits the factors that contribute most to system behavior over time. As per SDM theory, accumulation in stocks is what people can see, and it happens via delays and feedback structures[178]. Yearworth[116] recommends retaining only linkages involving delays and/or loops of three or more links. In this step, pruning was relaxed to also retain loops of two links (even when they did not contain delays). In this dissertation, this is referred to as *mild pruning*.

Figure 2.10 The Purposive Text Analysis example of arguments 1 to 4

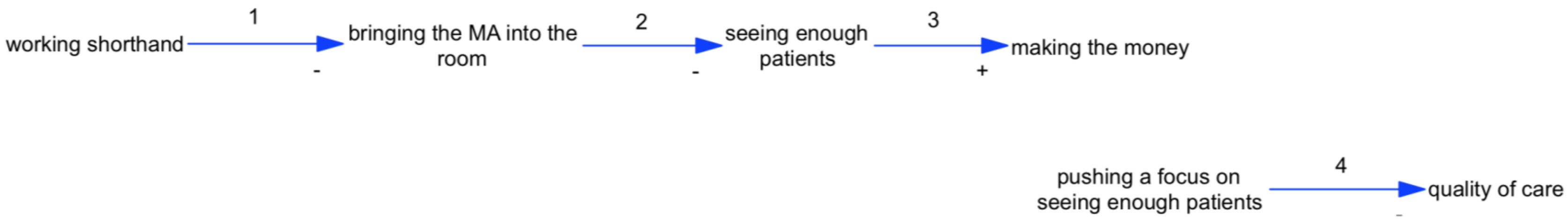
Quotation:

So we pretty much do it all except for staying in the room with the clinician. And [when] this is brought out, Dr. XXX will start doing it again for a week or so and then she'll stop again. I think that we get so involved with what we're doing ... we work shorthanded a lot like we've been very shorthanded today.[1] ... [Plus, the message we get so often is: you're] not seeing enough patients, not seeing enough patients, not seeing enough patients[2], and, you know, we're not making the money.[3] ... and unfortunately sometimes I think that its pushed too much. And the quality of care is not there. [4]

Main argument: we don't bring the MA member into the room because it slows us down when we are already short-staffed

		Argument 1	Argument 2	Argument 3	Argument 4
Quotation	MA03	*I think that we get so involved with what we're doing ... we work shorthanded a lot ...*	*[Plus, the message we get so often is: you're] not seeing enough patients...*	*and, you know, we're not making the money.*	*... and unfortunately sometimes I think that it's pushed too much. And the quality of care is not there.*
Causal Structure	Causal Variable	working shorthanded	bringing the MA into the room	seeing enough patients	pushing a focus on seeing enough patients
	Effect Variable	bringing the MA into the room	seeing enough patients	making the money	quality of care
	Relationship Type	Negative	Negative	Positive	Negative

Drawing:



2.4.3PHASE 2 – DEVELOP CONCEPTUAL MODEL

Preparing CLDs for CLD Combination involved: coding 20 interviews to create 20 corresponding CLDs and *mildly* pruning them into 20 individual CLDs (Phase 1).

In Phase 2, CLD Combination was implemented as individual CLDs were combined to develop a shared mental model. Specifically: the 20 mildly pruned individual CLDs are combined into 10 Team CLDs, which are further combined into 5 Clinic CLDs, and then finally combined and fully pruned into 1 shared mental model for all participants in the *model development set* (this model is called SMM1).

That shared mental model was then validated. Validation methods used were: SMM-S Test, CM-S Test, and Stakeholder Dialogue (see Section 2.4.5.4 for descriptions). After validation, the new model version was referred to as the Conceptual Model (also known as SMM3).

2.4.4PHASE 3 – DEVELOP SIMULATION MODEL

SMM3 was used to guide the development of the draft simulation model (SIM1). Quotations from the *model development set* and operational thinking[157, 179] (see Section 2.2.2.2) were used to generate more detailed structures behind the concepts identified in the Conceptual Model. The literature was also consulted for quantifying parameters in the model.

That said, I began the simulation model using SMM2 as a blueprint (i.e., model development began during the early part of Phase 2, just after SMM-S). Specifically, I began with the personnel sector by translating "care team capacity" into *Number of Clinicians* and *Number of MAs* and then added structure, step by step, as I followed the causal links fanning out from this variable in SMM2. Finally, the model revisions generated in Phase 2 (which result in SMM3), took precedence over prior formulations[27][113] (p. 80-89,104)[114] (p. 6), resulting in SIM1.

Throughout this phase, I employed various validation methods and made model revisions as needed (Section 2.4.5.5). Once these validation methods were completed, the model that resulted was referred to as SIM2.

[27] It should be noted that SDM is fundamentally a creative process, which is iterative and is not a cookbook recipe. In this process, researchers are intended to develop draft qualitative and simulation models which are continuously and iteratively improved. There are many drafts which are created and discarded along the way.

2.4.5 CROSS-PHASE STEP: VALIDATION

2.4.5.1 CESSATION

So far, this chapter has presented the various types of model development and model validation methods (SDM standard method ones and newly-developed ones) as well as the model development methods used in this dissertation. Here I consider the question: **when should the researcher feel sufficiently confident in their modeling to stop developing (and validating) the model and begin using it?**

Having observed that the SDM literature did not explicitly address this question, and having observed that in other disciplines the researcher's discretion has been used to answer it, Groesser & Schwaninger[161] propose a qualitative method to estimate the *validation cessation threshold* which should be used as a guide for answering this question. A conceptual model is presented for how to make the cessation choice by using relevant attributes of researchers, stakeholders and the project.

The descriptive rule for the threshold considers: model size (e.g., the total count of variables[161, 180]), relative importance/risk of decision, target group's experience with modeling; target group's expectations, data availability, and the researcher's level of expertise. The *conceptual framework* describes how these factors interact to determine cost and other relevant factors. The researcher describes these factors as they are characterized within the individual research project and uses the framework to determine the validation cessation threshold.

I used the framework as follows: I reflected on each factor and ascribed a score (range = low, medium, high) with a corresponding explanation. Then I used the framework to determine the *validation cessation threshold*. I used this score to guide my validation work – as a qualitative assessment of the level of effort I would need to make for validation.

2.4.5.2 DATA SUITABILITY

I reflected on the three types of data that are possible, the data available, and their limitations; in order to answer the following question: are the data upon which the model is based used within the scope of their limitations? I then created the *Data Suitability* table described in Appendix E.

2.4.5.3 METHODS SUITABILITY

I reflected on the modeling process and the modeling languages used in the research process, in order to answer the question: what are the trade-offs of the specific modeling approach used in this research? I then created the *Methodological Tradeoffs* table described in Appendix E.

2.4.5.4 VALIDATION METHODS AS USED IN PHASE 2

2.4.5.4.1 SHARED MENTAL MODEL SATURATION

I read the problem statement and attempted to tell stories using SMM1. Specifically, these stories involved the elements and behaviors identified in the problem statement as expressed in the dynamics visualized in SMM1. I then reflected on the stories told to assess whether the variables identified in SMM1 meet the model's purpose (as expressed in the problem statement).

Then, I created SMM-S Curves and SMM-S Diagrams as described in Appendix E.

First, I reviewed the SMM-S Curves for saturation (i.e., flattening curves means no new concepts are emerging). I started with the variables SMM-S Curve. Then, I reviewed the links, feedbacks and delays SMM-S Curves for saturation. (I used the same process for CM-S Curves.)

Second, I reviewed SMM-S Diagrams for relationships that were less well-established in those diagrams, meaning that fewer *clinics* mentioned them and/or fewer *clinics* mentioned them explicitly.

Third, I focused on each of those relationships that were less well-established, considering whether the relationship distorts aspects of the data on which the diagrams are based (the clinic CLDs), and whether it is plausible in terms of cognitive limitations and is feasible in terms of the culture (roles, goals, policies). Modifications were made as needed.

As the SMM-S Test was passed, additional interviews from the saturation reserve were not consulted.

2.4.5.4.2 CONCEPTUAL MODEL SATURATION

The *model validation set* of interviews (n=10) consisted of four clinician interviews, four MA interviews, and two clinic management interviews. The management personnel interviewed were the clinic's center manager and nurse manager (a staff member with clinical and managerial duties). Interview transcripts were coded and *rigorously-interpreted quotations for causality* (Causal RIQs) were generated from the coded transcripts. This process also generated CM-S Curves and CM-S Diagrams. The design and techniques for creating coded interviews, Causal RIQs, CM-S Curves and CM-S Diagrams are described in Appendix E.

CM-S Curves and CM-S Diagrams were used as in SMM-S. This time, after establishing that saturation had been reached (using CM-S Curves), I reviewed the coded model validation set transcripts and created Causal RIQs – this process tested how capable SMM2 was of exposing the variables in participants' mental models. Model modifications were made as needed. Elements that were revised during CM-S were checked against the problem statement to make sure that they fall within the model's purpose.

Then, I reviewed the CM-S Diagrams, focusing on the relationships that were marked as less well-established in those diagrams, meaning fewer *participants* mentioned them and/or fewer *professions* mentioned them. I considered whether the relationship distorts aspects of the *model validation set* participants' statements, and whether it is plausible in terms of cognitive limitations and is feasible in terms of the culture.

As the CM-S Test was passed, additional interviews from the saturation reserve were not consulted.

2.4.5.4.3 STAKEHOLDER DIALOGUE REVIEWING SHARED MENTAL MODEL

In this step, stakeholders were shown the shared mental model CLD. The shared mental model permits a high-level discussion of the emergent themes which enables stakeholders to think creatively and identify similar concepts, including qualitative theoretical constructs, which can be of value in making more useful both the conceptualizing of the simulation model and the development of theory.

Stakeholders engaged were: local, national and international experts in primary care, problem owners at the HSDO, primary care clinicians and experts in health services research, health systems research and SDM in health. Stakeholders were encouraged to question the shared mental model structure and point out flaws (based on their level of experience and proximity to the system in question).

These stakeholders were engaged at separate times as opportunities arose and they were shown the version of the shared mental model that was most up to date at that time. Opportunities for these dialogues included: informal one-off meetings, HSDO meetings as well as meetings scheduled during local and international conferences. Sometimes stakeholder discussions were one-on-one, other times they involved a group of stakeholders. In all cases, the effort was made to have an open dialogue where all stakeholder viewpoints were considered.

2.4.5.5 VALIDATION METHODS AS USED IN PHASE 3

2.4.5.5.1 SYSTEM DYNAMICS MODEL STANDARD VALIDATION METHODS IMPLEMENTED

Table 2.17 below lists the SDM standard method model validation tests (in the first column), the procedures I used to implement those tests (in the second column) and the test definition (in the third column).

Table 2.17 Validation Methods Implemented

Name of test	What I did (see Chapter 3 Section 3.5 for results)	Definition[123] (p. 189-195)} and[105] (p. 123)
Parameter Confirmation - Theoretical	I reviewed literature for numerical estimates of the constants in the model.	"Evaluating the constant parameters … [against generalized knowledge about the system that exists in the literature] both conceptually and numerically."[123]
Parameter Confirmation - Empirical	*Implemented using the following methods: SSM-S, CM-S, SIM-S and Stakeholder Dialogue.*	"Evaluating the constant parameters against knowledge [obtained directly from] the real system, both conceptually and numerically."[123]
Structure Confirmation - Empirical	*Implemented using the following methods: SSM-S, CM-S, SIM-S and Stakeholder Dialogue.*	"Comparing the model structure with information (quantitative or qualitative) obtained directly from the real system being modeled."[123]
Structure Confirmation - Theoretical	Related models were identified but addressed different problems, prohibiting formal comparison.	"Comparing the model structure with … [generalized knowledge about the system that exists in the literature]."[123]
Behavior Reproduction	I checked the model to see if it was able to reproduce the reference mode. I also checked that the model was able to reproduce both failure and success modes as described in the problem statement.	"Does the [simulation model's] behavior match any historical data and/or the reference mode?"[105]
Extreme Policy	I implemented values for each policy that represented extreme conditions and checked that the model behavior was reasonable.	"When policies are pushed to extremes are the [simulation model's] behaviors reasonable?"[105]
Mode Reproduction Ability	I implemented several past policies in the model to see if it was able to reproduce behavior consistent with historic behavior.	"With different past policies, does the [simulation model] yield behaviors consistent with other [examples] of the system"[105]
Behavior Prediction	I checked that the model reproduces the anticipated behavior for future/hypothetical situations.	"Does [the simulation model] reproduce the anticipated behavior in future/hypothetical situations?"[105]
Anomalous/Surprise Behavior	I simulated the model under various conditions and found anomalous behaviors. I studied these causes of these behaviors in the model. When these behaviors were indeed anomalous, corrections were made to remove them. Other times, they were surprise behaviors that led to insights.	"Have odd behaviors been studied to show that either: 1) they are anomalous, needing [simulation model] corrections to remove them? 2) the [simulation model] yields insights into a previously unrecognized mode?"[105]
Perspectives Boundary Adequacy	*Implemented using the following methods: SD-S, Stakeholder Dialogue Suitability and Methods Suitability.*	"Do the models support debate on different perspectives in the [Problem Statement] concerning: 1) choice of modelling approach used? 2) System dynamics issue addressed? 3) Goals to be achieved? 4) Policies for doing so?"[105]

Name of test	**What I did (see Chapter 3 Section 3.5 for results)**	**Definition**[123] (p. 189-195)} and[105] (p. 123)
Structure Boundary Adequacy	*Implemented using the following methods: SSM-S, CM-S, SIM-S, Methods Suitability and Stakeholder Dialogue*.	"Do the models contain sufficient and appropriate variables, policies and feedback loops to address the issue that they are being built to study?"[105]
Dimensional Consistency	Used the "check units" feature in Vensim to confirm that the right-hand side and left-hand side of each equation passed the dimensional consistency test.	"Checking the right-hand side and left-hand side of each equation for dimensional consistency"[123]
Extreme conditions in equations	I implemented extreme values for constants in the model and checked that the outputs of each policy were reasonable.	"Are the outputs of policies reasonable if the inputs take extreme values?"[105]
Behavior Sensitivity	I implemented sensitivity analysis without policies turned on and checked that the model output was plausible. Specifically, I performed a Monte Carlo analysis with 200 runs for each of the constants in the model, each in a range of $\pm$50% their original values.	"Are the previous behavior tests compromised by the plausible changes in parameter values?"[105]
Behavior Boundary Adequacy/Structure Sensitivity	I implemented multiple equally-likely structures in the model during model development. As I tested each one, those producing behavior that did not match the behavior described by interviews or operational thinking or those where a simpler model produced the same behavior were abandoned.	"Does the [simulation model] contain sufficient and appropriate variables, policies and feedback loops to address the issue when this is tested by adding new pieces of relevant structure and examining the resulting behavior?"[105]
Policy Sensitivity	I implemented sensitivity analysis for the model with various policies turned on.	"Are the suggested [policies and recommendations] robust to plausible parameter changes?"[105]
Policy Boundary Adequacy	I added possibly relevant structure and tested its impact on behavior, under different policy scenarios.	"Does the addition of more possibly relevant structure change the [policies and recommendations]?"[105]
Norms/Values Boundary Adequacy	*Implemented using the following methods: SSM-S, CM-S, SIM-S and Stakeholder Dialogue.*	"Do the models support debate concerning, and represent the behavior of the relevant actors': 1) goals (are the desired states acceptable?) 2) Policies (are the actions based on discrepancies between goal and actual conditions acceptable within the culture?)"[105]
Roles Boundary Adequacy	*Implemented using the following methods: SSM-S, CM-S, SIM-S and Stakeholder Dialogue.*	"Are the feedback links in the models consistent with the abilities of current actors in the system to access, interpret and employ information?"[105]
Italics are used to denote the use of newly-developed validation methods described in Appendix E.		

2.4.5.5.2 SIMULATION MODEL SATURATION

This test consulted SIM1, the Causal RIQs, the problem statement, and the dynamic hypothesis (i.e., the Conceptual Model) as described in Appendix E.

The problem statement and dynamic hypothesis were consulted to see: how consistent was SIM1 with the structure and behavior that it was meant to represent.

Also, variables and relationships in SIM1 were checked, that they make sense in terms of the descriptions provided in the quotations and interpretations (i.e., proper time frame, proper designation as stock/flow/constant, proper scope). Then, the full quotation was reviewed with SIM1 in hand, to make sure that the causality described in the quotation is covered in SIM1. When goals and policies were discussed, I checked that they were also found in the simulation model and that they were within scope of the cognitive limitations apparent in the quotations. Finally, the model was run using the structural aspects described in these quotations and the run results were compared with behavioral expectations mentioned in those quotations. Modifications to SIM1 were made, as needed in this test. Once complete, the model was called SIM2.

2.4.5.5.3 STAKEHOLDER DIALOGUE REVIEWING SIMULATION MODEL

In this step, stakeholders interacted with the simulation model. The simulation model permits a high-level discussion of the structural theory in the model and the behavior patterns it produces. Stakeholders interacted with the model as they would with a rough draft flight simulator – 1) checking for bugs (i.e., looking at anomalous simulation runs and discovering causes from flawed formulations, whether conceptually or mathematically), and 2) discussing how best to organize the dashboard and outcome graphs.

Stakeholders engaged were: national experts in primary care, primary care clinicians and experts in health services research and SDM in health. One-on-one meetings were held with each stakeholder. Notes were taken during each discussion and model modifications were made after reflecting on those notes.

2.4.5.6 VALIDATION METHODS AS USED IN PHASE 4

2.4.5.6.1 STAKEHOLDER DIALOGUE REVIEWING POLICY ANALYSIS

In this step, stakeholders interacted with the simulation model and discussed policy results. Policy discussions permitted one more review of the simulation model. This review considered model’s structure and behavior as well as the scenarios and policies incorporated into the simulation model.

Stakeholders engaged were: national experts in primary care, primary care clinicians and experts in health services research and SDM in health. One-on-one meetings were held with each stakeholder. Notes were taken during each discussion. After reflecting on those notes, the comments on structure and behavior resulted in model reformulation and policy analysis was performed again. Comments on policy and future research items were incorporated into the conclusions chapter.

2.4.5.7 SYSTEM DYNAMICS SATURATION

First, Cognitive RIQs were generated as described in Appendix E. Causal RIQs, CM-S Diagrams and Information Accumulation Graphs previously generated in CM-S were also consulted. All of these were reviewed to assess the extent to which the validation set interviews verified that the target group sees SDM research as a useful way of addressing the issue.

Causal RIQs verified that participants have mental models with causality and time delays. Information Accumulation Graphs visualized the extent to which participants' interviews contained causal statements. CM-S Diagrams visualized those causal statements. Finally, the Cognitive RIQs verified that participants see themselves as having mental models which are characteristic of Mental Models of Dynamic Systems, that they recognize that changing mental models is part of the perceived solution to the problem, and finally that the needed mental model changes will require engaging the target group on a sub-conscious emotional level. The SD-S Test was passed. Additional review of findings to date was not needed.

2.4.5.8 STAKEHOLDER DIALOGUE SUITABILITY

I reflected on the extent to which stakeholders have been engaged in this research project. Specifically, I considered which stakeholder groups were engaged, as well as how and when they were engaged. I also considered the extent to which there was opportunity for iterative dialogue with individual stakeholders.

2.4.6 PHASE 4 – POLICY ANALYSIS

Now that model validation tests have been completed, the next phase of research involved identifying, creating the structure for, and running the model with various policy changes and environmental conditions. Policies are features of the system that target audience stakeholders can change. Environmental conditions are contextual features of the system that are not changed within the system boundary (i.e., alternative realities).

First, I identified environmental conditions that could change how the system develops over time (i.e., resulting in the variation described in the problem statement) as well as policies that could be implemented. Environmental conditions and policies were documented as they were mentioned by participants and stakeholders. The environmental conditions mentioned focused on certain agents' preferences which, for purposes of understanding my problem, are assumed to remain unchanged over the course of the model's timeframe.

Second, I created model structure to incorporate these system changes into the model (SIM3)[114] (p. 110, 442). For each new model structure, I used SDM standard method tests to verify that formulation errors (conceptual or mathematical) were not introduced and that policies functioned as they were intended to function.

Third, I ran *what-if* scenarios to see the effect of each policy in isolation. I also performed sensitivity analysis to test the robustness of each policy recommendation to different environmental conditions and uncertainties. Finally, I ran policies in tandem and under different environmental conditions to see the impact of policy interactions (e.g., synergy)[113] (p. 86).

Fourth, from this learning, I developed a system policy-structure diagram[114, 180] (p. 162-163) (referred to in this dissertation as the Theoretical Model). This diagram is a simplified visualization of the Simulation Model that summarizes the understanding gained in policy analysis.

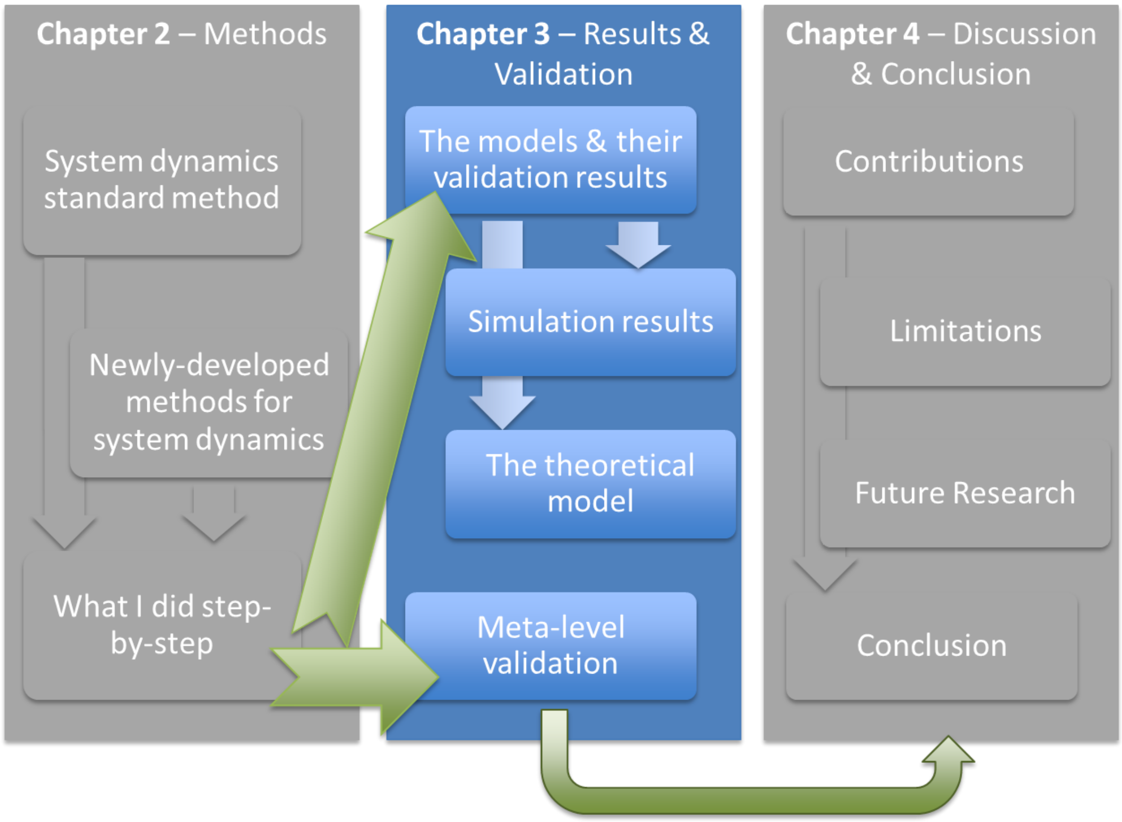
Validation in this phase consisted of Stakeholder Dialogue (Section 2.4.5.6).

2.5 METHODS AIM SUMMARY

My methodological aim was to select *adequate* methods, and create new ones when needed, in order to meet the theoretical aim undertaken in this dissertation. I used a mixed methods process to select the system dynamics methodology for this research (see Appendix C). In this chapter, I began with an overview of system dynamics and then presented my newly-developed model development and validation methods along with accompanying rationales (see Appendix E for details on these methods). While I created these new methods to complement the existing principles[108] of system dynamics methods, the process occasionally required new theory (specifically with respect to validity types and subtypes, see Appendix D).

In Section 2.4, I briefly described the methods I used in this dissertation, including ones from the SDM standard method and my newly-developed methods. The next chapter presents results from throughout the iterative model development and validation process. This format is used because insights were gained all along the way and to demonstrate the contributions of my newly-developed methods using the case of studying PCT.

CHAPTER 3 RESULTS & VALIDATION

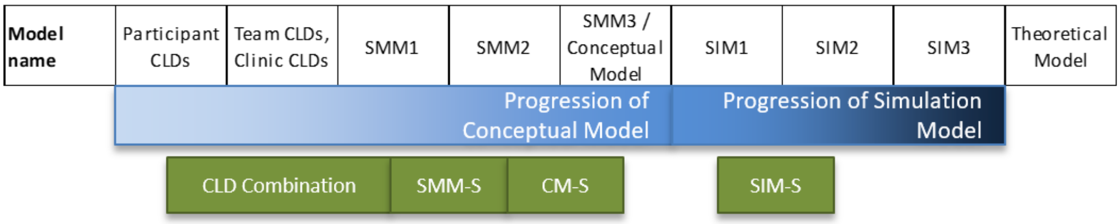


3.1 INTRODUCTION

This chapter presents how my theory of primary care transformation (PCT) iteratively improved. As I used newly-developed methods along the way, this chapter presents an example using the methods I developed.

Figure 2.1 will be used periodically in this chapter: each time that I transition to a new model iteration showed in gray. This is to help the reader orient themselves to where the results fit in the theory building process. Figure 3.1 below combines Figure 2.1 and Table 2.2 to present a more detailed orientation for the reader.

Figure 3.1 Model Progression Sequence & Summary



Model name	Role in model development
Participant CLDs	The CLDs that were generated; one specific to each individual interview.
Team CLDs, Clinic CLDs	The CLDs that were generated in implementing CLD Combination; one specific to each team and clinic in the model development set of interviews.
SMM1	Shared mental model 1 is the result of CLD Combination. It is the first model that is assumed to be a shared mental model (i.e., it is the first draft of the dynamic hypothesis).
SMM2	Shared mental model 2 is the version of the model after modifications made during the **Shared Mental Model - Saturation (SMM-S) Test:** Do the different clinics agree on the structure of the system?
SMM3 / Conceptual Model	Shared mental model 3 is the version of the model after modifications made during the **Conceptual Model – Saturation (CM-S) Test:** Does an additional clinic agree with SMM2 on the structure of the system? (SMM3 is also referred to as the Conceptual Model).
SIM1	Simulation model 1 is the first quantitative version of the model. It is produced using SMM3 as the blueprint.
SIM2	Simulation model 2 is the version of the model after modifications made during the **Simulation Model –Saturation (SIM-S) Test:** Does an additional clinic agree with SIM1 on the structure *and behavior* of the system?
SIM3	Simulation model 3 is the version of the model after policy analysis structures were added to SIM2.
Theoretical Model	It is a visualization of the simulation model, bringing together the policy structure diagrams that describe the simulation model.

This chapter describes the theory development process in 5 phases as below:

Section 3.2 **mental model elicitation** (Phase 1)

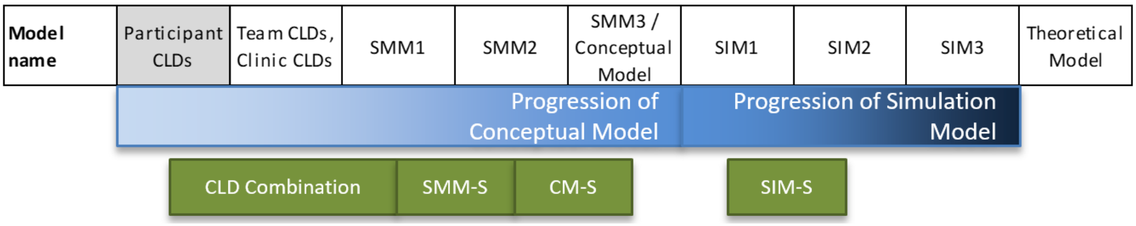
Section 3.3 the **Conceptual Model** (Phase 2)

Section 3.4 the **Simulation Model** (Phase 3)

Section 3.5 **validation** results (Cross-phase step)

Section 3.6 results of **policy analysis** as well as the **Theoretical Model** (Phase 4)

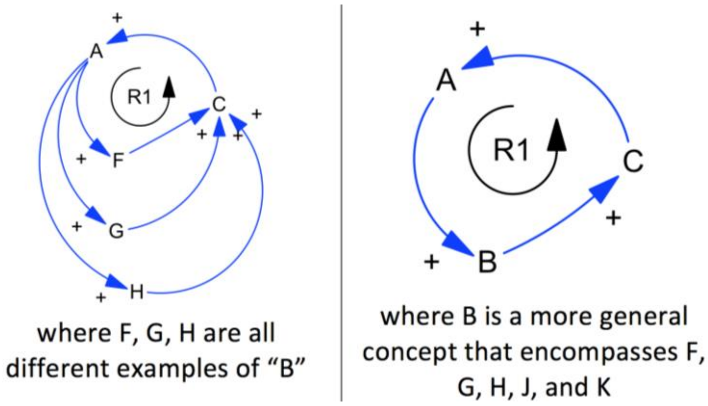
3.2PHASE 1 – MENTAL MODEL ELICITATION



This section presents the individual-level variation observed in CLDs produced from mental model elicitation. Visual inspection showed wide variation among the 20 individual CLDs. Clinicians and clinical staff members frequently mentioned a subset of variables. Clinicians and clinician staff members varied as much with others in their profession as they did with members of the other profession on the detail and dynamic complexity[28][181, 182] of their CLDs. Overall, these CLDs contained:

- Variables (some had only a few variables, others had very many)
- Causal links (by definition, variables in CLDs have causal links)
- Causal chains (sometimes very short, sometimes long but without feedback loops)
- Time delays (fairly common)
- Feedback loops (few and often very short)
- Unlinked portions (sometimes one structure was not linked to another structure in the CLD; i.e., not all variables were linked in some way to all of the other variables)
- Detail complexity (Figure 3.2 below provides two CLDs showing A to "something" to C. The CLD on the left has detail complexity. Instead of B, it uses F, G and H which are conceptually equivalent to the more aggregate concept of B).

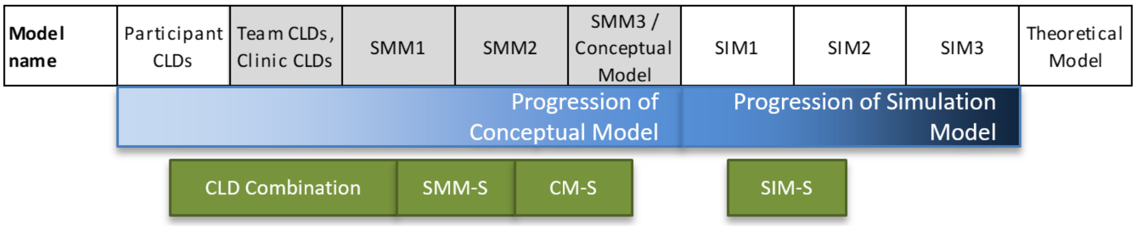
Figure 3.2 Detail Complexity



Mild pruning of these CLDs resulted in simpler diagrams that continued to show wide variation among individual CLDs[183-192].

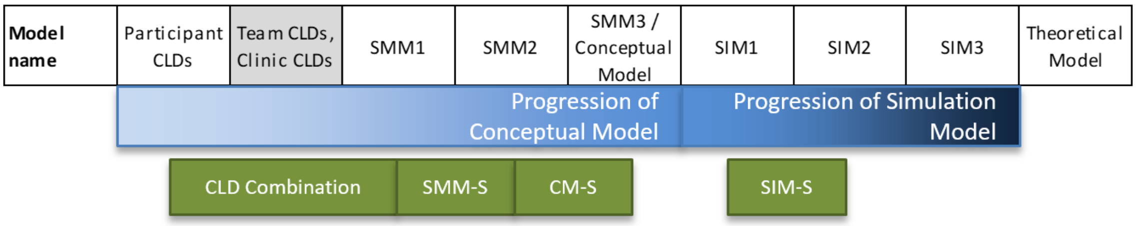
[28] See Hopper & Stave references for measures of CLD dynamic complexity.

3.3 PHASE 2 – CONCEPTUAL MODEL



Here I present the results obtained while building the Conceptual Model. Section 3.3.1 discusses the variation observed among these CLDs. Section 3.3.2 presents SMM2. Section 3.3.3 presents the Conceptual Model (SMM3) with a summary of description of its balancing and reinforcing feedback loops.

3.3.1 VARIATION WITHIN TEAM AND CLINIC CAUSAL LOOP DIAGRAMS



The 20 mildly-pruned individual CLDs were aggregated using CLD Combination into 10 Team CLDs, then into 5 Clinic CLDs, and finally combined and fully pruned into a *shared mental model* called SMM1. The CLDs of clinician and clinical staff members on any one team complement each other – many variables were discussed by both professions and new linkages emerge when creating Team CLDs. Visual inspection of Team CLDs showed[183-192]:

- Common variables appear (e.g., *flow*, *MA follows the patient throughout the visit/in room with patient and clinician*).
- Detail complexity could be pared down (e.g., in Team CLDs, several variables were used to describe the specific tasks delegated to MAs including: *charting*, *prescriptions*, *referrals*, *returning phone messages*, and *following the patient*. These are referred to as *Task-shifting* in the Conceptual Model. They are referred to as *MA Advanced Tasks* in the Simulation Model).
- Substantial variation persists (e.g., Team 09 has 20 variables and Team 06 has 3, where only one variable overlaps).
- While common variables are considered across teams, the causal links drawn to and from a given variable will be different across CLDs (a brief example is presented in Table 3.1 below).

Table 3.1 below identifies the teams that discussed causes of MA turnover (other teams did not mention it). Each team mentioned only 1 to 2 of the 5 causes mentioned across teams.

Table 3.1 Causes of MA Turnover in Team CLDs

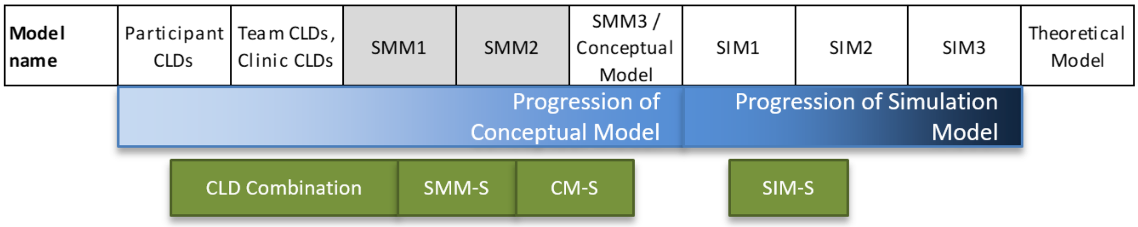
	Team ID						
Causes of MA Turnover (variables)	**03**	**04**	**05**	**07**	**08**	**09**	**10**
MA as a transitory job (a stepping stone)		x		x	x		
MA as a dead end job			x				x
MA satisfaction	x			x	x		
MA capabilities (increased role, like a medical student)						x	x
MA workload impacting MA satisfaction						x	

Visual inspection of Clinic CLDs also showed some common elements, with variation in content and complexity[183-192].

Table 3.2 below presents the variation among Clinic CLDs, alongside SMM1. Clinic E identifies the most elements in each category, yet even this clinic did not mention all of the elements that are found in the shared mental model. Clinic A identified the fewest elements in each category; however one variable, two links, one delay and one loop identified by Clinic A were missed by Clinic E.

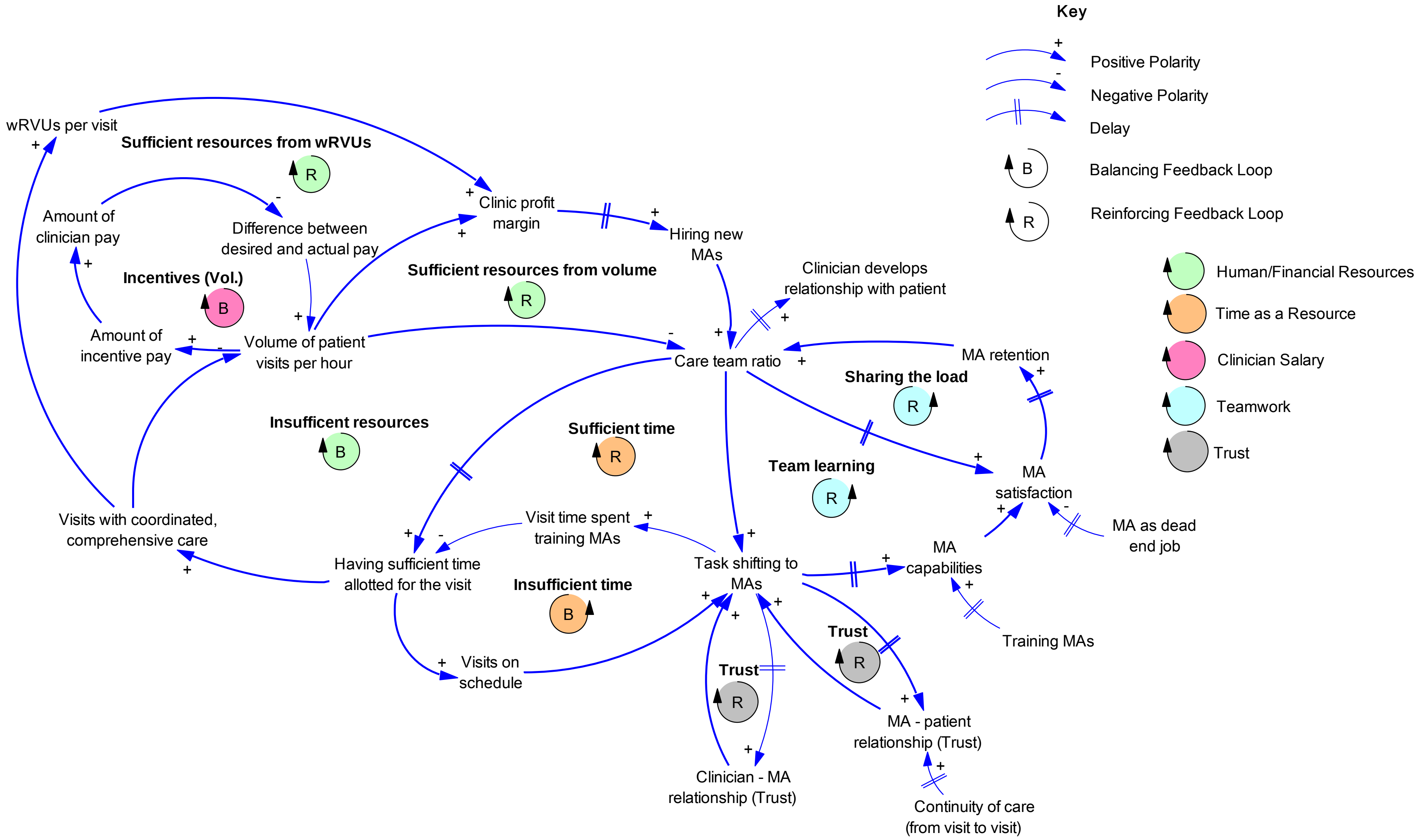
Table 3.2 Variation in Clinic Mental Models' Match to Shared Mental Model

	Elements	**Shared Mental Model (SMM1)**	**Clinic A**	**Clinic B**	**Clinic C**	**Clinic D**	**Clinic E**
Model	*Variables*	22	10	15	13	11	19
	Links (& Polarities)	31	10	18	18	13	26
	Delays	10	3	7	7	6	8
Specific to *Task-shifting* Variable	*Loops*	8	1	4	5	3	6
	Reinforcing	6	1	4	3	3	4
	Balancing	2	0	0	2	0	2
Other significant loops	*Loops*	2	0	0	1	0	2
	Reinforcing	1	0	0	1	0	1
	Balancing	1	0	0	0	0	1

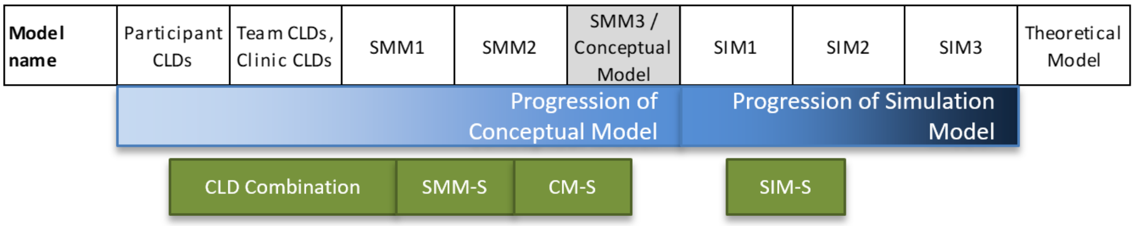


Clinic CLDs were merged to produce one model, referred to as SMM1. Figure 3.3 below presents SMM2 (which is equivalent to SMM1 as no modifications were made during the validation step between SMM1 and SMM2). I will not describe SMM2 as the next section presents the final Shared Mental Model – SMM3. Here, I show SMM2 only to present the result of this step in the process.

Figure 3.3 SMM2 (After SMM-S Test was Applied to SMM1)



3.3.3 THE CONCEPTUAL MODEL

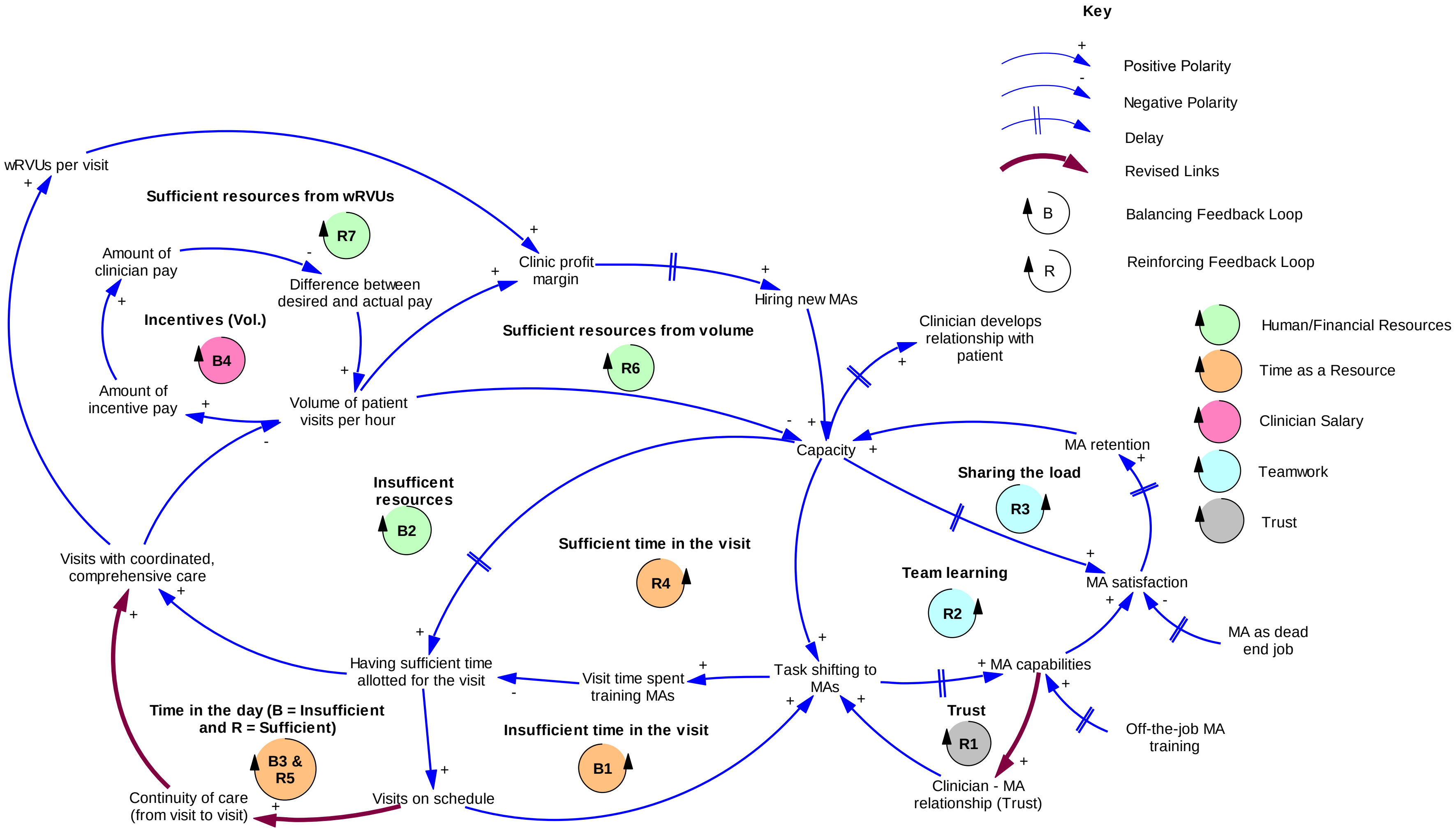


SMM3 is presented in Figure 3.4 below. Revised links from validation of SMM2 (producing SMM3) are identified in maroon. The reasons for these revisions are presented in Appendix F. No new variables were added, some were renamed. Removed links are not shown.

This model is the Conceptual Model. It is described, loop by loop, after the figure; first focusing on balancing loops, then on reinforcing loops. In the visuals following the Conceptual Model, some variables have been moved around for better visualization but the variables and links remain intact (as presented in SMM3).

SMM3 represents the dynamic hypothesis of the participants whose mental models were elicited in producing it. This hypothesis is that: in the short-term, the balancing loops dominate and, after various time delays, the reinforcing processes have the potential to overcome the balancing loops.

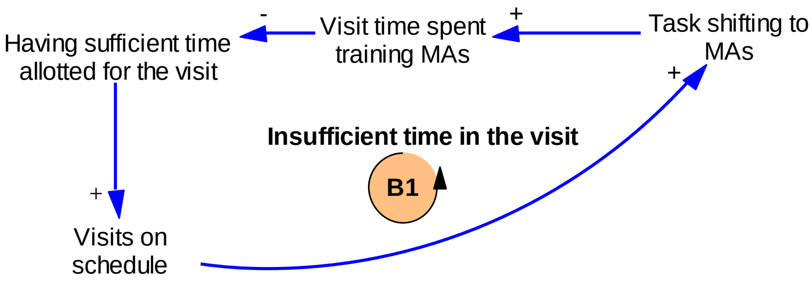
Figure 3.4 SMM3 – The Conceptual Model (After CM-S Test was Applied to SMM2)



3.3.3.1.1 BALANCING LOOPS

Loop B1 (Figure 3.5 below) shows how task-shifting, in the short term, causes teams to have an insufficient amount of time for each visit as this requires on-the-job training for MAs. On-the-job training expands the time associated with each visit, making it more difficult to keep visits on schedule. In order to address the issue of falling behind schedule, clinicians decrease the amount of tasks shifted to MAs.

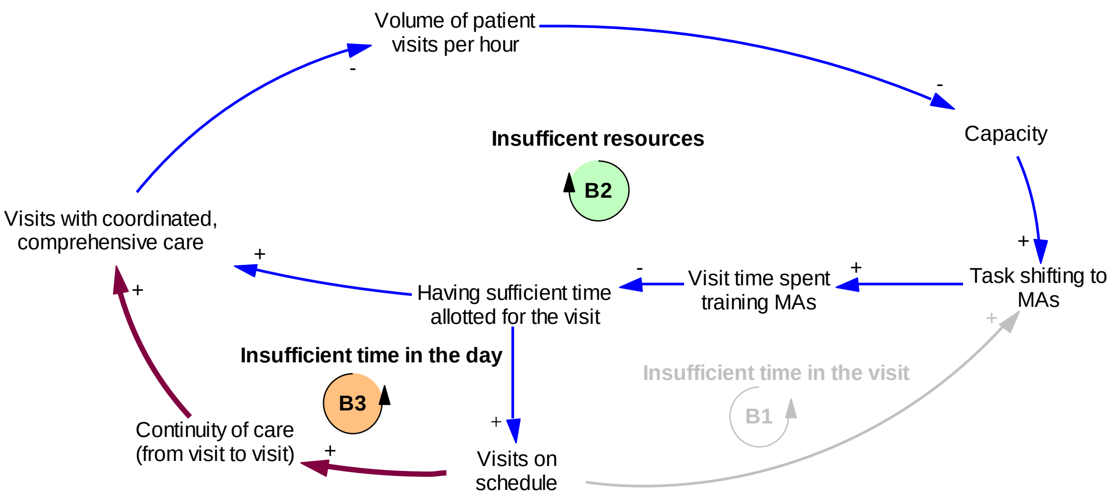
Figure 3.5 Insufficient Time in the Visit



Similarly, Loop B2 (Figure 3.6) shows how training pushes the clinician to shift fewer tasks. Having less time in the visit, pushes the team to provide less comprehensive/coordinated care in order to keep up with the patient volume, given their current level of capacity and task-shifting. Additional capacity would be required to allow the team to shift tasks, with sufficient time to train on those tasks, while at the same time continuing to provide comprehensive and coordinated care to all their patients.

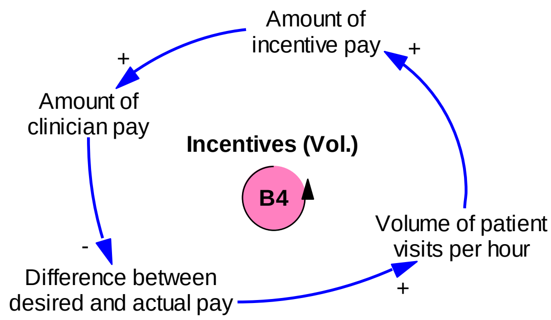
Loop B3 shows that there is another way to deal with this time and resources strain. What can clinicians whose visit time is under pressure from time spent training, resulting from task-shifting, do in the absence of additional capacity? They can allow other clinicians to see their patients, decreasing continuity of care. In so doing, they are able to remain on schedule; having sufficient time to train and to provide comprehensive/coordinated care to the patients that they see.

Figure 3.6 Insufficient time in the Day & Insufficient Resources



Loop B4 (Figure 3.7) shows that the way in which clinician salary is calculated influences the volume of patient visits. The higher the volume of visits, the higher the incentive pay (paid per encounter); thus, the higher the clinician pay. Clinicians compare their pay to their desired pay. The bigger the gap, the more they want to increase the volume of patients seen.

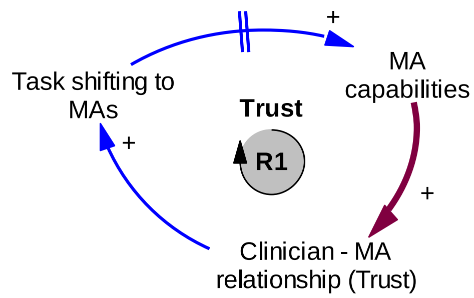
Figure 3.7 The Incentive Structure



3.3.3.1.2 REINFORCING LOOPS

Loop R1 (Figure 3.8) shows that, as tasks are shifted and MAs eventually become more capable at delivering the tasks that are shifted to them, trust increases. That trust is developed over time as *MA capabilities* increase. As that trust is developed, clinicians are more willing to shift tasks to MAs.

Figure 3.8 Trusting the MA

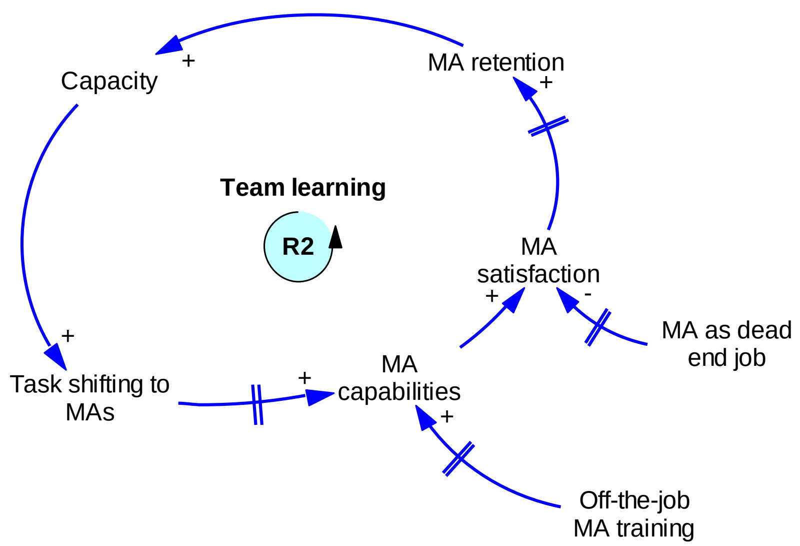


In loop R2 (Figure 3.9), participants indicated that team learning was crucial to the implementation of sustained task-shifting. Over time, as tasks are shifted to MAs, MAs become more capable in performing the requested tasks, increasing their job satisfaction. This source of satisfaction combats the feeling of *MA as a dead end job*. Over time, improved job satisfaction improves *MA retention*.

Having a consistent set of members on one's team (i.e., a high level of *MA retention*) increases the capacity of the care team; not by the physical addition of more MAs, but by the virtual addition of capacity given that the existing MAs become more capable. This increase in capacity increases the team's ability to engage in task-shifting just as adding more MAs would do.

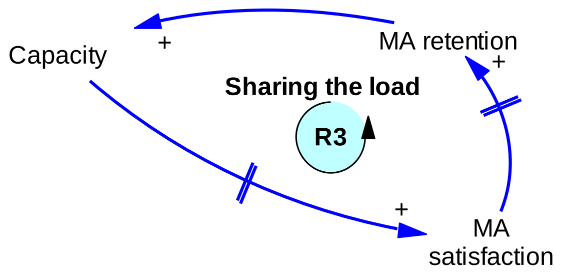
In addition to performing newly-shifted tasks, *MA Capabilities* can also be increased using *Off-the-job MA training*. However, this variable is only a small contributor to team learning (given that it is not directly involved in the learning loop). Also, it takes time for these trainings to sink in.

Figure 3.9 Team Learning



It is important to note that having the capabilities to function successfully in task-shifting is not the only contributor to *MA satisfaction*. Loop R3 (Figure 3.10) shows another contributor: having sufficient capacity to keep up with the work given. MAs are cautious about feeling more satisfied with their job such that increases in capacity need to be sustained for long periods of time in order to improve their satisfaction, hence the operative delay.

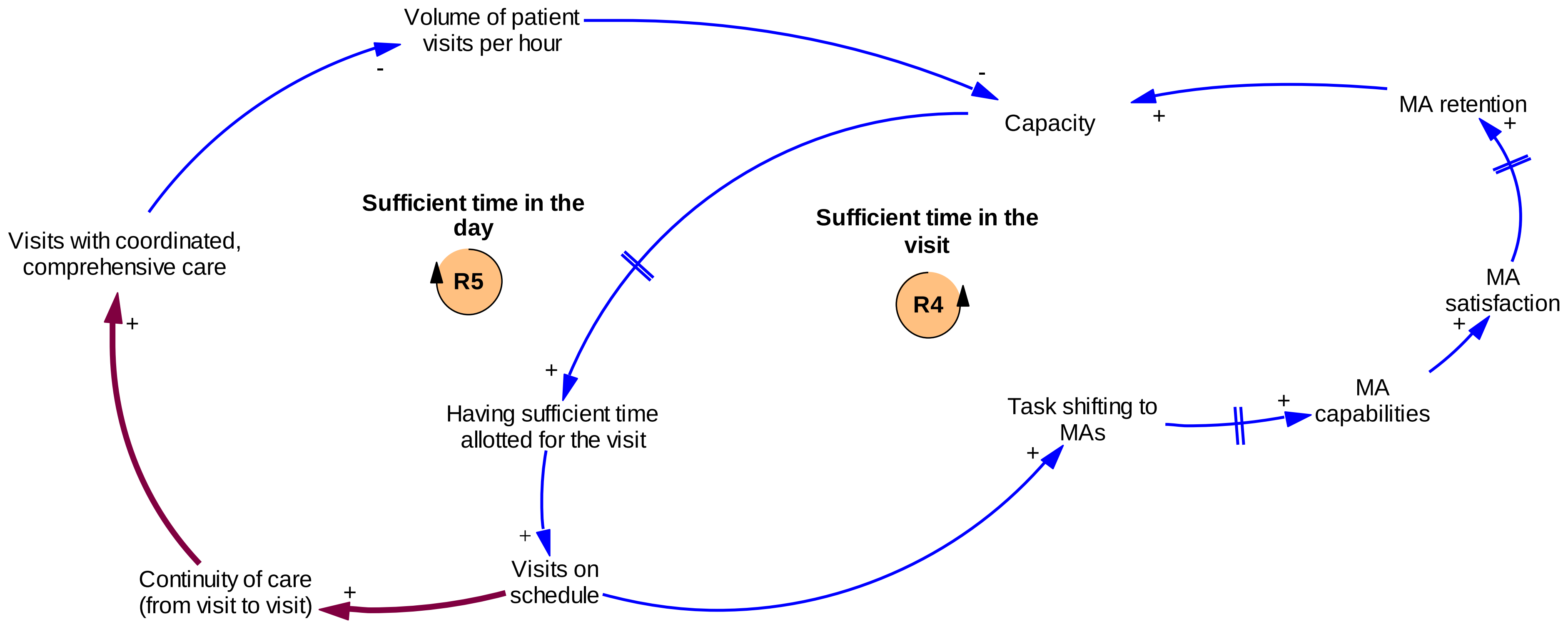
Figure 3.10 Sharing the Load



Loop R4 (Figure 3.11) shows how, over time, task-shifting results in MAs that are capable of performing with increased task-shifting (i.e., increased *Capacity*) such that there is sufficient time to provide the care that is needed in the visit and to keep visits on schedule. Just as *decreasing Visits on schedule* results in decreased *Task-shifting* (B1), increasing *Visits on schedule* results in increased *Task-Shifting* (R4).

Loop R5 indicates that increased capacity also gives the care team sufficient time to provide more comprehensive and coordinated care to their patients. They are able to provide this care to all of their scheduled patients – retaining continuity of care for their patients. Just as decreasing *visits on schedule* results in decreased continuity (B3), increasing *visits on schedule* results in teams that are able to deliver *more* comprehensive, coordinated visits for their own panel of patients (R5).

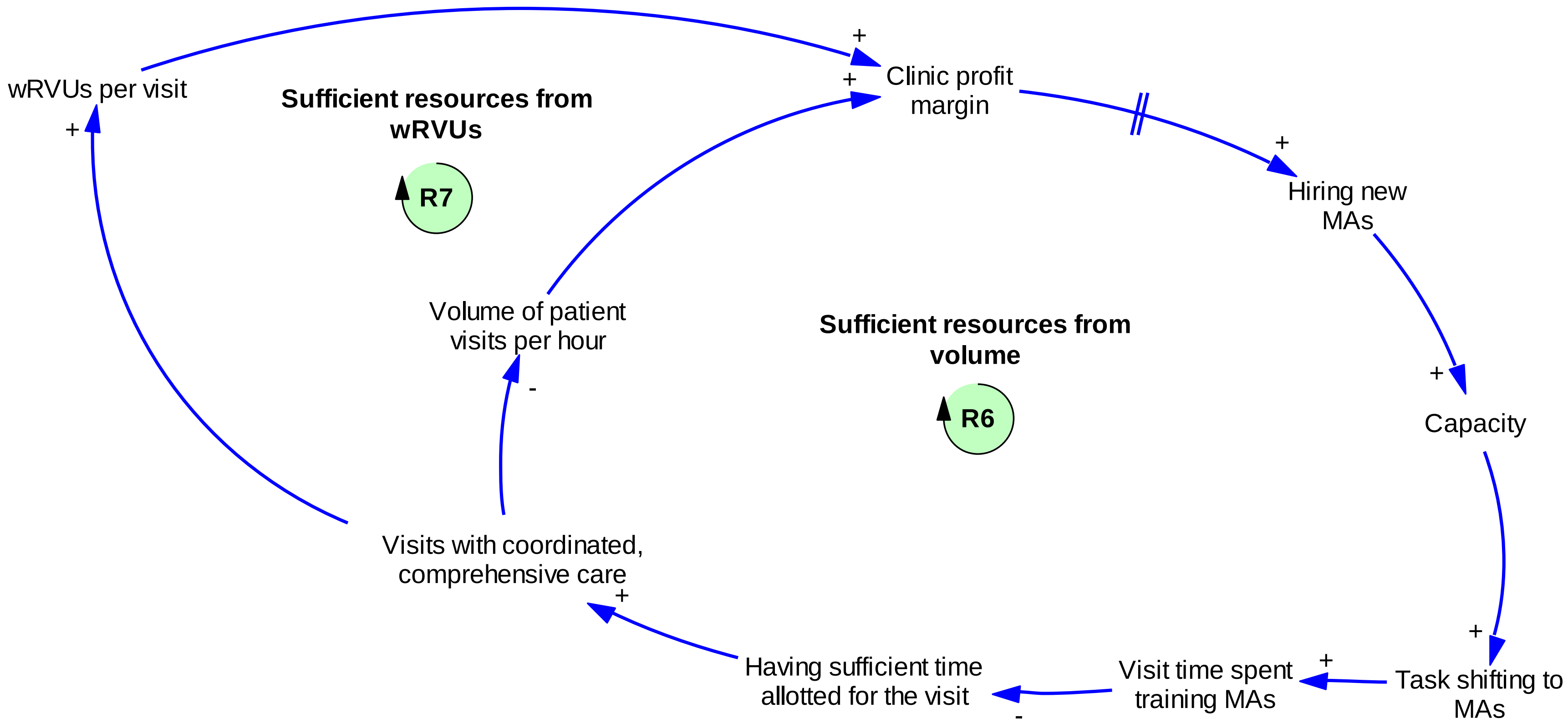
Figure 3.11 Sufficient Time



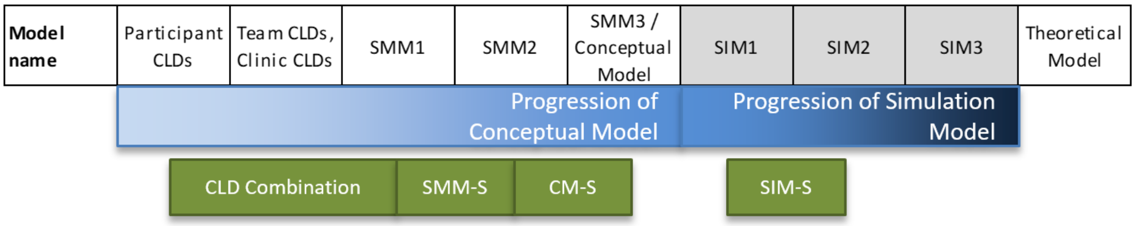
Loops R6 and R7 (Figure 3.12) below show participants' understanding of how resources are generated to justify hiring new MAs to increase capacity. There are two things that generate revenue (and thus contribute to the clinic profit margin): volume (R6) and wRVUs[29][193] (R7). If wRVUs are constant, the volume of patient visits must be kept sufficiently high such that there is a high enough clinic profit margin to hire new MAs. However, the clinic profit margin also increases as more coordinated and comprehensive services are provided (increased wRVUs).

[29] An alternative to paying clinicians for the number of patient visits (or encounters), essentially paying by volume, is paying by what happens in the visit. This is done using what are known as *work relative value units* (wRVUs). These are a task-based incentive where each patient care task (service) delivered is assigned a certain number of *relative value units* such that all tasks can be assigned to the same scale. This scale is then used to calculate how much to pay a clinician for the services provided, as follows: the number of patient care tasks (services) delivered in terms of wRVUs is multiplied by the compensation per wRVU.

Figure 3.12 Sufficient Resources



3.4PHASE 3 – SIMULATION MODEL



SIM1 is the first quantitative version of the model. It was produced using the Shared Mental Model as a blueprint, starting with SMM2, and evolving as understanding grew in developing iterations of SMM2 and finally SMM3. SIM1 was also iteratively improved via SDM standard method validation tests as well as via the **SIM-S** validation method. When these validation tests were passed, the simulation model was referred to as SIM2. Policy analysis structures were added and tested, finally resulting in SIM3. Here, I present the Simulation Model (SIM3). Section 3.4.1 provides a model overview and Section 3.4.2 describes the model structure in greater detail.

3.4.1 MODEL OVERVIEW

This overview presents the main assumptions, time horizon, model boundary, and model sectors.

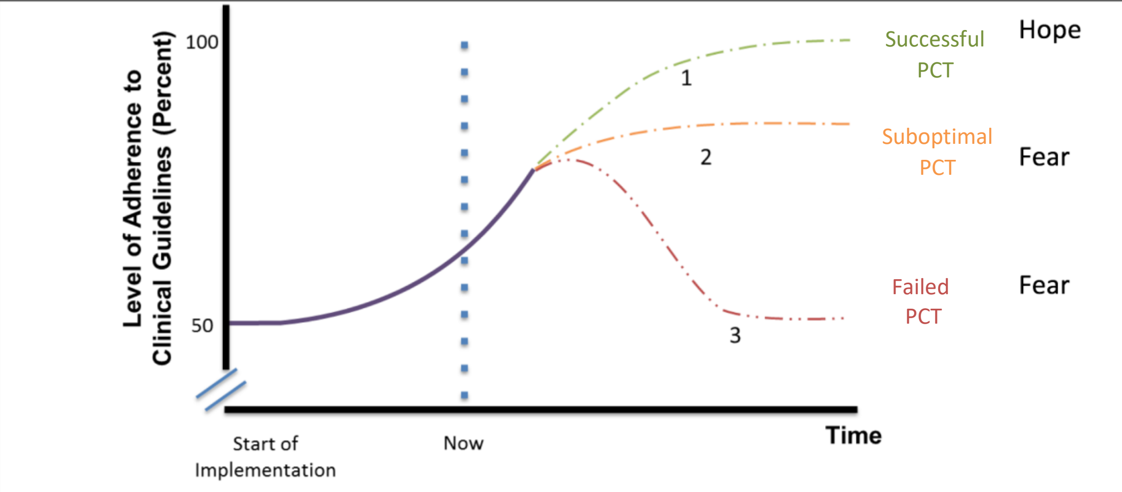
3.4.1.1 PROBLEM STATEMENT & REFERENCE MODE

Figure 3.13 below reproduces the problem statement and presents the associated reference modes.

The Conceptual Model variable which was selected as the reference mode for the PCT model is *visits with coordinated/comprehensive care* (i.e., actual adherence to clinical guidelines, or, *comprehensiveness* for short). This variable figures prominently in participants' descriptions of the 'tensions', combines the goals of two key aspects of primary care, and adequately represents the hopes and fears of system participants. The reference mode shows the start-up behavior and three potential futures: one *hope* and two *fears*. "Hope" is sustained, optimal implementation. The first "fear" is sustained, sub-optimal implementation. The second "fear" is reaching a sub-optimal level of implementation, and then regressing (failure).

Participants described a feeling of frustration with implementation; indicating that they had expected a "step change" to occur where a change would quickly result in the desired outcome. The reality was that, at the "now" point on the reference mode, much less progress has been made than was expected despite the level of effort. Looking toward the future, respondents feared that all the work would be for nothing and implementation would fail (#3 – fear "Failed PCT"). Respondents also feared that the progress might not ever reach the highest level of comprehensiveness (#2 – fear "Suboptimal PCT"). In this scenario, the team would settle into a new normal level of comprehensiveness beyond which they would feel unable to progress. Finally, respondents hoped that with time, sustained optimal implementation would be reached – resulting in full comprehensiveness (#1 – hope "Successful PCT").

Figure 3.13 Reference Mode & Problem Statement



Based on the findings of the scoping study, this dissertation conceptualizes the problem as follows:

> *Primary care transformation has been and continues to be an elusive target. In the short term, it is hard and failure abounds. In the long term, some reach successful implementation. We lack sufficient understanding of the structure of the 'tensions,' which contributes to failure and success modes, and of the policies that can impact this structure.*

This dissertation seeks to develop a grounded, dynamic theory of PCT; in order to build understanding of the key structures generating these hoped-for and feared observed behaviors.

3.4.1.2 TIME STEP & TIME HORIZON

The model time step is one month. This choice is supported by model development and validation interviews where participants describe shifting tasks for a matter of months before they feel they have sufficiently experienced the consequences and decide whether to continue shifting those tasks. It also allows the model to focus on the longer-term trends and accumulations involving PCT rather than on the shorter-term oscillations that even out over the longer period. It also allows the model to focus on aggregate variables rather than, for example, each type of capability and each type of task.

The model's time horizon is 100 months (a little over eight years). The time horizon should be long enough to extend "far enough back in history to show how the problem emerged and describe its symptoms"[113] (p. 90). This time horizon was selected based on existing information and modeling needs. The time horizon needs to be at least two to five years long to match documented experiences regarding how long this transformation has taken at the HSDO and at other organizations implementing the Patient-Centered Medical Home (PCMH). Program documents described the transformation as having begun eight years prior to data collection. Respondents interviewed described wrestling with the transformation for the previous three plus years and they were still in the process of implementing. Recent studies exploring the cost of PCT indicated that a time frame greater than two years is necessary to observe the effects of a *transformation*[194, 195]. As two to five years is the minimum or average, then a longer time horizon is needed to explore the start-up dynamics and the full potential effects. This longer time horizon allows for exploration of the dynamics involved in stakeholder decisions and uncertainties in the Simulation Model.

Without task shifting, the PCT Simulation Model is in equilibrium: normal conditions are not perturbed for the entire time horizon and standard clinic policies remain in effect (i.e., MDs are paid via a fixed salary and benefits, and the clinic hires personnel when needed).

In task-shifting runs, the PCT Simulation Model retains normal conditions until month 10 when a small pulse of task-shifting is introduced. This allows for verification that initial conditions are equivalent between the task-shifting base run and sensitivity analysis as well as policy analysis and scenario runs.

Furthermore, this longer time horizon allows time delays to play out prior to the end of the time horizon. Any one time delay is no longer than a few months, but it is their interaction within the complex feedback structure in the model that creates the longer-term impacts the model intends to capture.

3.4.1.3 BOUNDARY

SIM3 models PCT at the level of a single, generic primary care team consisting of two clinicians, their 2,000-person patient panels and several MAs. This is the smallest possible team where capacity is shared across clinicians. The number of clinicians is held constant and their patient panels are held constant. This model allows the user to more readily see the impact of policy changes on PCT as clinicians decide how much to shift tasks. The number of MAs depends on this decision.

This boundary is within the system scope set out in the problem statement, which refers to "practices" or clinics. PCT can occur in practices ranging in size from a stand-alone clinic operated by one clinician up to large HSDOs which also operate hospitals and insurance companies. In all cases, the basic unit of service delivery is the clinician-led team. In the model, personnel changes are a fraction of a person at a time. This aspect makes the model more like a mid-size group practice or HSDO where the model team represents the average experience of all the teams in the clinic. Thus, one person may leave a specific team at a given point in time, but the average is that a fraction of a person has left when looking across teams at the organization.

Interviews and relevant literature also support this modeling choice. Both describe PCMH as a *clinician-led, team-based intervention*[27]. The interviews provide three motivations for considering a two-clinician team. First, this is a sufficiently accurate representation of the number of clinicians per team in the real system at the HSDO. Second, this allows for ambiguity about whether one of the clinicians is a non-physician clinician (i.e., a physician assistant or nurse practitioner). Third, the HSDO policy of having two-clinician teams was found to work well at bringing some stability to the workload. With this buffer in place, care teams can pay attention to task-shifting long enough for it to succeed. Without it, task-shifting will amplify, rather than attenuate the fluctuations in the daily workload. This greatly increases the risk that the team will judge PCMH a failure. As a two-clinician team brings some stability, it allows clinicians in the model to assess whether or not *Task-Shifting* is working for them without conflating it with workload spikes generated for other reasons.

Sensitivity analysis of table functions explores effects of having a clinician team with varying preferences (e.g., *MD Caution* – degree to which clinician hesitates in shifting tasks). Clinicians in the model have "middle of the road" preferences regarding their practice. As clinicians lead the team, MA preferences are focused on how they internalize their capabilities and capacity (impact on satisfaction) and their desire to learn.

This model is initialized in equilibrium, with a *full patient panel* for both clinicians. Patients are held constant. Clinicians are at capacity for the number of patients they see and they are responsible for what happens to that panel of patients. Clinicians' comprehensiveness is at 50% (*Task-shifting* is at zero). The team's choices regarding comprehensiveness and task-shifting are endogenous and the main focus of this model.

Model runs are not assumed to predict future timelines, but are intended to capture the dynamics which lead to success and failure modes. This is in the tradition of the World Dynamics study where study authors make clear[30][196] that the model's power to predict the results of policies diminishes as soon as the failure modes begin. The goal is to improve understanding of the dynamics preventing teams from reaching successful PCT, without claiming to also know what teams would do to reverse a failed implementation[31].

---

[30] Because their goal was "'prediction' only in the most limited sense"(p. 92) their research was "primarily concerned with the correctness of the feedback loop structure …. that can be exactly analyzed"(p. 122) and tested in order to build understanding of "the system's response to alternative policies"(p. 122). Therefore, they clarified, "what validity our model has holds up only to the point in each output graph at which growth comes to an end and collapse begins"(p. 142).

[31] Research on this problem would require a more detailed structure and adequate numerical data.

The majority of the model's structure is endogenous. These are concepts participants described as being part of feedbacks in the Conceptual Model. Stocks are tightly-coupled. Important MD- and MA-related policies (i.e., those related to MDs' willingness to shift tasks and MA's turnover) are formulated as endogenous feedbacks.

A smaller portion of the model's structure relies upon exogenous variables. These variables are derived from operational thinking about how the feedback loops operate. For example, the Conceptual Model does not show the number of MDs as being part of the tensions but it is necessary to determining capacity of the team. Similarly, with patient panel size, participants described feeling burdened by the workload but not by changes in or the size of the patient panel; however, to estimate the workload we need to estimate the size of the patient panel.

There is another set of variables: those related to external policies impacting the system. These are policies that are turned on and off exogenously by clinic management and payers (i.e., clinic incentive policy and health insurance reimbursement). Participants described these as outside the team's control.

Listed below are components that were touched upon in the interviews but that are excluded from the model:

- MD capability (e.g., systems thinking, problem solving, team leadership, team management)
- MD satisfaction and turnover
- Competition between MDs for visits (e.g., when paid by the encounter)
- MA staffing policies (e.g., the on-call reserve pool, having capable MAs be reassigned to work for specialty clinicians in the primary care clinic either permanently or temporarily when specialty is short staffed)
- Alternative financing instruments used to support PCT (e.g., grants, pilot funding from the larger HSDO or partners or payers)
- MD practice variation on the team (i.e., when clinicians practice vastly differently, they each take "their MA" or they work with all the MAs but require them to do things "their way")
- Scheduling polices (e.g., double-booked and triple-booked appointments)
- Team members that are not there 100% of the time (e.g., part-time clinicians, residents)
- Impact on quality / rework when overworked

- Other aspects of the health system (e.g., hospital readmissions – rework, and specialty visits – seeing a specialist for services that could be done in primary care because they are not currently done in primary care)
- After-hours work done by MD and over-time work done by MA
- Physical context (technology, clinic facility design, co-location with specialty and / or pharmacy)

Broadly speaking, reasons for their exclusion are: (1) the impact of the variables are on a shorter or longer time horizon than the current model (e.g., minutes within the day or on the order of years rather than months) and (2) the concepts are context-specific and can be broadly understood via the variables already in the model; including them as specific variables is not needed to understand the problem. Future research looking at similar problems (e.g., at the organization or community level) may need to consider these elements as well as other time horizons.

Table 3.3 below presents a summary of key model variables or variable sets that are *endogenous* as well as ones that are *exogenous* (i.e., constants in the model). Furthermore, it presents variables that were considered and decided to be outside to the problem statement – labeled *excluded* (i.e., not found in the model).

Table 3.3 Model Boundary Chart

Endogenous	Exogenous	Excluded
Number of MAs	Number of MDs	MD capabilities
Number of encounters	Patient panel size	MD satisfaction
Number of tasks	Clinic hiring policies	MD turnover
MD task-shifting	Payer policies	MA staffing policies
Willingness to shift tasks	Clinic incentive policies	Variations in financing specific roles on the team
Comprehensiveness (adherence to clinical guidelines)	MA external preferences (e.g., minimum acceptable level of satisfaction)	MD practice variation
Training	MD external preferences (e.g., desired salary)	Scheduling policies
MA capabilities	Patient complexity	Part time clinicians and staff
Productivity	Normal productivity	Other health system components
Employee compensation	Normal workload	Impact on quality / rework
Patient satisfaction	Clinic overhead	After-hours work
MA satisfaction	Patient external preferences (e.g., initial patient satisfaction)	Clinic physical context
Facility revenue		MD / MD relationships
Clinic net profit		MA / MA relationships
Actual workload		
Actual productivity		
Patient satisfaction		

Model constants are presented in Appendix F. That said, SIM3 is a small policy model rather than a detailed calibrated one. Its goal is to better understand the problem and identify policies that appear to successfully overcome the forces bringing about this problem – therefore, behaviors are important but specific parameter predictions are not. The impact of uncertainty in these constants was tested using sensitivity analysis.

3.4.1.4 MODEL SECTOR SUMMARY

This section and the next both use sector maps[112] to orient readers to the sector, or portion of the model being described. Following are three figures presenting: a simple diagram of the sectors and their flows (Figure 3.14), a more complex view (Figure 3.15) and a view focusing on where policies and preferences intervene (Figure 3.16).

The PCT Simulation Model has six sectors: personnel, work generation, tasks, capabilities, satisfaction and accounting. There are three flows within and/or between sectors: information, tasks and money. Information flows both within and between sectors. Tasks flow from work generation into the tasks sector. Task-shifting occurs as non-technical tasks are shifted from MDs to MAs; thereby converting them from *MD non-technical tasks* to *MA-Advanced tasks*. Tasks also flow within the tasks sector as some get completed and others are not ever completed (also referred to here as “not completed” or “shed”). Money is generated as money is received for tasks completed and money is spent on personnel and facility cost. MDs, MAs, and clinic management make decisions based on information received from the various sectors (grey circles) – these are called *information inputs* and they inform the *policy functions*[32] that then inform the decisions in the model.

[32] The term *decision function* is used interchangeably with the terms *policy function* and *operating policies* in system dynamics and in this dissertation.

Figure 3.14 PCT Model Sector Map – simple

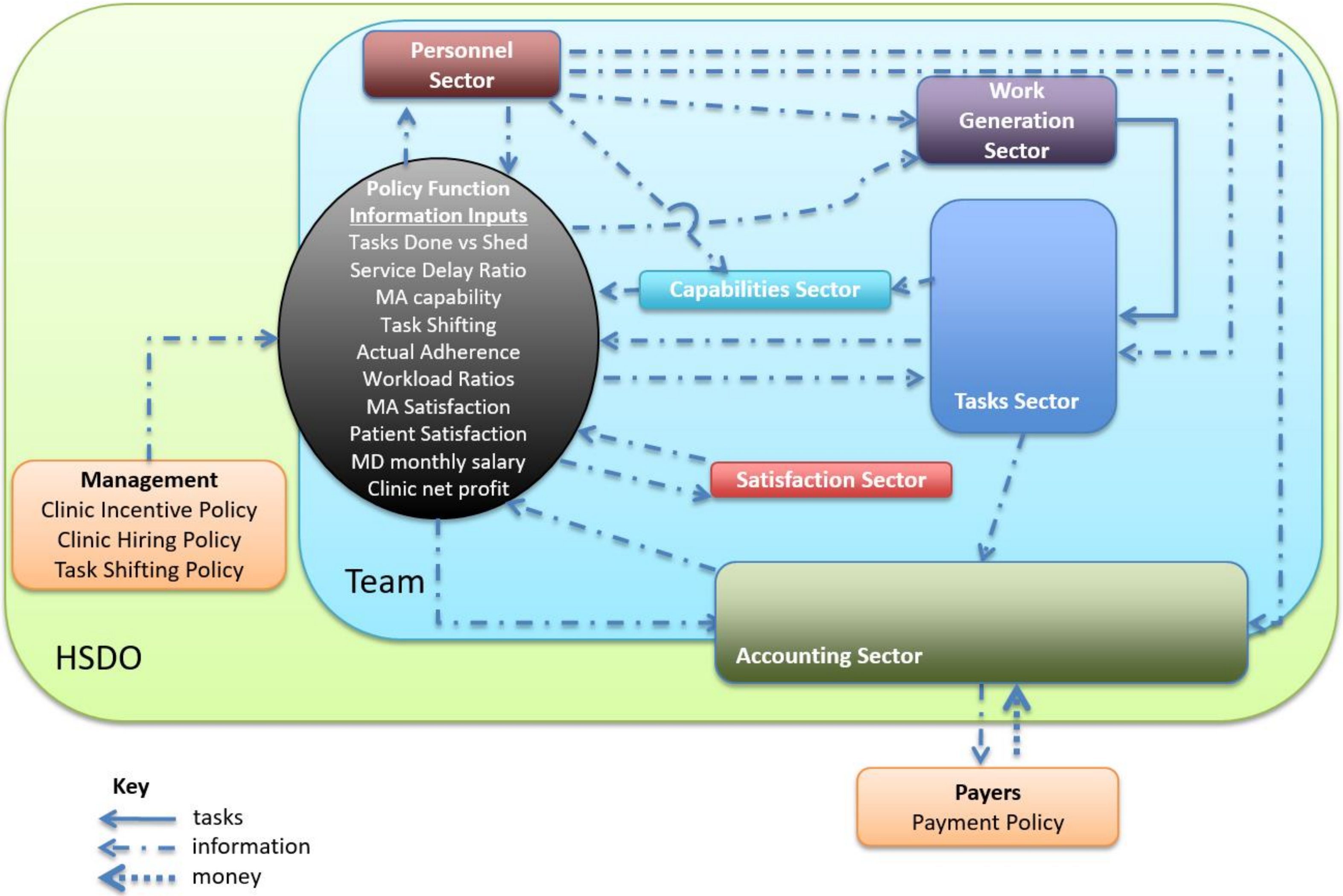


Figure 3.15 PCT Model Sector Map – more detailed

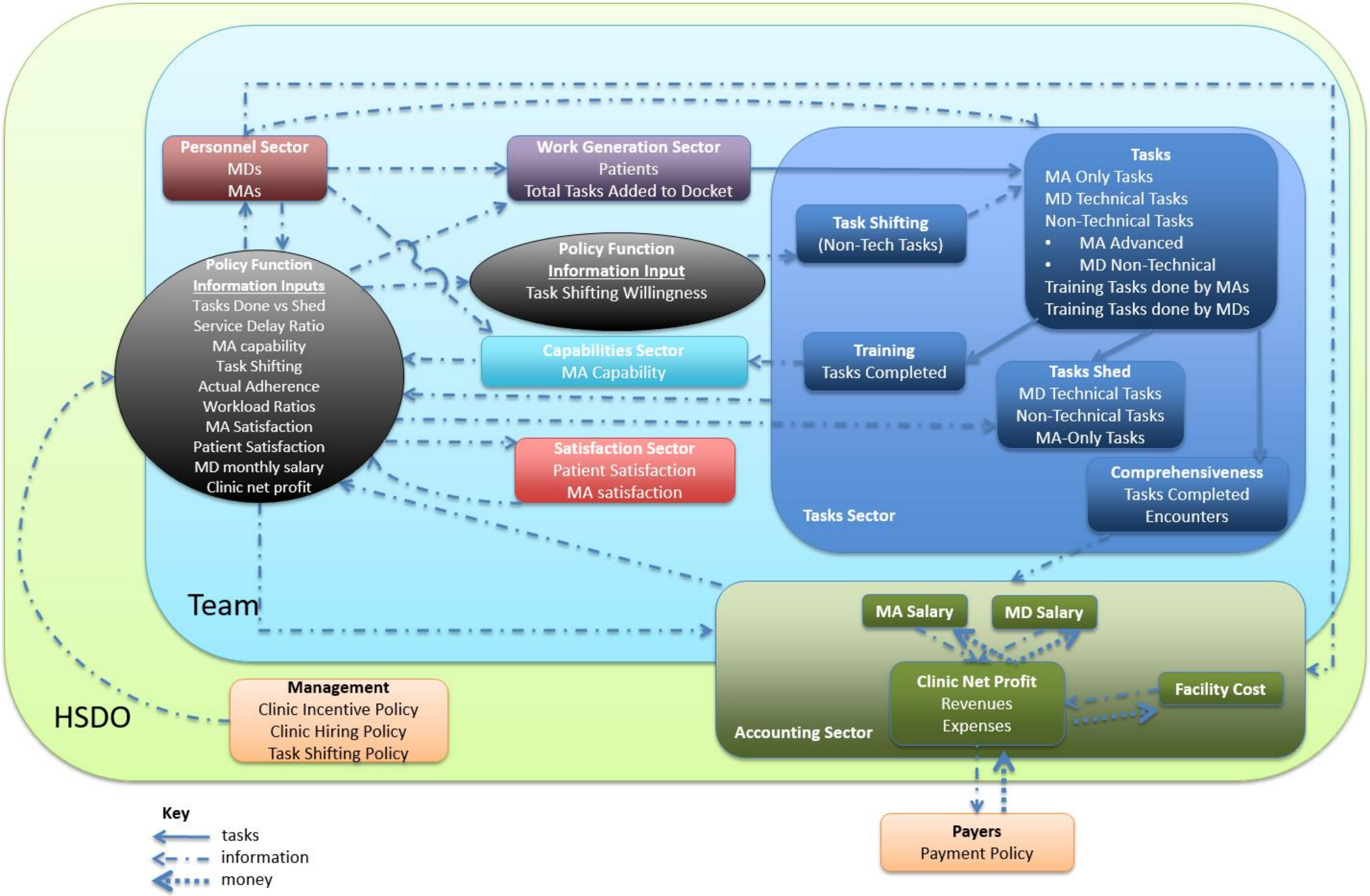
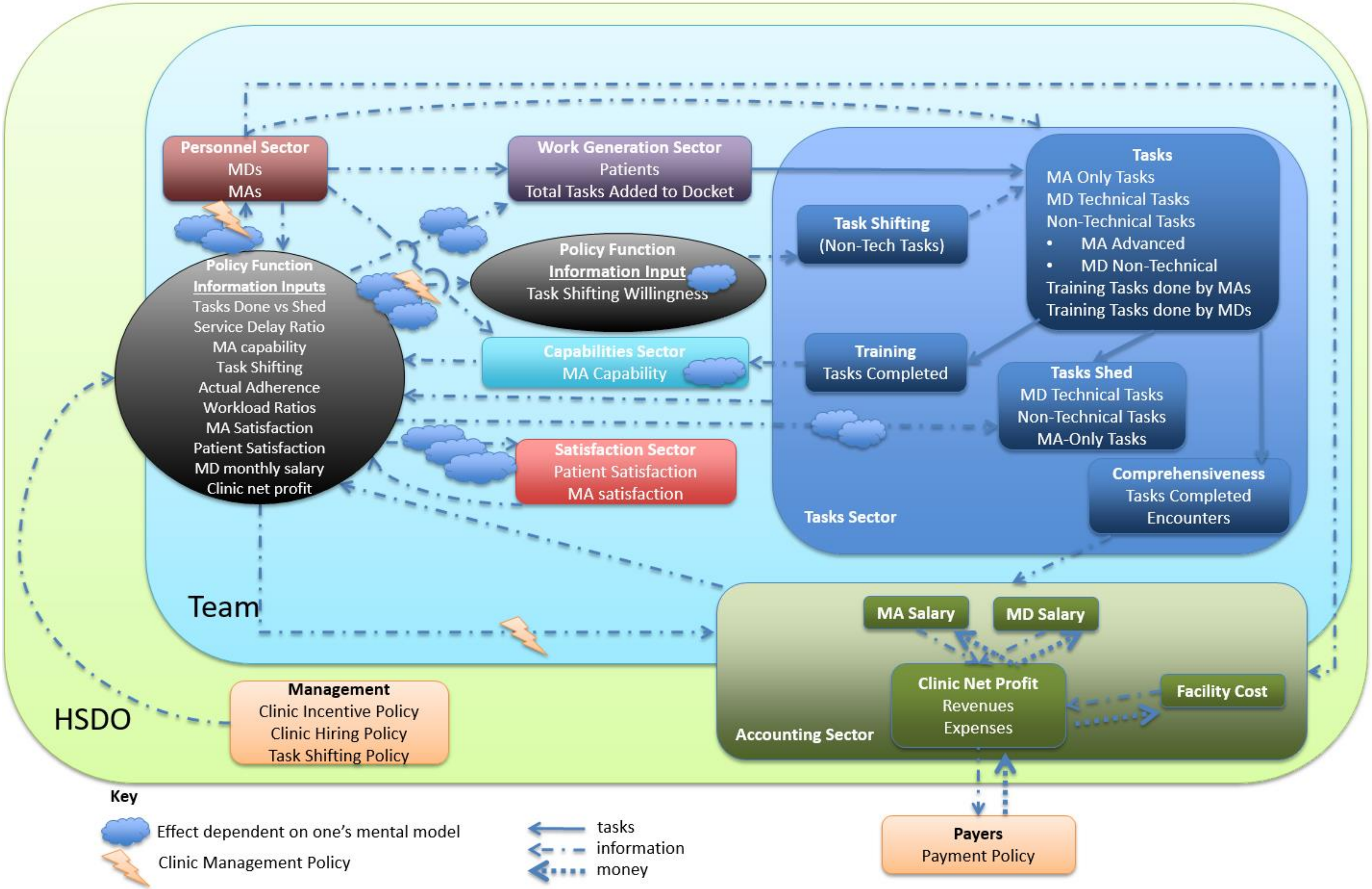


Figure 3.16 PCT Model Sector Map – where policies and preferences intervene



3.4.2 A WALK THROUGH THE SIMULATION MODEL

This section describes the model structure in detail across five model sectors, as follows:

- For each model sector, a summary of how it fits into the overall model is presented using **sector map(s)**[114] (p.113)[113] (p. 99-102)[112].
- Following this, the contents of the sector are described.
- **Policy structure diagrams**[114] (p. 162-163)[113] (p. 102) are used to communicate the way stakeholders make key decisions in the context of PCT.
- **Equations** are then provided for the stock and accompanying flows shown in the corresponding diagrams. Appendix F provides all model equations and **constants**.

The task-shifting decision (*Task-Shifting Willingness*) integrates information from several system sectors. While this is not a model sector, it is described in Section 3.4.2.6. Following this, calculated process and outcome measures in the model are identified in Section 3.4.2.7.

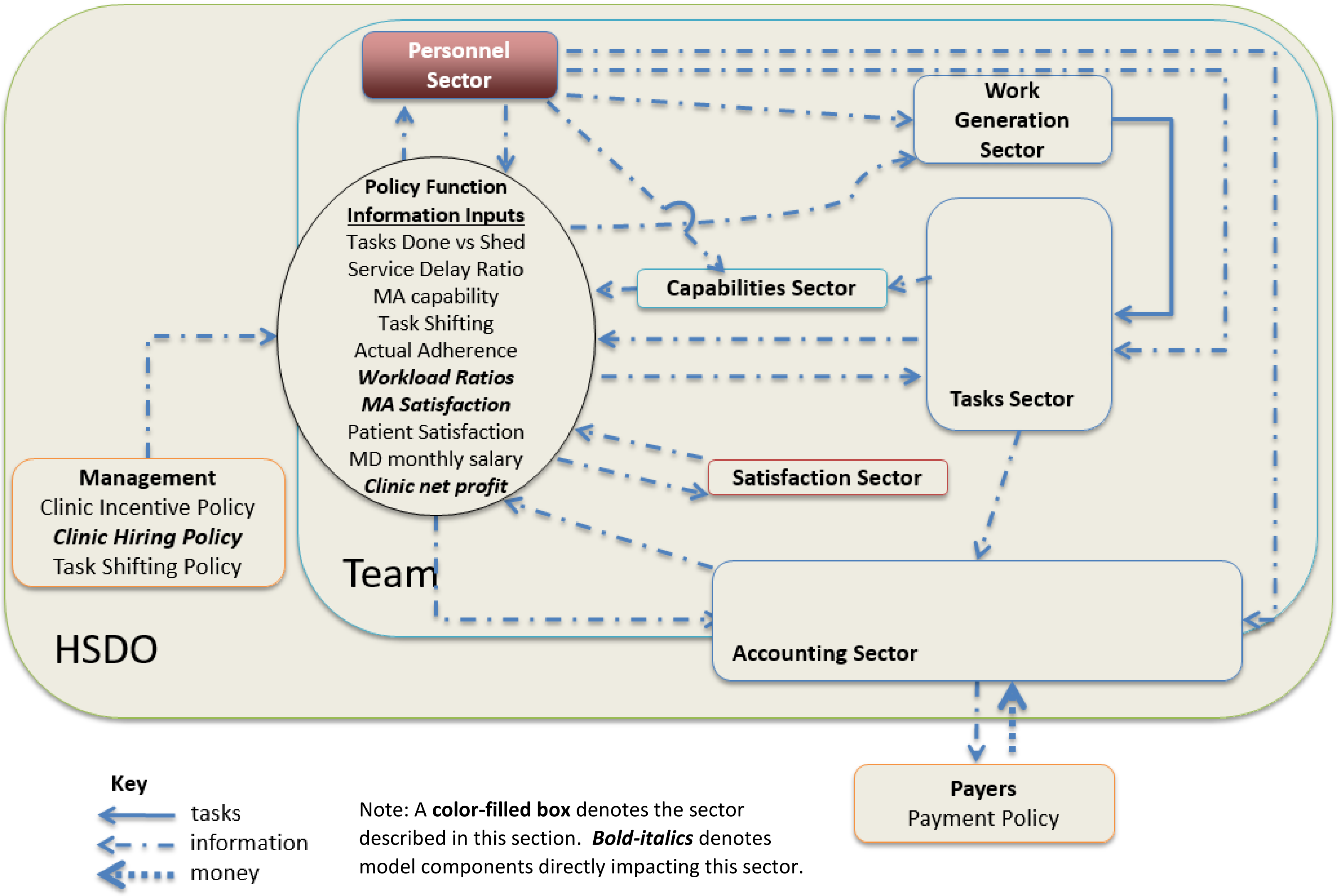
Where decisions involve a non-linear relationship between two variables, "**table functions**" are used. These decisions could also be represented using stock and flow relationships. Instead, the assumed nonlinear relationship between those two variables is represented via a table function. I consider the shape of the table function, for the purposes of this problem-specific model, to be static and to be a sufficient representation of the realism of this decision. These general assumptions apply to all table functions in this model.

Table functions are derived from the literature and/or interview descriptions and operational thinking. The latter process involved beginning with the s-shaped *universal table function*[197]. Table functions, like model constants, are also subjected to sensitivity analysis (see Section 3.5.5.1.9).

3.4.2.1 PERSONNEL SECTOR

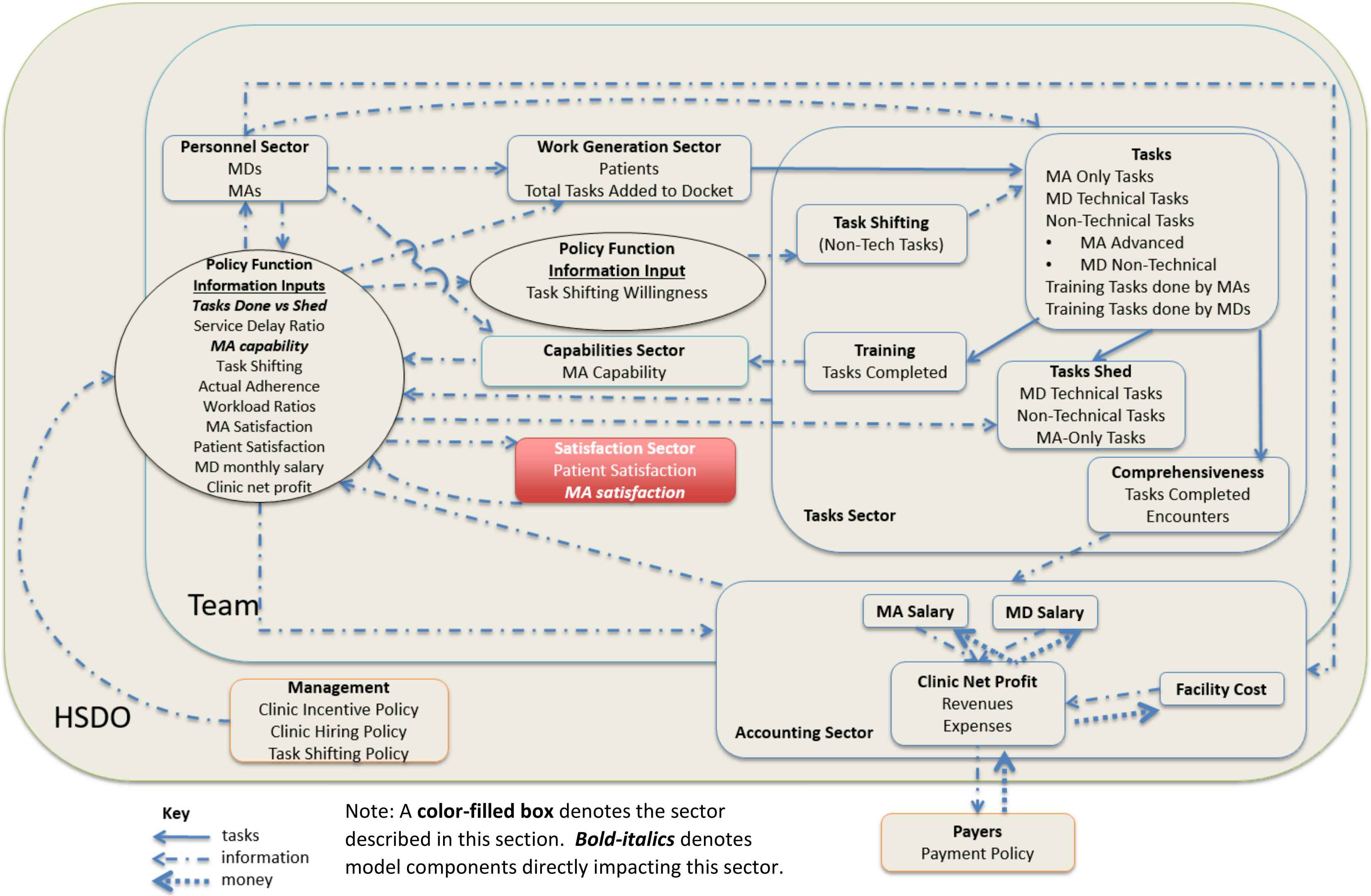
Figure 3.17 highlights the personnel sector on the sector map. The words in ***bold-italics*** identify the information inputs and policies directly relevant to this sector. Figure 3.18 highlights the satisfaction sector and related information inputs on the detailed sector map, since MA satisfaction informs MA turnover (and MA satisfaction is not described elsewhere). The personnel on the team are the clinicians and MAs.

Figure 3.17 Sector Map - Personnel Sector



Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Figure 3.18 Detailed Sector Map – MA Satisfaction



Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

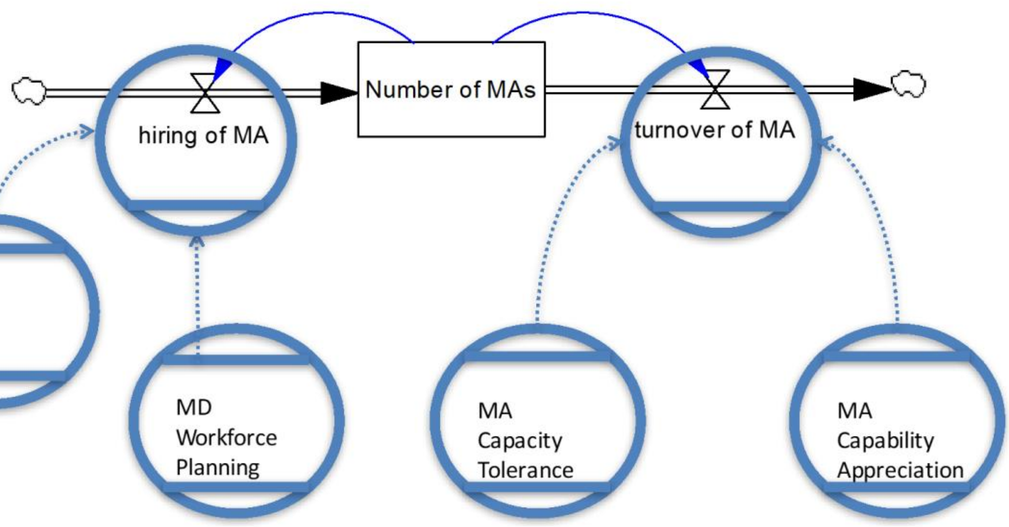
In a primary care clinic, clinicians lead the care team. They decide the pace of the work flow and the level of comprehensiveness delivered. Staff members are hired to support the work of clinicians on staff. For all of these reasons, clinicians are exogenous in the model. We assume that there are two clinicians on the care team – the smallest care team. In model variables and graphs clinicians are referred to as *MDs* (an artifact of the naming convention when the model was built).

MAs work with clinicians to facilitate patient care during the visit (e.g., rooming the patient), to deliver care (e.g., give shots, draw blood, do medicine reconciliation, administer clinical surveys, patient education), and to do out-of-visit work (e.g., following up on patient messages to the clinician, processing referrals, gathering patient reports from other sources)[33].

Under pre-PCT conditions, the ratio between MAs and clinicians is 1:1, therefore the initial number of MAs is 2, corresponding with the 2-clinician team in the model. That number changes with MA hiring and MA turnover (see Figure 3.19 and Box 3.13). Two decisions influence the inflow: the clinic hiring policy and clinicians' preference regarding when MA capacity is such that new hires are needed. Two decisions influence the outflow as well: MA perspectives regarding the level of capacity and capability experienced. MA firing is not a part of the model as it was not reported to be an important aspect of clinic operations and PCT.

[33] In the model, clinical staff members are referred to as MAs. This was initially the case in the clinics studied. However, during the transformation, not only did MAs grow in their capabilities (sometimes even attaining added certifications), but additional staff members were also added to the team (e.g., care managers with varying backgrounds including nursing, social work and pastoral care; clinical pharmacists; and transition navigators). For simplicity, the model refers to all of these individuals as MAs, but it is important to understand that as MAs grow both in number and capability in the model, they (in real life) grow in capability to a point, after which additional growth is attained by hiring people with the different job titles above.

Figure 3.19 Policy Structure Impacting the Number of MA



Box 3.1 The Level of MAs

Stock

Number of MAs (MAs) = $\int$ [hiring of MA - turnover of MA] *dt*

Initial value = 2

Flows

hiring of MA (MAs/Month) = MAX(MA workforce sought - Number of MAs) / adjustment time to hire MA,0)

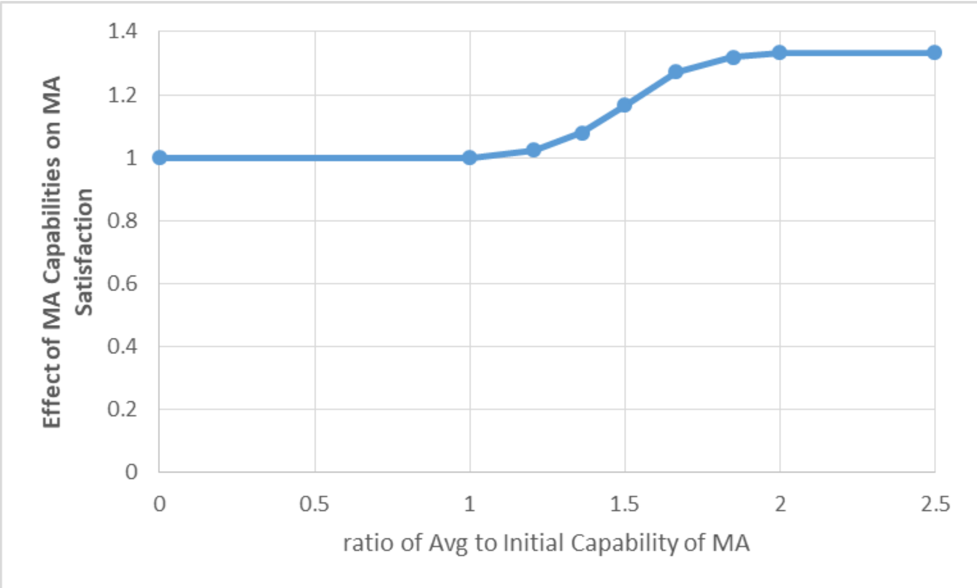
turnover of MA (MAs/Month) = (Number of MAs - MA workforce wanting to stay) / adjustment time for MA to leave

The rate of MA **turnover** is calculated as the difference between the number of MAs that currently want to stay in their job and the total number of MAs. As fewer MAs want to stay, turnover increases. The *MA Workforce wanting to stay* variable depends on their satisfaction, which depends on *MA Capability Appreciation* and *MA Capacity Tolerance*. Both influence satisfaction via *table functions*.

As this is the first time that a table function[197] is presented, I will briefly explain how it is used in the model. The key is found in the name of the variable. Of the influences on MA Satisfaction, a given change to MA Capabilities (x-axis, independent variable) has this effect (y-axis, dependent variable) on the overall MA Satisfaction – it is sometimes also referred to as a look-up table. At time t, the model takes the x-axis value and uses that to look up (find) the y-axis value corresponding to that x-axis value on the table function. This y-axis value is then used to impact the outcome (usually by multiplying the value this outcome would have otherwise by this y-axis value) (see Sterman[113] (p. 551-563).

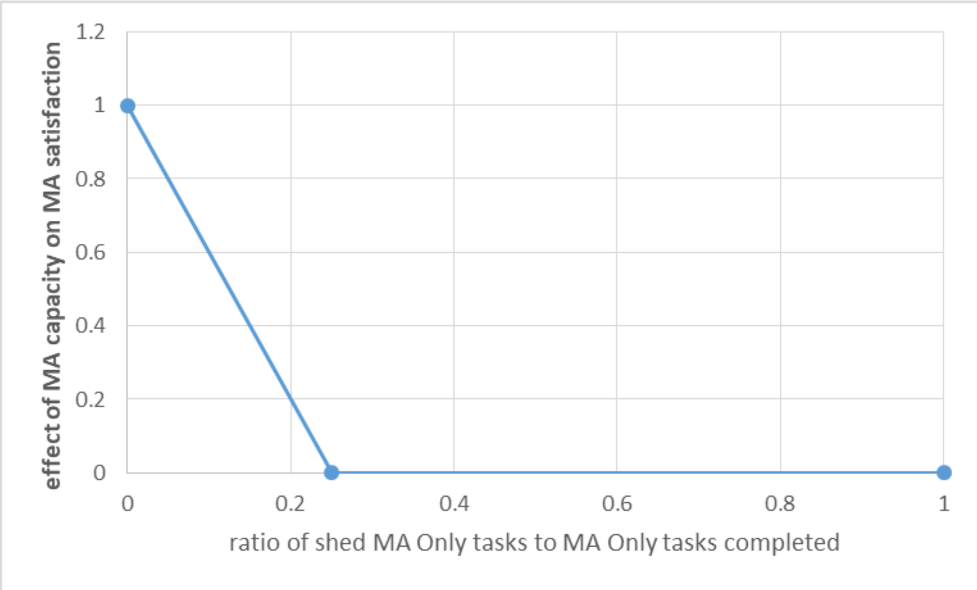
MA Capability Appreciation: MA satisfaction increases with increased capabilities (Figure 3.20). The information input is the ratio of current to initial MA capabilities. There are 50 capability units initially, and the maximum is 100 capability units. Thus the initial ratio is 1 (50 current units divided by 50 initial units). This ratio increases MA satisfaction if it is above the value under normal conditions (1,1) and decreases it when below. When this ratio is at its maximum, it is expected to induce the maximum satisfaction value (it does so at (2, 1.33)). S-shaped growth, the universal table function[197], is assumed between these values, where the midpoint is half-way in between (1.25, 1.1666).

Figure 3.20 Effect of MA Capabilities on MA Satisfaction



MA Capacity Tolerance: MA satisfaction increases with increased capacity (Figure 3.21). The information input is the ratio of *MA-only tasks not completed* divided by *MA-only tasks completed*. When MAs keep up with the work, satisfaction from capacity is maximized (0,1). As it gets harder to do so, their satisfaction decreases – I assume a linear decrease to zero when a quarter of their MA only tasks are being shed (i.e., not ever completed).

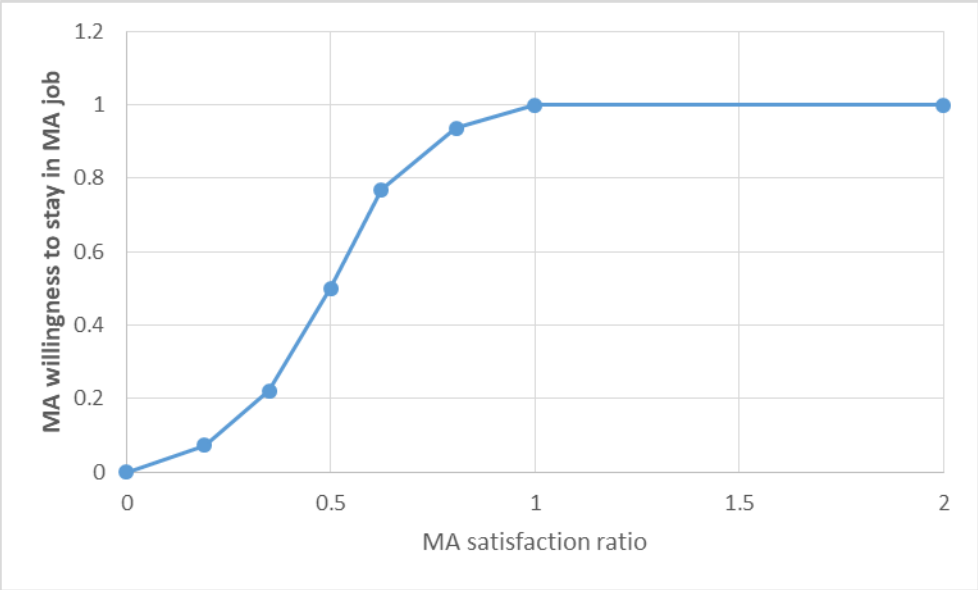
Figure 3.21 Effect of MA Capacity on MA Satisfaction



Both effects (Figure 3.20and Figure 3.21) are multiplied to each other and to the *MA satisfaction acceptable minimum* (beyond which the MA would choose to leave the clinic; 0.75). This value then modifies the overall MA satisfaction via a goal-gap formulation with a delay of 2 months (i.e., it takes 2 months for the MAs to adjust their level of satisfaction to the new conditions that they experience) – this value is their *perception of MA Satisfaction*. It influences turnover via a table function describing how their relative level of satisfaction (relative to acceptable minimum) impacts their desire to stay.

MA willingness to stay in MA job: MA willingness to stay increases with increased MA satisfaction (Figure 3.22). The information input is the *MA satisfaction ratio*; the *perception of MA Satisfaction* divided by *MA satisfaction acceptable minimum*, where the max ratio is 1. This ratio sustains MA willingness to stay when it is above the value under normal conditions (1,1) and decreases it when below. When this ratio is at its minimum, it is expected to induce the minimum value for MA willingness to stay (it does so at (0,0)). S-shaped growth is assumed between these values, where the midpoint is half-way in between (0.5,0.5).

Figure 3.22 Effect of MA Satisfaction Ratio on MA Willingness to stay in Job

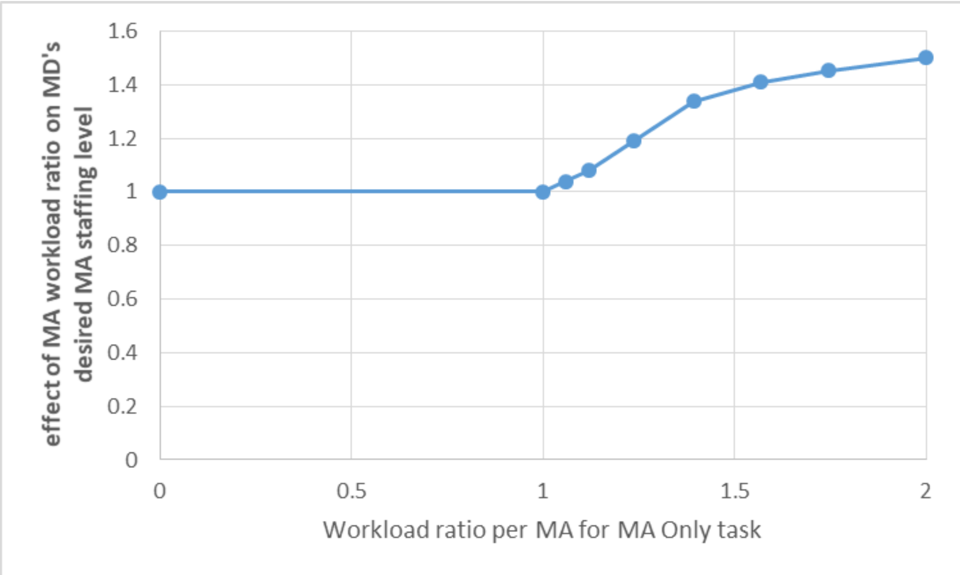


The rate of MA **hiring** is calculated as the difference between the number of MAs that are sought and the total number of MAs. As more MAs are sought, hiring increases. However, MA hiring only happens as long as the clinic manager is willing to hire them. If the clinic manager decides not to hire even though MAs are needed, then the *MA workforce sought* is 0. Without the MAX function in Box 3.1 above, this would result in *hiring of MA* being negative (incorrect). Instead, if the clinic decides not to hire even when MAs are needed, then the *MA workforce sought* is 0, causing *hiring of MA* to be 0 (correct).

The *MA Workforce sought* variable depends upon *MD Workforce Planning* and *Clinic Hiring Policy*. The former is based on how many MAs the MD feels are needed in order to keep up with the work. The latter is based on whether the clinic manager feels they have sufficient funds to hire the additional MAs requested.

MD Workforce Planning: Clinicians seek to increase MA workforce in response to strained capacity (Figure 3.23). The information input is the *workload ratio per MA for MA only tasks* (the actual divided by the normal workload for MA-only tasks). As long as the MA-only tasks are at or below the total number of tasks that the MA can normally handle, there is no need to hire more MAs – the current number of MAs suffices (0,1) (1,1). When there are more MA-only tasks than can be done in the normal load of work, clinicians advocate for hiring more MAs. The model assumes that MDs feel they cannot advocate for hiring more than 1.5 times the current MA workforce (2,1.5); where initially they are hesitant but as the need increases they request a higher level of hiring and then ultimately hesitate again, more this time, given the difficulty of hiring close to the max. This hesitation is shown by S-shaped growth.

Figure 3.23 Effect of MA workload Ratio (MA-only Tasks) on MD's Desired MA Staffing Level

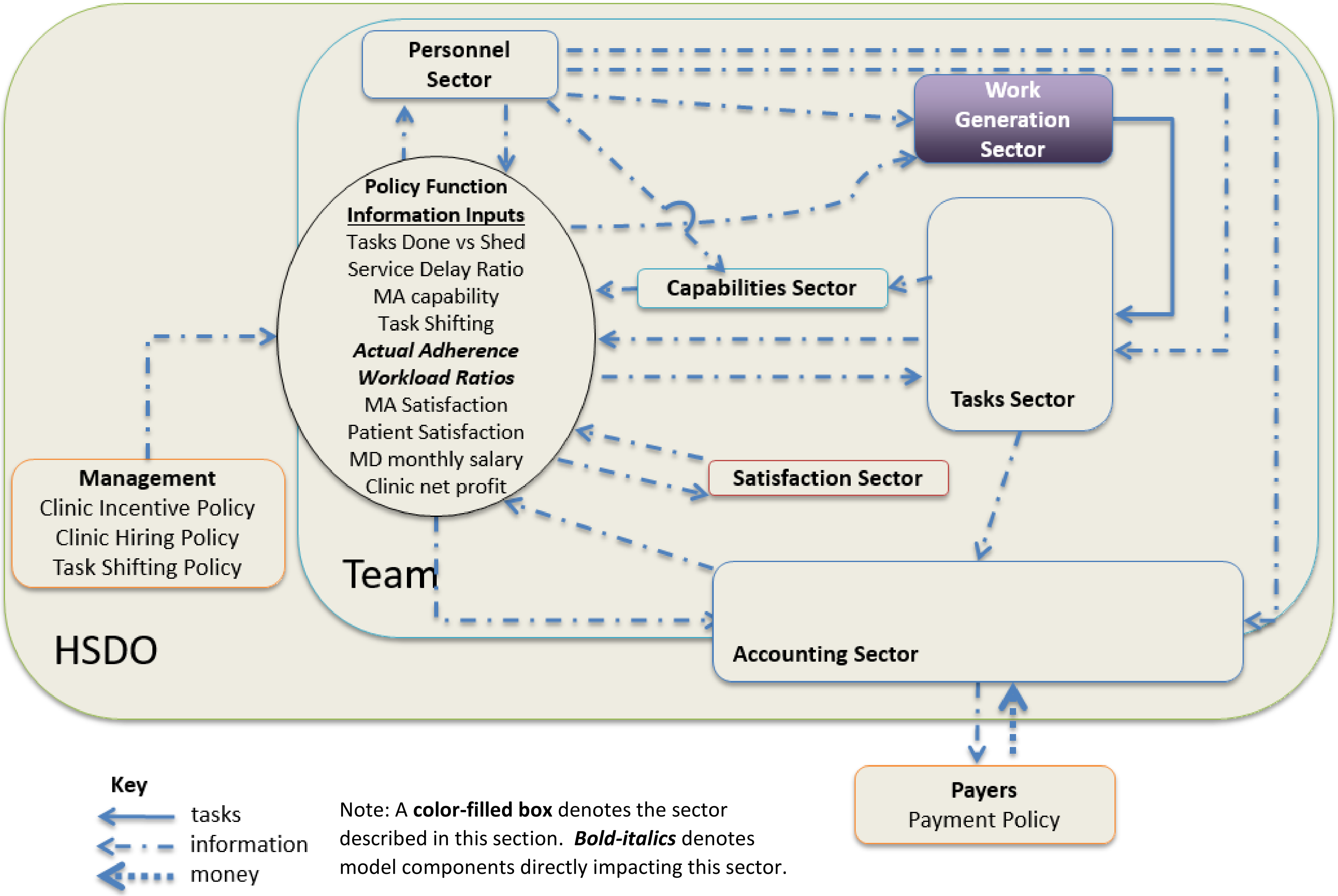


Clinic Hiring Policy: This policy determines whether the MAs that clinicians request are hired when needed (on) or only when the clinic net profit is at or above zero (off).

3.4.2.2 WORK GENERATION SECTOR

Figure 3.23 highlights the work generation sector on the sector map. Patients are "empaneled" – that is, each patient is assigned a particular clinician or team responsible for overseeing their care. The number of patients in the model is calculated based upon the number of MDs and the average patient panel size per clinician.

Figure 3.24 Sector Map – Work Generation Sector



Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Work is generated as patients seek care from this team. The total number of *potential* MD tasks per month (100,000 technical and non-technical tasks) is calculated based upon the number of patients and the average number of visits per patient per month, as well as a conversion factor of 100 MD tasks per visit. The total number of potential MA tasks per month (100,000 MA-only tasks) is calculated similarly, with a conversion factor of 100 MA-only tasks per visit. Completing all potential patient tasks means a team completes 200,000 tasks per month.

The clinicians determine the proportion of these tasks that are added to the docket – the proportion of the 200,000 tasks that they intend for the team to complete. A month's worth of work equates to 25,000 tasks per clinician (12,500 technical and 12,500 non-technical tasks) and another 25,000 tasks per MA (25,000 MA-only or non-technical tasks).

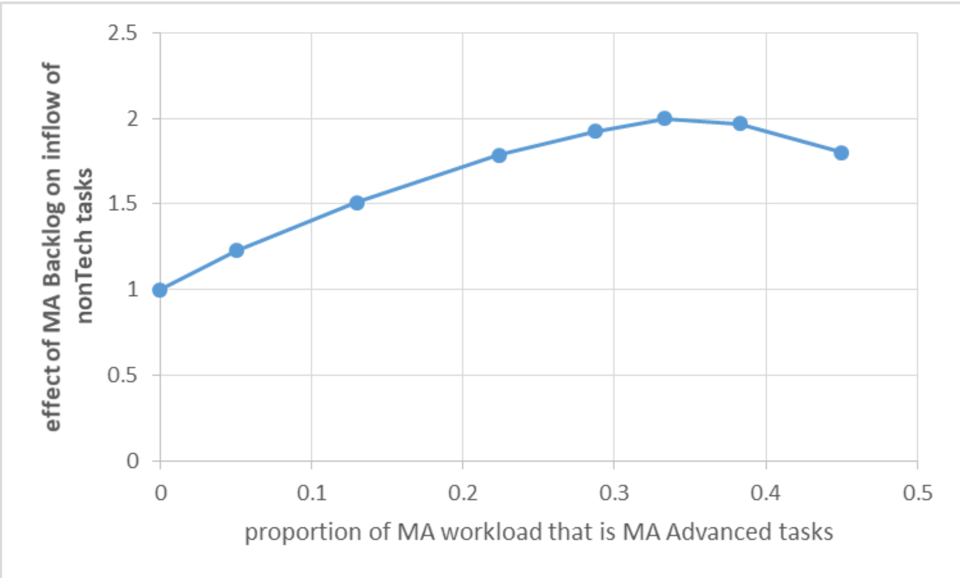
Under initial conditions, the size of the docket equates to the team completing 50% of the potential MD tasks (where half of them are technical at 25,000 tasks and the remaining half are non-technical at 25,000 tasks), and 50% of the potential MA-only tasks (50,000 MA-only tasks). Each month, clinicians add these 50,000 MD tasks and 50,000 MA-only tasks to the docket and the team completes them.

Clinicians manage the size of the docket (the rate of task inflow) to ensure that the team is keeping up. Clinicians monitor two workload ratios in deciding when to trigger a change: (1) for technical tasks, "proportion of MD workload that is technical" and (2) for non-technical tasks, "proportion of MA workload that is MA-advanced tasks". The first ratio begins at 50%, the second ratio begins at 0%. An increase in either ratio signals to clinicians that the team is ready to improve their comprehensiveness or "actual adherence" to clinical guidelines. As the team demonstrates the ability to complete these tasks without getting behind on everything else, then the clinicians allow the size of the docket to increase. For each of these two types of tasks, when one task is added to the docket it also adds one MA-only task to the docket.

To reach full adherence, tasks must be added to the docket. However, under normal conditions, clinicians and MAs are at capacity, and additional tasks cannot be added until some of that work is shifted, specifically, from clinicians to (additional) MAs. Thus, with a small task-shifting policy kickstart at month 10, MAs are given a small amount of the non-technical tasks on the clinicians' docket. This allows clinicians to see that MAs are starting to do them (on the outflow side) and that MAs have capacity to do more MA-Advanced tasks (i.e., "proportion of MA workload that is MA-Advanced tasks"). This can only happen successfully as more MAs are hired since the existing MAs are already at capacity. As this proportion is small, the MD decides to 'sign the team up' to do more Non-Technical Tasks. In so doing, he is increasing the fraction of Non-Technical tasks, and thus the “actual number of Non-Technical tasks” that can flow in to the MA’s stock of MA-Advanced tasks to be completed.

Comprehensiveness (MA-advanced tasks): Clinicians add non-technical tasks to the team’s docket with an increase in the corresponding workload ratio (Figure 3.25). The information input is the “proportion of MA workload that is MA-advanced tasks”, starting at 0%, with a goal of having 33% of MA tasks being non-technical ones. At this point, clinicians add all non-technical tasks to the docket (twice the initial fraction; (0.33, 2)). Between these two points, a weak exponential approach to the value 2 is assumed – clinicians are eager to become more comprehensive at first and gradually become more concerned with the ability of MAs to keep up, the more the workload is taken over by MA-advanced tasks. Beyond 0.33, clinicians pull back to 0.33 because passing it means that MAs are less able to perform their MA-only tasks.

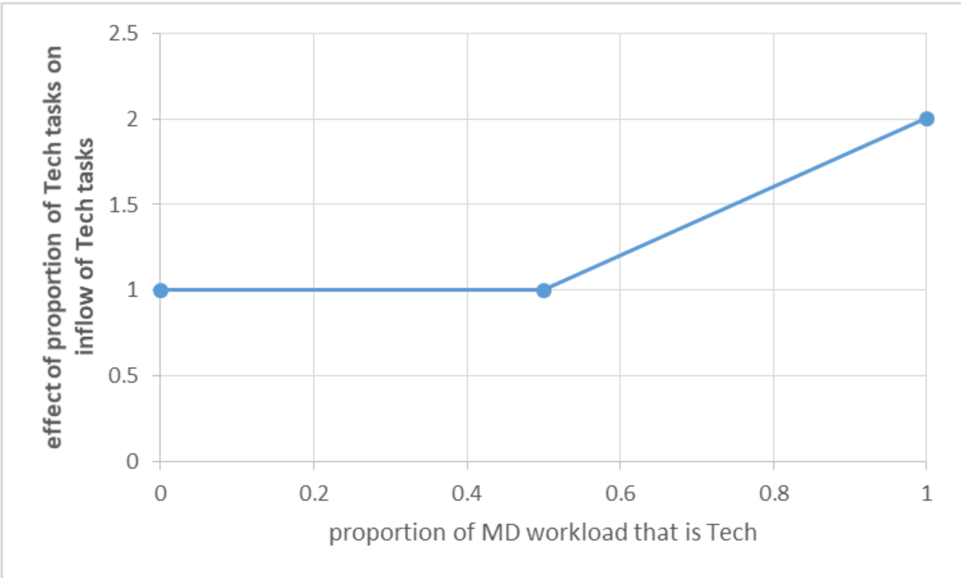
Figure 3.25 Effect of MA Backlog on inflow of nonTech Tasks



Similarly, for MD Technical Tasks – the effect table depends upon the “proportion of MD workload that is technical” because this is the trigger to adding more tasks for the MD.

Comprehensiveness (Technical Tasks): Clinicians add technical tasks to the team's docket with an increase in the corresponding workload ratio (Figure 3.26). The information input is the "proportion of MD workload that is technical", starting at 0%, with a goal of having 100% of clinician tasks being non-technical ones. Clinicians do not add technical tasks to the docket unless technical tasks make up 50% or more of their workload ((0,1) up to (0.5, 1)). They don't take any away since that would mean less than average care and malpractice. Beyond this point, clinicians add technical tasks to the docket until their workload consists entirely of technical tasks and they have added all possible technical tasks to the docket (1,2). A straight-line increase is used.

Figure 3.26 Effect of proportion of Tech tasks on inflow of Tech tasks



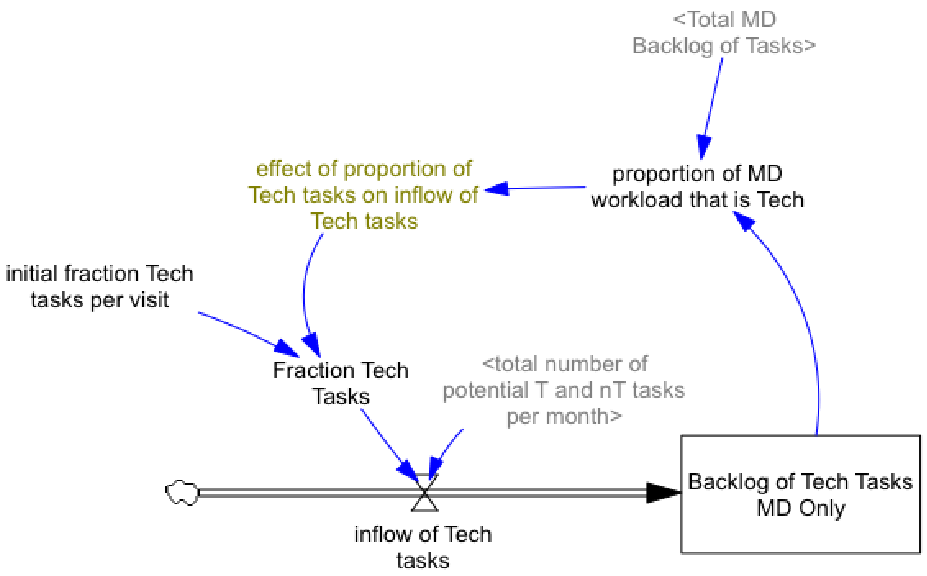
Under full task-shifting, 100% of the potential tasks are added to the docket and completed. Each month, clinicians add 50,000 technical tasks, 50,000 non-technical tasks and 100,000 MA-only tasks to the docket. The two clinicians complete the technical tasks (a full load of 50,000 tasks for 2 clinicians) and MAs complete the non-technical tasks as well as the associated MA-only tasks (a full load of 150,000 tasks for 6 MAs). The number of tasks per person remains 25,000 tasks.

Box 3.2 and Figure 3.27 below point out the structure for changing the inflow of Technical Tasks. A similar structure is used for the inflow of non-technical tasks (the difference is the addition of MD willingness as a modifier determining whether these tasks go to clinicians or MAs).

Box 3.2 Inflow of Technical Tasks

Flow
inflow of Tech tasks (Tasks / Month) = total number of potential T and nT tasks per month*Fraction Tech Tasks
Auxiliary Variable
Fraction Tech Tasks (dimensionless) = initial fraction Tech tasks per visit*effect of proportion of Tech tasks on inflow of Tech tasks

Figure 3.27 Structure for Modifying Technical Tasks



3.4.2.3 TASKS SECTOR

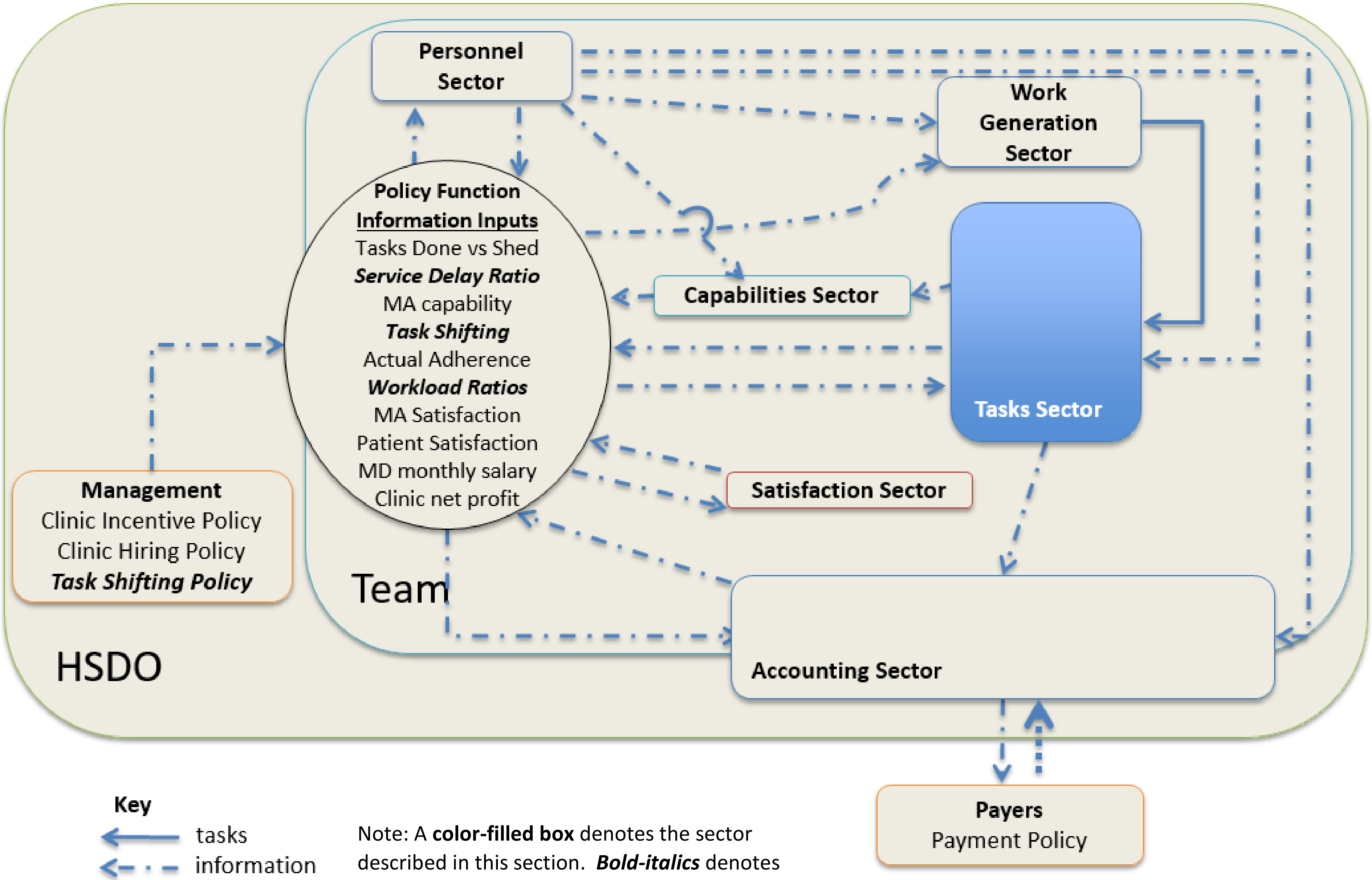
Figure 3.27 and Figure 3.29 below highlight the tasks sector on the simple and more-detailed sector maps respectively. Tasks generated in the work generation sector flow into this sector. Clinicians' decision[34] regarding task-shifting determines how many non-technical tasks they shift to MAs. This decision also generates training tasks for both clinicians and MAs. In this sector, tasks are completed, and (sometimes) tasks are shed (i.e., never completed). There are two penalties for task shedding: 1) patient satisfaction decreases since patients are not receiving what they have come to expect[35] and 2) patient health may decrease[36].

[34] This decision is discussed in detail in Section 3.4.2.6 Task-shifting Willingness.

[35] See Section 3.4.2.6.3 Patient Satisfaction Ratio

[36] Patient health depends on how task shedding impacts overall adherence to clinical guidelines. It would only be worse if shedding caused less than 50% adherence (the initial value).

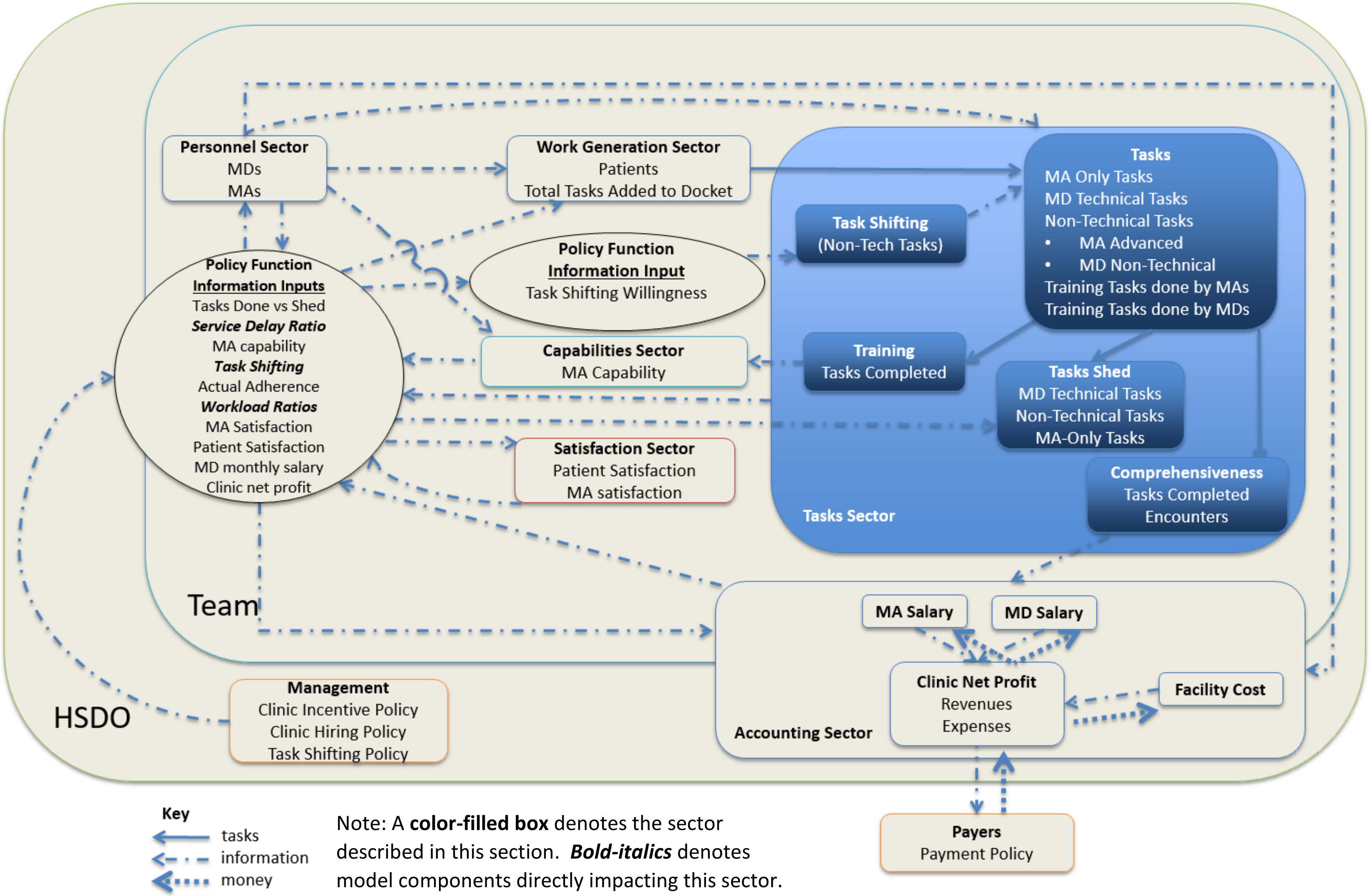
Figure 3.28 Sector Map – Tasks Sector



Key
- tasks
- information
- money

Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Figure 3.29 Detailed Sector Map – Tasks Sector



Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Comprehensiveness desired or "MD willingness to adhere to clinical guidelines" is equal to the sum of two proportions: (1) technical task added to docket over technical tasks possible and (2) non-technical tasks added to docket over non-technical tasks possible. The maximum value is 1, indicating that clinicians are allowing all technical and non-technical tasks onto the docket. The value under normal conditions is 0.5 (0.25 fraction Technical tasks and 0.25 fraction Non-Technical tasks).

Comprehensiveness delivered or "actual adherence to clinical guidelines" is equal to the sum of three proportions: (1) technical task completed over total possible, (2) non-technical tasks completed over total possible (whether completed by clinician or clinical staff member), and (3) MA-only tasks completed over total possible. The maximum value is 1, indicating that all the care needed to be comprehensive is being delivered. The value under normal conditions is 0.5 (0.125 fraction Technical tasks, 0.125 fraction Non-Technical tasks and 0.25 fraction MA-only tasks).

Initially, when the team is delivering 50% of the tasks patients need, there are 800 patient encounters per month. Under 100% comprehensiveness (with 4 additional MAs and full task-shifting), there are 1600 patient encounters per month. The length of encounters does not change (20 minutes each). Encounters are made up of in-visit patient care tasks (they do not include training tasks or out-of-visit work).

Box 3.3 below presents the equations associated adherence to clinical guidelines.

Box 3.3 Adherence to Clinical Guidelines

Comprehensiveness Desired

MD willingness to adhere to clinical guidelines (dimensionless) = Fraction nonTech Tasks + Fraction Tech Tasks

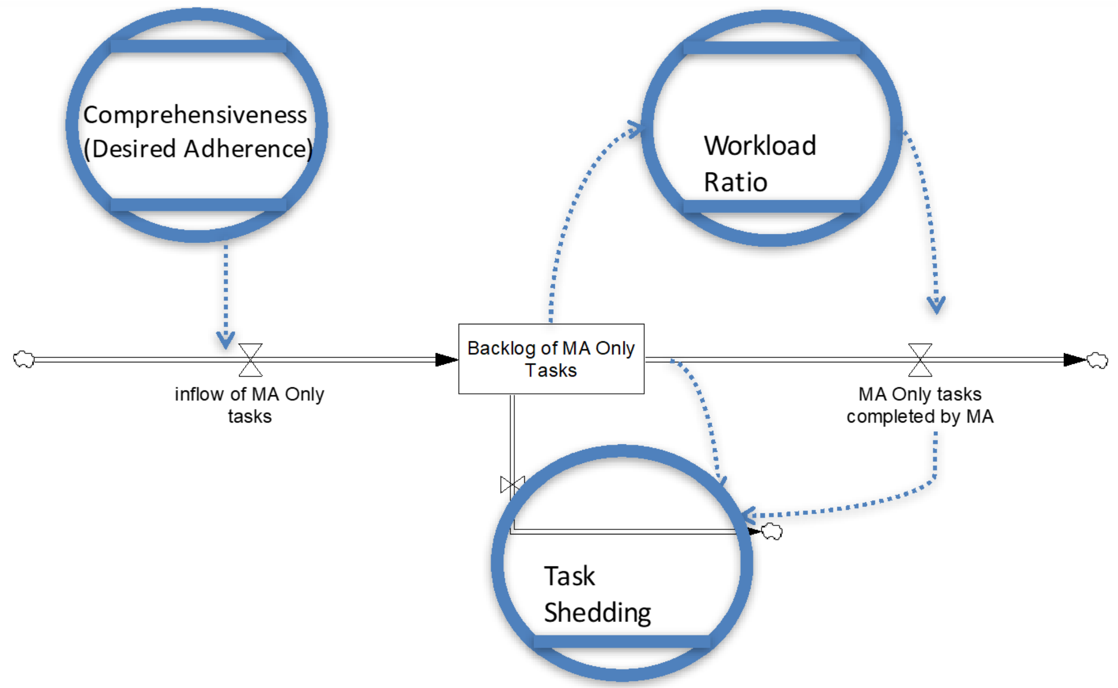
Comprehensiveness delivered

actual adherence to clinical guidelines (dimensionless) = total wRVU tasks completed / total number of potential wRVU tasks per month

3.4.2.3.1 MA-ONLY TASKS

Figure 3.30 below presents the operating policies influencing MA-only tasks. As mentioned just above, for both technical and non-technical tasks, when one of these tasks is added to the docket it also adds one MA-only task to the docket, thus "inflow of MA-only tasks" is determined by clinicians' desired comprehensiveness. There are two outflows: (1) tasks completed and (2) tasks that are shed (never completed). This structure for tasks completed is used for the other task types as well. Task shedding operates in the same way for MA-only and MA-advanced tasks; however, it operates differently for clinician shedding of tasks.

Figure 3.30 Policy Structure Impacting Backlog of MA Only Tasks

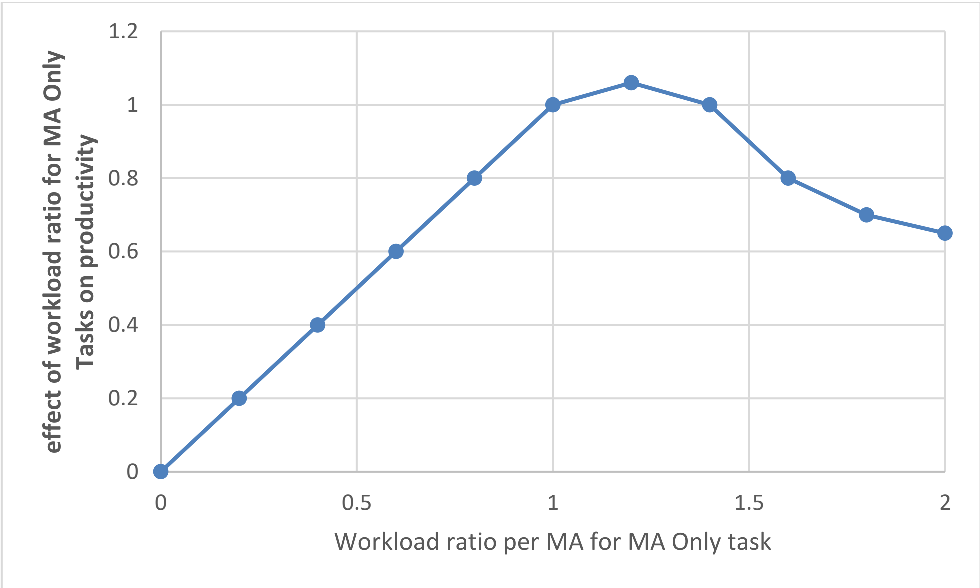


Task shedding is influenced by the desire to remain on schedule – when the actual completion rate is longer than the desired completion rate, the additional tasks are shed. MAs need to shed these tasks to keep up with clinicians on their team. MA-only tasks are shed only when MAs are behind schedule; thus, a MAX function is used.

The normal workload is three months' worth of tasks. A table function is used to express how MAs react to the level of work they have.

Workload Ratio (MA-only tasks): When the backlog is significantly higher than the normal workload for this task (beyond (1.4,1) MAs are stressed) or lower (below (1,1) MAs are bored), tasks are completed more slowly (lower productivity) (Figure 3.31). Before (1,1), we assume a linear increase from (0,0) – as they become busier, they become more productive. Between (1,1) and (1.4,1), MAs are a bit more stressed such that they are slightly more productive at these tasks. As the ratio continues to increase, the productivity falls (more slowly via exponential approach). This is because the MAs are experiencing a high level of stress as they are more overworked. This table function shape is based on previous work on productivity[113, 198] (p.577-578,582)[199].

Figure 3.31 Effect of Workload Ratio for MA-Only Tasks on Productivity



Box 3.4 below presents the equations associated with this stock.

Box 3.4 The Level of MA Only Tasks

Stock

Backlog of MA Only Tasks (tasks) = ∫ [inflow of MA Only tasks - MA Only tasks completed by MA - MA Only tasks not completed] *dt*

Initial value = 150,000

Flows

inflow of MA Only tasks (tasks/Month) = total number of potential MA Only tasks per month*MD adherence to clinical guidelines

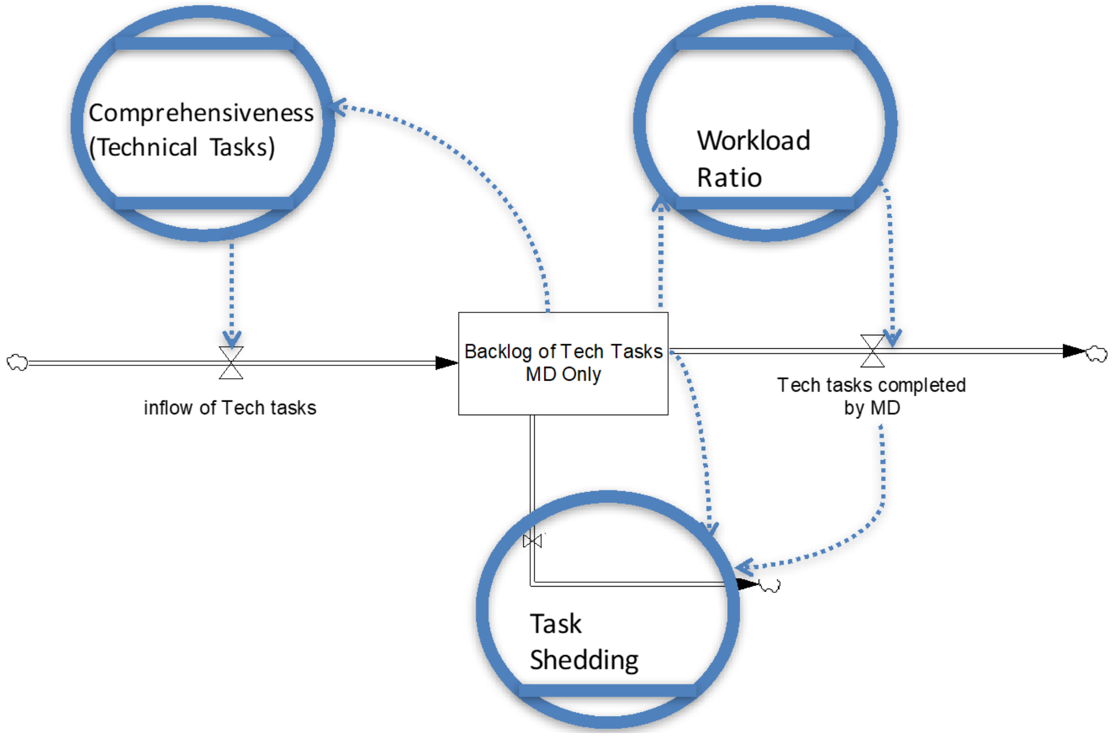
MA Only tasks completed by MA (tasks/Month) = Number of MAs*productivity per MA for MA Only Tasks

MA Only tasks not completed (tasks/Month) = MAX(desired MA completion rate of MAonly tasks-MA Only tasks completed by MA,0)

3.4.2.3.2 MD TECHNICAL TASKS

The operating policies influencing MD Technical tasks and associated equations (Figure 3.32 and Box 3.5) are similar to those operating for MA-only tasks.

Figure 3.32 Policy Structure Impacting the Level of MD Technical Tasks



Box 3.5 The Level of MD Technical Tasks

Stock **Backlog of Tech Tasks MD Only** (tasks) = ∫ [inflow of Tech tasks - shedding Tech Tasks - Tech tasks completed by MD] *dt* Initial value = 75,000 **Flows** **inflow of Tech tasks** (tasks/Month) = total number of potential MD tasks per month*Fraction Tech Tasks **Tech tasks completed by MD** (tasks/Month) = Number of MDs*Productivity per MD for Tech Tasks **shedding Tech Tasks** (tasks/Month) = effect of time to complete backlog Tech tasks on shedding*Backlog of Tech Tasks MD Only

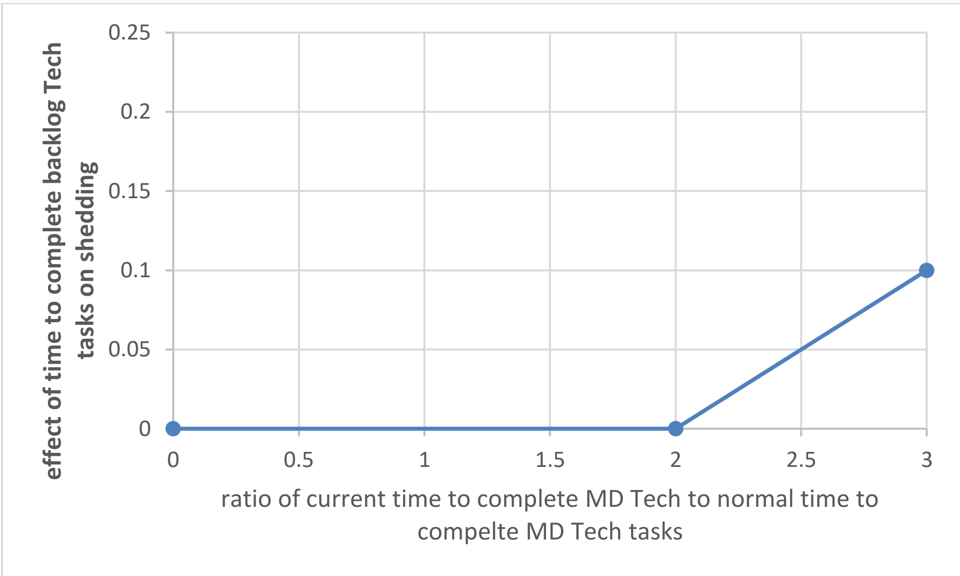
Technical Tasks are generated when clinicians see patients. The number of Technical tasks generated depends on their desired comprehensiveness.

Differences between the stock of MA-only tasks and that of Technical tasks is that, for clinicians: (1) the inflow is only influenced by the desired comprehensiveness specific to Technical Tasks and (2) the shedding policy involves a table function. Clinicians also shed tasks when they are behind; however, they determine how many tasks to shed differently from

MAs. As the leaders of the team, they determine what service delivery delay is acceptable and how to deal with it when the delay is beyond that point. A table function represents clinicians' choice regarding when to shed tasks. This same table function is also assumed for non-technical tasks performed by clinicians.

Task Shedding (for clinicians): When clinicians are within the normal three-month delivery delay, they do not shed clinician tasks (thus (0,0) and (1,0) (Figure 3.33). Shedding occurs only when clinicians are beyond the service deliver delay value that they deem acceptable. The amount of shedding increases beyond this point. The table function assumes that clinicians are willing to accept a delay twice that of the normal delay (2,0). Beyond this point they shed some tasks. When they are three times the normal delay, they shed 10% of their backlog of technical tasks. When they are four times the normal delay, they shed 20% of their backlog of technical tasks.

Figure 3.33 Effect of Time to Complete Backlog Tech Tasks on Shedding

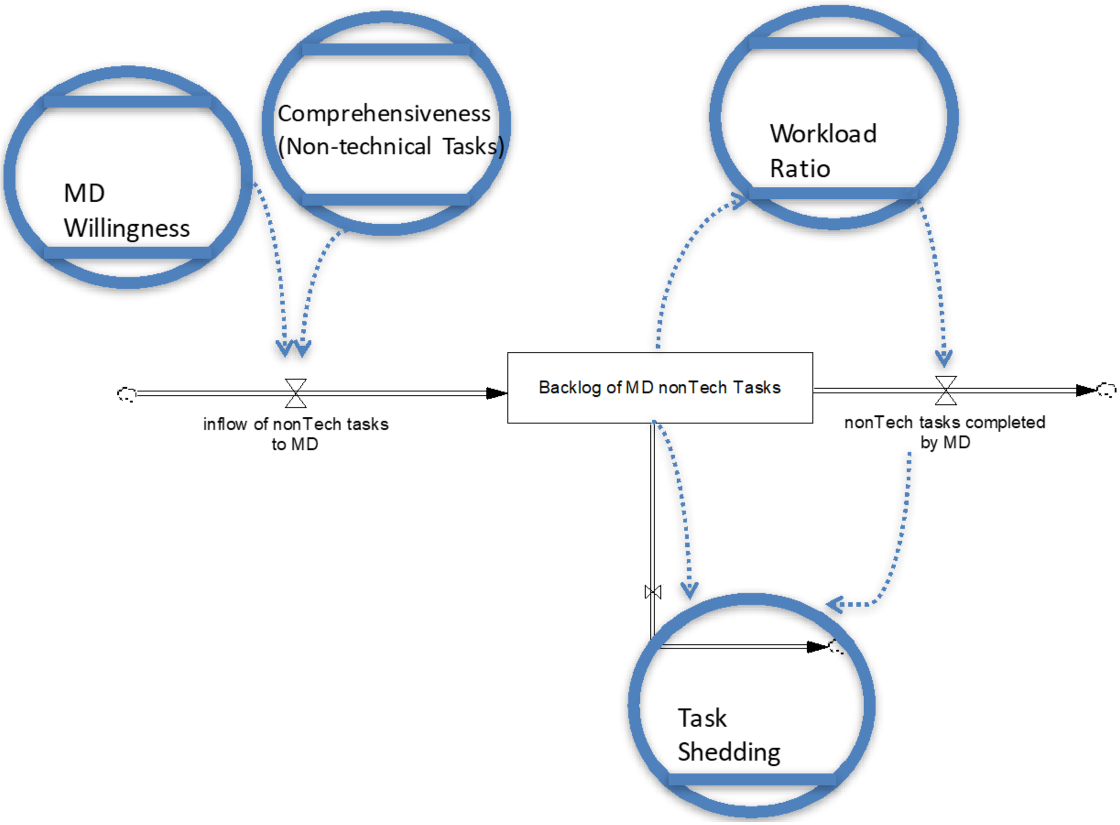


3.4.2.3.3 MD NON-TECHNICAL TASKS

The operating policies influencing non-technical tasks to be performed by clinicians[37] (Figure 3.34 and Box 3.6) are similar to those described just above. Non-Technical Tasks are generated when clinicians see patients. The number of Non-technical tasks generated depends on the desired comprehensiveness of clinicians. Then, from those tasks placed on the docket, a portion (between 0 and 100%) flow to the MA depending on the clinicians' willingness to shift those tasks (Section 3.4.2.6). Once on the MA's docket, these tasks are now called "MA-advanced tasks". Also, when behind, a portion of MD Non-technical tasks are shed (using the same shape table function as that for Technical tasks, not shown).

Differences are that, here, the inflow is influenced by the desired comprehensiveness specific to non-technical tasks. Also, when MD willingness to shift tasks is above zero, the inflow is reduced by the proportion of tasks that are shifted to MAs.

Figure 3.34 Policy Structure Impacting the Level of MD Non-technical Tasks



---

[37] When these tasks are performed by MAs, they are called *MA-advanced tasks*.

Box 3.6 The Level of MD Non-technical Tasks

Stock

Backlog of MD nonTech Tasks (tasks) = ∫ [inflow of nonTech tasks to MD-nonTech tasks completed by MD-shedding nT tasks] *dt*

Initial value = 75,000

Flows

inflow of nonTech tasks to MD (tasks/Month) = (1-willingness to Task Shift)/max willingness*actual number of nonTech tasks per month

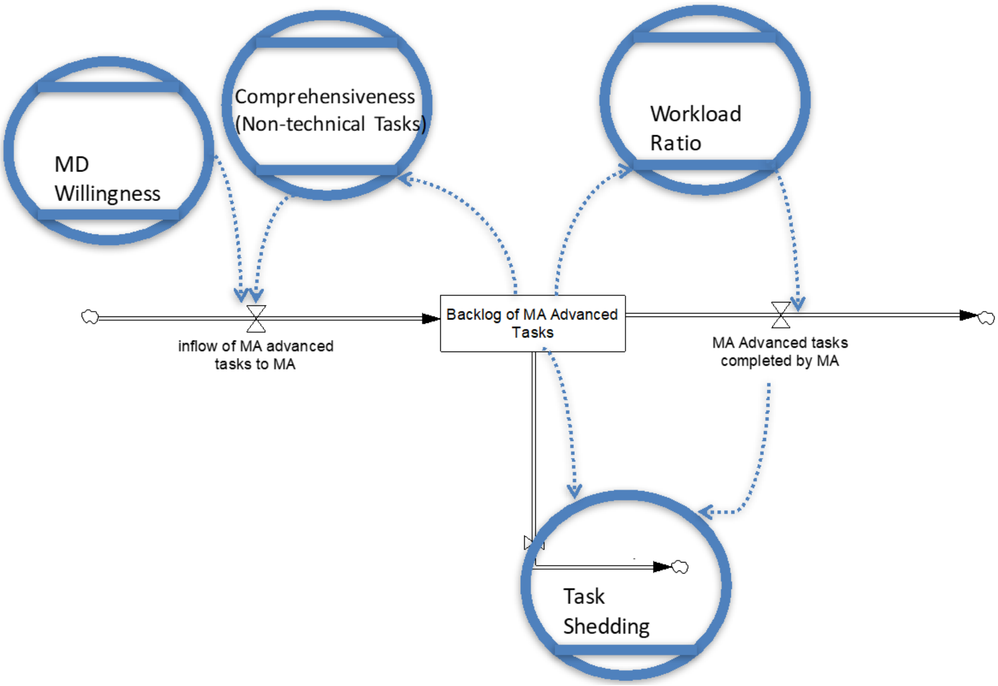
nonTech tasks completed by MD (tasks/Month) = Number of MDs*productivity for nonTech Tasks

shedding nT tasks (tasks/Month) = effect of MD time to complete nT tasks ratio on shedding of those tasks*Backlog of MD nonTech Tasks

3.4.2.3.4 MA-ADVANCED TASKS

The operating policies influencing MA-advanced tasks (the non-technical tasks to be performed by MAs) are similar to the three policies described for the non-technical tasks performed by clinicians (Figure 3.35; Box 3.7)). The difference is that, when MD willingness to shift tasks is above zero, the inflow is *increased* by the proportion of tasks that are shifted to MAs. Also, as mentioned for MA-only tasks, MAs shed tasks when their current task completion rate is lower than the desired completion rate. In other words, there is no table function through which they interpret and adjust how much to shed like there is for clinicians. While clinicians have the authority to use their judgement in making this choice, MAs do not – they can only choose to never complete tasks in order to keep up with the team's desired pace of work.

Figure 3.35 Policy Structure Impacting the Level of MA-advanced Tasks



Box 3.7 The Level of MA-Advanced Tasks

Stock
Backlog of MA Advanced Tasks (tasks) = ∫ [inflow of MA advanced tasks to MA - MA Advanced tasks completed by MA - MA Advanced tasks not completed] *dt*
Initial value = 0
Flows
inflow of MA advanced tasks to MA (tasks/Month) = actual number of nonTech tasks per month*willingness to Task Shift / max willingness
MA Advanced tasks completed by MA (tasks/Month) = productivity per MA for MA Advanced Tasks*Number of MAs
MA Advanced tasks not completed (tasks/Month) = MAX(desired MA completion rate for MAadvanced tasks - MA Advanced tasks completed by MA, 0)

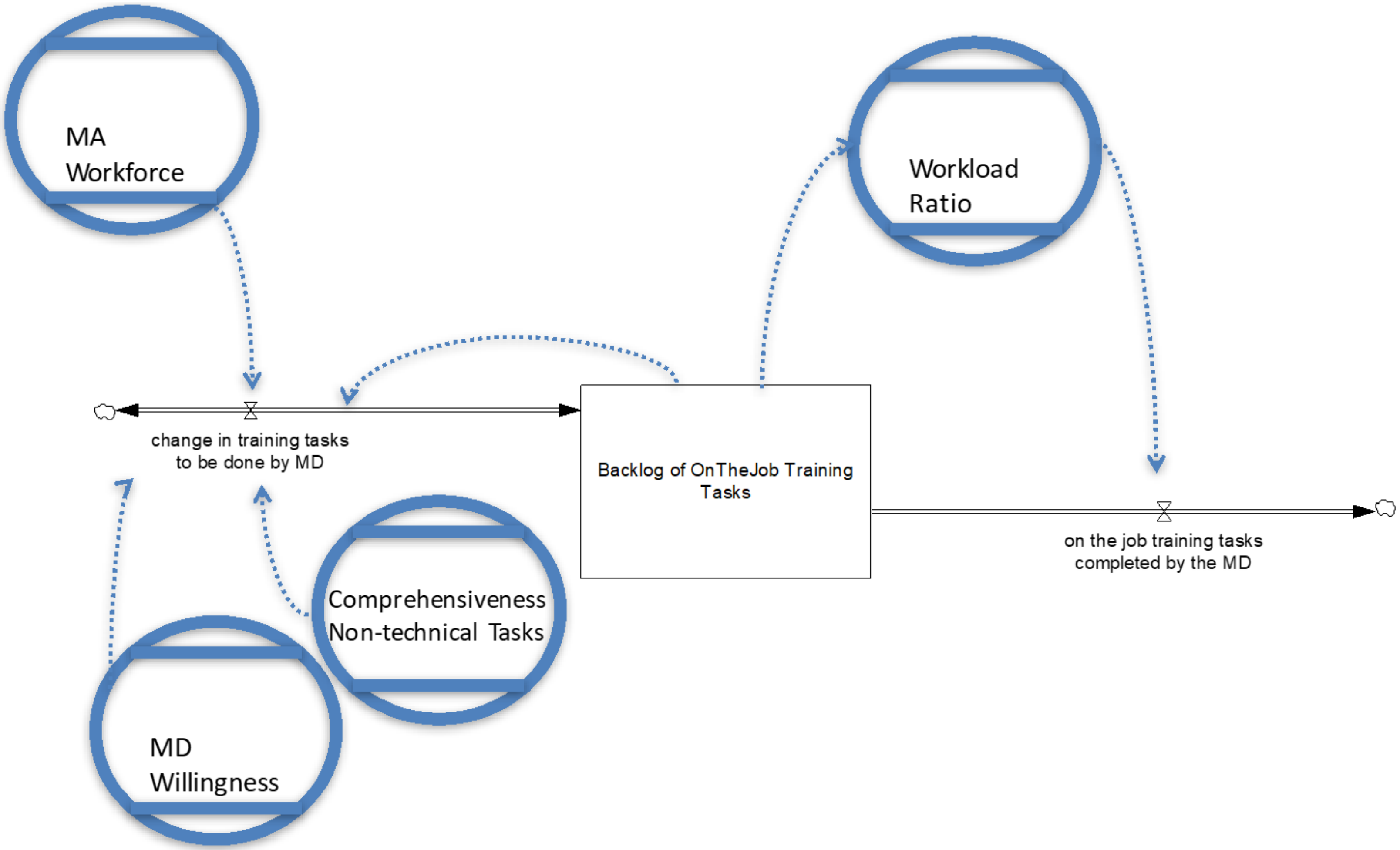
3.4.2.3.5 TRAINING TASKS

MA-only Tasks do not require on-the-job training, while MA-advanced tasks do. This training prepares the clinical staff member to perform MA-advanced tasks as a member of the care team. This training involves problem solving as a team, learning how to perform new tasks, as well as learning how that team specifically prefers the task done. When MA-advanced tasks can only be performed legally by a licensed professional who is not an MA, there would be a new clinical staff member of that profession (e.g., a PharmD) joining the team for a fraction of their time. These are all called MAs in the model as (in the model) “MA” refers to all clinical staff members.

MA training is influenced by hiring as well as turnover. Training for those newly-joining the team is added to the backlog. Training for those who have left is removed from the backlog. MA training is completed as the team performs these tasks, using the same Workload Ratio structure described for the tasks above. When clinicians choose to increase or decrease task-shifting, the training associated with that change is added to/removed from the backlog.

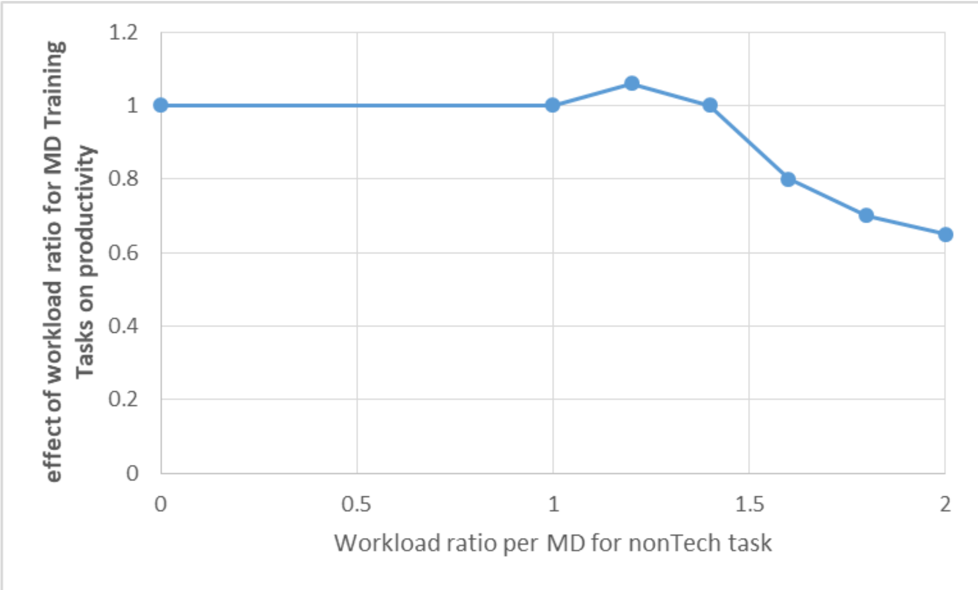
Clinicians spend time training MAs (therefore they have a backlog of training tasks). MAs are also spending time being trained by clinicians (therefore they also have a mirrored backlog of training tasks). Figure 3.36 presents the policy structure impacting the backlog of training tasks for clinicians (note that the same structure impacts training tasks for MAs).

Figure 3.36 Policy Structure Impacting Backlog of Training Tasks (MD)



Workload Ration (training tasks): There is a slight modification for the Workload Ratio policy for completing training tasks (Figure 3.37). The assumption changes such that for all values below 1, the effect on productivity is 1. In other words, whereas with their regular tasks clinicians and MAs are assumed to be less productive if they have less to do, with training tasks, they are 100% productive even if they only have a few to do. This is because they are assumed to be engaged, actively involved in wanting to and completing the training as soon as possible so that task-shifting can be successful. After the (1,1) point, the effect table shape remains the same as for other tasks because being overburdened is assumed to have the same effect regardless of the type of task.

Figure 3.37 Effect of workload ratio for MD Training Tasks on productivity



Box 3.8 below presents the equations associated with this stock. Note the first flow equation matches the outflow in Figure 3.37 above, the other flow equations below are all collapsed in the figure into the “change in training tasks to be done by MD” flow. In the model, they are separate flows. The first two are co-flows which relate to the Personnel Sector (i.e., MA Workforce). The second two relate to changes in task-shifting.

Box 3.8 The Level of Training Tasks (MD)

Stock

Backlog of OnTheJob Training Tasks (tasks) = ∫ [change in training needed from MD for existing staff of MAs - on the job training tasks completed by the MD - Training tasks shed due to MA turnover + training needed due to MAnewhire for MD to do] *dt*

Initial value = 0

Flows

on the job training tasks completed by the MD (tasks/Month) = Number of MDs*productivity of training tasks per MD

training needed due to MAnewhire for MD to do (tasks/Month) = Fraction nonTech Tasks shifted to MA*training per unit fraction change per MA*hiring of MA

Training tasks shed due to MA turnover (tasks/Month) = average backlog of training tasks per MA*turnover of MA

inflow of training for existing staff due to upshifting of tasks to be done by MD (tasks/Month) = Number of MAs*training per unit fraction change per MA*MAX(change in Fraction nonTech Tasks shifted to MA, 0)

outflow of training for existing staff due to downshifting of tasks to be done by MD (tasks/Month) = MIN(change in Fraction nonTech Tasks shifted to MA, 0)*Backlog of OnTheJob Training Tasks for MD

Auxiliary Variables

change in Fraction nonTech Tasks shifted to MA (Dimensionless / Month) = (Fraction nonTech Tasks shifted to MA - Cumulative Fraction of nonTech Tasks shifted to MA) / time to change fraction of nT tasks shifted to MA

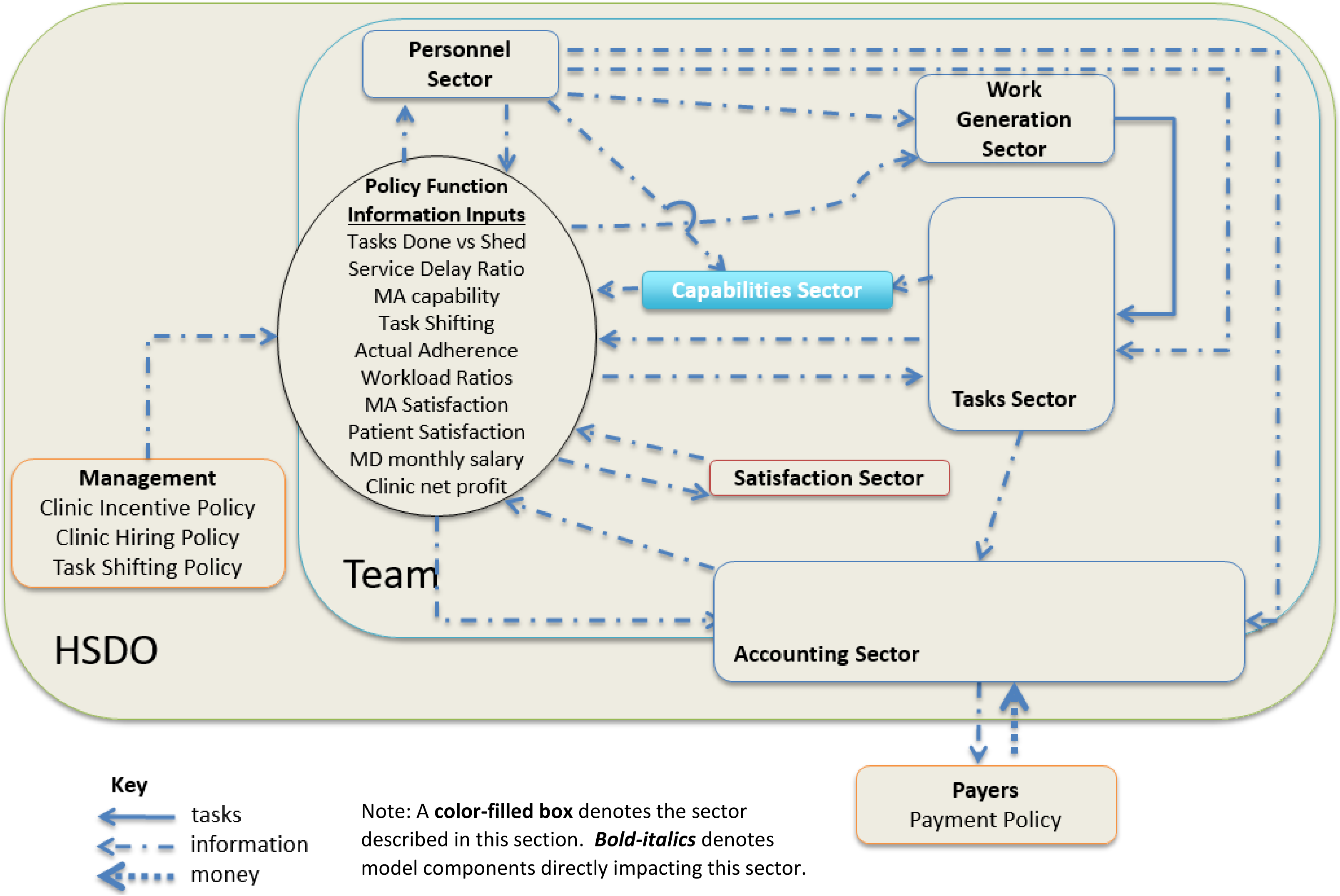
Constant

training per unit fraction change per MA (tasks/MAs) = 50000

3.4.2.4 MA CAPABILITIES SECTOR

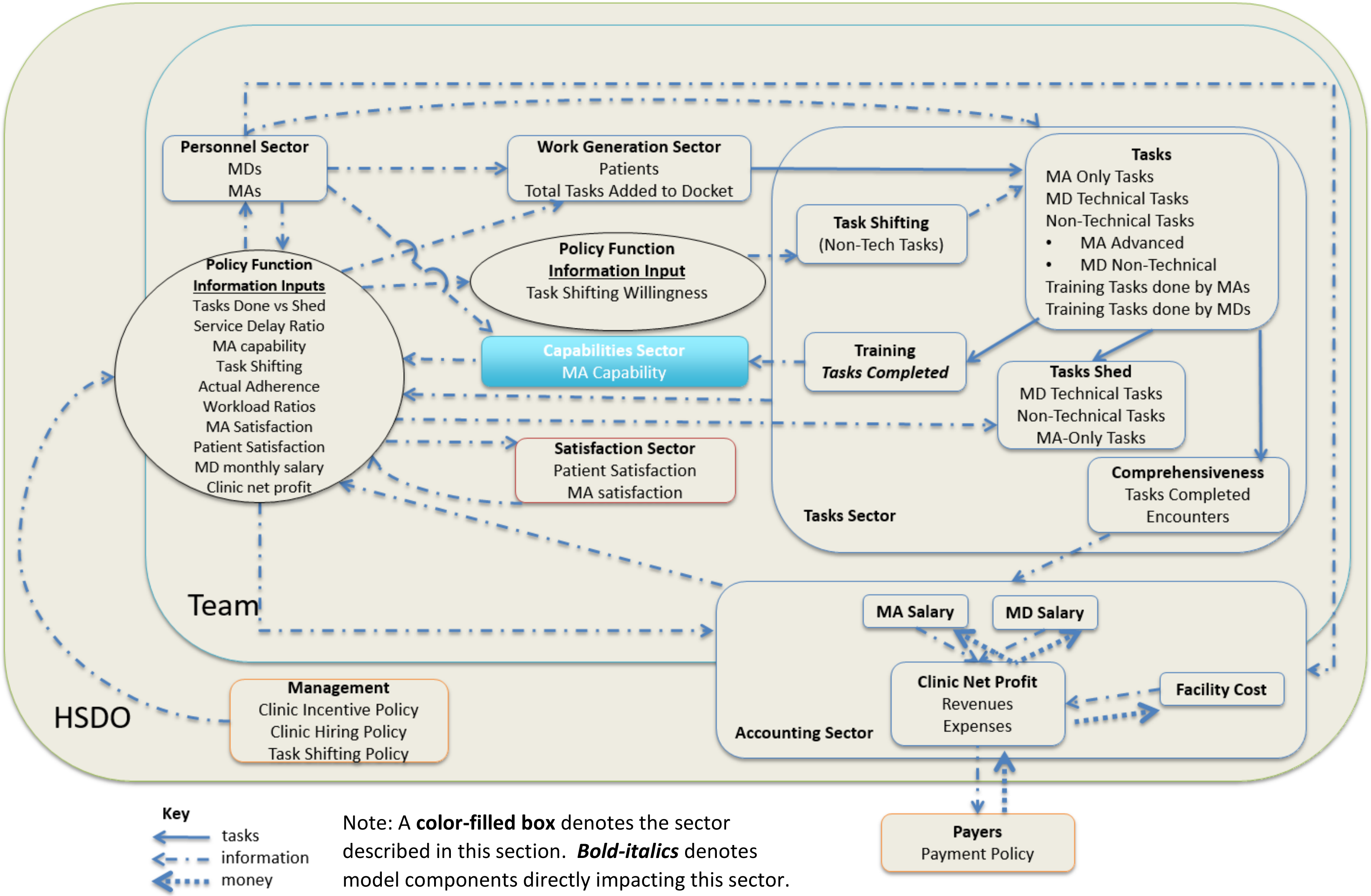
Figure 3.38 and Figure 3.39 highlight the MA Capabilities sector on the sector map. There are no bold/italics items since capabilities are gained from the training tasks that are completed in the tasks sector. Decisions influence MA capabilities indirectly through assignment and completion of training tasks (described in the previous section). Once completed, the on-the-job training tasks are translated into capabilities. The overall level of MA capabilities for the team is also influenced by changes in the MA workforce. Figure 3.40 below presents the operating policies influencing MA Capabilities: MA workforce, speed of uptake and capabilities learning curve. Box 3.9 presents the associated equations.

Figure 3.38 Sector Map – MA Capabilities Sector



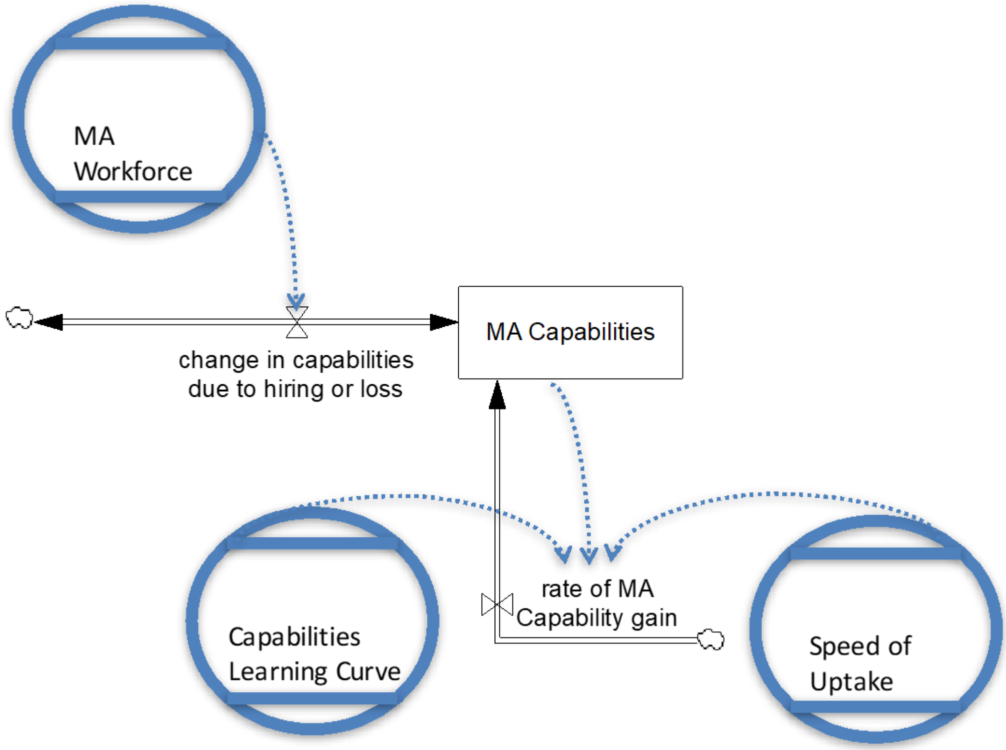
Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Figure 3.39 Detailed Sector Map – Capabilities



Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Figure 3.40 Policy Structure Impacting MA Capabilities



Box 3.9 The Level of MA Capabilities

Stock **MA Capabilities** (Capabilities) = ∫ [new hire capabilities + rate of MA Capability gain - turnover capab loss] *dt* Initial value = initial capability of MA new hire * Number of MAs **Flows** **new hire capabilities** (Capabilities / Month) = initial capability of MA new hire*hiring of MA **rate of MA Capability gain** (Capabilities / Month) = (on the job training completedMA / training tasks needed to gain capability)*effect of MA capab ratio on change in MA capab **turnover capab loss** (Capabilities / Month) = average capability per MA*turnover of MA

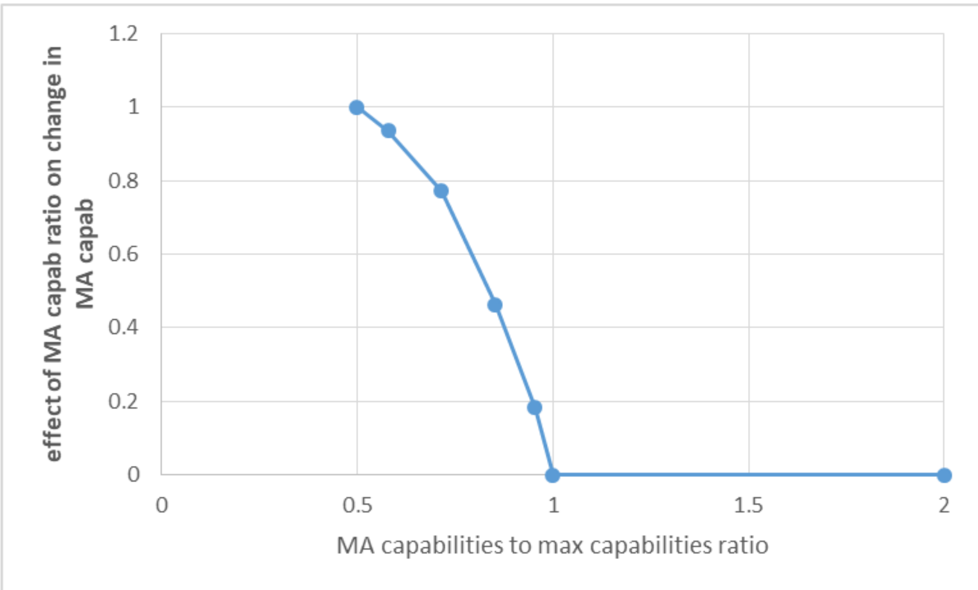
Under normal conditions, MAs have 50 capability units which correspond to the 50% of the potential MA-only tasks being performed. When new MAs are hired, they too come in with 50 capability units. When they leave, they take all their capability units with them.

Under task-shifting, as MAs are given MA-advanced tasks, on-the-job training is also being performed. As training is completed, MAs are learning and thus becoming more capable at performing these tasks. Under 100% task-shifting, once the MAs have completed all the training tasks, they are 100% capable (reaching the max 100 capability units/MA). Any new-hire MAs come in with the standard 50 capability units and training tasks are triggered for them so that they too gain in capability to match the current level of task-shifting.

This capability gain is calculated by converting the training tasks completed into units of capability. This conversion is modified by the speed of uptake (learning). In the model, this is set to a constant value – "training tasks needed to gain capability". A second modifier is the "Capabilities Learning Curve" which represents the growing difficulty of gaining new capabilities at higher levels of already-gained capabilities. Initially, easier tasks will be shifted (needing fewer training tasks). As task-shifting continues, harder tasks will be shifted (needing more training tasks). The conversion factor between training tasks completed and capabilities gained is a constant. This modifier adjusts the inflow of capabilities due to training tasks completed to account for the difference in difficulty of tasks shifted over time.

Capabilities Learning Curve: Initially, MAs have 50% of the total capabilities possible, and training translates to new capabilities at the fullest amount possible (0.5,1) (Figure 3.41). As MAs approach 100% of the total capabilities possible, it takes more training to gain each additional unit of capability. Finally, as MAs reach 100% of the total capabilities possible, training no longer increases capability (1,0).

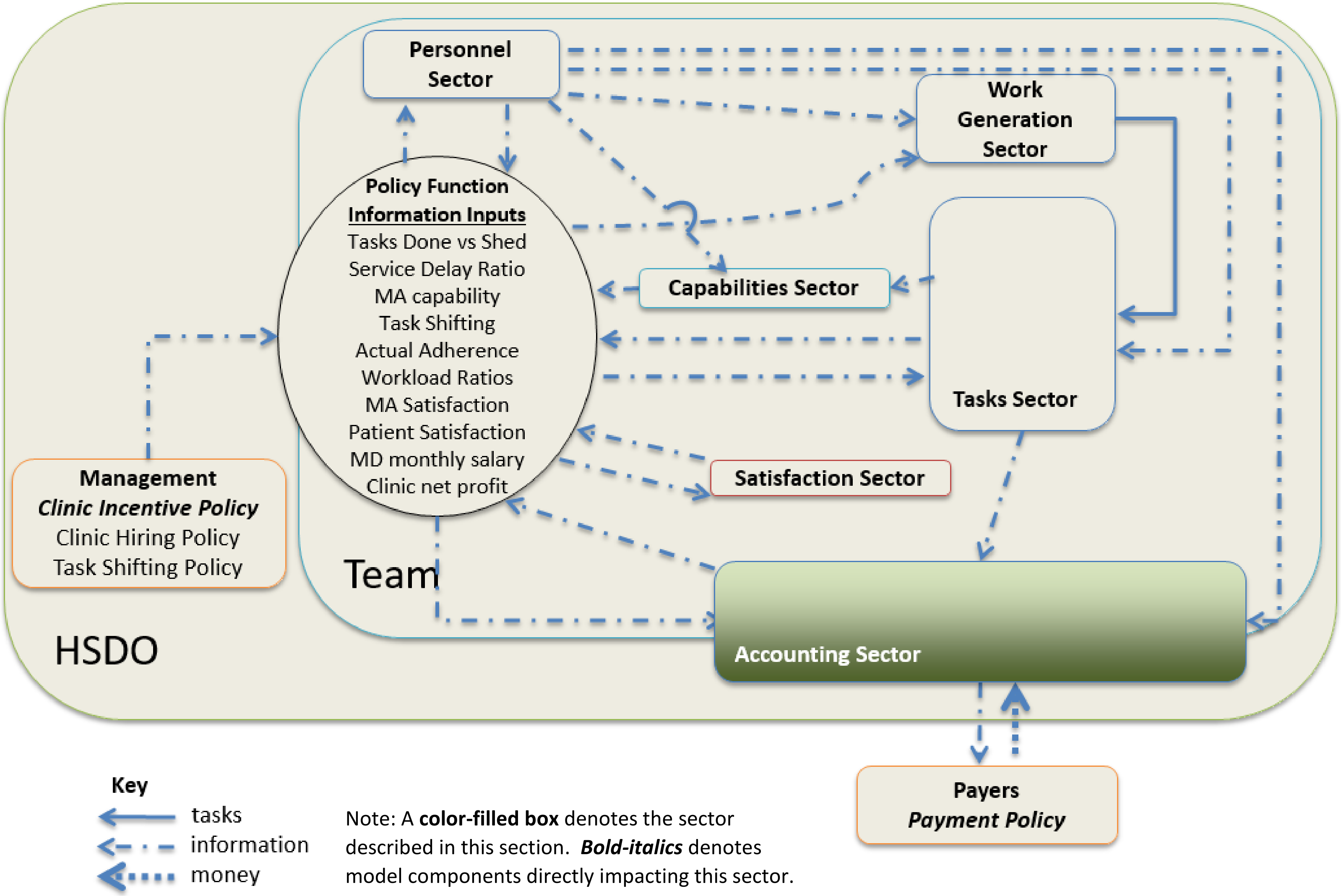
Figure 3.41 MA Capabilities Learning Curve (effect of MA capab ratio on change in MA capab)



3.4.2.5 ACCOUNTING SECTOR

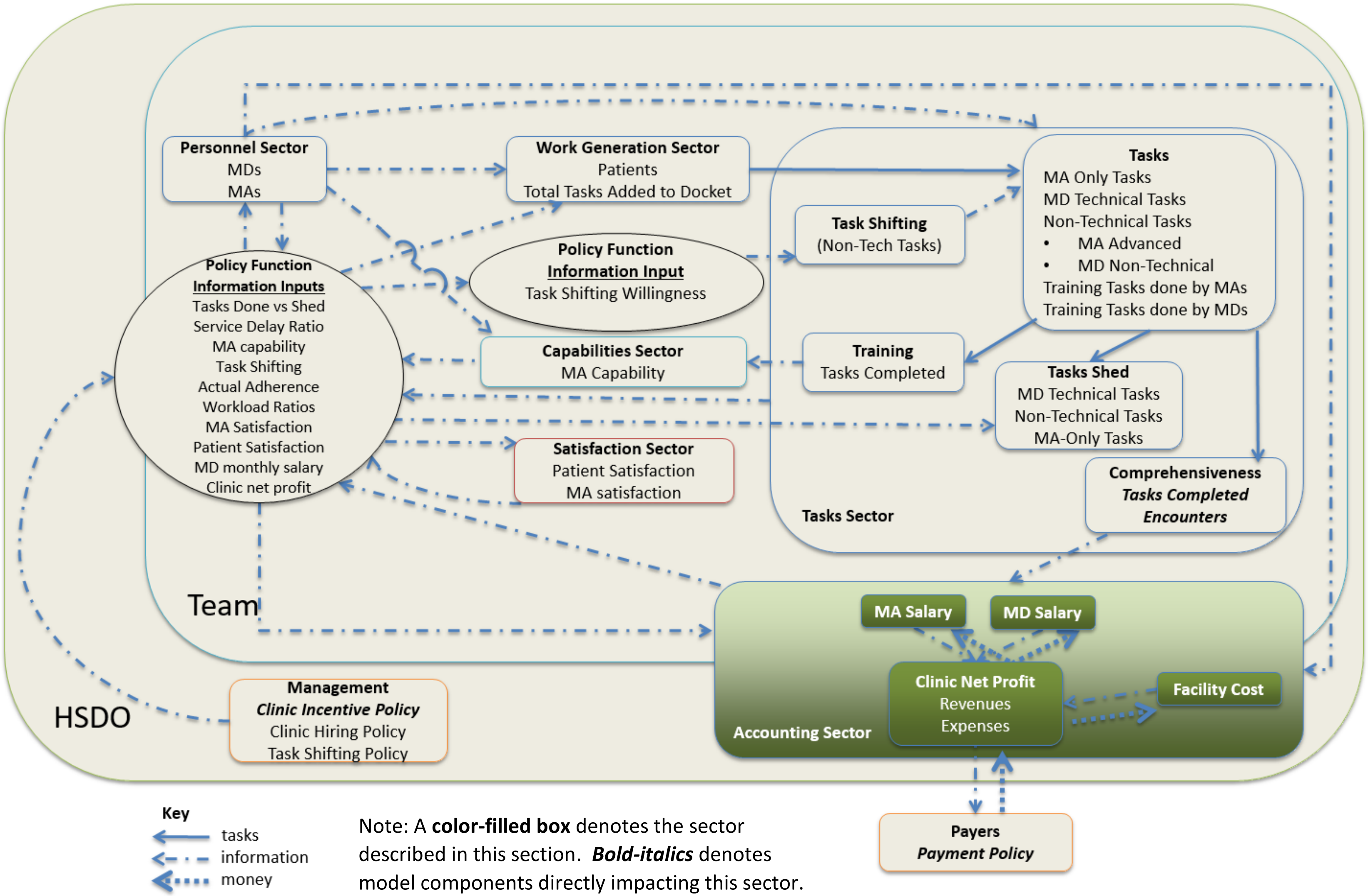
Figure 3.42 and Figure 3.43 highlight the Accounting sector on the sector map. The revenue from patient tasks performed (the "services delivered") as well as the associated facility costs determine the amount of payment received from payers. The clinic incentive policy determines the payment made to clinicians for their work. The net of revenue and expenses drives the clinic profit margin, which is summed over time to calculate the cumulative revenue. Figure 3.44 below presents the operating policies influencing monthly clinic net profit and cumulative revenue. Box 3.10 presents associated equations.

Figure 3.42 Sector Map – Accounting Sector



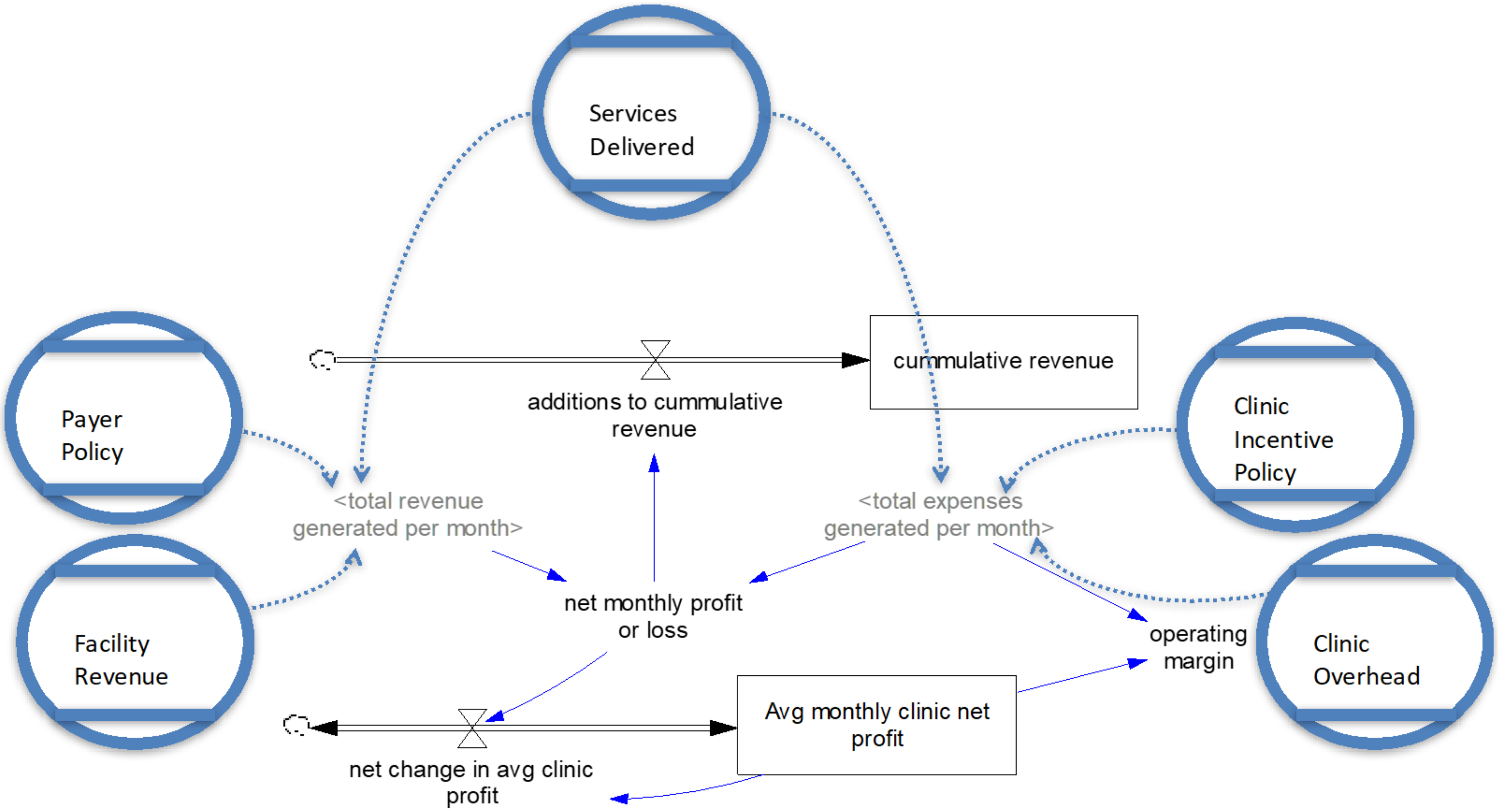
Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Figure 3.43 Detailed Sector Map – Accounting Sector



Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Figure 3.44 Policy Structure Impacting the Accounting Sector



Box 3.10 Clinic Finances

Stocks
Avg monthly clinic net profit (USD/Month) = ∫[net change in avg clinic profit] *dt*
Initial value = net monthly profit or loss
cumulative revenue (USD) = ∫additions to cumulative revenue dt
Initial value = 0

The net change in average monthly clinic net profit is calculated by subtracting the total monthly expenses from the total monthly revenue. A three-month running average is then calculated as this is the perceived monthly clinic net profit (as perceived by the clinic manager determining whether to hire more MAs).

The total monthly revenue is calculated by adding up all the sources of revenue – the revenue generated by the team and the facility revenue. The revenue generated by the team is a function of the number of patient care tasks (services) delivered, the number of wRVUs associated with each task type, and the compensation per wRVU (as determined by the payer and the average visit relative complexity). The revenue generated by the facility charges is calculated in a similar manner, this time with the facility compensation per wRVU (as determined by the payer and the average visit relative complexity).

Each facility negotiates payment contracts with payers. The US Centers for Medicare and Medicaid Services determines payment policy for Medicare and Medicaid insurance plans. Their negotiated rates influence rates in the rest of the market. This level of complexity is not central to the problem statement and thus not included in the model. One amount was used as an approximation: the Centers for Medicare and Medicaid Services payment for a complex visit located in an office ($208). This amount is then adjusted down in the model, to represent a more typical primary care practice (where not all visits are complex visits).

I account for random variations in the environment causing uncertainty in the total monthly revenue. The total monthly revenue is multiplied by "pink noise in revenue generation"[38]. Pink noise is a type of noise using a number representing a random underlying process that is generating a disturbance that has some inertia such that "future values of the disturbance depend on its history"[200] (p. 68). A detailed description of pink noise is provided by Sterman[113] (p. 395,913-924). I used the template structure provided in Vensim[177]. The

[38] It is a stock where the accompanying flow, the *change in pink noise*, is determined based on an exogenous constant – the *standard deviation of revenue generation*.

need for this structure was uncovered in Phase 4 stakeholder dialogue and is described in greater detail in validation.

The total monthly expense is calculated by adding up all the sources of expenses – personnel and clinic overhead cost.

Clinician cost is calculated based on the clinic's incentive policy (salaried, wRVU-based or encounter-based) and the associated measure (number of clinicians, wRVUs completed, or number of MD encounters). Of course, there is also the opportunity to blend these incentives. Box 3.11 below presents the equations for clinician salary under each policy.

Box 3.11 Clinician Payment Policies

Resulting Costs

MD cost per month based on fixed salary (USD/Month) = salary and benefits per MD per month * Number of MDs

MD cost per month based on wRVUs (USD/Month) = MD compensation for Tech tasks per month + MD compensation for nonTech tasks completed by MD per month + (adjusted benefits per MD per month*Number of MDs)

MD cost per month based on encounters (USD/Month) = (number of MD encounters per month * adjusted incentive pay per encounter) + (adjusted benefits per MD per month * Number of MDs)

MA cost is calculated in a similar manner. MAs were salaried in the HSDO. For a period of time, they were paid extra for additional capabilities. Encounter-based compensation was not reported an option for MAs. As actual adherence increased, teams required clinical staff members of new professions, with higher salaries than MAs. To account for this, the model uses wRVU-based compensation for MAs.

Clinic overhead cost was calculated based on the assumed overhead for a clinic with two clinicians treating the most complex patients multiplied by the average visit relative complexity.

The number of MD encounters (Box 3.12) is a measure of the amount of time that clinicians spend with patients delivering technical or non-technical tasks. This number is calculated based on the number of patient tasks completed (i.e., not training tasks), the time per task, the length of each encounter, and the proportion of staff work that is in-visit work. The values are calibrated based on the literature and such that the MD salary under normal conditions is the same whether paid via the salary, MD encounters or wRVUs clinic incentive policy.

Box 3.12 The Level of MD Encounters

Stock

Cumulative number of MD encounters (encounters) = ∫[number of MD encounters per month] *dt*
Initial value = 0

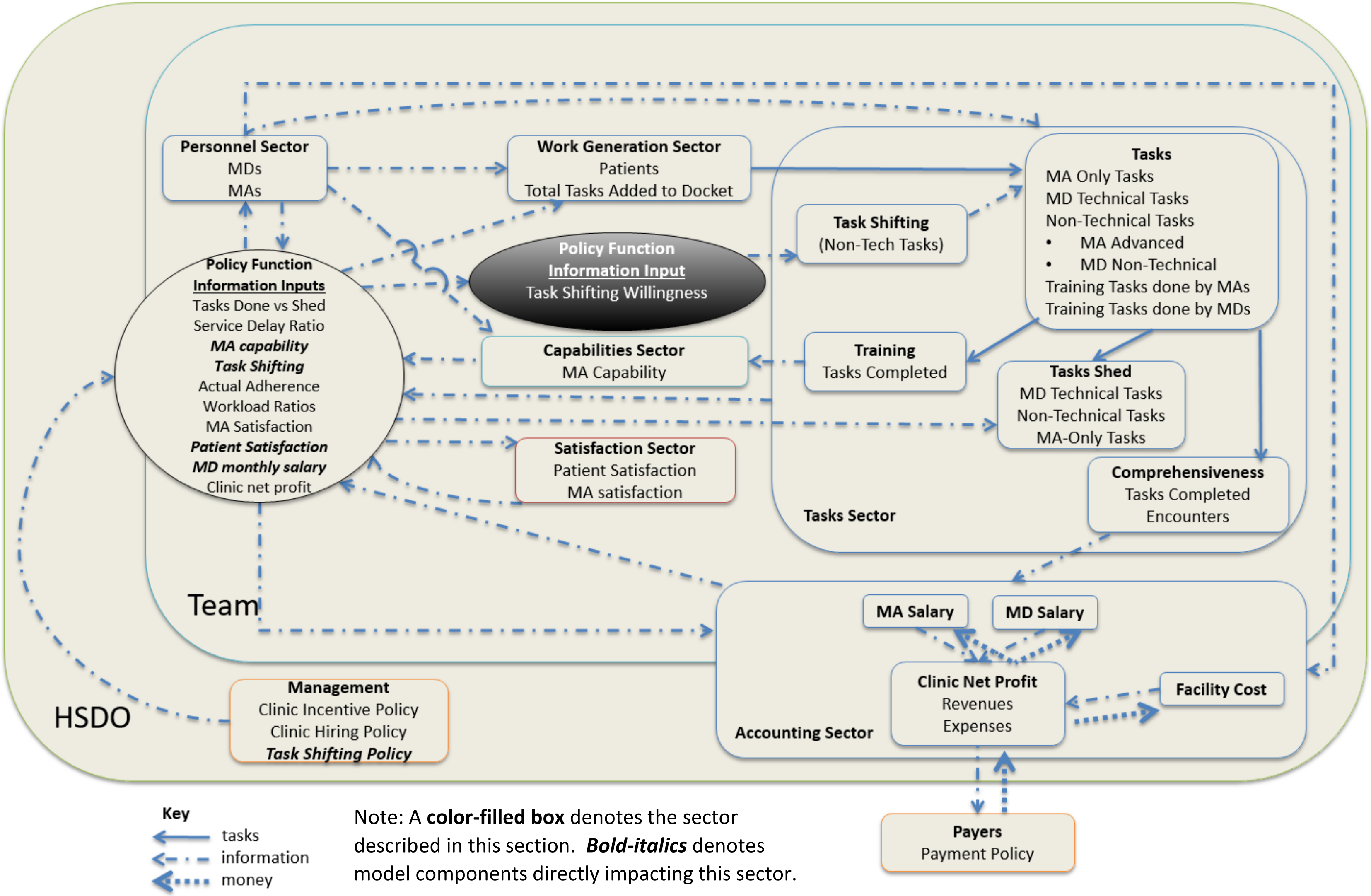
Flow

number of encounters per month (encounters/Month) = time for all patient tasks per month/encounter length*proportion of staff work that is in visit work

3.4.2.6 TASK-SHIFTING WILLINGNESS

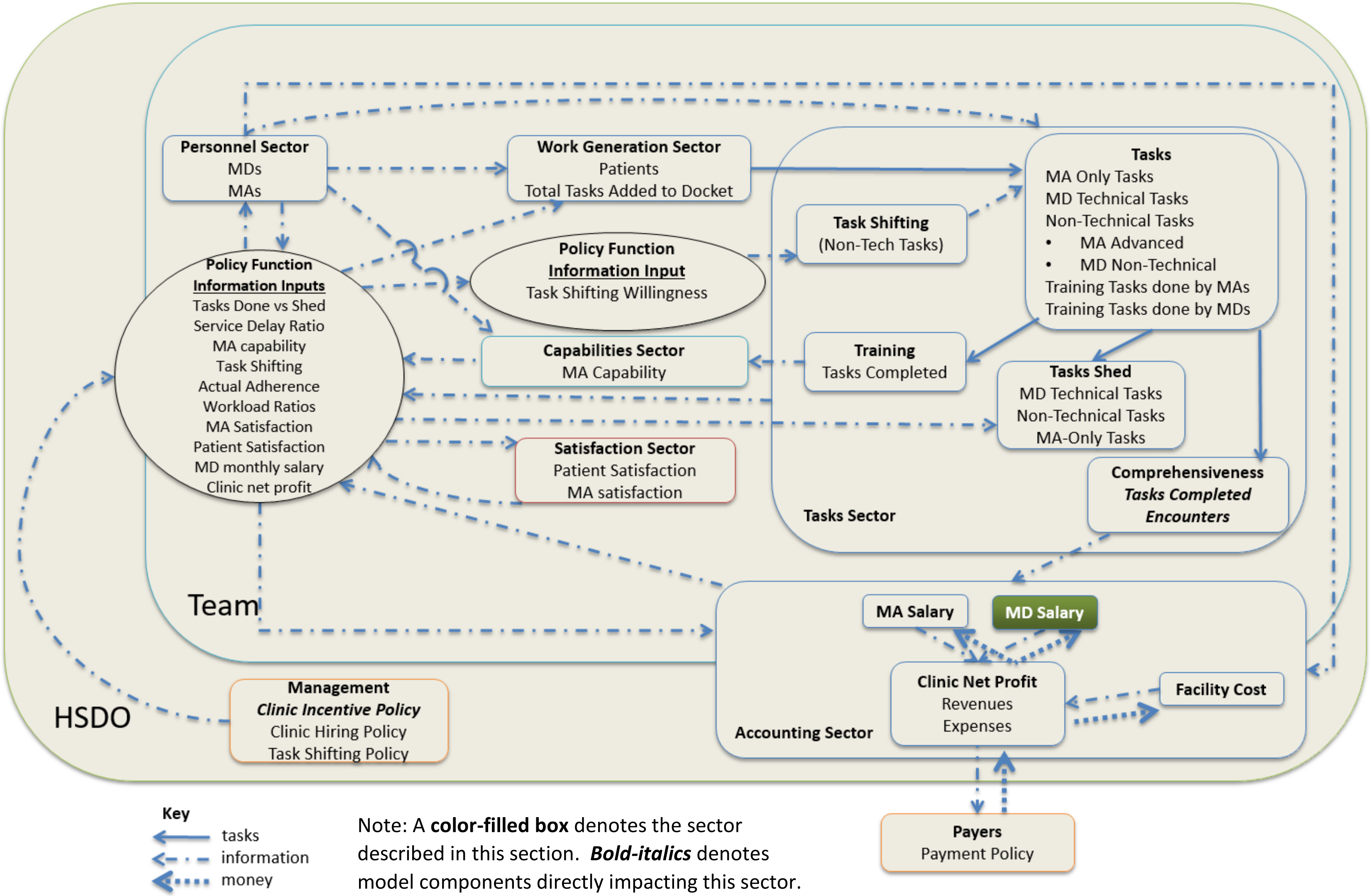
Figure 3.45 highlights the location of task-shifting willingness as well as its information inputs (in ***bold-italics***) on the detailed sector map. Figure 3.46 and Figure 3.47 do the same for MD salary and patient satisfaction, respectively.

Figure 3.45 Sector Map – Task-shifting



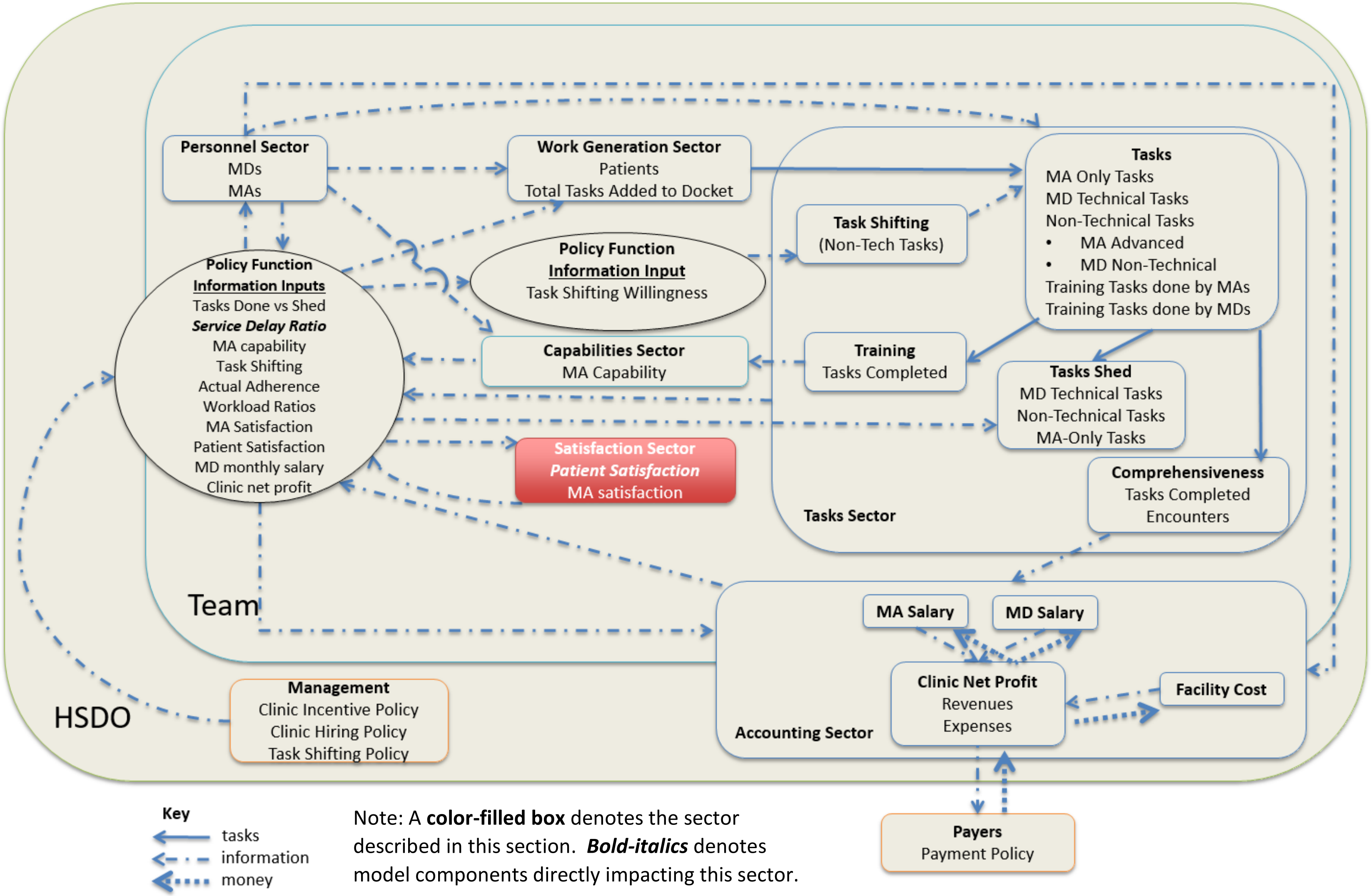
Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Figure 3.46 Sector Map – MD Salary



Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Figure 3.47 Detailed Sector Map – Patient Satisfaction



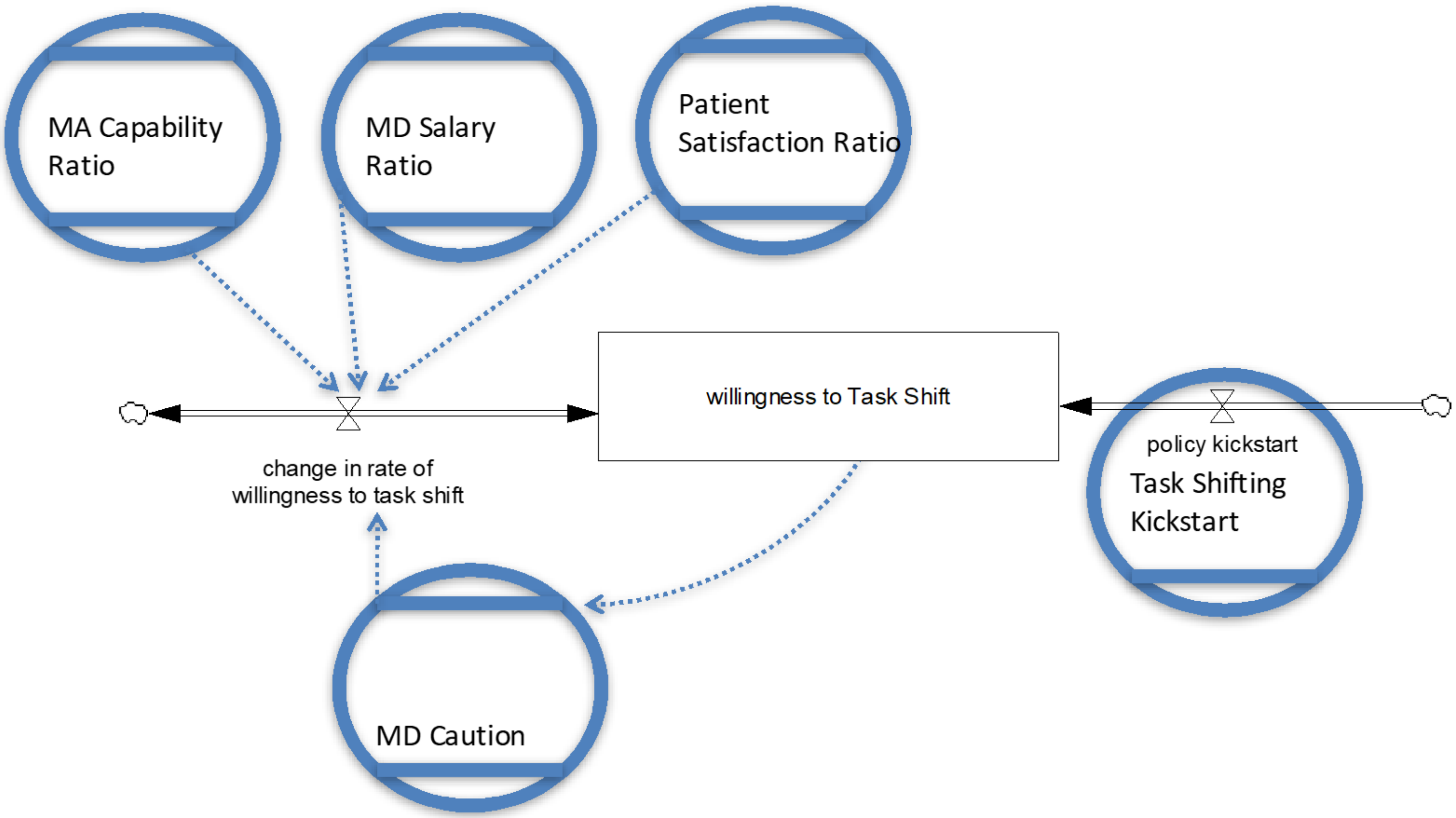
Note: A **color-filled box** denotes the sector described in this section. ***Bold-italics*** denotes model components directly impacting this sector.

Clinicians' willingness to shift tasks changes depending on four policies: Perceived MA Capability Ratio, MD Salary Ratio, Patient Satisfaction Ratio, and MD Caution (on the left of Figure 3.48; Box 3.13). However, the model is initiated in equilibrium. The model is disturbed by implementation of the Task Shifting Kickstart policy (on the right of Figure 3.48).

As a brief summary, this initial one-time pulse of willingness or "task-shifting kickstart" is caused by management: requesting that clinicians try a little task-shifting. They may do this by simply requiring the pulse of task-shifting that is desired and/or, for example, by communicating the PCT aims, structural changes involved, and the worse-before-better behavior (see Radzicki[201]) that is anticipated, touting benefits to patient health and clinic efficiency. Management may do this, and clinicians may agree, regardless of whether other supportive policies are in place. In all cases, the clinicians on the team are sufficiently convinced about the merits and feasibility of task-shifting that they are now somewhat willing to try it.

Simplistically speaking, task-shifting generates training tasks, as well as patient care tasks. As training is completed, those tasks are replaced by more patient care tasks. This does not mean that more patients are seen, but that the same patients are receiving more services – more adherence to clinical guidelines (see Section 3.4.2.2 Work Generation).

Figure 3.48 Policy Structure of the MD's Willingness to Task Shift



Box 3.13 The Level of MD Willingness to Task-shift (TS)

Stock

willingness to Task Shift (willingness) = ∫ [change in rate of willingness to task shift + policy kickstart] dt

Initial value = 0

Flows

change in rate of willingness to task shift (willingness/Month) = willingness to Task Shift * effect of MA Capability on MD willingness * effect of MD monthly salary on MD's willingness * effect of willingness ratio on further changes to willingness * effect of perceived patient satisfaction on MD willingness / time to develop willingness

policy kickstart (willingness/Month) = PULSE(TS start date, 1) * kickstart amount

The 1-month-long duration of this pulse could be thought of as the initial enthusiasm around the beginning of PCT which endures after the first announcement, but wears off as work returns policy-making back to normal conditions. Clinicians are enthusiastic to begin PCT because of its promise to allow them to practice closer to their own self-image as providing the ideal primary care for their patients. The stock of willingness to task-shift adequately represents this enthusiasm for PCT because, given they are at full capacity already, the only way a team can become more comprehensive is to increase the involvement of clinical staff members.

This pulse sets off the rest of the system; as System Dynamics founder Jay W. Forrester recommends:

> "…our first investigations should be designed to divulge *the inherent internal characteristics of the system itself*. One way is to start from a condition of steady-state balance, then to provide an initial disturbance, and to observe the ensuing interactions within the system."[107] (p. 200)

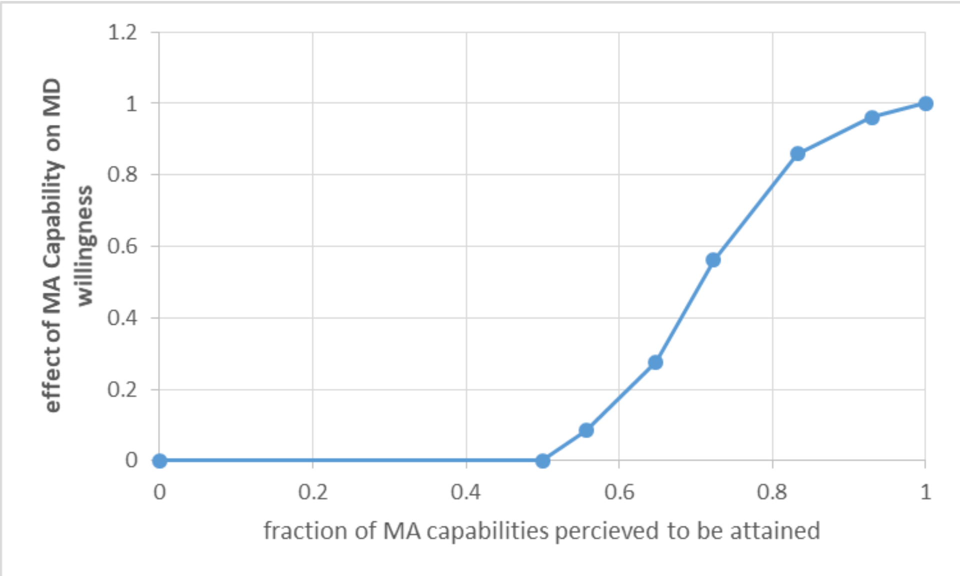
How clinicians react to this initial pulse depends on the structure of the system, which includes their preferences (left side of the figure above). Each policy contains a table function that identifies how the information input is interpreted and acted upon. Each is described in a separate sub-section below.

3.4.2.6.1 PERCEIVED MA CAPABILITY RATIO

The Perceived MA Capability Ratio is the MD's perception of the MA's capability divided by the maximum capability that the MA can attain. The MA may be performing at a higher capability level, but clinicians take time to realize that the MA's capabilities have increased. This is modeled via a goal-gap structure where the perception delay is assumed to be two months.

The clinicians' willingness to shift tasks changes relative to this ratio (see Figure 3.49). Under normal conditions, this ratio equals 0.5 and the clinicians' change in willingness is zero (0.5, 0), otherwise the MD would be shifting tasks under normal conditions (and we know that the MD is not doing so). We also know that, at max task-shifting, the MA capability ratio is 1 and the MD is 100% willing (1,1). In between these points, we assume S-shaped growth because clinicians may initially be more hesitant even when an increase in the ratio is observed, then as that increase persists, they become more willing, but then toward the end they again become more hesitant because, for example, the tasks are more complex.

Figure 3.49 Effect of MA Capability on MD willingness

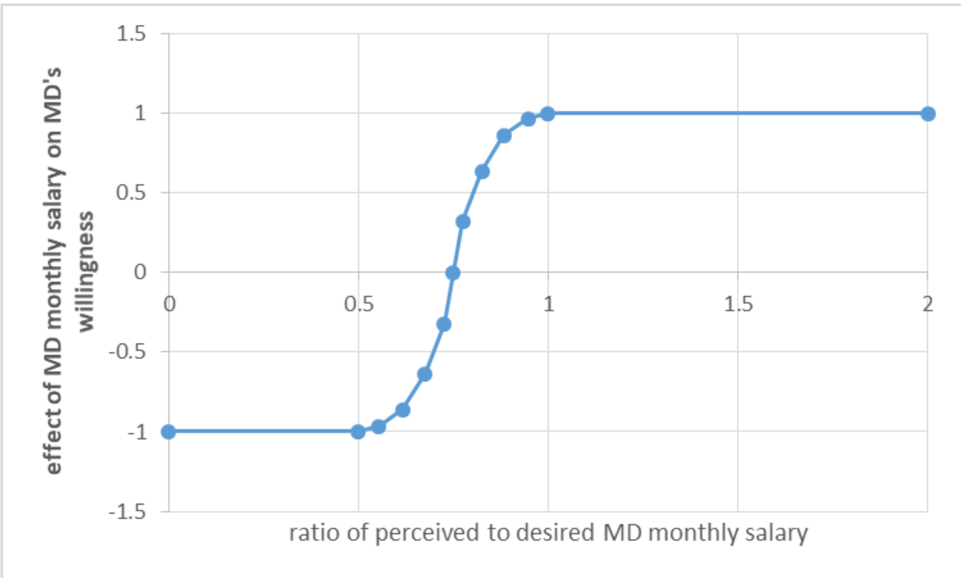


3.4.2.6.2 MD SALARY RATIO

The MD Salary Ratio is the perceived monthly salary divided by the desired monthly salary. Again, a goal gap formulation is used with a perception delay of 2 months.

Under normal conditions, we assume that this ratio equals 1 (clinicians receive their desired monthly salary) (Figure 3.50). As long as clinicians are receiving the desired salary, we assume that they are satisfied and therefore willing to entertain the idea of task-shifting (1,1). We also assume that clinicians would be pleased with receiving more than the desired salary (2,1). However, we know that clinicians are sensitive to decreases in salary. We assume that clinicians are unwilling to accept less than 75% of their desired monthly salary (0.75,0). Between these points, clinicians are initially more hesitant to stop task-shifting, when an initial decrease in the ratio is observed – it is to be expected given the uncompensated time spent training MAs. Therefore, clinicians' hesitance to grow their willingness increases. Below 75% of their desired salary level, clinicians actively resist task-shifting, pulling back on past task-shifting (i.e., a negative change to willingness) until reaching 50% of their desired salary when they pull back on all task-shifting (0.5,-1). Following the graph the other way, as clinicians see their salary getting a little better than the 50%, their willingness increases slowly (moving from active resistance back toward full willingness). As the salary ratio continues increasing, willingness increases more quickly.

Figure 3.50 Effect of MD monthly salary on MD's willingness



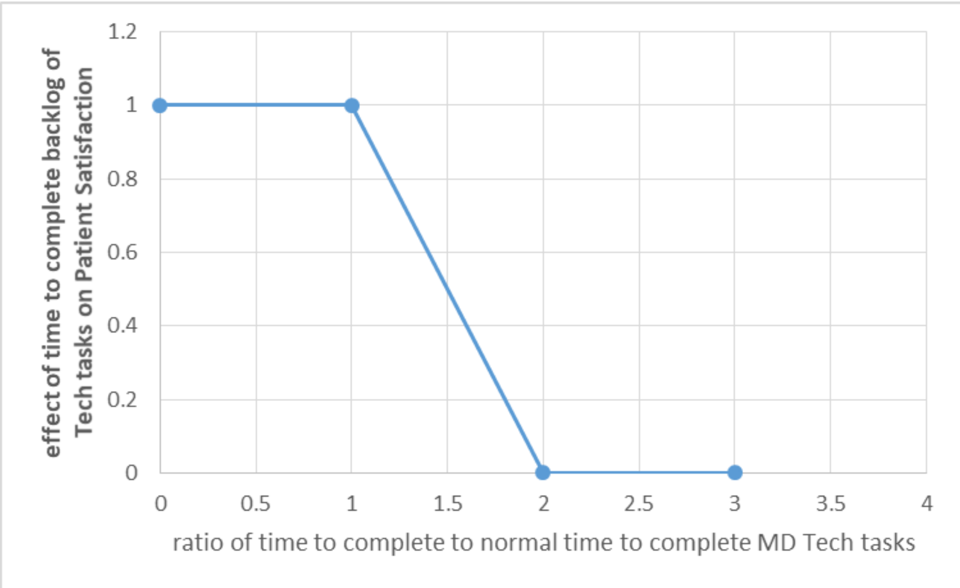
3.4.2.6.3 PATIENT SATISFACTION RATIO

The Patient Satisfaction Ratio is made up of clinicians' perception of their patients' satisfaction divided by the initial patient satisfaction. A goal-gap formulation is used, with a perception delay of 3 months.

In interviews, clinicians often reported that they were concerned with the increased wait times as the team was struggling to implement PCT (referred to as the service delay). Therefore, in the model, patient satisfaction is derived from the "time to complete backlog of technical tasks" as these are the tasks that patients are assumed to care most about.

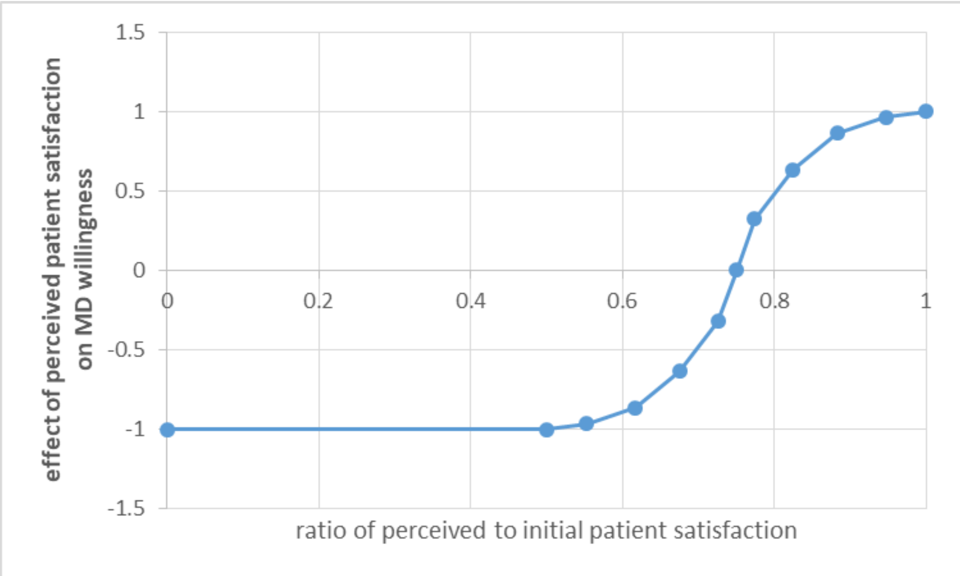
Patient satisfaction is derived from this delay using the table function below. A three-month delay (assumed to be normal conditions) has no effect on patient satisfaction (3,1) (Figure 3.51). Having a smaller delay would keep patients equally happy (0,1). A longer delay impacts patient satisfaction linearly, where a 6 month backlog results in a zero patient satisfaction score (6,0).

Figure 3.51 Effect of time to complete backlog of Tech tasks on Patient Satisfaction



Clinicians perceive and react to this patient satisfaction via the table function in Figure 3.52 below. When patients are perceived as satisfied, clinicians are satisfied with the way things are and would therefore entertain the idea of task-shifting (1,1). As the ratio decreases, clinicians would be less willing to continue task-shifting, at first giving more of the benefit of the doubt (being more hesitant to decrease willingness) but then decreasing it more quickly until losing the willingness all together at a ratio of 0.75 (0.75,0). Any further decrease in the ratio will result in clinicians pulling back on the task-shifting already given (i.e., active resistance). When the ratio reaches 0.50, clinicians become completely unwilling to shift tasks.

Figure 3.52 Effect of perceived patient satisfaction on MD Willingness



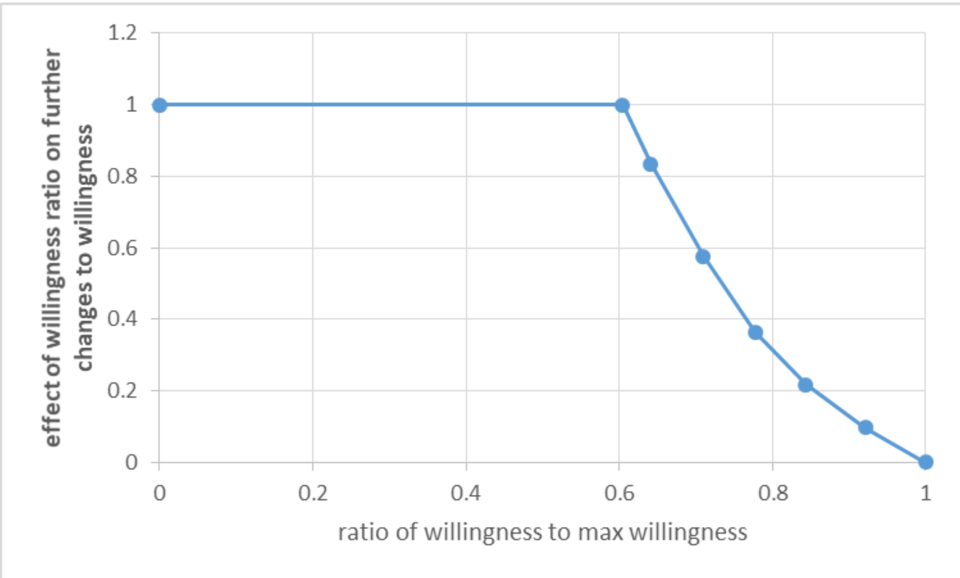
3.4.2.6.4 MD CAUTION

For MD Caution, a ratio is used to gauge how close clinicians are to their maximum level of willingness to shift tasks. This willingness ratio is the current level of willingness divided by the max level of willingness. It modifies the stock of the MD's willingness to shift tasks via a table function with a 4-month delay.

Clinicians are assumed to be inherently willing to trust others and therefore willing to try task-shifting (0,1) to (0.6,1) but their *actual* willingness (i.e., the value of the stock) does not increase under normal conditions because MAs have not shown themselves to be capable of task-shifting (i.e., the zero value from the effect of the Perceived MA Capability ratio on willingness results in no change to the zero level of willingness to shift tasks). With a kickstart from management, clinicians see improvement in MAs' Capabilities while also seeing that their salary and patients' satisfaction are not adversely affected – these values result in their absolute multiplicative positive change in the amount of willingness.

As clinicians' willingness exceeds 60% of their max level, they begin to hesitate to trust, knowing that the remaining non-technical tasks are close to the top of clinical staff members' licensure. Out of an abundance of caution, clinicians are less willing to shift tasks as they pass this point, even though the other effect values may be encouraging. This is shown in the table function via a curved decrease (from (0.60, 1)) all the way to (1,0) when the max willingness has been reached and the effect on further willingness is zero (i.e., no further inflow of willingness) (Figure 3.53).

Figure 3.53 Effect of willingness ratio on further changes to willingness



3.4.2.7 MEASURES OF INTEREST

The primary measure of interest is adherence to clinical guidelines (referred to in the Simulation Model as "**actual adherence to clinical guidelines**" to distinguish it from the team's attempted level of adherence). It is calculated by dividing the total amount of patient care delivered divided by the total possible (needed to attain full adherence to clinical guidelines). This is the primary measure of interest because it could be considered a composite measure of the tenets of primary care – when one is delivering accessible, coordinated and comprehensive care in a continuous relationship with their patients, one is delivering primary care that is aligned with its aim.

The model also tracks other measures that were mentioned in participant interviews and of interest to stakeholders, such as: satisfaction for patients, clinicians and clinical staff members; productivity and salary for clinicians and clinical staff members; as well as clinic revenue and expenses.

3.5CROSS-PHASE STEP: VALIDATION

Given the high validation cessation threshold observed for this research (Section 3.5.1 below), an extensive set of validation tests is employed. This set is comprehensive of SDM standard method model validation tests of structure, behavior (and culture), except those involving longitudinal numeric data. Given the gaps in SDM standard method validation tests described in Chapter 2, new tests which use mental data were developed for both model and meta-level validation. These tests are used in concert with the SDM standard method model validation tests.

Table 3.4 below presents the validation methods, in the order that they are presented in this section. Model-specific validation tests were **used in the various phases of model development (i.e., Phases 2 - 4)**. Meta-level validation tests were referred to throughout and **present my reflection on the *meta-level* validity of this work** – they are aspects of validity that relate to the overall project. The first column lists the test, the second column identifies the validation level, and the third column identifies the issue addressed.

Table 3.4 Validation Methods

Method	Level	Issue addressed
Cessation	Meta	When should the researcher feel sufficiently confident to move from developing the model to using it?
Data Suitability	Meta	Are the data upon which the model is based used within the scope of their limitations?
Methods Suitability	Meta	What are the trade-offs of the specific modeling approach used in this research?
Phase 2 • Shared Mental Model Saturation (**SMM-S**) • Conceptual Model Saturation (**CM-S**) • Stakeholder Dialogue	Model	Is model development ready to continue? SMM1 => SMM2 => SMM3
Phase 3 & 4 • SDM Standard Method Validation Tests • Simulation Model Saturation (**SIM-S**) • Stakeholder Dialogue	Model	Is model development ready to continue? SIM1 => SIM2 => SIM3
System Dynamics Saturation (**SD-S**)	Meta	Does the target group see SDM research as a useful way of addressing the problem?
Stakeholder Dialogue Suitability	Meta	Who, when, and how were stakeholders engaged in validation?

3.5.1 CESSATION

The *validation cessation threshold* is a framework which permits researchers to consider, in a comprehensive way, how much effort they *should* place in validation. Traditionally researchers using modeling report that budgets largely determine how much effort is placed in validation[161]. On the other hand, researchers using other social science methodologies can complement this important consideration with more data-driven techniques to assess the point at which they should stop analyzing data (developing a theory or model) and begin using what they found.

I drew from one of these techniques – the Grounded Theory concept of saturation – to justify and design my newly-developed model development tests. Three of these *saturation* tests (SMM-S, CM-S and SIM-S) helped me to know when I could feel confident that a stage of model development was completed, and a fourth test (SD-S) helped me to consider how appropriate my target audience might judge my choice of SDM to have been. Like Grounded Theory saturation, these tests used my data as a guide.

Groesser & Schwaninger's *validation cessation threshold* framework[161] (see Section 2.4.5.1) helped me to estimate how others might evaluate the research as a whole using the broader features of the research itself as a guide. These included features of my data as well as of my research's stakeholders, and of my own expertise with SDM. Considering these features, I observed that this research would need to pass a high Validation Cessation Threshold. The factors upon which this determination was made are described in Table 3.5 below.

Table 3.5 Description of Factors for Assessing the Validation Cessation Threshold

Factor	Description	Determination
Relative Importance/Risk of Decision	See Chapter 1 description.	High
Target Group's Expectations	As this dissertation research is one piece of the Mixed Methods Project, and one that was added after initial grant submissions, expectations for this piece are low.	Low
Target Group's Experience with Modeling	Stakeholders in the HSDO do not have experience with this type of modeling.	Low
The Researcher's Level of Expertise	I learned SDM over the course of this dissertation.	Low
Model Size	The Simulation Model is labeled as mid-range in size because it is bigger than a typical small policy model[202] and smaller than a typical detailed calibrated model[203, 204].	Mid
Data Availability	Interview data (transcripts focusing on PCT, oversampling of those stakeholders involved in PCT) were deemed to be very useful (high); Other data sources were less enlightening on system structure and dynamics (low).	Mid
Validation Cessation Threshold		High

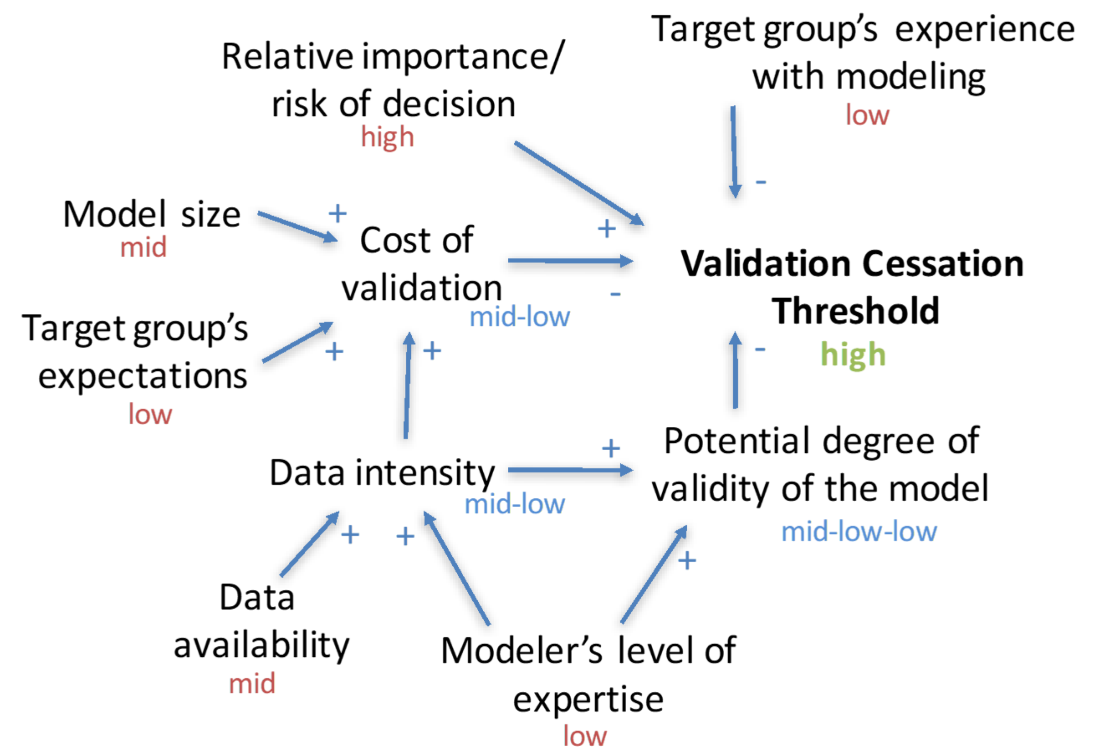
The first type of determinations were purely subjective judgments influenced by my own reflections and discussions with stakeholders in various phases of this research. I judge my research subject to be relatively important and to entail a relatively high amount of risk (see section 1.2.2). I judge my research audience's expectations of my research and experience with SDM, and my own level of expertise in SDM all to be low.

The second type of determinations were partially subjective and were informed by my data. I considered that a mid-range size simulation model would be appropriate for my work: a small model would have been too small for my qualitative models and a large predictive model was beyond my problem statement. I considered my data availability to be of a mid-range (mental data high, written and numeric data low).

Figure 3.54 below shows the *validation cessation threshold* framework. The input variables on the outside of the framework were assigned the values in Table 3.5 above (determination column). I applied the Groesser & Schwaninger framework using the following approach. I considered the input variables to have three possible values (high, mid and low, shown in red) and that their interaction would be linear and of equal proportion. In other words, if I judged one variable to be *high* and another to be *low* and both influenced a third variable with positive polarity links, the result would be *mid*, the average of the two inputs. The resulting values (blue text) were then also used in like manner to determine the value of their resulting variable – the validation cessation threshold (green text).

For example, the *Potential degree of validity of the model* depends on the availability of good data for modeling (upon which the model is based and/or validated – *Data intensity*) and upon the expertise of the modeler in modeling from the data and using the data for validation purposes – *Modeler's level of expertise*. A "low" level of expertise and "mid" level of data availability were averaged to be a "mid-low" level of *Data intensity*. With a low level of expertise this led to a "mid-low-low" level of *Potential degree of validity of the model*.

Figure 3.54 Applying Groesser & Schwaninger's Framework to this Dissertation Research



This figure is adapted from Groesser & Schwaninger's framework[161]. For each factor, I assign the determination made in Table 3.5 above (red text). Then, I use the arrow polarity to determine the value of the intermediate variables (blue text), finally leading to the determination for the validation cessation threshold (green text).

Since **the *validation cessation threshold* was found to be high**, an extensive application of validation tests is indicated. As described in Chapter 2, the validation tests I employ comprise all aspects of validity (conceptual, formulational, experimental and data validity) considered in SDM. **The results of this thorough assessment (presented throughout Section 3.6.2.1); together, providing a high level of evidence for the potential usefulness of my research.**

3.5.2 DATA SUITABILITY

This test considers whether the most important model assumptions are being made on the right kind of data, and whether that data is suitable. Like the other validation tests, it is not pass/fail but rather a reflection exercise such that the modeler is confident in and reports the uses of available data.

Table 3.6 below presents my analysis of data suitability within this research. The first column lists the data types. The second column lists the data sources. In the third column, I describe the data validity ascribed to these data (upon reflection on being reliable/appropriate, accessible, and sufficient). The fourth column documents the determination that I made with respect to the use of these data in model development. In the fifth column, I document the way(s) in which each data type was used in model development. During model development and upon model completion, columns four and five are compared.

For each type of data, I consider two types of datasets. For mental data, I differentiate between data I created based on my subjective observations and raw transcript data from the semi-structured interviews. For written data, I accessed archival meeting notes as well as system dynamics models available in the relevant literature. For numeric data, I accessed data managed by the Mixed Methods Project team (whether administrative data from the HSDO or data gathered by the Mixed Methods Project team) as well as parameter estimates in peer-reviewed publications.

Each type of data source contributed to model development. I relied primarily on mental data to build and validate the model. Written data provided background information. Numerical data aided in assigning parameter values. Each type of data was also used in preliminary work (including the scoping study and preparing for conducting the interviews).

Datasets were available in all three types of data (mental, written and numerical). Comparing the determination of how data could be used (column four) with the way in which each dataset was actually used in the research (column five) indicated that data were used within the scope of their limits.

Table 3.6 Data Suitability in this Research

	Specific Data Available	Description of Data Validity	Determination	Use in Model
Mental Data	Field notes & reflective journal	•It was not the purpose of these notes to document instances of causality	Useful for developing the problem statement. Also, useful for background knowledge.	Used in the scoping study
	Recorded, transcribed semi-structured interviews	•Interviews were conducted such that candid descriptions could be given, and discussions focused on experienced and behavioral responses to system changes •Participants were selected for their first-hand experience, different perspectives, and length of time experiencing the problem •Sufficient data was provided to understand cultural acceptability and cognitive limits as goals and policies were described in first-person accounts	Useful for model development and model validation	Used to build and validate the models Also used in the scoping study
Written Data	HSDO management meeting notes	•These notes did not contain causal statements	Useful for developing a timeline of PCT at the HSDO	Used to prepare for interviews as background (conversant in terminology as used at HSDO)
	Other models[198, 205-208]	•These models contained some generic information about system dynamics	Useful in broad understanding	Used in considering alternative model structures
Numerical Data	Administrative data (e.g., productivity, satisfaction, quality) and data gathered specifically for the research (e.g., implementation, burnout, teamness, services utilization data)	•While datasets were accessible, each one had at least one issue; for example: 1) measurement at a different level of analysis than that in the model, 2) irregular and infrequent measurement, 3) changing variable definitions, or 4) numerous process and outcome measures specific to what the model treats via one aggregate variable.	Useful for developing the problem statement. Also, useful for background knowledge.	Used in the scoping study
	Data in peer-reviewed publications	•Papers provided a numerical description of the average one-clinician primary care practice	Useful for providing parameter values	Used for assigning model parameter values

3.5.3 METHODS SUITABILITY

The trade-offs of the specific modeling approach (i.e., modeling methods and modeling languages) used in this research are presented in Table 3.7 below. At the top of the table, the methods are described briefly in three categories: elicitation, aggregation and validation. Elicitation includes methods used in capturing individuals' mental models. Aggregation includes methods used in making assumptions about shared mental models. The lower portion contains descriptions of the key trade-offs (in terms of pros and cons) for the methods used in each of the three categories. With this information and upon further reflection throughout the research process, I made the determination that the modeling approach used was suitable; specifically that: 1) it addresses cultural acceptability and cognitive limitations and uses a non-coercive approach and 2) the methods used in the modeling process did not omit or distort important elements or relationships.

Table 3.7 Methodological tradeoffs

<u>Sampling design:</u> Purposive sampling was applied to select participants based on characteristics important to the problem (to capture the *different* perspectives). Oversampling was used to ensure that the various perspectives on the problem were captured comprehensively (to *fully* capture those different perspectives).

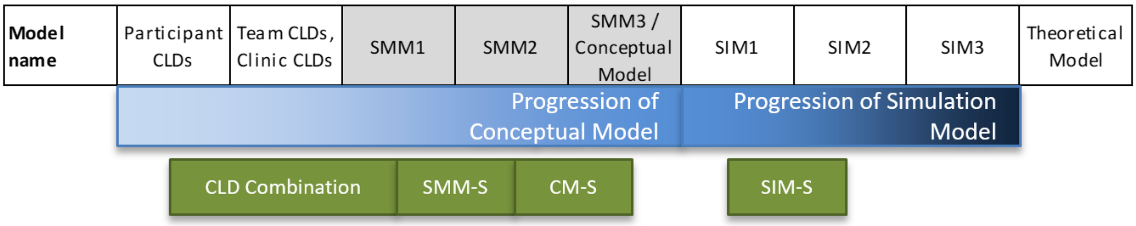
<u>Elicitation methods:</u> Participant semi-structured interviews were conducted and coded for causal statements. Purposive text analysis exposed the variables and relationships from descriptive causal statements resulting in individual CLDs. Standardized variable names and mild pruning were then applied.

<u>Aggregation methods:</u> CLD Combination combines individuals' CLDs to produce a set of CLDs based on parallel real groups (teams and then clinics), which are subsequently combined into a shared mental model and quantified in a *stories with numbers* simulation model. Theoretical Model summarizes the key structures in the Simulation Model.

<u>Validation methods:</u> SDM standard methods for validation as well as newly developed methods for validation are used. Internal consistency is assessed using the SMM-S test; CM-S, SIM-S and SD-S tests compare the Conceptual Model and Simulation Model to an external validation set; and stakeholders are engaged throughout. Confidence in the models is based on having met the validation cessation threshold, across conceptual, formulational, experimental and data validity.

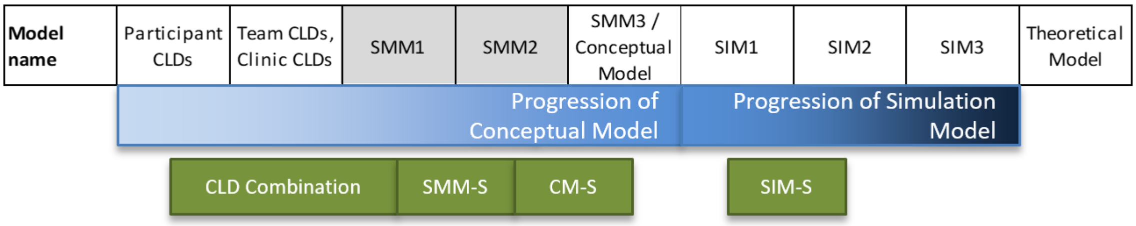
Methods	**Pros**	**Cons**
Sampling Design	•Sample came from participants with first-hand experience with the problem as well as from those with responsibility for addressing the problem at higher administrative tiers •Purposive sampling captured diverse perspectives on this experience (based on judgment of HSDO management, evidenced by representation from all clusters in Scoping Study) •Oversampling attempted to capture many instances of each type of perspective; this approach provided sufficient interviews for model development and validation	•Oversampling is costly for the researcher as well as for the HSDO (compensating participants for their time in the interview)
Elicitation Methods	•Clinician and MA descriptions of causality are treated equally (regardless of individual characteristics, i.e., profession, tenure, affinity toward PCT) •Applying methods (purposive text analysis, pruning and CLD Combination) based on a rigorous, theory-based definition of mental models[118, 119] ensures that the shared mental model is equal to the combination of individuals' mental models	•Others are treated as secondary to front-line participants. They contributed to understanding in the scoping study and validation but not to model development •The investigator plays a large role in exposing and combining variables, especially in development of individual CLDs
Aggregation Methods	•Allows for the representation of how problem x arises from system structure y, as perceived by system participants •Facilitates stakeholders' learning about how and why their policies and mental models create the dynamics of problem x •Applies an integrative methodology to produce equity between qualitative and quantitative methods, resulting in a simulation model capable of telling *stories with numbers* •Simulation and theoretical model engage with stakeholders' mental models and emotions •Develops a model to guide empirical research as well as policy-making (intangible variables are identified, given units and quantitative relationships are theorized)	•Direct participant description was used to expose variables and relationships. The Simulation Model requires more than this. Stakeholder dialogue and operational thinking provided this additional insight
Validation Methods	•Validation cessation threshold was set for the research and met across conceptual, formulational, experimental and data validity types •Tests considered the structure, behavior and cultural aspects of the problem and models •Tests are performed throughout model building, such that confidence is built step by step (and revisions are made when issues are discovered) •Descriptive data is utilized to its maximum effect (for elicitation, aggregation and validation)	•Interview participants were engaged in an early phase of research and were unable to see or use the Simulation Model •Model confidence relies on extensive coding of interviews (time intensive)

3.5.4VALIDATION IN DEVELOPING THE CONCEPTUAL MODEL (PHASE 2)



This section presents the validation results obtained while building the Conceptual Model: Shared Mental Model Saturation (SMM-S), Conceptual Model Saturation (CM-S), and Stakeholder Dialogue.

3.5.4.1 SHARED MENTAL MODEL SATURATION



The following sub-sections present the SMM-S findings demonstrating that the SMM-S Test has been passed, that: the variables identified meet the model's purpose (Section 3.5.4.1.1) and conceptual saturation has been reached (Section 3.5.4.1.2 - Section 3.5.4.1.3).

3.5.4.1.1 VARIABLES MEETING MODEL'S PURPOSE

I was able to read the problem statement and tell stories using SMM1 (results not shown here). I did this telling stories by walking around the feedback loops. These stories involved the elements and behaviors identified in the problem statement as expressed in the feedbacks and delays visualized in SMM1. This was also done for SMM3 (see Section 3.3.3).

Clinic CLDs identified key objectives of PCT: visits with coordinated comprehensive care, task-shifting, teamwork (active in "Trust", "Sharing the load", "Team learning" loops), and financial viability (active in "Sufficient resources" and "Incentives (Vol.)" loops). The Clinic CLDs also surface causal mechanism behind these key objectives that participants had previously identified as in tension. Resource loops, which are necessary for PCT, are constrained by the incentive structure, the scheduling structure, and staffing structure (money, time, and people). For each of these resources, there is a short-term and a long-term loop; where the short-term constraints can be overcome in the long term.

The problem statement focuses on "primary care". While it does not explicitly mention these CLD variables, these variables are understood to be part of primary care as they expose variables related to the primary care tenets as well as to clinic context. These CLD variables and relationships are able to describe causal pathways leading to PCMH implementation failure (e.g., through the short-term loops) as well as causal pathways leading to success in PCMH implementation (e.g., through the loops with delays). In so doing, these variables also hint at potential policies that may be useful in facilitating successful transformation (e.g., don't tie hiring MAs to profit margin).

No portion of the problem statement was found to be omitted from SMM1; indicating that the boundaries of SMM1 appear to be well-developed.

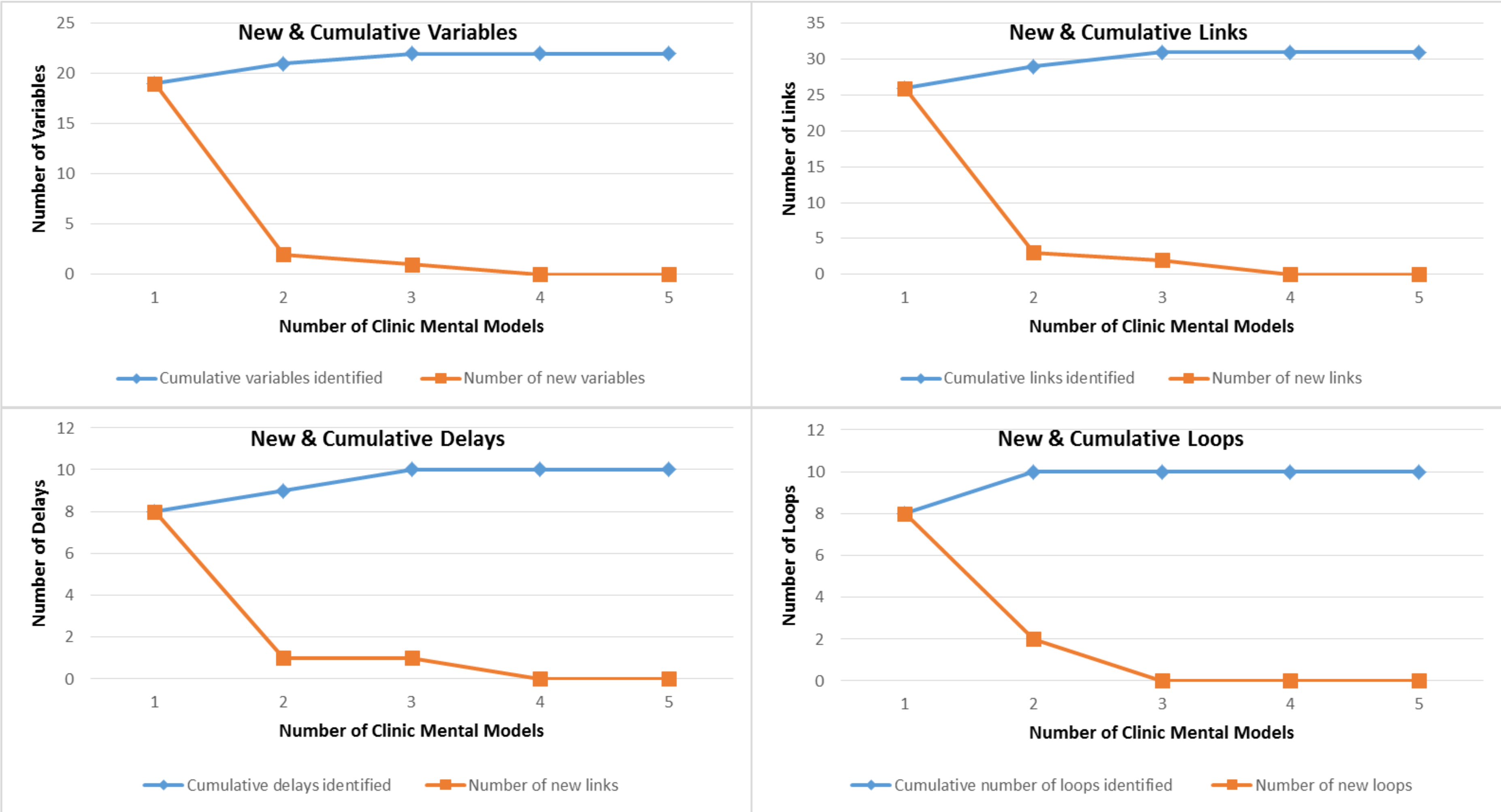
The successful completion of this activity (where no discrepancies were identified) indicated that SMM1 is comprehensive of the dynamics expressed in the problem statement.

3.5.4.1.2 SHARED MENTAL MODEL SATURATION CURVES

Clinic CLD pair-wise comparisons were used to generate SMM-S Curves. The x-axis is the number of clinic CLDs. The y-axis is the number of elements: in blue, clinic-specific cumulative element counts are sorted in ascending order; and, in orange, clinic-specific new element counts are sorted in descending order.

In all four graphs, SMM-S Curves demonstrate that saturation has been reached (see Figure 3.55). For variables, links and delays, saturation is reached after three clinics. For loops, it is reached after two clinics. Therefore, overall, saturation has been reached with the five clinics that were used to generate SMM1. In other words, this finding indicates that new and relevant data regarding the elements in SMM1 have ceased to emerge. Variables, links, delays and feedback loops in SMM1 appear to be well-developed.

Figure 3.55 SMM-S Curves



3.5.4.1.3 SHARED MENTAL MODEL SATURATION DIAGRAMS

In the previous section, the SMM-S Curves visualize saturation for specific types of elements in the model. In this section, I use SMM-S Diagrams to visualize the agreement between clinics about the structure of the system. These diagrams show: 1) the percentage of clinics identifying each relationship (Figure 3.56) and 2) the percentage of clinics explicitly identifying each relationship (where the denominator is the number of clinics that mentioned each relationship) (Figure 3.57). Each figure uses darker shades of color and thicker lines to indicate a higher percentage of clinic mentions of (i.e., level of saturation in) these relationships.

Figure 3.56 shows that some relationships were mentioned more than others. A few relationships were mentioned by only one clinic (i.e., 20% arrows). Figure 3.57 shows that key relationships that close loops are sometimes not explicitly mentioned in clinic CLDs (i.e., dotted lines). Four of the labeled loops could not be closed without implicit mentions. Another four of the labeled loops are closed by a minority of clinic CLDs. Only two of the labeled loops could be closed relying on a majority of clinic CLDs identifying them explicitly.

These results indicate that most relationships in SMM1 are well established (i.e., many clinics mentioned them; and, of clinics mentioning them, many mentioned them explicitly). I reviewed the relationships that were less well established and concluded that each relationship is plausible (i.e., in terms of bounded rationality and cultural acceptability[105, 108-110, 209-212]), that it had not been distorted, but that it was one that was just less understood by participants and thus mentioned less often.

The SMM-S Test is passed, and the Shared Mental Model is now referred to as SMM2. Variables identified meet the model's purpose (as expressed in the problem statement) and conceptual saturation has been reached. Therefore, I concluded that there was no need to consult interviews in the *saturation reserve for model development* at this time.

Figure 3.56 SMM-S Diagram – Clinic Mentions

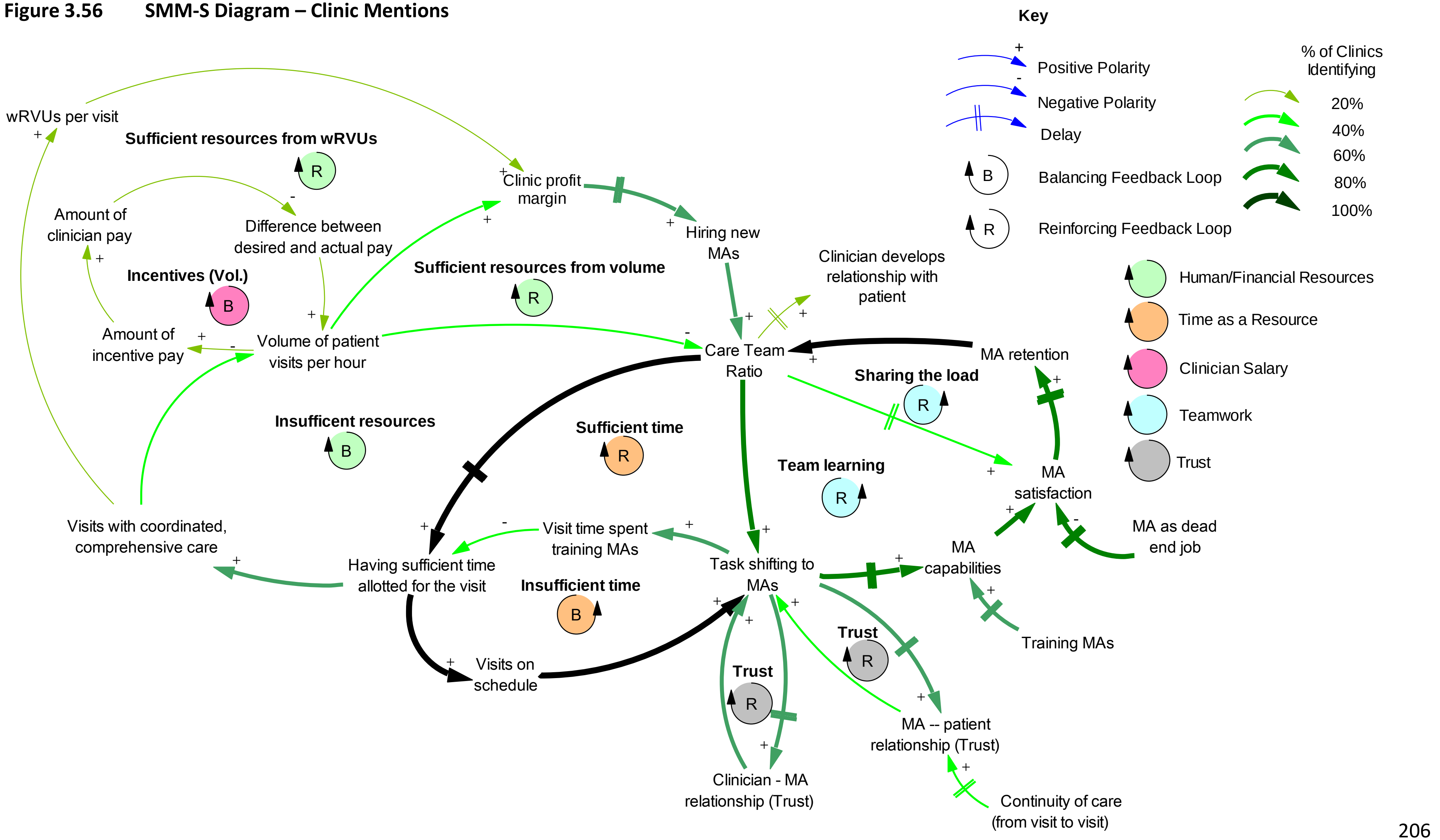
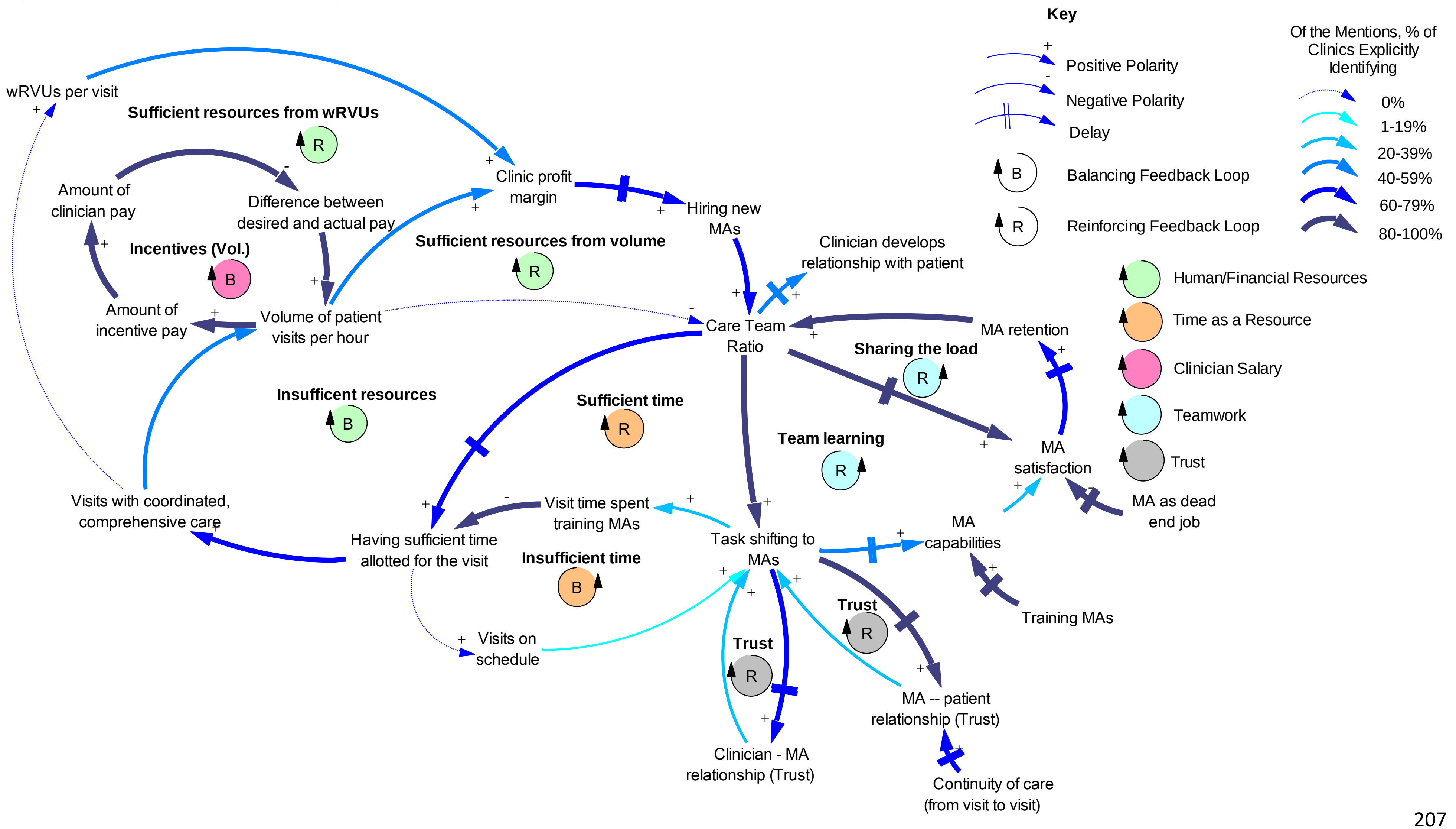
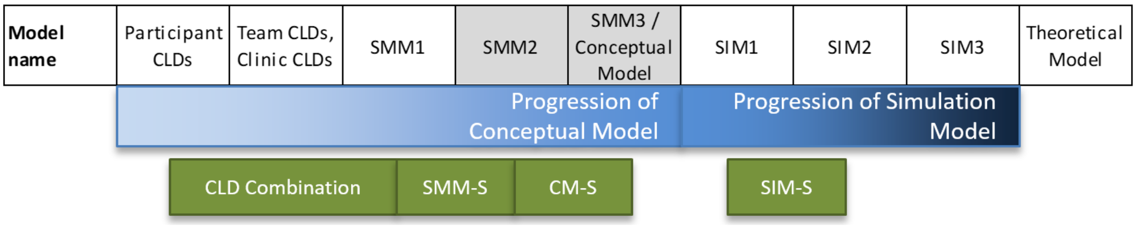


Figure 3.57 SMM-S Diagram – Explicit Clinic Mentions



3.5.4.2 CONCEPTUAL MODEL SATURATION



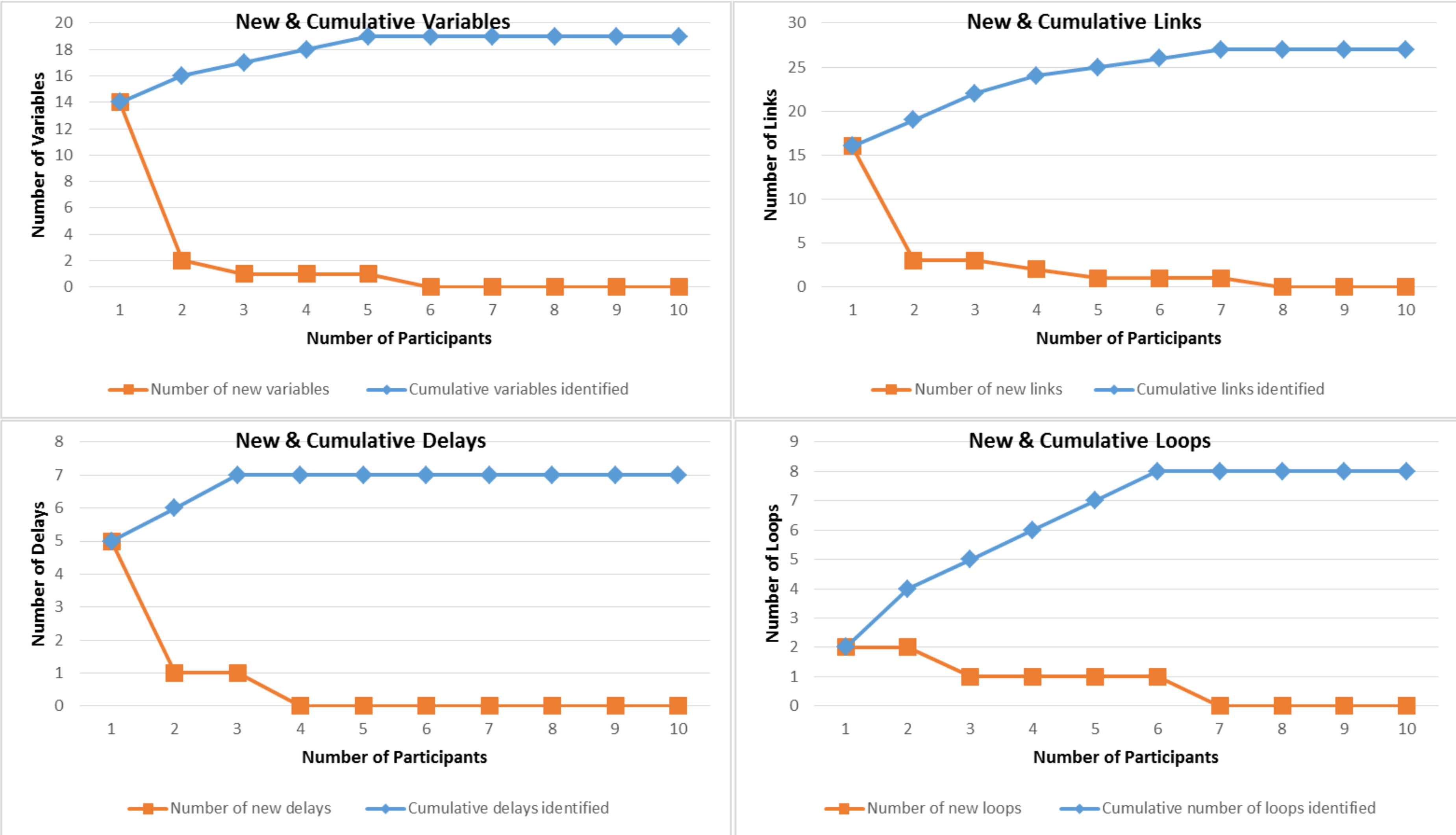
SMM2 was then validated using the CM-S Test and the result is the Conceptual Model (SMM3). The following sub-sections present the CM-S findings demonstrating that the CM-S Test has been passed: 1) CM-S Curves showing saturation, 2) CM-S Diagrams showing how well-established each relationship is and 3) a brief note on the Rigorously-Interpreted Quotations – for Causality (Causal RIQs) demonstrating how capable SMM2 is of exposing elements in participants' mental models (quotations are found in Appendix F). In so doing, Causal RIQs provide a detailed description of each variable and link in SMM3 –the nervous system for the Simulation Model and Theoretical Model. For a discussion of the model improvements made during CM-S, please see Appendix F.

3.5.4.2.1 CONCEPTUAL MODEL SATURATION CURVES

CM-S Curves show the number of elements in SMM2 that were identified in validation-set interviews. The x-axis is the number of participants in the *model validation set*. The y-axis is the number of elements identified in validation-set interviews: in blue, participant-specific cumulative element counts are sorted in ascending order; and, in orange, participant-specific new element counts are sorted in descending order.

All four CM-S Curves demonstrate that saturation has been reached (see Figure 3.58). Saturation was reached for variables, links, delays, and feedback loops after 3 to 7 participants. Therefore, overall, saturation has been reached within the ten validation-set interviews from Clinic 6. In other words, this finding indicates that new and relevant data regarding the elements in SMM2 have ceased to emerge. Elements in SMM2 appear to be well-developed.

Figure 3.58 CM-S Curves



3.5.4.2.2 CONCEPTUAL MODEL SATURATION DIAGRAMS

CM-S Diagrams highlight the relative level of conceptual saturation for each relationship of SMM3 identified in model-validation-set interviews; using a thicker line to indicate a higher percentage of participants identifying a relationship.

Figure 3.59 visualizes the percentage of participants who identify each link. 80% of participants mentioned the new “Trust” loop. 20% of participants mentioned the new links involving “continuity of care (from visit to visit).” 16% of the relationships were each mentioned by only one participant. Three links were not mentioned by the validation-set interviews (blue dashed arrows); these were kept as explained in Section 3.5.4.2.1.

Figure 3.60 visualizes the number and type of professions identifying a particular link. It shows how understanding of relationships is shared across professions: 16% of relationships were each mentioned by participants from only one profession, 26% by two professions, 46% by all three and 10% by none. In total, MAs did not mention 36% of relationships, MDs and Managers did not mention 26%.

The validation-set has **perceived 90% of the same system** as the *model development set*. These results indicate that most relationships in SMM2 are well established. I reviewed the relationships that were less well established via CM-S (thin lines on Figure 3.59) using Causal RIQs presented in the next section and concluded that each relationship is plausible, that it had not been distorted, but that it was one that was just less understood by participants and thus mentioned less often.

Figure 3.59 CM-S Diagram – Respondents Identifying

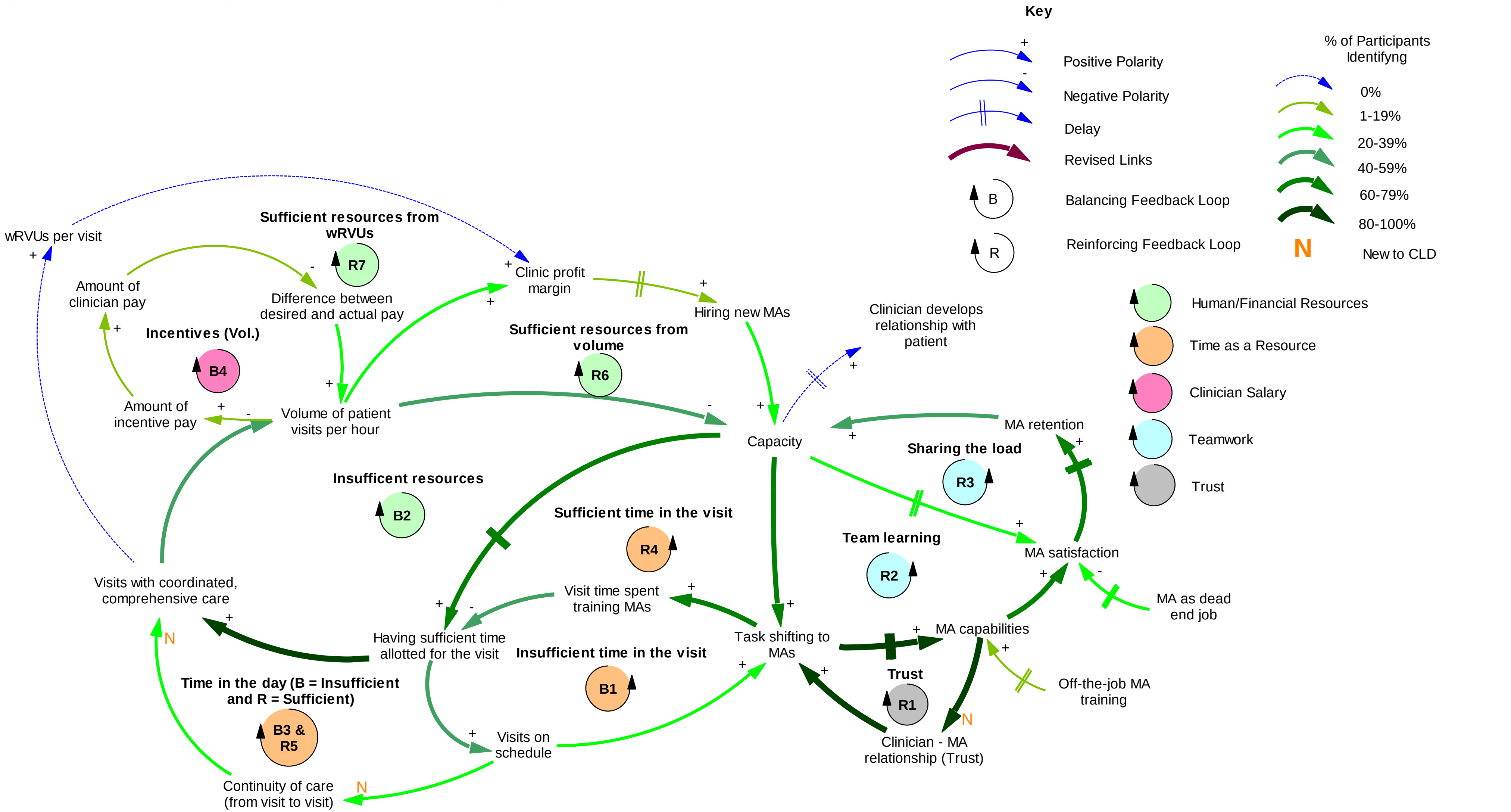
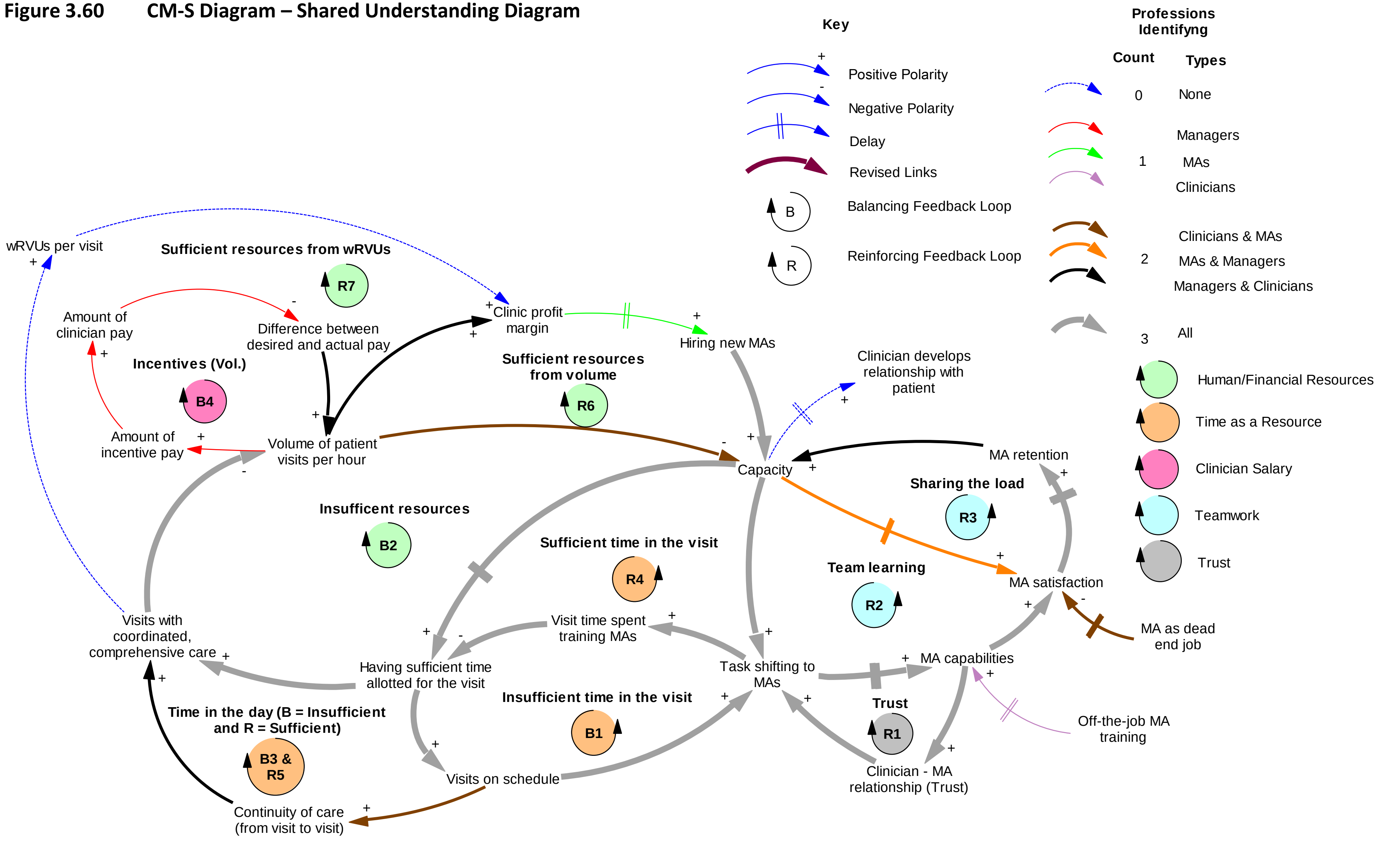


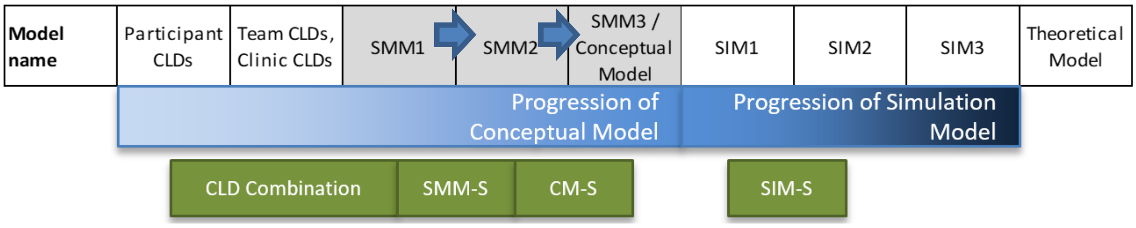
Figure 3.60 CM-S Diagram – Shared Understanding Diagram



3.5.4.2.3 RIGOROUSLY-INTERPRETED QUOTATIONS FOR CONCEPTUAL MODEL SATURATION

Rigorous analysis of causal statements consisted of using Causal RIQs (*Rigorously-Interpreted Quotations – for Causality*) to subject the *variables and relationships* in SMM2 to interview transcripts set aside for validation. Detailed results for this step are presented in Appendix F. Here, I only note that the Causal RIQs demonstrate that SMM2 is capable of exposing variables and relationships in validation-set participants' mental models**.** In so doing, this evidence builds confidence in the causal structure of SMM3.

3.5.4.3 STAKEHOLDER DIALOGUE

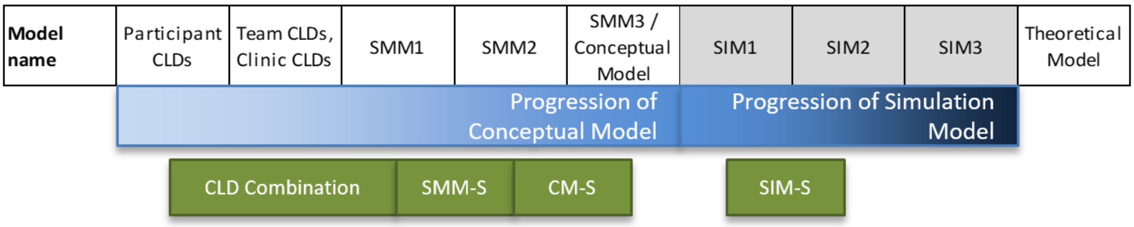


This section reports the cumulative results of stakeholder dialogues to this point in modeling. Stakeholders were shown the shared mental model (the version of the shared mental model that was in existence when each discussion was held). Stakeholders were encouraged to question the CLD structure and point out flaws (based on their level of experience and proximity to the system in question). Three examples of what was learned are presented below:

- **HSDO management personnel pointed out that** from a clinician perspective, everyone who needed to be seen that day would be seen that day "PERIOD" – in other words, getting behind doesn't mean that patients don't get seen. For managers, this would mean that clinicians never get behind, but in actuality they do get behind. It's just that they find a way to circumvent the delay before the day is over, for example: clinicians are able to do less during the visit for each person, they stay late to finish all their patients, and/or continuity with the clinician is broken (the patient ends up seeing a different clinician instead). This stakeholder recommends changing the structure related to "Visits on schedule" to clarify this point. This structure was verified during CM-S.

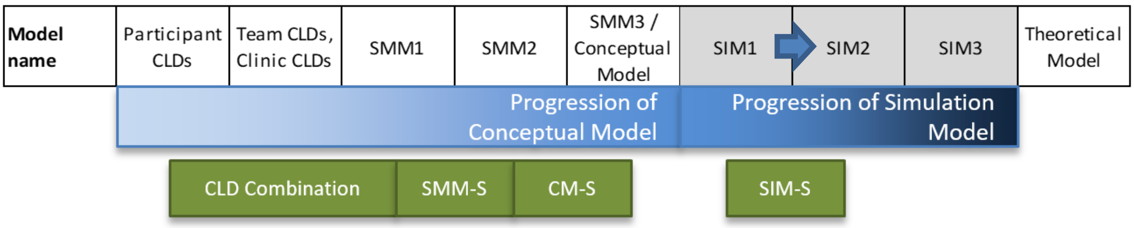
- **SDM experts in the health domain area helped revise the variable names used in the SMM.** For example, the variable "Capacity" had previously been named "Care team ratio", as that is how it was labeled by participants. This revision did not change the structure of the model in terms of the relationships between variables, or the conceptual definition of the variable. Although participants appreciated that a dynamically-important aspect of capacity is that each person has differing capabilities as they learn over time, the way of talking about team capacity emphasized the number of MAs per clinician. In mathematical language, this ratio of personnel did not capture the full participant descriptions. These stakeholder discussions helped to revise the variable name to capture the larger conceptualization (in this case, capacity, due to people and capabilities). This choice was verified during CM-S.
- **International health systems expert helped to identify "task-shifting" as the appropriate term for what used to be a more** complicated, verbose term that listed several aspects of task-shifting all in one variable. This expert pointed out that this sounded like "task-shifting" – a common term used in international health systems work but one that was not commonly used at the HSDO. I renamed the variable.

3.5.5 VALIDATION IN DEVELOPING THE SIMULATION MODEL & POLICY ANALYSIS (PHASES 3 & 4)



This section presents the validation results obtained in building the Simulation Model: SDM Standard Method Validation Tests, Simulation Model Saturation (SIM-S), Stakeholder Dialogue.

3.5.5.1 SYSTEM DYNAMICS MODELING STANDARD VALIDATION METHODS RESULTS



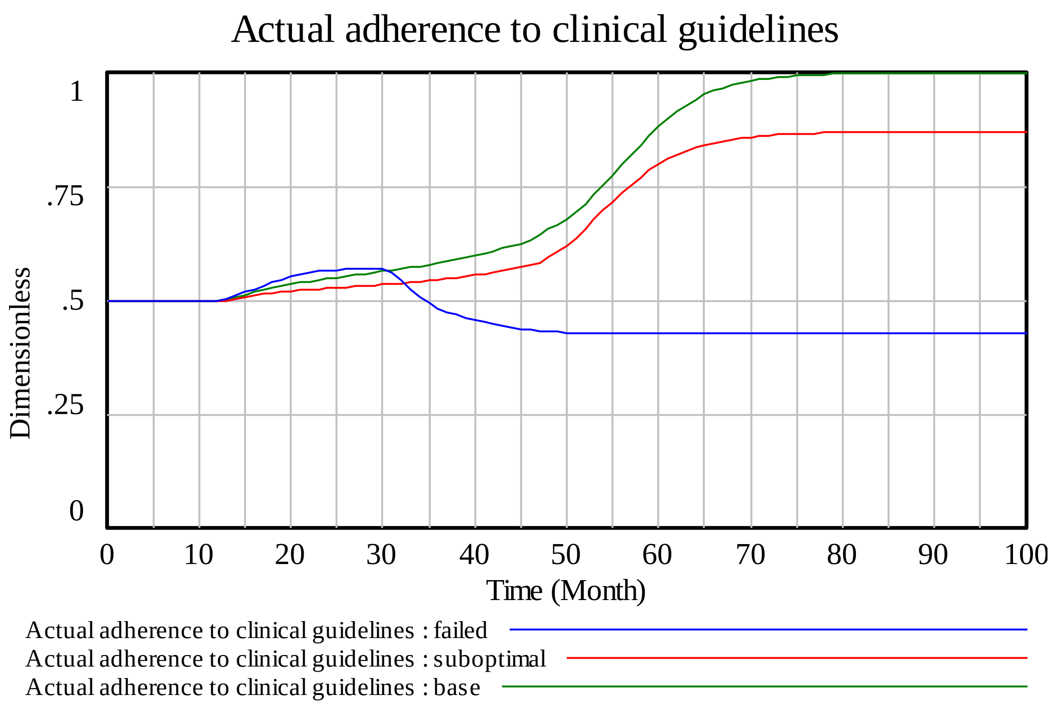
SDM standard validation method were applied to the simulation model and their results are presented in this section.

3.5.5.1.1 BEHAVIOR REPRODUCTION (REFERENCE MODE)

Figure 3.61 presents simulation runs for the "level of adherence to clinical guidelines" outcome measure. As evident in this figure, the simulation model is capable of reproducing the reference mode. The green line displays behavior that corresponds to the "Successful PCT" trajectory on the reference mode. The red line displays behavior that corresponds to the "Suboptimal PCT" trajectory. The blue line displays behavior that corresponds to the "Failed PCT" trajectory.

There are many combinations of policies and environmental conditions (i.e., preferences) that reproduce some variant of the three reference mode trajectories. For example, Figure 3.61 below shows how the three modes can be reproduced using the base case scenario for task-shifting and one of three different settings[39] for the clinicians' preference for how to increase comprehensiveness in terms of non-technical tasks when clinical staff members are perceived to be getting behind on their work. The green line was observed using the "original" setting. The red line was observed using an alternative preference where clinicians are more hesitant to adjusting the inflow of non-technical tasks as the MA's Backlog of tasks changes relative to the "original" table function shape. The blue line was observed using an alternative preference where clinicians are quicker to make this adjustment. Being too eager, too hesitant and somewhere in the middle on this variable makes a big difference in this scenario.

Figure 3.61 Simulation Model Reproducing the Reference Mode



[39] The specific preference is "effect of MA Backlog on inflow of nonTech tasks". The green line is generated using the "original" setting, the red line uses "ALT3" (more hesitant) and the blue line uses "ALT1"(more eager). See Appendix G for the original and alternative preference table functions.

3.5.5.1.2 EXTREME POLICY

For policies that had a range of possible values, I implemented values for each policy that represented extreme conditions and checked that the model behavior was reasonable. Take the "kickstart amount" policy – in other words, the amount of task-shifting initially required by clinic management and implemented by clinicians) at month 10 in the model. The model uses a value of 0.1 or 10% increase in the amount of task-shifting. The range for this policy is 0% (no task-shifting requirement) to 100% (where maximum task-shifting is immediately implemented). 0% leads to the equilibrium *No Task Shifting Scenario* (the amount of adherence to clinical guidelines does not change). A value below 0.1 delays the trajectory but 100% adherence is still attained. The lower the value, the more delayed the outcome. Figure 3.62 below presents the *Successful Primary Care Transformation* scenario trajectory (blue) as well as trajectories for Kickstart equal to 0.06 (green) and 0.01 (red).

Figure 3.62 Extreme Conditions Test: Policy of Low Kickstart Amount

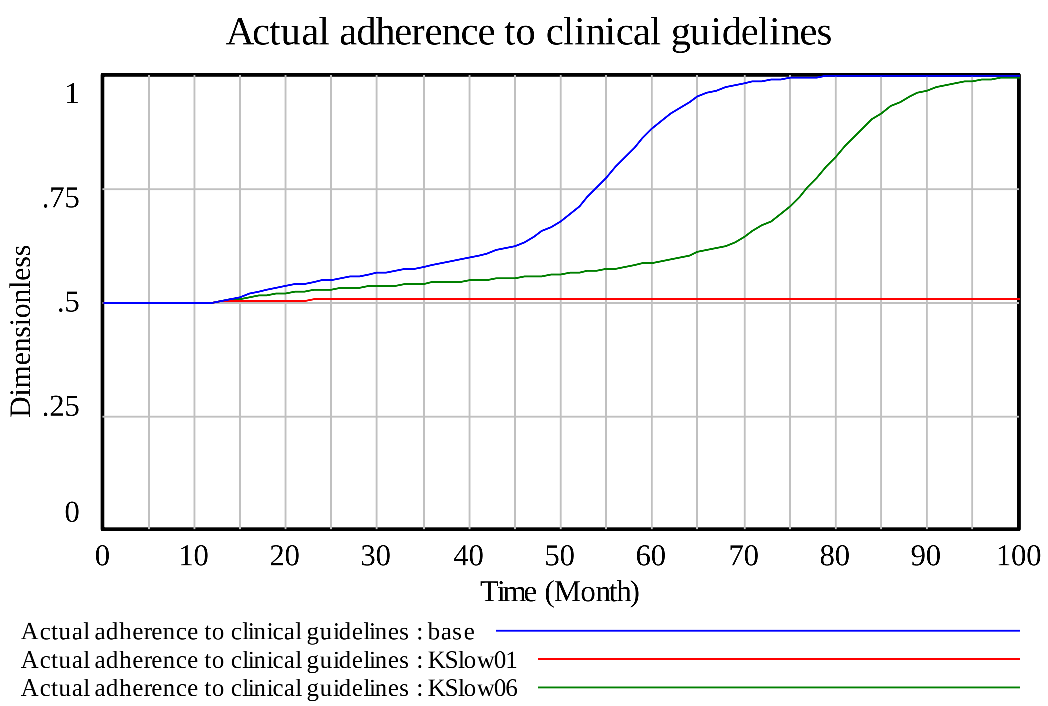
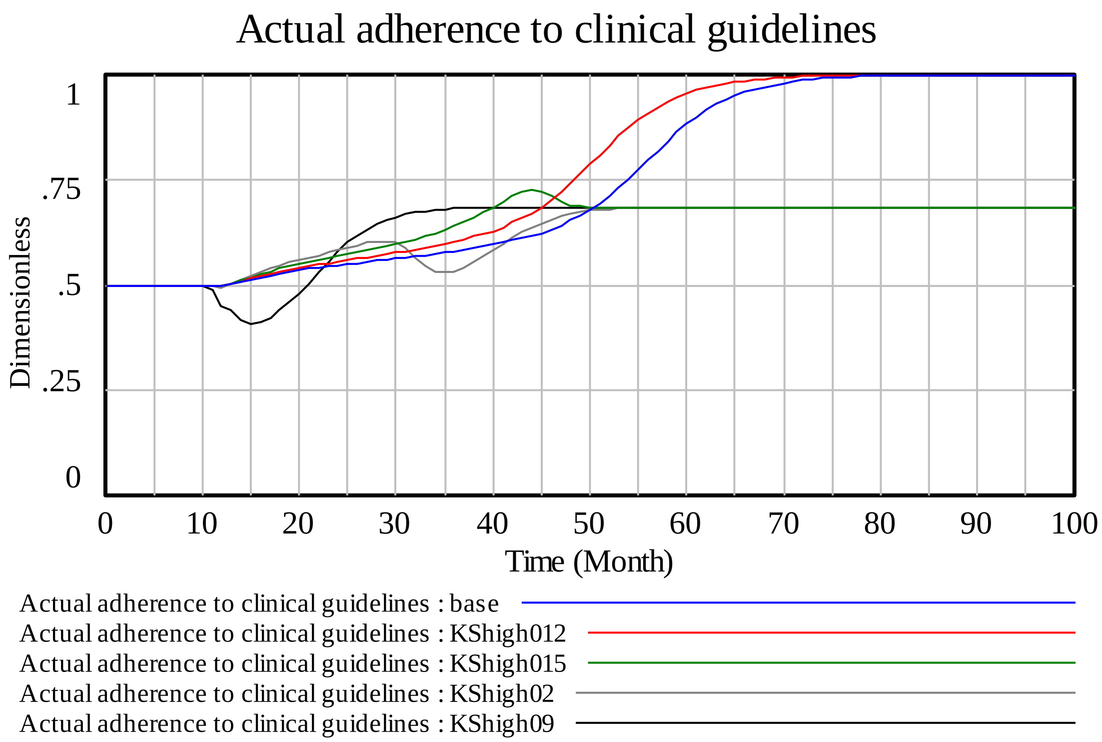


Figure 3.63 below presents the simulation runs for kickstart amounts that are higher than the one used in the *Successful Primary Care Transformation* scenario, with subsequent runs representing more extreme conditions for this policy. The higher the kickstart amount, the quicker the successful outcome is obtained (see red run, kickstart amount= 0.12). However, beyond a certain point, a suboptimal outcome is reached (see green line, kickstart =0.15 and grey line, kickstart = 0.20). Under extreme conditions, when the care team starts implementing a high amount of task-shifting, the team experiences worse-before-better behavior in this outcome, where the final result is suboptimal (see black run, kickstart = 0.90).

Figure 3.63 Extreme Conditions Test: High Policy of Low Kickstart Amount



For policies that are on/off, there were no extreme conditions tests (just 0 = off and 1 = on).

3.5.5.1.3 MODE REPRODUCTION ABILITY

I implemented several past policies in the model to see if it was able to reproduce behavior consistent with historic behavior. The model was able to reproduce them as shown in Section 3.5.5.2.

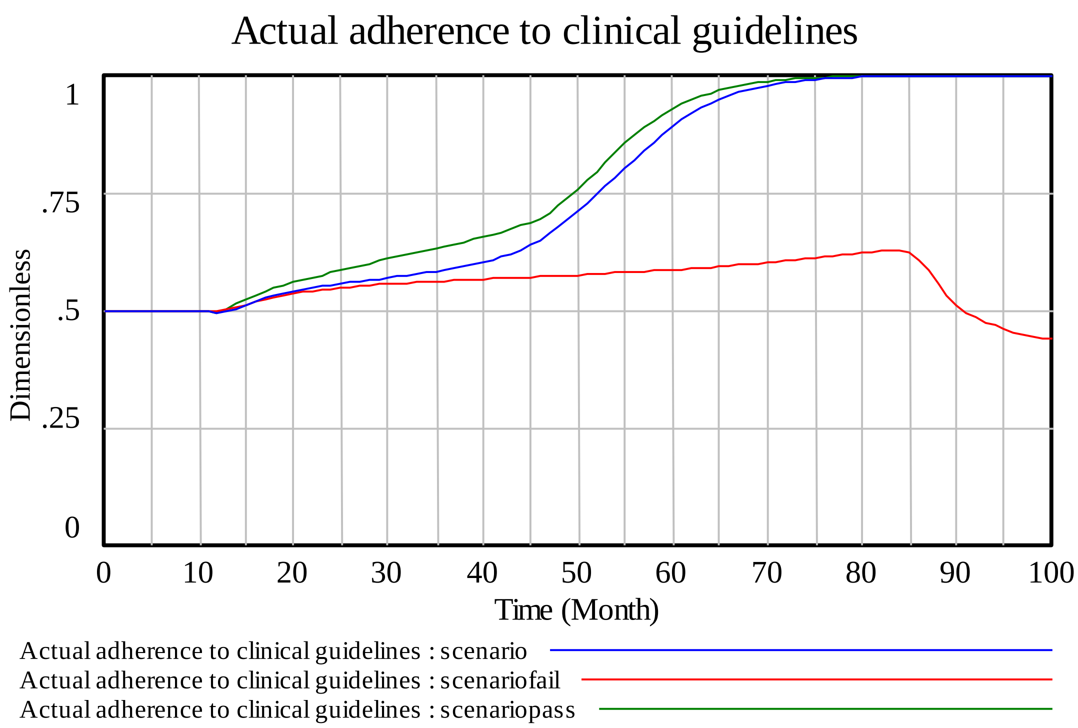
3.5.5.1.4 BEHAVIOR PREDICTION

The model is able to reproduce the anticipated behavior for future/hypothetical situations. Here is one example. This scenario involves having clinicians that are more hesitant to shift tasks (table function "Alt2" is used for "effect of MA Capability on MD willingness"). The predicted behavior would be sub-optimal or even failure of PCT (depending on the degree of hesitation).

Clinicians perceive the increase in MA capabilities but rewards it with a smaller increase in the amount of willingness given for the observed increase in capabilities (smaller over the base-case with task-shifting). This results in some improvement over time but, ultimately, the system collapses into a failure mode (red line). This is because clinicians are getting more and more backed up as they shift some tasks and train MAs on those tasks but do not benefit enough (in increased capacity) from the meager amount of tasks that MAs have taken over. Their patients perceive clinicians' increasing backlog (in technical tasks) and become less and less satisfied. As the level of patient satisfaction becomes less acceptable to the clinicians, they reduce the amount of tasks that they shift. This increases their backlog (as they now are to complete these tasks for the patients), negatively impacting their productivity as they are overworked at this point. This exacerbates the situation and the result is a level of adherence to clinical guidelines that is lower than the initial.

In this situation (see Figure 3.64), if the clinic management requires a higher level of initial task-shifting (for example, kickstart amount = 0.2), the successful trajectory is obtained (and even quicker than in the base-case; see green line). While, in the base-case, this high a kickstart results in Suboptimal PCT; in this scenario, this increased amount of task-shifting at the beginning compensates for the clinician's hesitation, resulting in Successful PCT.

Figure 3.64 Behavior Prediction Example

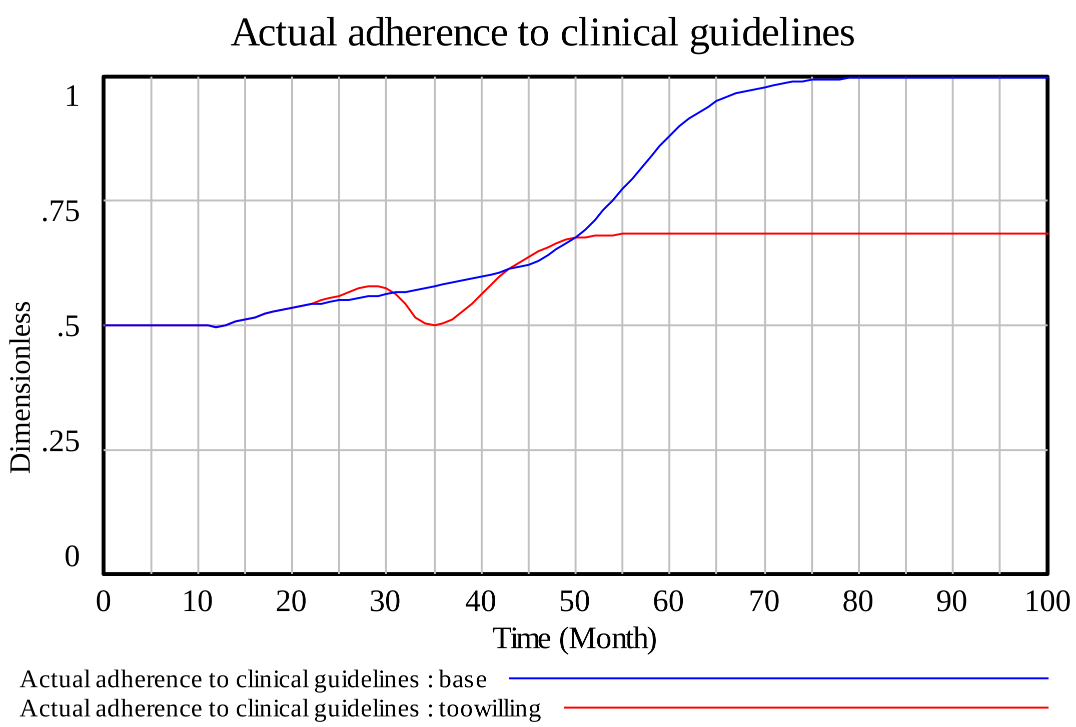


3.5.5.1.5 ANOMALOUS/SURPRISE BEHAVIOR

In the course of building the model and in model analysis, the model was run under various conditions. At times, *surprise* behavior in certain model variables was generated in that it was behavior that I did not expect. Upon closer examination of the variables feeding into the variable(s) that appeared to have surprise behavior, the causes of that behavior were identified. The causes were the behavior of those other variables feeding in as directed by the model structure. The surprise behavior was explained and consistent with what is expected given this closer examination – it was no longer surprising.

One surprise behavior (for me) was that it is possible to have clinicians that shift tasks too quickly (that are too willing to shift) such that PCT results in suboptimal adherence to clinical guidelines (see red line in Figure 3.65 below). The team becomes overwhelmed and dissatisfied; the MA turnover increases such that the team does not reap the full benefits (e.g., increase of capacity) offered by task-shifting, because the team spends that excess time training (the new MAs hired due to the higher turnover previously mentioned).

Figure 3.65 Surprise Behavior



3.5.5.1.6 DIMENSIONAL CONSISTENCY

All of the variables in the SD simulation model have been labeled with units. When variables are used in an equation, Vensim computes the expected units of the outcome and compares it to the unit that the modeler has given the outcome (the units label). If this does not match, Vensim returns an error message. Dimensional consistency requires that the units match. This model passes dimensional consistency as confirmed by the Vensim "check units" feature [177].

3.5.5.1.7 EXTREME CONDITIONS IN EQUATIONS

Unlike the "extreme policy" test which uses extreme values for policies to test model behavior, the extreme conditions test uses extreme values for other variables in the model. The simulation model passes this test. For example, I entered extreme values for the "initial ratio of MD to MA" (which governs the number of MAs initially on the care team). I entered 0.33 (meaning that there were 6 MAs working with the 2 clinicians from the beginning). This is possible but unlikely in initial conditions. Running the model produced the successful trajectory for adherence to clinical guidelines and more quickly than the base-case (see red line, Figure 3.66). No hiring was needed in this scenario. The MAs were less productive in initial conditions (due to having excess capacity) and as the clinicians shifted more tasks to them, not only were they able to keep up but they were able to complete training tasks more quickly – more quickly leading to more capable MAs (see three productivity graphs in Figure 3.67 below).

Figure 3.66 Extreme Conditions – Initial MD to MA Ratio – Actual Adherence

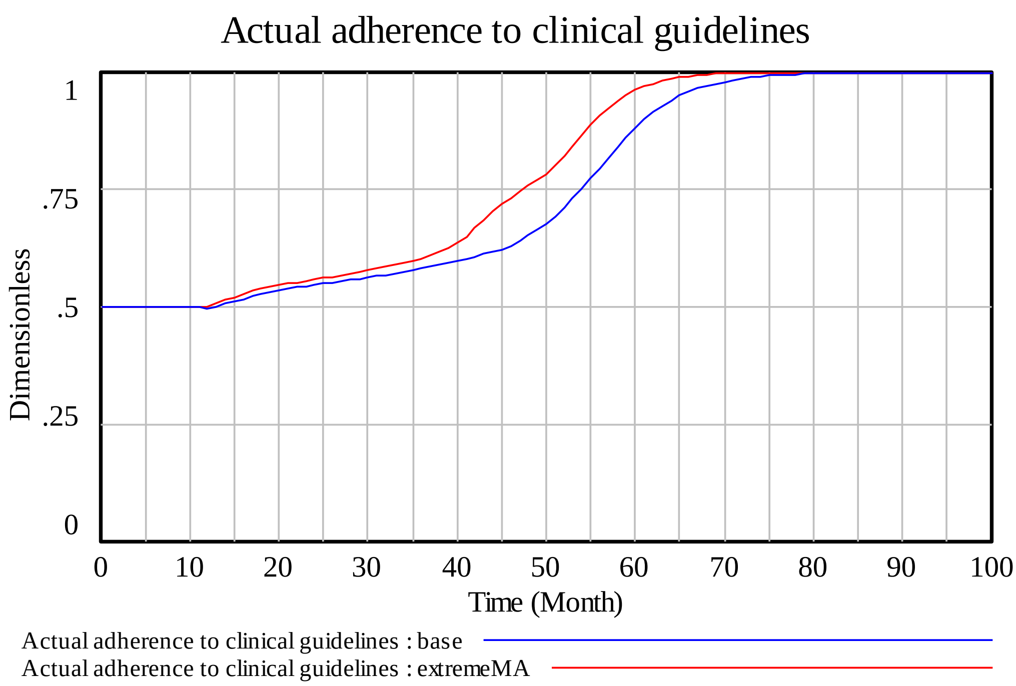
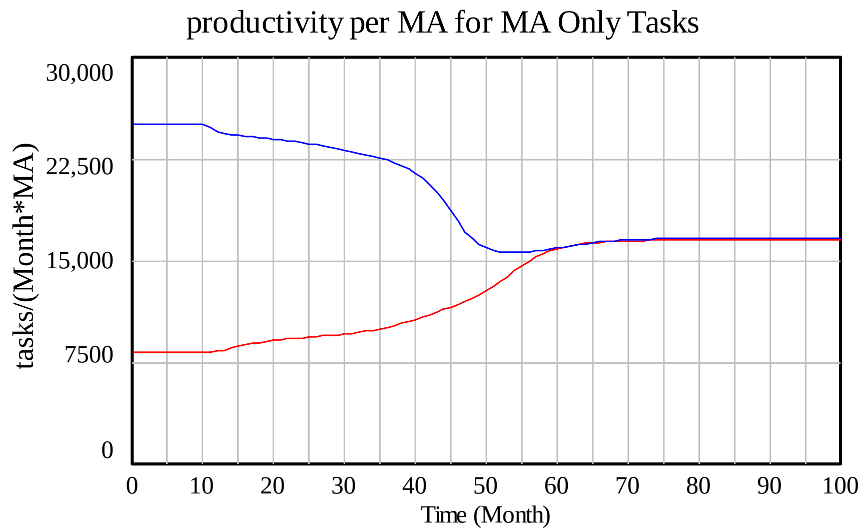
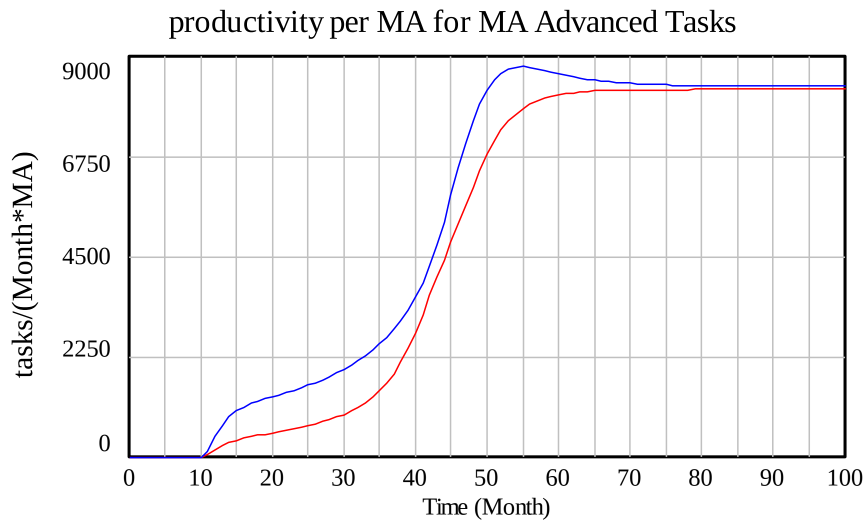
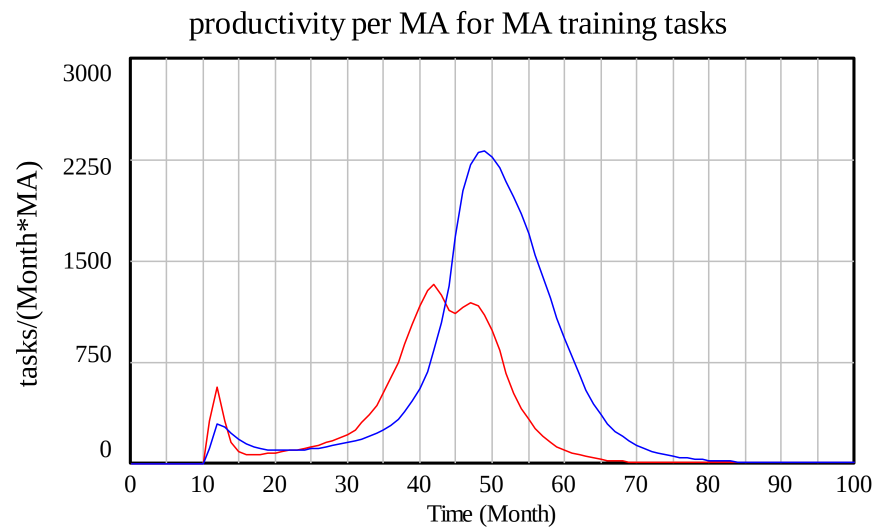


Figure 3.67 Extreme Conditions – Initial MD to MA Ratio – Productivity







Key:

productivity per MA for MA Only Tasks : base

productivity per MA for MA Only Tasks : extremeMA

3.5.5.1.8 BEHAVIOR SENSITIVITY

I implemented sensitivity analysis without policies turned on and found that varying most of the constants did not significantly change model behavior. For these variables, the estimate used in the model should be adequate.

The constants that did change the model output fell into two categories: (1) behavior sensitivity was observed, as expected and (2) behavior change was observed because the model was not placed in analytic equilibrium. In this latter case, changing the parameter value would require changing the initial values for other parameters. Because an analytic equilibrium model was outside of the scope of this dissertation work, a systematic testing and documentation of this issue was not performed. See Appendix G for more detailed results.

In all variables with behavior sensitivity, additional data collection would be useful to narrow their uncertainty.

3.5.5.1.9 BEHAVIOR BOUNDARY ADEQUACY/STRUCTURE SENSITIVITY

I implemented multiple structures thought to be equally-likely in the model during model development. As I tested each one, those producing behavior that did not match that described by interviews or operational thinking were abandoned. Also, those structures that produced an overly-complex model were abandoned for simpler ones when the simpler structures resulted in the same behavior. (Model iterations not shown.)

I modified the model to include plausible additional structure as I relaxed boundary assumptions and I observed the resulting behavior. One example is when I added the feedback between service delay and clinician willingness to task shift (via the impact on perceived patient satisfaction). Patient satisfaction was not found in the Conceptual Model. The closest variable in the Conceptual Model was “Clinician develops relationship with patient”. This variable was not modeled until this test as interviews were unclear what structure it would have.

When this test was performed, I put the question: Are there “potentially important feedbacks omitted from the model” that were “important given the purpose of the model”?[113] (p. 861). I realized that this variable could feed back into the model and created the structure to test this hypothesis. PCT could have negative impacts on the patient-clinician relationship and clinicians could feel justified in resisting PCT for this reason. Specifically, I looked at the Conceptual Model and realized that the link between “Capacity” and “Clinician develops relationship with patient” is closely related to the chain of links between “Capacity” and “Visits on schedule”. It is the on-time delivery of services that pleases or displeases patients (i.e., patient satisfaction) – influencing the clinician-patient relationship. Descriptions of these links were often framed in terms of efficiency, rather than in terms of the clinician-patient relationship. Being inefficient could impact the clinician's reputation, I realized. What was missing was a variable showing patient satisfaction and how clinicians perceive that and what they do about that.

Indeed, introducing such a structure altered the dynamics of the model to better reflect the problem statement, in relation to the reference modes. This structure is a key mechanism whereby clinicians determine when they should not only stop but pull back on the task-shifting already made.

3.5.5.1.10 POLICY SENSITIVITY

Management policies were the focus of Policy Sensitivity. I assessed the sensitivity of management's kickstart policy (0.1) to varying the constants in the model (note this is with fixed salary since this is the default payment plan in the base-case scenario). The same methods were used as in Behavior Sensitivity, only this time the task-shifting policy was turned on. I also assessed the sensitivity of management's clinician payment policies, considering the question: under a kickstart of 0.1, what if clinicians were compensated using each payment policy.

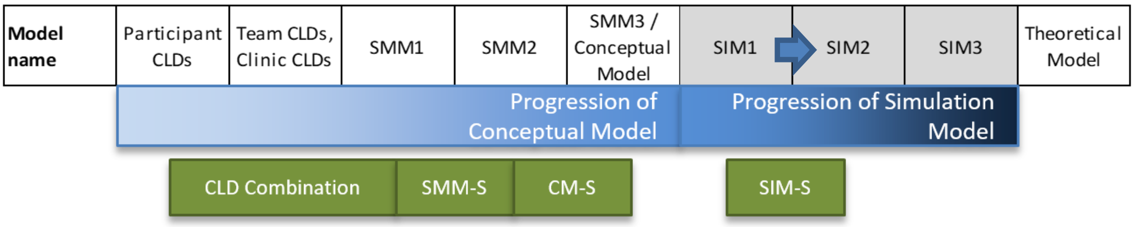
I found that varying most of the constants did not significantly change model behavior in these policy runs. The constants that did change the model output, fell into three categories: (1) no behavior change due, (2) policy sensitivity was observed, as expected, (3) behavior change was observed because the model was not placed in analytic equilibrium. As described in behavior sensitivity above (Section 3.5.5.1.10), for the second category, additional investigation is warranted. See results in Appendix G for detailed results.

Future policy design might consider ways to influence sensitive variables.

3.5.5.1.11 POLICY BOUNDARY ADEQUACY

In Section 3.5.5.1.9 above, I describe the addition of patient satisfaction to the model. This structural change also tested the sensitivity of *policy recommendations* to changes in the model boundary. I found that the addition of this structure did change model behavior such that policies had a different effect than they did previously. The reason they had a different effect was that adding the feedback of patient satisfaction to clinician willingness further slowed down the increase in task-shifting. This slow-down allowed the team to keep up with training and patient care tasks (thus keeping clinician salary above what it would otherwise be), reducing the feeling of being overwhelmed (and thus MA turnover) while continuing to grow capabilities (and thus not losing MA capabilities gained due to what would have been higher turnover).

3.5.5.2 SIMULATION MODEL SATURATION



SIM-S has been reached when it has been demonstrated that the simulation model does not contain radical departures from the Causal RIQs (*Rigorously-Interpreted Quotations – for Causality*). This question was considered from the perspectives of structure, behavior and culture (as described in Appendix E). The following sub-sections present SIM-S findings demonstrating that the SIM-S Test has been passed.

The simulation model contains the elements identified in the Conceptual Model. While writing equations for the links involves specifying each element in greater detail, the simulation model contains the same essential causal chains and tells the same stories as the Conceptual Model. The simulation model was found to be within-scope of the problem statement. Figure 3.61 above demonstrates that the simulation model is capable of reproducing reference modes identified in the problem statement and drawn in Figure 3.13.

Structures inside SIM1 were compared to the structural concepts referred to in the Causal RIQs. Original Causal RIQs are presented in Appendix F. I found that, in all cases, the variables and equations in the simulation model are in line with the descriptions in Causal RIQs. While at times more or less aggregated than the quotations, the level of detail in the simulation model was sufficient to capture the different dynamics described in the quotations. The time frame was sufficient to capture dynamics described in quotations. The causality expressed in quotations is evident in the simulation model. Omissions or distortions were not found.

In the Causal RIQs, participants describe real-world scenarios and policies. These are also found in the simulation model. When these scenarios and policies were run in the simulation model, as they were described in the quotations, the model was able to generate the expected results (as described in these quotations). However, these results were obtained under specific circumstances for parameters that the individual quotation did not address. Varying these latter parameters produced results other than those expected in the quote.

Finally, in performing the tasks described above, I found that the goals and policies described in the quotations are also found in the simulation model, and that they are true to the cognitive limitations inferable from the quotations. Goals are in terms of the ones that the target group would actually desire. Policies are in terms of those that the target group could actually implement. Decision functions[40] are based on information the target group actually sees.

Following are *examples* of how concepts in specific quotations are represented in the model accompanied by the model-generated behavior. Quotations are separated into the following four sub-sections (and are presented using the same quotation numbers as for the Causal RIQs presented in Appendix F):

> Section **3.5.5.2.1** presents quotations for which the identified structure is found in the model and the identified behavior is reproduced by the model
> Section **3.5.5.2.2** presents quotations for which the identified structure has pieces that are found in the model as well as pieces that were deemed outside the model and the identified behavior is reproduced by the model
> Section **3.5.5.2.3** presents quotations for which the identified structure is found in the model, but was incomplete in the Conceptual Model, and the identified behavior is reproduced by the model
> Section **3.5.5.2.4** presents a general discussion of the ability of the simulation model to represent the reference modes in Causal RIQs.

[40] The term decision function is used interchangeably with the terms policy function and operating policies in system dynamics and in this dissertation.

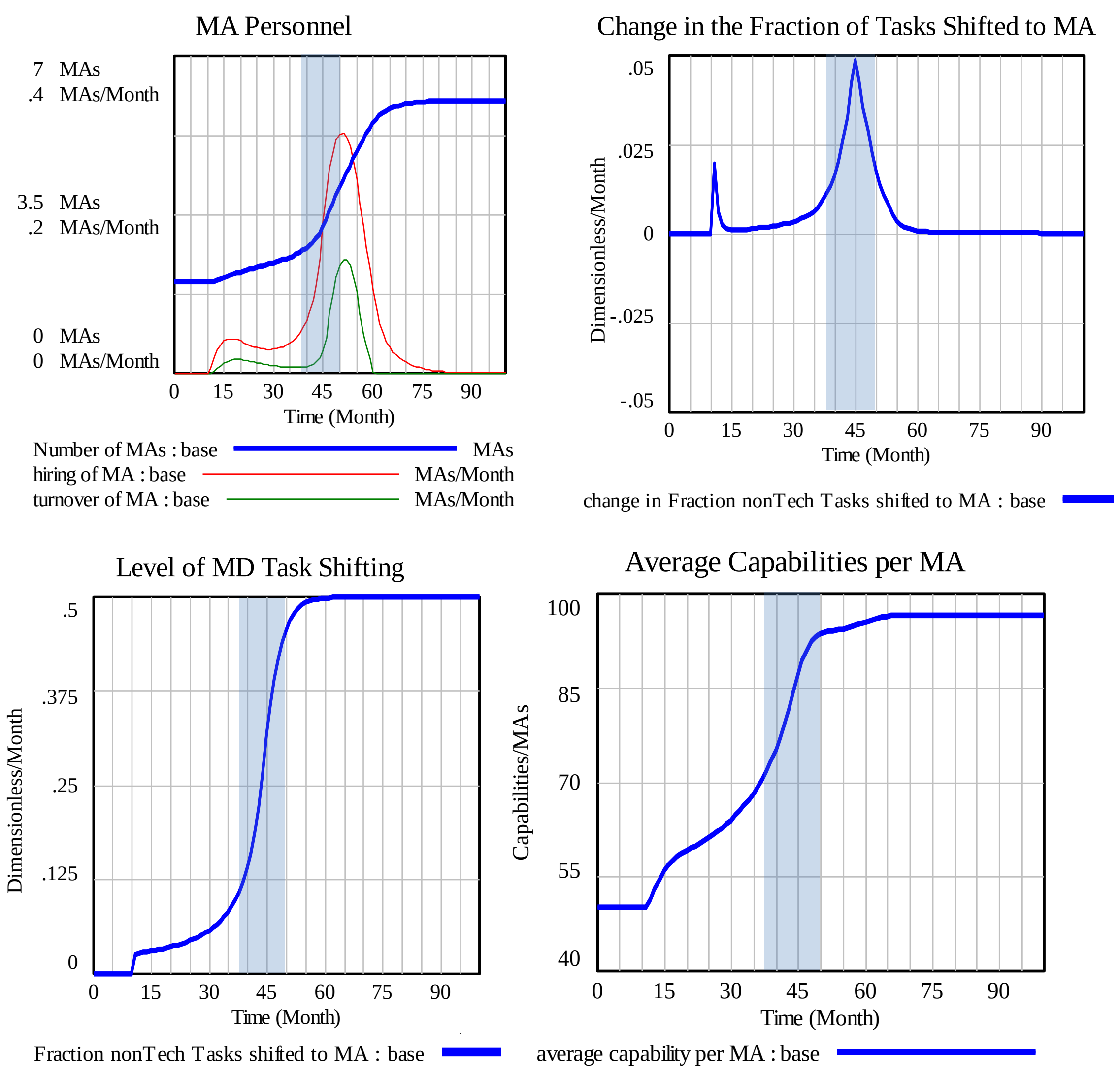
3.5.5.2.1 SIMULATION MODEL REPRODUCES SCENARIOS IN QUOTATIONS

3.5.5.2.1.1 QUOTATION CL03-20

In this quote, *CL03* describes the positive impact of an increase in the MA to clinician ratio on task-shifting. CL03 also affirms that their relationship has noticeably improved in this timeframe – "*huge improvement*" and "*a lot better*". Furthermore, CL03 ascribes the staffing level to be around 2MDs:4MAs and 2MDs:3MAs. In the base-case, the simulation model shows that the tipping point for task-shifting is in the center of the staffing range that CL03 had identified; between months 37 and 50 (see blue shaded regions in Figure 3.68).

In this same quote, CL03 then goes on to say that the MAs are "able to accommodate", they know what to do. In other words, they have reached a high level of capability. The simulation model also matches this described behavior as the "MA Capabilities" complete their growth during this period, reaching close to the maximum level (Figure 3.68).

Figure 3.68 Simulation Model Saturation for CL03-20

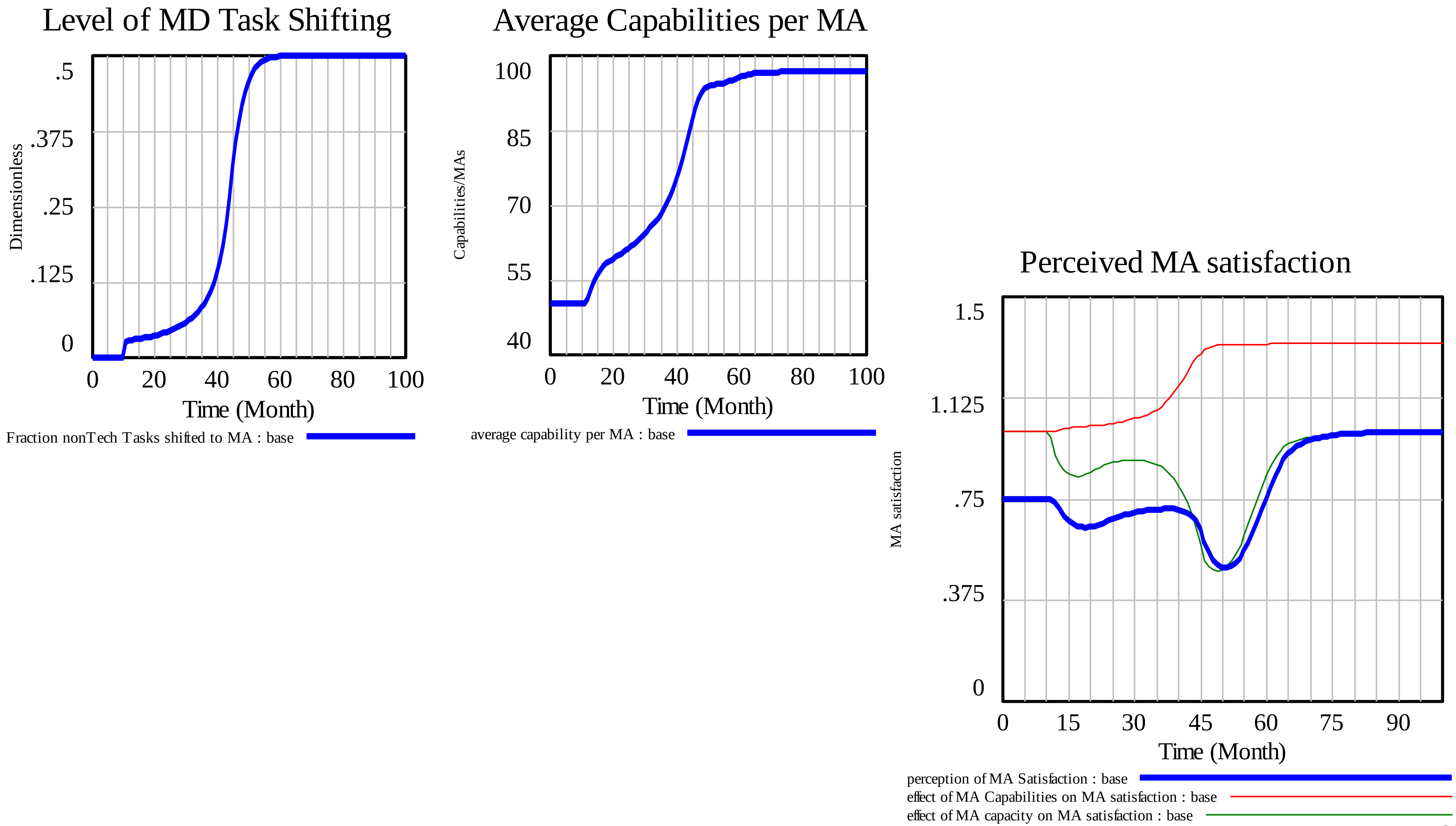


3.5.5.2.1.2 QUOTATION MA03-07

This quote describes the increase in MA satisfaction that comes with task-shifting, for MAs that desire to take on this expanded role. The Simulation Model is able to represent elements described in this quote as follows:

- Structure
  - This MA desires to take on this expanded role. MAs in the model base-case also desire to take on this expanded role as expressed via the "effect of MA Capabilities on MA satisfaction" (a table function shown in Figure 3.20 above). Alternate forms of this table function represent scenarios where the MAs have other preferences about their expanded role.
  - This MA also indicates that "Task-shifting" increases "MA satisfaction". The simulation model also has this link. In the "MA Satisfaction" view of the simulation model, the capabilities gained via task-shifting influence the MA's satisfaction via the "effect of MA Capabilities on MA satisfaction".
- Behavior (Figure 3.69)
  - The base-case scenario in the simulation shows the behavior described in this quotation: as task-shifting increases, the average capability of each MA grows, eventually improving MA satisfaction.
  - The bumps in "Perceived MA satisfaction" are not described in this quotation. They are due to an aspect not discussed in this quotation: the effect of the strain on MA Capacity on MAs' level of satisfaction. This is evident in the graph below. The green line represents this effect in the base run. The effect of MA Capacity on MA satisfaction starts at 1 (not increasing or decreasing MA satisfaction). As MA capacity is strained, the green line dips down. As MA capacity is improved, the green line returns to the original level of 1. The red line represents the effect of MA capabilities on MA satisfaction. This effect also starts at 1. As MA Capabilities increase, the red line increases until it plateaus. This plateau occurs because MA capabilities have reached their maximum level under the base-case scenario.

Figure 3.69 Simulation Model Saturation for MA03-07



3.5.5.2.2 SIMULATION MODEL REPRODUCES THE ESSENTIAL ASPECTS OF SCENARIOS IN QUOTATIONS

3.5.5.2.2.1 QUOTATION MA03-9, CL03-19, 22 THROUGH 24, CM01-57, 60 & 62

Clinicians, MAs and the center manager all reference the need for a shared understanding of PCT in order to succeed at implementing it. While "shared understanding" is not explicitly found in the model, there is structure that captures what these participants describe as such.

CL03 describes how, as clinicians invest time in training their MAs for task-shifting, they build capabilities. While the tasks *are 'X, Y and Z'*, the capabilities are '*how to do X, Y and Z'* – the specific way in which they are done on any given team *are specific to that team* – the way that clinicians on the team prefer to get it done. CL03 states: "I have certain ways that I like doing things" (quote 19). In the successful trajectory, all tasks that can be shifted are shifted; however, how they are accomplished is different on each team. This fact requires team members to have a shared understanding.

Having this understanding requires that it be built. Participants describe this need as "getting it to the point where everybody is on the same page and everybody understands… pieces that they need to do, in order to make it work " (CM01-57); "you have to coordinate and get them all on the same page… [and] change that in their minds" (CM01-60); "to have it … be a team-generated set of goals" (CL03-22).

All of these participants describe how such shared understanding and team-specific capabilities are built:

- CM
  - "resolve concerns and frustrations that they go through on a daily basis to keep the ball moving… communication" (CM01-60)
  - "negotiation to find better ways to deliver care… It might be painful a little bit, but look what could be in the end if we could get there" (CM01-62)
- CL
  - "training to make it a team approach... honest about giving their feedback, about what's working and what's not working..." (CL03-22)
  - "enough time... how to actually implement this on the ground... trying to figure out how to add these things... you need to have a concentrated effort, and training, and have some administrative support..." (CL03-23)

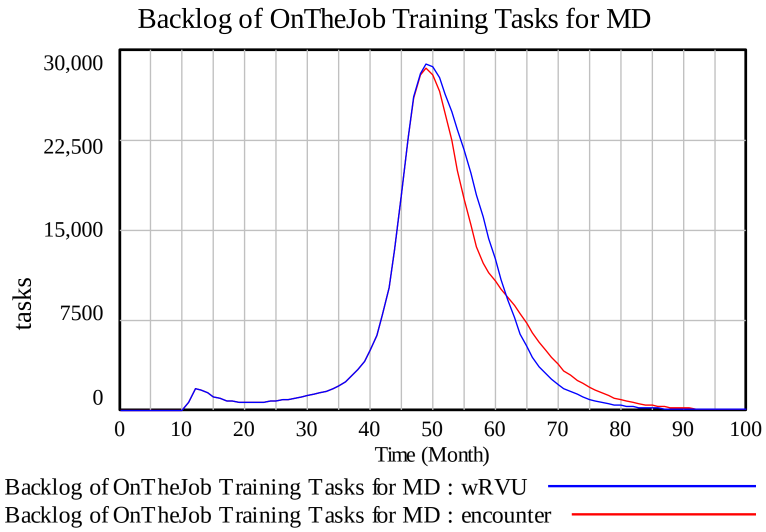
    - “open communication... feel really accountable... and get rewarded when they do well and get really appropriate feedback when things don’t go well.... communication process... huddles and … team meetings and they are time consuming... exercises where people really openly say what they think is working and isn’t working, and how they want, like setting personal goals for what they want to do, and how they want to make things better, and then having someone sort of help and support them with those goals, and doing that all together as a team...” (CL03-24)
- MA
    - “clinicians are very open… to hear suggestions from the MAs, and… work with things that way... If there is a problem, and it’s brought about in a positive, open way… changes are made quickly… we all talk about it and we try and figure out a way to solve it” (MA03-9)

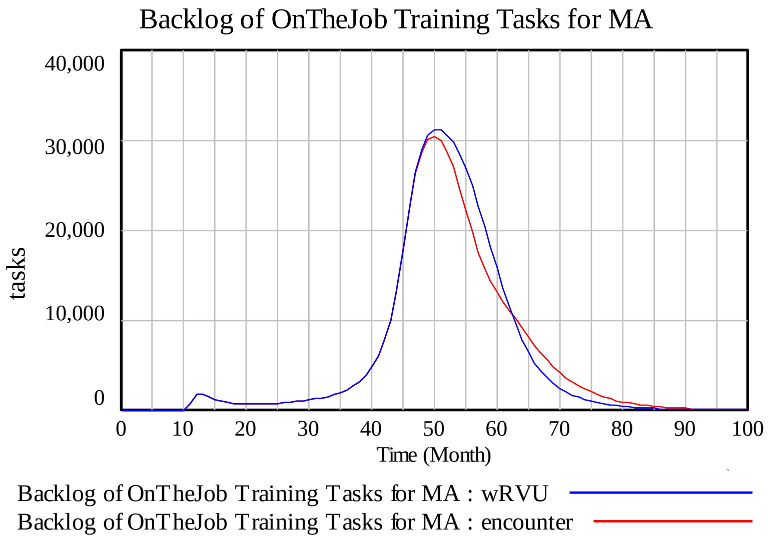
These participants identify the need for process improvement skills as well as problem structuring and dialogue skills to figure out how to successfully achieve PCT. MA03 makes also describes what happens when MDs do not have these skills and prefer to discuss issues with a supervisor instead of with MAs (not reproduced here). While these specific processes are not explicitly found in the model, there is structure that captures what these participants describe. The model behavior also corresponds to participant descriptions.

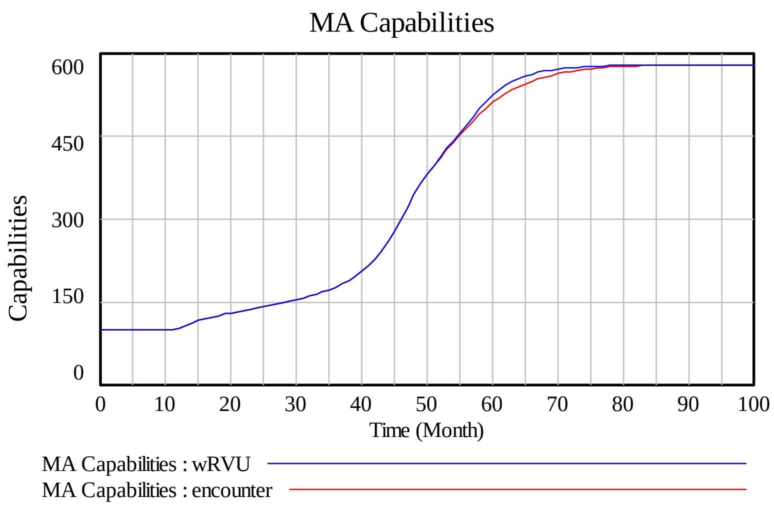
- Structure
    - On-the-job training tasks are generated when clinicians shift tasks. These training tasks are in addition to the tasks that are actually shifted. These training tasks take time. As the team completes these training tasks, the MAs become more capable and have more time to spend completing tasks for patients.
    - What do these training tasks consist of? All those things just outlined above by these participants. Skilled MAs are hired. Part of the on-the-job training is learning the specific skill not previously implemented on the team. Another part of it is figuring out how to do it in a way that works for that team. The model aggregates all of these into “training tasks”.
    - Shared understanding is captured via the stock of “MA capabilities”.

- Behavior (see Figure 3.70)
    - Training tasks accumulate as tasks are shifted. They accumulate both for clinicians and MAs as building shared understanding requires all team members to spend time together (out of the visit), wrestling with how to implement PCT on their team.
    - As training tasks are completed, “MA capabilities” are gained.
    - Worse-before-better behavior is observed for some outcomes of interest, under certain policy conditions; for example, for clinician salary using encounter-based pay or wRVU based pay.
        - Under the encounter based pay policy (red lines), there is a temporary decrease in clinician salary between month 43 and month 90. During this time, clinicians’ salary is reaching below their desired salary. This causes greater hesitation with respect to task-shifting, shown as willingness to shift tasks in the final graph below. This lagging willingness influences the trajectory of training and capabilities.
        - Under the wRVU based pay policy (blue lines), there is also a temporary decrease in clinician salary between month 43 and month 54. This decrease is less dramatic and more short-lived. Also, salary continues to increase after month 54 as salary is based on the higher level of comprehensiveness delivered.

Figure 3.70 Simulation Model Saturation for MA03-9, CL03-19, 22 through 24, CM01-57, 60 & 62







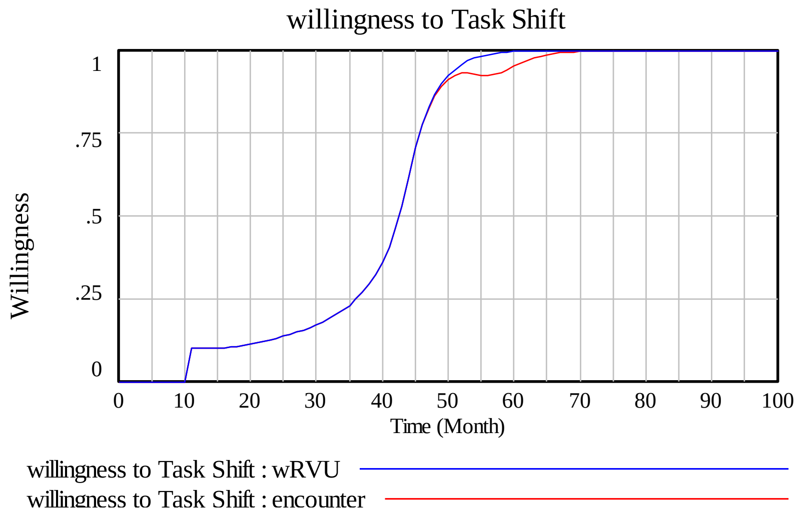


Figure 3.70 Simulation Model Saturation for MA03-9, CL03-19, 22 through 24, CM01-57, 60 & 62 (continued)

3.5.5.2.2.2 QUOTATION CM01-43

In this quote, CM01 describes several variables which are not explicitly-named variables in the simulation model. I now describe how each of these variables is handled by the simulation model.

"Clinician satisfaction" is not explicitly tracked in the simulation model, although its impact is seen in the clinician's willingness to shift tasks. Table functions are used in the simulation model that link three variables that impact clinician satisfaction to the clinicians' ultimate decision of how much to continue shifting tasks. The three variables are MA capabilities, monthly salary, and patient satisfaction. All else equal, if clinicians are dissatisfied with respect to one of these three aspects, that dissatisfaction manifests in a corresponding reduction of willingness to shift tasks.

"Clinician capabilities" is also not explicitly tracked in the simulation model, the various capabilities are considered in a disaggregated manner. These capabilities are visible in all of the clinician table functions' variables (Sections 3.4.2.2 and 3.4.2.6). Each scenario describes clinicians with differing aptitudes for a certain aspect (e.g., ability to perceive growth in MA capabilities). These relate directly to the capabilities which clinicians in PCMH desire (e.g., see description of desire for "team approach" in CL03-24).

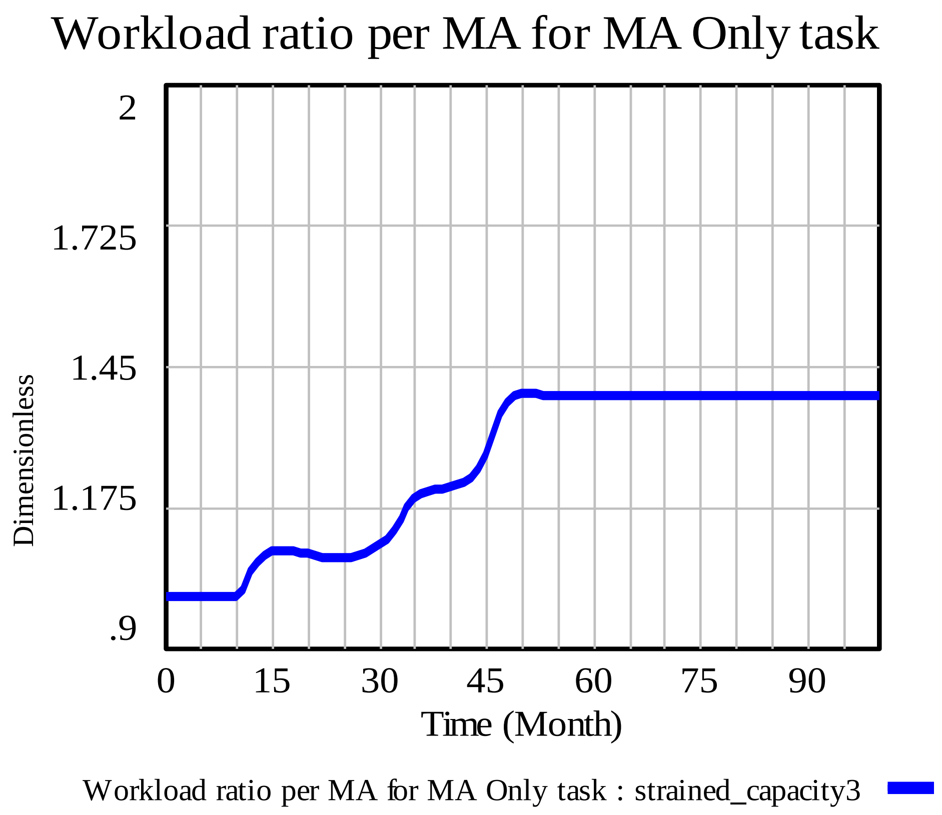
"Clinician turnover" is explicitly out of scope for the simulation model. The number of clinicians is held constant at 2. They are "in charge" of the care team. Clinicians would therefore change their practice before they left.

3.5.5.2.2.3 QUOTATION MA03-04, NM01-36

The MA quote describes the negative impact strained capacity has on task-shifting and on MA satisfaction. The nurse manager quote describes the opposite effect occurring when capacity is not strained, as well as the impact on the team's ability to keep up with their visits. The time frame in the model is longer than the time frame mentioned in the quotes (a day). Also, the system scope (a clinic consisting of several teams) mentioned by the MA is larger than that considered in the model (a single team). That said, the effect of strained capacity over longer periods of time can be explored in the simulation model. When a model scenario shows strained capacity, task-shifting and MA satisfaction decrease, and the team is unable to see the number of visits they could have. This is the case in the scenario below.

- Structure
    - Strained capacity can be introduced in the model in several ways. One way is by increasing the amount of task-shifting initially attempted (managers push for a higher initial level, implemented with a “kick-start” of 0.15 instead of 0.10 in the base-case). The capacity of the MAs to do their job is strained as a higher initial level of task-shifting places more work on MAs’ shoulders. As more tasks are shifted to them, they not only do more tasks but they also spend more time being trained. They are overworked in all the types of tasks that they are asked to do. Figure 3.71 is an example of this. A workload ratio of 1 means that they MAs are getting the amount of work that they are normally able to do (above 1 means they are asked to do more than they are normally able to). It shows the strain in MA only tasks as the workload ratio increases to 1.4 and remains chronically high.
- Behavior (Figure 3.71)
    - Both quotes describe MAs feeling stressed under strained capacity. Perceived MA satisfaction shows this phenomenon in the green and blue lines. The green line shows the decrease in satisfaction due to the strained capacity. The blue line shows that overall MA satisfaction also decreases due to this strain.
    - MA03 describes this as a “*kink in the day”*. Even though the level of MD Task-shifting reaches 0.5 (the maximum) the team’s actual adherence to clinical guidelines reaches an improved but suboptimal level due to this strained capacity. The actual adherence graph shows a kink around 45 months. The perceived MA satisfaction curve also shows the final downward trend starting at that time.
    - NM01 describes this as causing “real contention… when people get behind”. Under strained capacity, the number of encounters per month fails to keep up with what would have been required to reach 100% adherence to clinical guidelines. The MD attempts to achieve this, but is chronically unable to do so.

Figure 3.71 **Simulation Model Saturation for MA03-04, NM01-36**



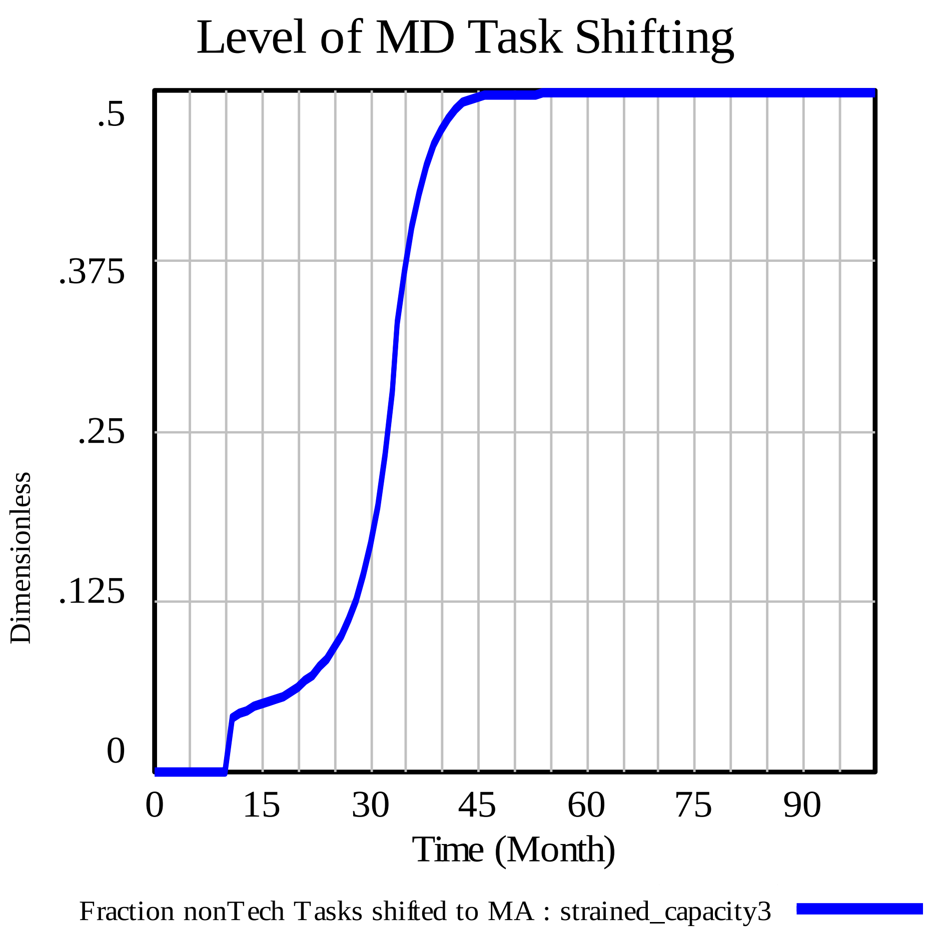
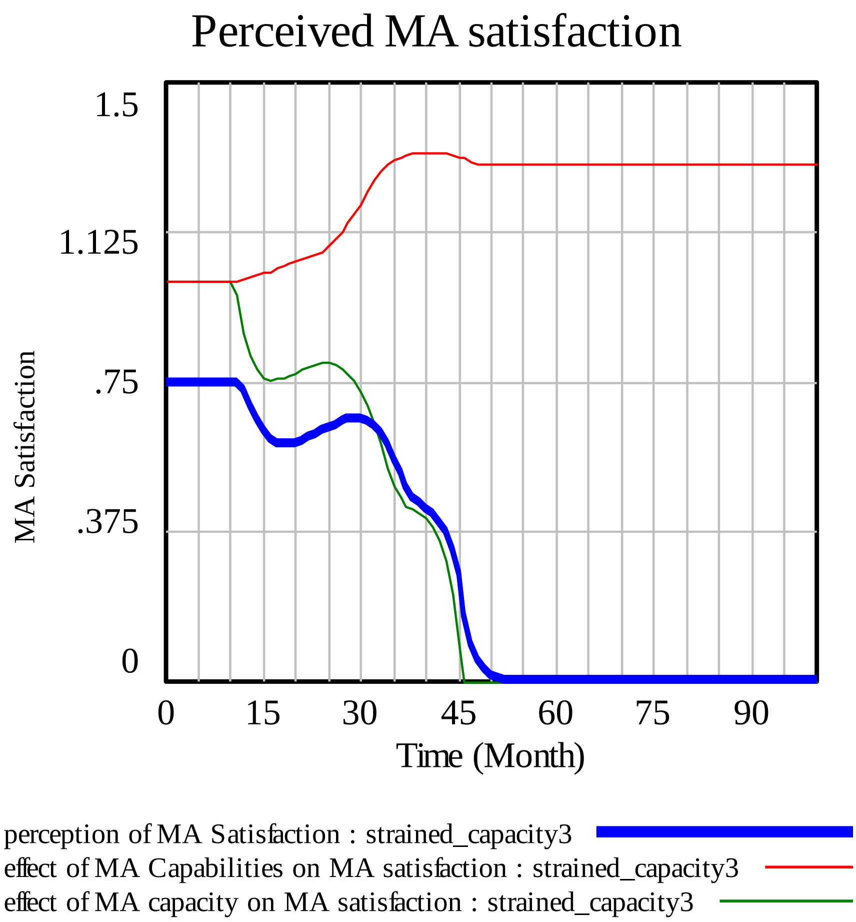


Figure 3.71 Simulation Model Saturation for MA03-04, NM01-36 (continued)



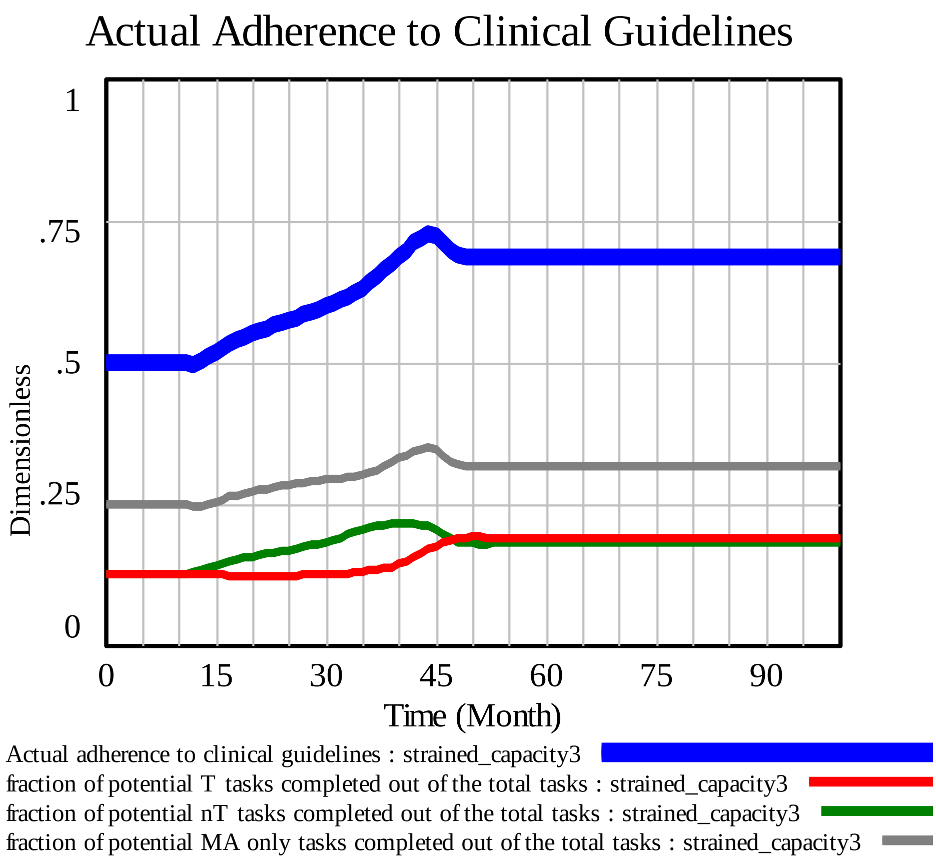
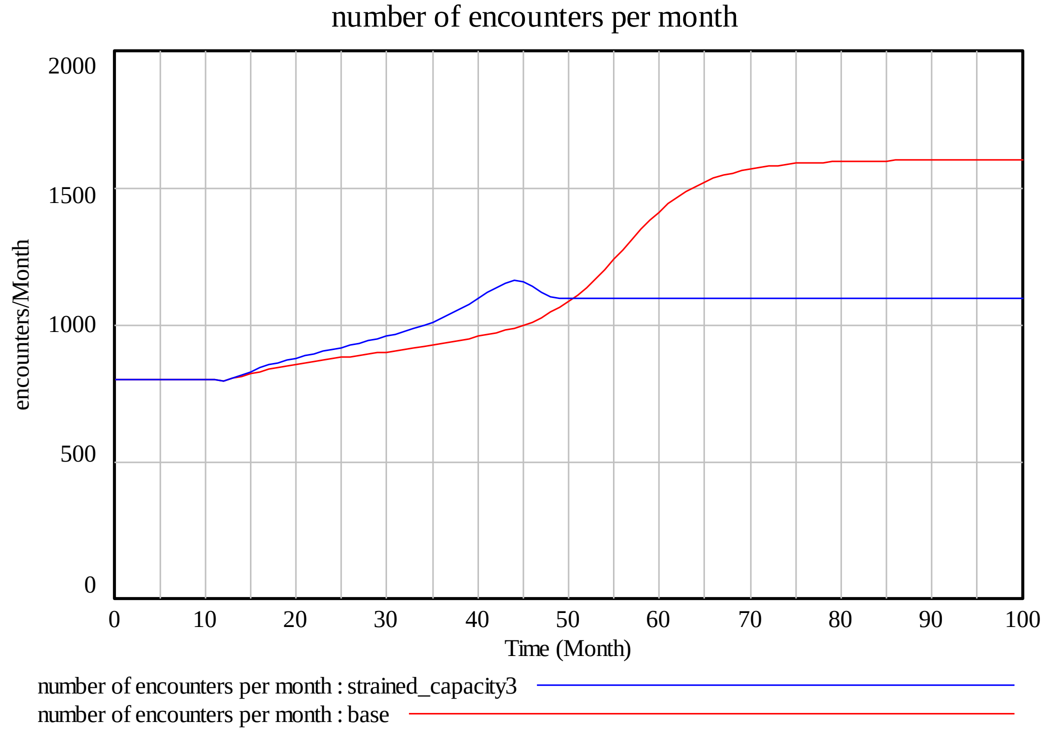


Figure 3.71 Simulation Model Saturation for MA03-04, NM01-36 (continued)



3.5.5.2.3 SIMULATION MODEL CLARIFIES ASPECTS OF CONCEPTUAL MODEL

3.5.5.2.3.1 QUOTATIONS MA03-9, 23 NM01-33 AND CL03-22&24

Quotes across model-validation-set participants described how clinicians seemed to hesitate to shift tasks until they felt confident that the MAs on their team would be capable of completing the new tasks without errors. This recurring story is represented in the Conceptual Model using the three variables of the Trust loop: “Clinician - MA Relationship (Trust)”, “MA Capabilities” and “Task-shifting to MAs”. The quotes reviewed in this section address the ways these variables are implemented in the simulation model, highlighting the importance of on-the-job training.

Quote MA03-9 describes that MDs lead process improvement on their team. This role involves an element of on-the-job training when MDs meet with MAs to discuss proposed changes to their routine (improving the way task-shifting is working). These discussions train MAs, making them more capable of correctly performing the tasks which have been shifted to them.

CL03-23 describes what on-the-job training means in practice. The time taken to train MAs is valued because once they are trained they will be more capable, but that does not mean it is an easy choice to make. In the short term, the training must be accomplished within the work day, therefore it takes time away from completing other tasks.

NM01-33 describes on-the-job training as “get them all functioning at the same speed” which relates to three elements of the Conceptual Model: capabilities development delay, “MA capabilities” and “Visit time spent training”. The Conceptual Model separates “Visit time spent training” from the other two elements, even though it is causally linked. This emphasizes the impact training has on efficiency in the short term, but it obscures its impact on capabilities development, which in the long term improves efficiency. This long-term effect is displayed in the Conceptual Model via a fourth element: the delay between “Capacity” and “having sufficient time allotted for the visit”. Once sufficient capabilities have been attained, the drag of “visit time spent training” on “having sufficient time allotted for the visit” is overwhelmed by the positive influence of the more capable MAs being able to do more. The simulation model explicitly defines the mechanisms involving on-the-job training as it relates to capabilities development, care team capacity and having time for visits.

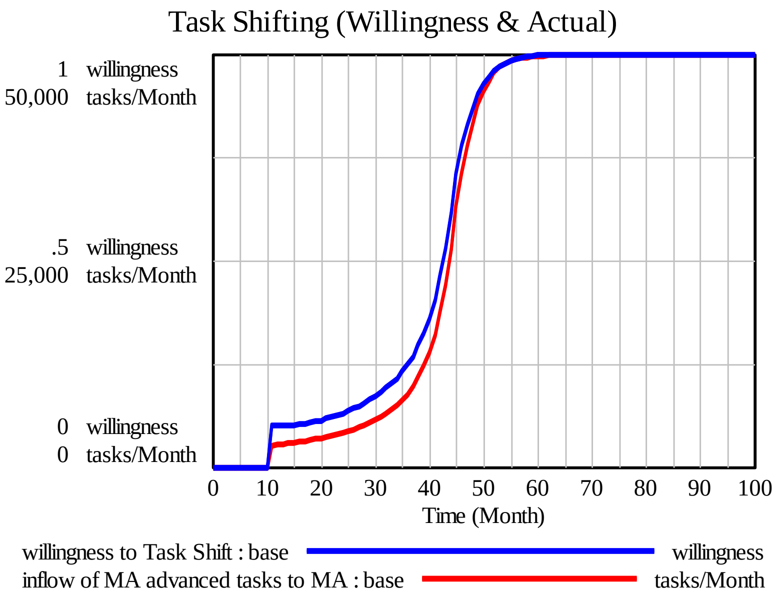
NM01-33 also describes MAs as being “high performers, medium and low performers”. At an earlier phase of PCMH implementation, it was decided that MAs would be paired with clinicians based on their personalities and shared desire to work in a certain way. The simulation model deals with the heterogeneity of these agents in the same way. The base run assumes an average set of clinicians and an average set of MAs. Scenarios allow for the exploration of different types of pairings.

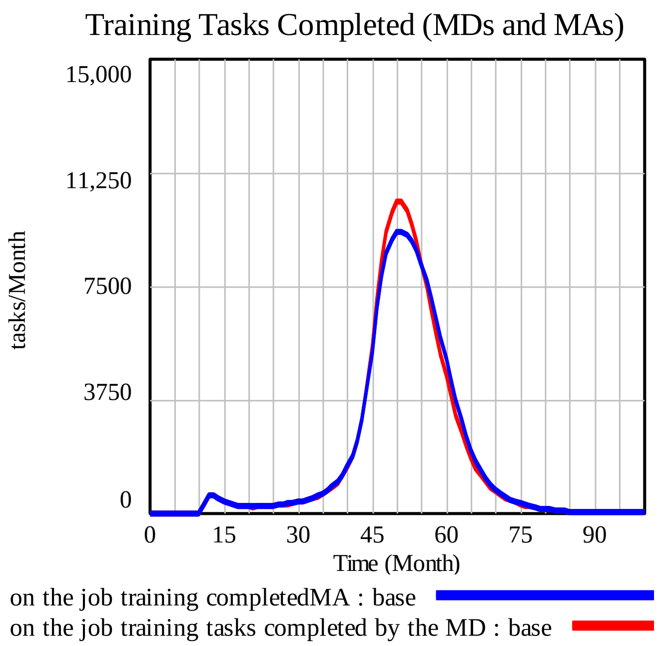
Finally, NM01-33 indicates that "when the doctor … wants to make sure [something] gets done, then they always go to the high performers"; in other words, capable MAs induce trust in clinicians such that tasks are shifted to them.

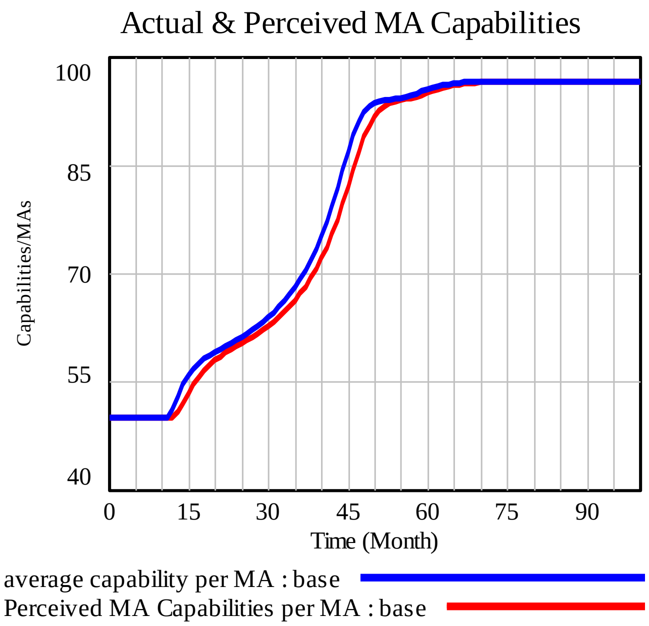
CL03-24 and CL03-22 describe that for task-shifting to reach a high level on the team, there must be a "team approach" where a high level of trust exists between team members. The clinician describes this trust as open communication about problems. Echoing CL03-23 above, team-based problem-solving is the training that results in increased capabilities.

- Structure
  - When a task is shifted to MAs, on-the-job training tasks are generated for MDs and MAs. Completing these training tasks increases MAs' capability as training tasks are converted to capabilities (via the constant "training tasks needed to gain capability").
  - These on-the-job training tasks take time in the day. They do not duplicate the tasks completed for patients.
  - In the short term, training tasks are being completed so less time is available for patient-generated tasks. In the long term, less training is done which increases the team's capacity to complete patient-generated tasks.
  - Clinical staff members' differing competence with respect to learning on the job are modeled in the scenario variables: "effect of MA capab ratio on change in MA capab" and "training tasks needed to gain capability". In so doing, one can explore the impact of having a team with MAs who are high performers, medium or low performers, on average.
  - The issue of how capabilities development relates to trust is a very important part of this model. The associated causal chain involves many variables to implement including perceived and actual MA capabilities, training, task-shifting, and willingness to shift tasks.
- Behavior (Figure 3.72)
  - The base-case scenario in the simulation shows the behavior described in these quotations: as willingness increases, task-shifting increases, training tasks are generated for both MDs and MAs (as the MA becomes more capable) the MD perceives the MAs as being more capable.

Figure 3.72 Simulation Model Saturation for MA03-9,23 NM01-33 and CL03-22&24







3.5.5.2.3.2 QUOTATION CM01-37

CM01 refers to "…retain[ing] the loyalty of the patient …". The Conceptual Model shows this concept as two separate variables: "clinician develops relationship with patient" and "continuity of care (from visit to visit)". The simulation model clarifies how the clinician and patient relationship comes into play in PCT and it does so in accordance with the description given in this quotation. Prior to task-shifting, clinicians are performing technical and non-technical tasks; however, their team is only able to provide 50% of the comprehensiveness that is needed. The goal of task-shifting is to move clinicians away from non-technical tasks by shifting those tasks to MAs. Successful task-shifting allows patients to receive 100% of the comprehensiveness that is needed; thereby greatly increasing the *quality* of care.

At the same time, this increases efficiency for the two clinicians. They are able to take better care of their patient panel. They perform the same number of tasks but the tasks that they are performing are tasks that only they can do. More MAs have been hired so that they can take on the non-technical tasks in addition to the tasks that only MAs can do.

Patients expect clinicians to perform technical tasks – those tasks that only clinicians can do – and to do so in a timely manner. Patients experience satisfaction when these expectations are met and dissatisfaction when these expectations are not met. All else equal, when clinicians are able to see their own patients in a timely manner *and* provide more of the care that their patients need, they positively impact patient satisfaction.

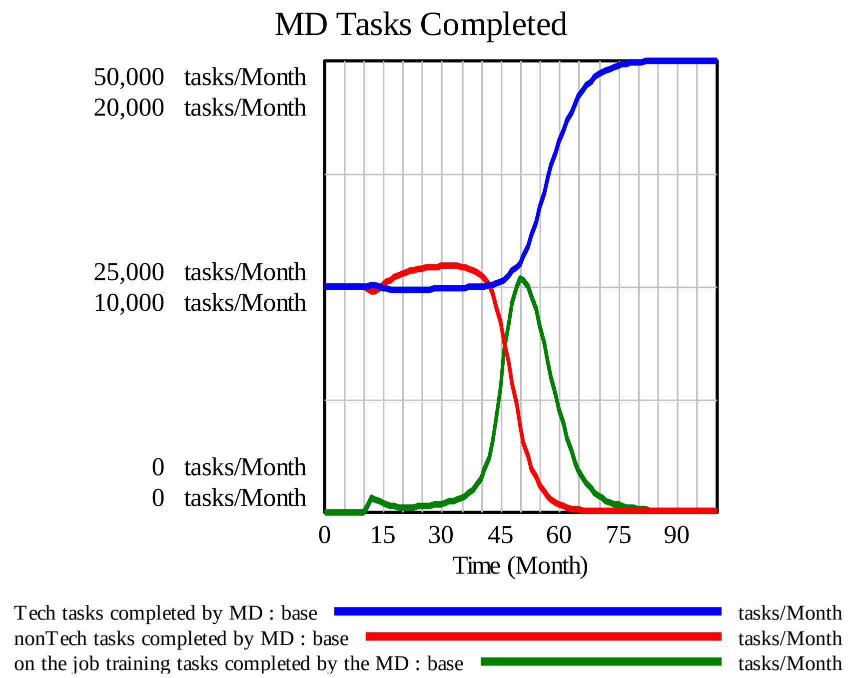
The simulation model structure and behavior are as follows:

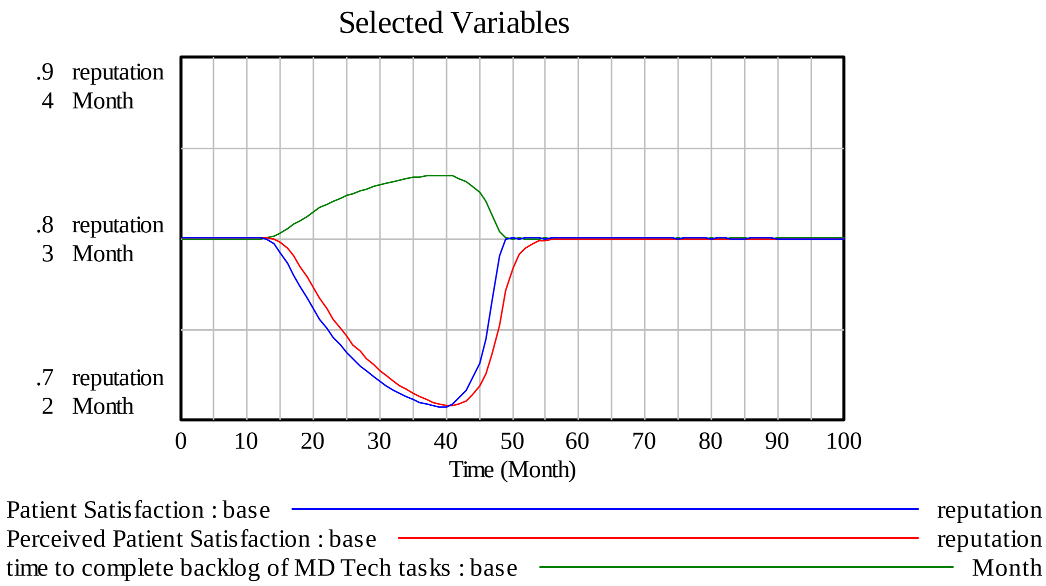
- Structure
    - In the model, clinicians are sensitive to how they are perceived by their patients, because this impacts their patients' loyalty.
    - At month 10, task-shifting begins. This introduces a need for training. Clinicians then work on completing three types of tasks: technical, non-technical and training tasks. This causes delay in completion of technical tasks.
    - While clinicians are behind in completing technical tasks, patient satisfaction is negatively impacted. Patients perceive this by monitoring the ratio between the expected and current time it takes clinicians to complete these tasks. The expected time is the normal time it took clinicians to complete these tasks (3 months) in the past. When patients perceive it is taking their clinician longer to complete their important technical tasks, they feel less satisfied with their care. As clinicians catch up with their work, patients go back to their original level of satisfaction.

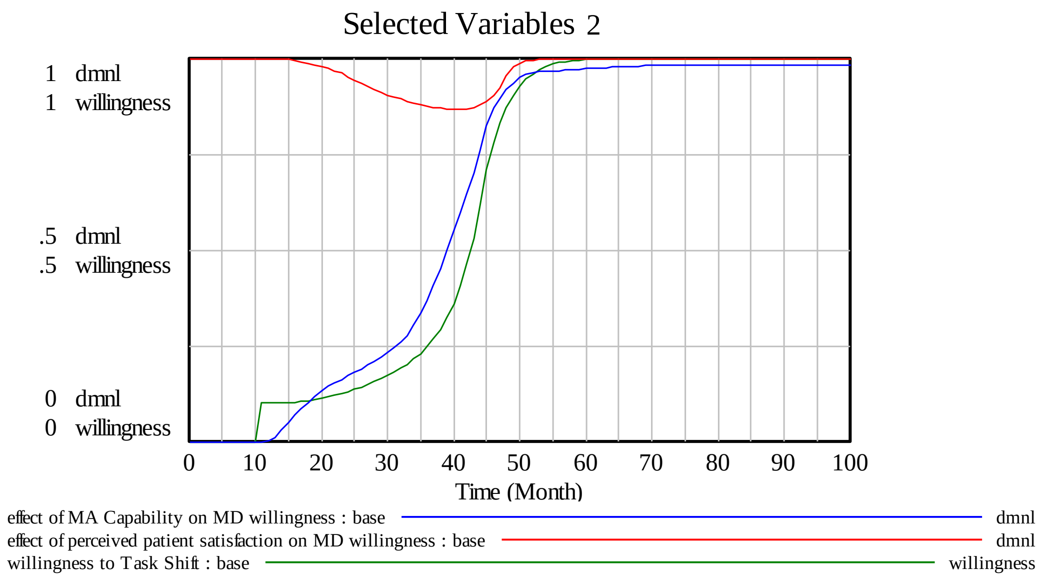
- Clinicians are less willing to shift tasks if they perceive their patients as less satisfied. As clinicians perceive changes to patients' satisfaction, they adjust their willingness to task shift accordingly.
- Thus, patient loyalty is retained. Clinicians reduce their level of task-shifting before allowing patients to become so dissatisfied that they become disloyal.

- Behavior (Figure 3.73)
  - Before task-shifting, the number of MD Tasks Completed per month is 50,000, which is the sum of 25,000 technical (blue) and 25,000 non-technical (red) tasks. In month 10, task-shifting begins. Initially, this has three effects on clinician workloads: 1) training tasks begin (green), with an initial spike and then moving into slow, then fast growth, 2) slightly more non-technical tasks are completed by clinicians, before steeply declining and 2) slightly fewer technical tasks are completed. After month 50, training and non-technical tasks decline, and the clinician's workload becomes increasingly made up of technical tasks, which increase until reaching 50,000 per month.
  - The decrease in the number of technical tasks completed during the initial period leads to an increase in the time it takes clinicians to complete these tasks (green line in "Selected Variables" graph). This increase makes patients less satisfied (blue line). Clinicians perceive this drop in satisfaction (red line).
  - As clinicians see their patients feeling less satisfied, they feel pressure to reduce their willingness to shift tasks (red line in "Selected Variables 2" graph). However, MAs' capabilities are growing quickly during this same period. This encourages the clinicians to be more willing to shift tasks (blue line). Together, these two effects result in a slower rate of increase in willingness to task shift (green line).

Figure 3.73 Simulation Model Saturation for CM01-37







3.5.5.2.4 SIMULATION MODEL REPRODUCES SYSTEM BEHAVIORS IN CAUSAL RIGOROUSLY INTERPRETED QUOTATIONS

The simulation model was able to reproduce the system behaviors described in Causal RIQs (*Rigorously-Interpreted Quotations – for Causality*). These behaviors include two of the reference modes described in Section 3.4.1.1. Causal RIQs described scenarios where PCMH implementation achieved “Successful PCT” and where it reached only a “Suboptimal PCT”. The simulation model reproduced both of these behaviors, and for the same reasons described in Causal RIQs. The model uses the same goals and policies found in the Causal RIQs, so I have confidence that these are the policies used in the HSDO. “Successful PCT” scenarios are shown in all of the Figures above, except Figure 3.71 above which shows a “Suboptimal PCT” scenario. A third reference behavior mode was not mentioned in Causal RIQs. See Section 3.5.5.1.1 for model runs showing “Failed PCT” implementation.

3.5.5.3 STAKEHOLDER DIALOGUE

Model name	Participant CLDs	Team CLDs, Clinic CLDs	SMM1	SMM2	SMM3 / Conceptual Model	SIM1	SIM2	SIM3	Theoretical Model

In this step, stakeholders were shown the simulation model. Again, stakeholders were encouraged to question the structure and point out flaws (based on their level of experience and proximity to the system in question). Four examples of what was learned are presented:

- **Health Services/Operations Management researcher using SDM discussed** the need for operational thinking regarding the clinicians' time – that units need to make sense for the amount of working hours clinicians actually have during the day. Another way of thinking about this is that there is a normal workload for the amount of time clinicians have in a day. This discussion helped me to think about the unit of time that the simulation model would use and pointed to relevant parameters that I hoped to find in the literature (e.g., a clinician would need to work 21 hours in a day to fully adhere to clinical guidelines for the average panel of patients).

- **Experts in health services research identified the need to reorganize** the "Output" page of the model was revised to make it more useful to stakeholders. They identified additional key output variables. Additional graphs were produced to make specific behavioral comparisons across these variables of interest.

- **Experts in SDM in health helped to think through the difference between scenarios** in this research. The "Policy & Scenario Variables" view of the model contains both variables representing actions that agents can choose to take (i.e., policies) and variables representing potential environmental conditions (i.e., preferences) where each constellation of these represents a different scenario. The SDM expert helped to clarify which variables were policies and which ones were scenarios.

- **Experts in SDM in health helped verify the adequately bug-free nature of the model**. They were presented the model sector by sector. For each sector, they asked questions about formulations, running live tests of the model, checking for possible bugs. Across multiple discussions, bugs were cleared out of the model.

At a later stage, stakeholders were shown the simulation model again, this time we also spent some time running the model together and discussing its output. One comment induced a model change (described below). Other comments were incorporated into my understanding of model findings and their congruence as well as added value to what these stakeholders had previously understood.

The main model change that came out of these discussions was the addition of pink noise (see Section 3.4.2.5). A primary care expert questioned the formulation for calculating revenue. The stakeholder, a practicing clinician and health services researcher, indicated that there are only a few CPT billing codes (generating revenue from payers) that clinicians generally choose from when coding the complexity of a visit. He raised doubts that the revenue would increase in step with the increase in comprehensiveness because clinicians would only use the next higher CPT billing code[41][23] once the threshold for what that type of visit entails has been passed. Until that point, clinicians are doing more, but clinicians are not seeing the financial benefit. He further clarified that individual clinicians are expected to bill somewhere in the middle of that CPT range, and that a "red flag" would be raised if clinicians started billing more in the upper range.

These comments caused some reflection. I realized that the second part of his concern was addressed by the fact that, with the added comprehensiveness, there were also more visits generated (so the individual visit was not "a lot" more comprehensive than before. However, reformulation was needed to account for the uncertainty in the payment that would be received on any given month (the first part of his concern).

There is **variation in the payment** that the clinic receives each month due to the mismatch between the lock-step increase in the cost of a visit with increase in comprehensiveness (i.e., visit complexity) and the stepwise increase in the reimbursement level attached to the visit complexity codes. Clinicians might be delivering more care but not enough to get that higher code for the visit, so the additional compensation is not received. In another visit, clinicians might be delivering less care but within the range of care for a visit of that code type, qualifying for the same CPT code (and thus the same amount of revenue). These variations do not necessarily average out within the month, causing fluctuations in monthly revenue, so they need to be represented in the model.

With noise, sometimes reimbursement is more than cost, sometimes it is less than cost. So, there are losses and profits. Sometimes, the losses are great (when the mismatch between cost and compensation is high), forcing policy response that the model could not previously induce (e.g., sharper regression in TS and an MA hiring freeze). Without noise, in the previous model, the revenue losses came so slowly that a net level of MA attrition over hiring and a gradual rate of regressing on TS could compensate (more drastic policies were never induced).

[41] CPT codes for primary care include 99211-99215 (office visit, established patient) as well as 99201-99205 (initial office visit); where the higher the number in the sequence, the more complex the visit. For a complete list of codes and example vignettes, see the American Medical Association CPT Knowledge Base.

3.5.6SYSTEM DYNAMICS SATURATION

This section presents how model-validation-set interviews demonstrate that the SD-S Test has been passed. The SD-S Test queries the validation-set interviews to verify that the target group sees SDM research as a useful way of addressing the problem under investigation. Specifically, it verifies participants are aware of their mental models, that these mental models contain causality, and that, together with system structure, mental models are perceived to be causing their problem. Evidence for these claims is presented one piece at a time over four sections.

While SMM-S, CM-S and SIM-S results are presented in their own sections above, results from these validation tests that contribute to the SD-S Test are briefly discussed here as well (Section **3.5.6.1)**. Then, I present visualizations of the extent to which causality forms part of validation set participants' mental models in the *Information Accumulation Graphs* (Section **3.5.6.2**). Finally, I present a sample of the *Rigorously-Interpreted Quotations* from the model-validation-set that demonstrate the passing of SD-S. Two types are presented. Results come from *all* model-validation-set participants.

The first type are Causal RIQs (*Rigorously-Interpreted Quotations – for Causality*). These are equivalent to the quotations presented in CM-S (Appendix F), with the difference being that this section only considers statements coded with a system dynamics-related item[42]. As this section uses a rich dataset of participants' perceptions of their system, several angles are explored: tensions, time delays and feedback loops (Section **3.5.6.3)**.

The second type are Cognitive RIQs (*Rigorously-Interpreted Quotations – for Cognition*). These consider *conceptual* statements coded with a system dynamics-related item[43] (Section **3.5.6.4**). These verify that participants understand that they have mental models as characterized in SDM, that these models are important, that these models can (and indeed must) change, and that doing so involves emotional engagement.

[42] In causal statements, these items include time delays and feedback loops.

[43] In conceptual statements, these items include descriptions of mental models, their own and those of others in the system, as well as the importance of mental models and emotional engagement to understanding and addressing the problem statement.

3.5.6.1 PRIOR SATURATION RESULTS CONTRIBUTING TO SYSTEM DYNAMICS SATURATION

The SMM-S Diagrams and CM-S Diagrams demonstrate that individual clinicians and MAs have mental models with causality and time delays, and that the Shared Mental Model has a structure with time delays and feedbacks. As these diagrams also visually display variation in the number of people (which also includes managers in CM-S) and clinics (SMM-S) recognizing individual relationships, these diagrams also demonstrate that the structure of the entire Shared Mental Model is not seen by individual participants[44]. SMM-S Diagrams are found in Section 3.5.4.1.3 and CM-S Diagrams are found in Section 3.5.4.2.2.

CM-S Diagrams also demonstrate that clinicians and MAs' have mental models as characterized in SDM. The Cognitive RIQs show how participants' understand this in a direct way (Section 3.5.6.4). The "CM-S Diagram – Respondents Identifying" (Figure 3.59) indirectly demonstrates the following mental model characteristics:

- Validation-set participants perceive 90% of links in the Shared Mental Model (SMM2), supporting the assumption that **the model's causal structure is *analogous to the perceived structure* of the real *external system***.
- Participants perceive only part of this system, supporting the assumption that **individual mental models are *limited***.
- 30% of links in SMM2 were mentioned by 60% or more of participants in the validation set, supporting the assumption that **mental models are *accessible*.**

I also wish to make special note of the "CM-S Diagram – Shared Understanding Diagram". This visualizes the variation in individuals' mental models when aggregated by the three professions directly accountable for PCT (Figure 3.60). The diagram indicates that all three professions have limits to the extent of their shared understanding. More than half of the links are not fully shared across professions, indicating important *blind spots* in understanding how PCT works. In other words, each profession is using narrow mental models.

In the CM-S Diagrams, "Task-shifting --||→+ MA Capabilities" is at the core of participants' shared perception of PCT. It is both the best understood and most-widely-understood link in SMM2. For example, one participant's interview (CL03) shows how this time delay and proximate causal links can form an important part of one's mental model of PCT. CL03 validated this link with a disproportionate share of statements (this person contributed 35% of

[44] SMM-S Curves and CM-S Curves also demonstrate this point. In showing accumulation of elements as more clinics (in SMM-S) or interviews (CM-S) are added, these curves show that the entire structure is not seen by individual participants.

all validation-set participants' statements discussing this link). CL03 describes this link in chains involving 7 other links and refers to it in 70% of words in all CL03's causal statements. Across these statements, CL03 explains this link using one of the most typical and dynamically-important aspects of time delays: they cause frustration and make organizational work difficult[213, 214].

The SIM-S Test reviews Causal RIQs to check that the simulation model does not contain radical departures from those statements (Section 3.5.5.2). In so doing, it is another check that participants do indeed see these causal structures and delays (at this stage, as they are represented in the simulation model).

3.5.6.2 INFORMATION ACCUMULATION GRAPHS

Information Accumulation Graphs visualize the inflow and accumulation of various types[45] of information over the course of an interview. One of these types of information is *causal information* relevant to the problem statement. While individual graphs display variation across participants, all of these graphs show some amount of causal information. In so doing, they demonstrate the extent to which model-validation-set participants see causality in the system and in the problem being described.

Figure 3.74 presents Information Accumulation Graphs for MAs, clinicians, and clinic management personnel. The y-axis shows the word count from the interview transcript. The blue and gray lines show how each individual statement (x-axis) contributes to the accumulation for each information type (extraneous, conceptual and causal); in other words, the inflow of words into the specific information type due to that statement. As concepts outside the boundaries of PCT are discussed, extraneous information grows – shown in green. As the concepts validate causal assumptions, causal information accumulates – shown in red. As concepts are discussed which validate or deepen understanding of variables, but not relationships, conceptual information accumulates – shown in orange.

An interview with a red line dominating the graph is rich in causal information about the problem statement. An interview with equal orange and red lines has this information, but

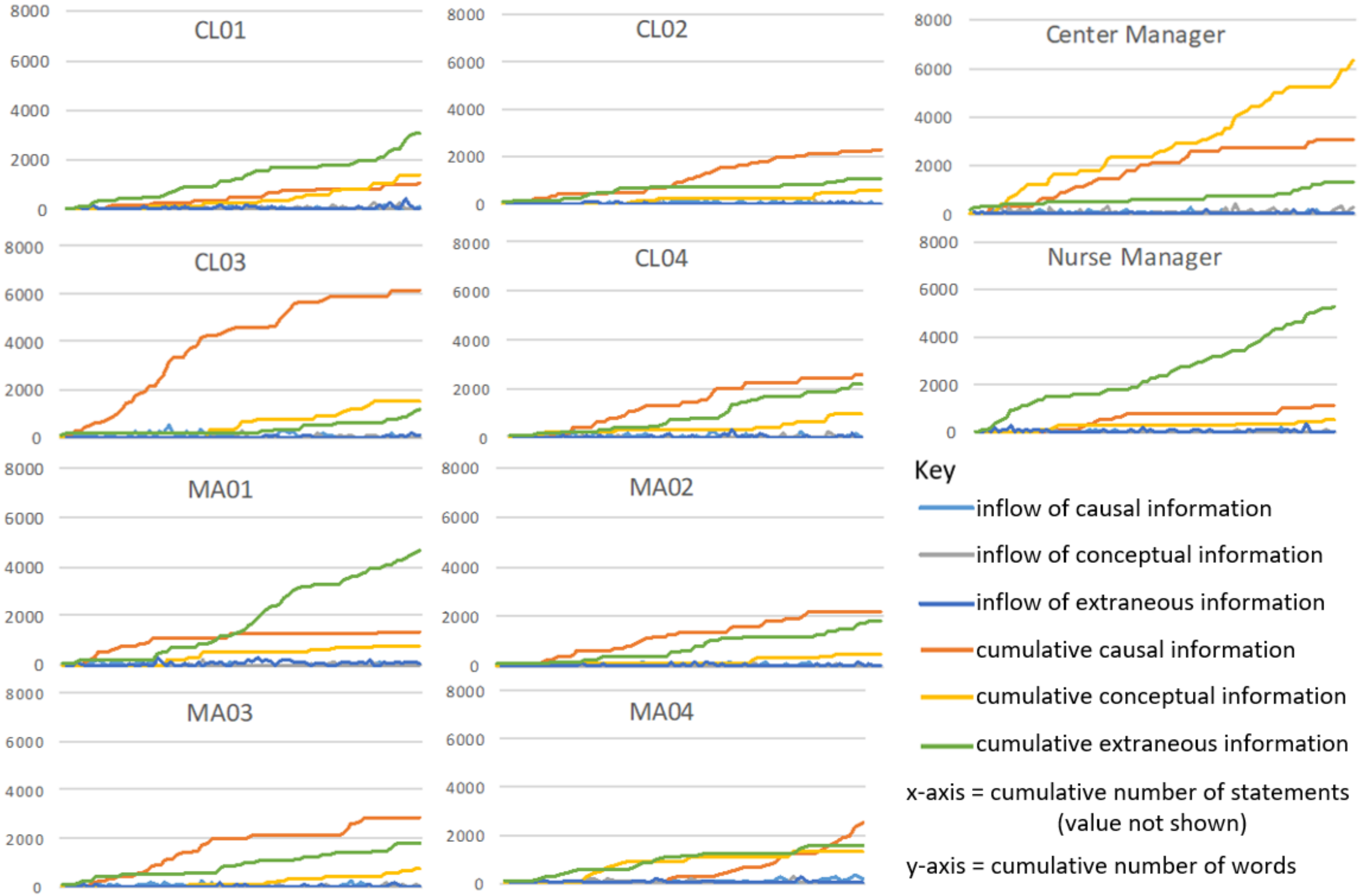
[45] These types are: causal information, conceptual information, system dynamics related items and extraneous information. *Causal information* identifies cause and effect between at least two variables. *Conceptual information* provides detailed definitions of meanings attached to concepts or assumptions (e.g., system boundary, time step, time horizon). *System dynamics-related items* can be found in causal information (showing awareness of complex elements of structure: delays, feedbacks, stocks, flows and nonlinearities) and conceptual information (showing that mental models and emotions are important to problem owners). *Extraneous information* is information that is not relevant to the problem statement. (See Appendix E for more details).

also spends time providing details regarding a variable. Finally, an interview which is dominated by a green line spends most of the time discussing issues which were not germane to the problem statement. It is the first of these three lines (red) which demonstrates SD-S.

SD-S is demonstrated as all of these participants provided numerous causal (and conceptual) statements – statements that are relevant to and elucidate the causal structure of the problem statement. 75% of both the MA and clinician interviews are dominated by red lines. While neither management interview is dominated by red lines, these interviews contain a similar amount of causal information (see word counts).

These graphs also point out that some participants spend a lot of (and sometimes the most) time discussing extraneous concepts. These statements were reviewed and confirmed to be extraneous as they were normative statements lacking depth of description of variables as well as links to other variables. The existance of extraneous statements does not detract from the existance of causal (and conceptual) statements. It merely demonstrates variation in the way in which participants described their mental models and overall feelings about the transformation underway.

Figure 3.74 Participant Information Accumulation Graphs



3.5.6.3 CAUSAL RIGOROUSLY-INTERPRETED QUOTATIONS FOR SYSTEM DYNAMICS SATURATION

This section uses quotes coded as containing **system dynamics-related items** from model-validation-set interviews to demonstrate perceptions of causality: (1) when tensions (interpreted as a system structure) are perceived to be causing the problem and (2) when participants' mental models contain causality, time delays and feedback loops – structural elements which are appropriately addressed using SDM. Demonstrating that participants perceive causality in this way is sufficient to pass this part of the SD-S Test.

3.5.6.3.1 EVIDENCE OF TENSIONS

This section presents one way[46] that participants verify **tensions**: participants can *directly* describe the tensions in terms of a trade-off and frustrating emotions. This section provides two typical examples.

In quote MA03-03 (analyzed in more depth in Appendix F), MA03 explicitly calls out the tensions that MA03 sees as "*conflicts*".

> MA03-03 "It seems like, like we are being told … 'get the patients through faster' but we need to sit down with them and have a conversation with them. But we need to be fast. And, we don't want to keep the patients waiting. But we need to double book these patients. And you know… ***so there are a lot of conflicts***" (emphasis added).

CL01 describes how tensions exist in the system which require clinicians to constantly balance quality and revenue. The tension is a "*reality*" which prevents either variable from being optimized.

> CL01-26 "***The same questions*** of delivering excellent care versus making money, which is a ***reality***. ***We are always trying to balance that out***." (Emphasis added).

The tensions are described by positing two variables which are crucial to achieving PCT. In these quotes, these variables are the primary care tenets of access (e.g., "*be fast*", "*double book*", "*making money*") and comprehensiveness (e.g., "*have a conversation*", "*excellent care*"). Next, the participants imply that the two variables interact with each other (e.g., "*there are a lot of conflicts*", "*balance that out*") in a system which governs participants' ability to raise and maintain the level **of both variables** simultaneously (e.g., "*we are being told … but*", "*the same questions … we are always trying*"). Direct participants see the tensions as crucial, misunderstood and taken-for-granted. Tensions are crucial because the variables

[46] Participants can also *indirectly* describe the tensions in terms of emotional strain in frustrating situations, without direct reference to a trade-off. Validations of this type are found in Section 3.5.6.3.2.1 Time Delays (Table 3.8) and Section 3.5.6.4. Cognitive RIQs.

often include at least one of the four tenets. Tensions are misunderstood in that they are always described in the form of a struggle with uncertain issues ("we are being told … *but … and* [x3] … conflicts", "questions … trying"). Tensions are also considered as taken for granted to be part of the system ("*same questions ... always*"). Tensions, therefore, are necessary for understanding PCT and, possibly, for understanding primary care.

Generally, participants perceived each tension as being a pair of variables (A,B) with simultaneous interactions between them (A→B→A) which result in each variable limiting the ability of the other to grow. Participants see this tension as having mechanisms and implications which are poorly understood (A→system→B→system→A). SDM is useful in researching problems of this type, as it uses systems of variables which interact via simultaneous differential equations to simulate the implications of the causal structure underlying a problem. This fully represents the tensions. These simulation models have a simple visual language to display these interactions and *friendly algebra* for calculating the interactions in equations. SDM also requires that decision functions consider the cognitive limitations of decision-makers (e.g., asking what information would they have and when[113] (p. 516-520]) and the cultural milieu of organizations[105, 108]. To achieve this objective, SDM seeks, where possible, to ground the conceptual basis of its models in the mental databases of direct problem participants[113] (p. 520). Demonstrating that participants perceive tensions, therefore supports the assumption that participants will see SDM as a useful way to study the problem.

3.5.6.3.2 EVIDENCE OF CAUSALITY

Some evidence for participants' perception of causality has already been presented[47]. This section provides specific examples for two types of causal relations: time delays and feedback loops. Time delays are a specific type of causal link. These links take longer for their effect to be seen than other links in the system. The relative difference between the length of time for some links (i.e., instantaneous links) and the length of time that constitutes a time delay depends on the system scope and time scale mentioned in the problem statement. Feedback loops are a set of causal links. Feedback loops occur when a set of links forms a circular chain of cause and effect (A→B→C→A). For the most part, feedback loops are considered to be

[47] The Causal RIQs developed in CM-S demonstrate how participants in the validation-set saw causal links between variables in the system surrounding PCT. These causal relationships were described in several ways. Some examples include using metaphors, passive description (i.e., A→B "B is caused by A"), in hypothetical if/then scenarios, or through telling behavioral event stories with first/then sequences of actions. The Information Accumulation Graphs above show that every participant in the model-validation-set spent significant portions of their interview using causality to describe PCT.

beyond the cognitive capacities of individuals[133]. There are many statements where delays and/or feedbacks are described in a single statement, but neither is considered to be a *perception* of delay or feedback unless the participant describes how that delay or loop functions over time.

3.5.6.3.2.1 EVIDENCE OF TIME DELAYS

This section presents quotations which demonstrate how participants perceived significant time delays. First, two quotations are presented which demonstrate detailed description of a time delay. Then, phrases demonstrating perception of time delays are presented, using excerpts from the Causal RIQs from this section and CM-S. These delays are significant because they were perceived by participants to be integral to understanding PCT.

In quote 11, MA02 describes the challenge of estimating the length of the time delay for achieving significant changes under PCT. One must consider the aspects of care ("*things*") that need to change as well as the different rates of change for each one ("*real fast… a little longer*"). MA02 uses the analogy of getting an object "*rolling*" to describe the effort which is required to overcome the inertia in the traditional system of primary care, and cites an approximately two-year duration until the momentum of PCT can be perceived by direct problem participants. (Note: underlined text describes the time delays.)

> MA02-11 "Well, it's… ***it takes a while to change things***. I think people have to discuss things-- the best way to do things. So ***I don't know how long it would take to change things***. I don't know. ***Some things we have changed real fast in our group and some things may take a little longer. It is just hard to say***. Well, ***it took quite a long time*** to get into (PCMH) <<laughs>>. ***To begin with***, we had a lot of doctors that were, 'Oh no, we aren't doing this.' And a lot of people that were saying, '(No), we don't like this at all.' ***It just kind of just kept evolving*** … ***into what we are doing now and it seems to be fine.*** So it kind of ***took a long time to get that rolling***, in this department …***Maybe a couple years***." (Emphasis added for phrases referring to time delays).

In quote 13, MA02 describes how a time delay can result in participants experiencing emotions which are worse-before-better. This occurs when the initial period of change is "worse" meaning it causes one to feel even more frustration than the original state, but once this period is passed, this feeling is resolved and one enters a new state of the system which is "better", or less frustrating, than the original state.

> **MA02-13)** "Well, I think ***anything is frustrating when you first start doing it.*** And you are not quite sure what you are supposed to be doing and … But it hasn't been that frustrating, I don't believe. ***It was a little frustrating*** with people trying to get along together ***at first,*** deciding what their role was, or how we [were] going to do things. ***But that seems to be resolved***." (Emphasis added for phrases referring to time delays).

Time delays can induce negative emotional responses when expectations are not met. In quote 13, MA02 describes having had a higher tolerance to the frustrating aspects of the change when compared to others on the team. MA02's cites the expectation that the changes would be frustrating "*when you first start*" (in quote 13) as well as the recognition that estimating the length of the delay would be very difficult (in quote 11).

Table 3.8 below presents phrases demonstrating perception of time delays, using excerpts from the Causal RIQs from this section and CM-S (see Appendix F). Participants demonstrated perception of significant time delays in the following ways:

- **Things take time** – describes an activity or set of activities which must be accomplished before a desired action can be implemented in practice
- **Progression** – describes an activity which progresses more slowly over time than expected
- **Before and after** – describes a significant difference in conditions which is observed in retrospect as being the result of a specific activity.

Table 3.8 is organized as follows: the first two columns contain the phrases demonstrating the perception of a delay and their respective quote numbers (for location of the full quote), the next three columns are used to assign each quote[48] to any of the delay perception types mentioned above. The final column for emotion calls out when participants described delays in terms of an emotional struggle, or that the delay caused them to experience emotional strain.

Demonstrating delay perception supports the claim that significant delays exist in the system, which indicate the usefulness of an SDM research approach. Participants' attaching emotional strain to time delays also supports their significance to PCT. More quotes refer to "*progression*" than to "*before and after*" or "*things take time*". More quotes used just one type of delay description, than two or three. All quotes using more than one type involved the "*progression*" type. Emotions were attached more often to "*things take time*" and "*progression*", than to "*before and after*". Finally, clinicians attached emotion to their perception of time delays more often than MAs or managers.

[48] Appendix F presents the full quotes for the following participants: CL03, MA03, NM01 and CM01. The remaining quotations, for CL01, MA02 and MA04 are presented in this section. Other participants in the model-validation-set perceived significant delays, but are not shown here.

Table 3.8 Participant Perceptions of a Time Delay

Quote Number	Phrases interpreted as "Time Delay"	Type of description for delay			Emotion
		Things take time	Progression	Before and after	
CM01-40	• and then end up leaving • Another thing here and there and then eventually they just say		x		
CM01-42	• Some [aspects of PCMH] were harder than others and [for] some we are still struggling		x		
MA02-11	• it takes a while to change things • I don't know how long it would take to change things • Some things we have changed real fast in our group and some things may take a little longer. It is just hard to say • it took quite a long time • To begin with • It just kind of just kept evolving into what we are doing now • took a long time to get that rolling • Maybe a couple years	x	x	x	
MA02-12	• it took a little bit • it took a little while		x		
MA02-13	• anything is frustrating when you first start doing it • It was a little frustrating … at first • But that seems to be resolved			x	x
MA03-06	• Honestly since • there used to be • since we… were aware of what was going on that way			x	
MA03-08	• After you have been here for a while and you have seen the same type of thing over and over again			x	
MA03-09	• If there is any issue then … and then … and then … and then … and if	x			
MA04-10	• When I first started, I was a little bit intimidated • I would try to go into [it] a little slow • next time it would progress itself to the next level where now I know		x		
NM01-32	• a long time ago • it took a while • to learn • but it worked really well		x	x	
NM01-33	• it's kind of hard to get them all functioning at the same speed • In the beginning		x	x	
CL01-25	• it has become a more organized process • it has become more efficient		x	x	
CL01-31	• after a couple of times • It normally takes a day or two • it takes MAs a little bit	x			
CL03-19	• challenging • for me it was hard		x		x
CL03-21	• it's actually how to make it work that is challenging		x		
CL03-22	• I just struggle. I just find it very hard. There is a lot of things that I feel like we could be doing better • I get a little bit frustrated • It's just hard		x		x
CL03-23	• haven't had enough time • I've become a little bit cynical • concentrated effort • trying to figure out how to add these things	x	x		x
CL03-24	• more and more • it just feels like a lot more work to do than that		x		x
CL04-29	• it seems like that takes a lot of work • it does take a lot more time and patience	x	x		x

3.5.6.3.2.2 EVIDENCE OF FEEDBACK LOOPS

This section presents quotations which demonstrate how participants perceived feedback loops. These quotations are presented using Causal RIQs, tables showing *rigorously-interpreted quotations for causality*. A comparison is made between a single causal statement in the transcript (left column) and the model variables and causal links which it describes in SMM2 (blue areas) generating an *interpretation* (right column lower half). This comparison involves producing a *quotation* (by truncating the statement) which tells a story represented by a portion of SMM2. When a specific phrase in the quotation refers to a model variable it is underlined and copied to the *Phrases needing interpretation* section (light blue area). In this section, phrases are placed in bullet lists adjoining the model variable(s) they refer to, in descending order along the feedback loop's causal chain in SMM2 (dark blue area). This causal chain is then presented (e.g., Model Variable →+ Model Variable) (yellow area).

The description above applies to direct perception of model variables. Indirect perception exist when a variable is part of the loop but not explicitly mentioned in the causal statement being analyzed. In such instances, the variable is placed in the dark blue area but a "0" is placed in the light blue area (indicating that it was not mentioned in this quotation). In the yellow area, this variable and associated links are still listed as part of the chain but they are grey instead of black. This is also done for links that are not mentioned. Finally, the quotation is interpreted (green area).

When perception of information-feedback is mentioned, *Feedback Loop* is identified in the right-hand side cell, with the corresponding phrases on the left-hand side cell shown in bullet points. The same process is followed for perception of a *Time Delay*.

Sometimes, a quotation will have more than one causal chain. In this case, the process above is repeated for new phrase-variable pairs (i.e., without duplicating phrase-variable pairs). Before delving into these tables, the reader may wish to first review Appendix F as it contains more basic applications of the Causal RIQ tables.

The quotes from CL01 and MA04 demonstrate perception of a feedback loop and the quote from MA02 demonstrates a passive description of the stages in a process which occurs with changing dominance across feedback loops. In all cases presented, while feedback loops are closed, not all variables and links inside the loops are mentioned. Sometimes, one link is omitted, other times a variable is jumped over. In these cases, what is omitted is represented in Table 3.9 using grey text.

In quote MA04-10, MA04 describes learning to work on a care team as a set of changes which occurred over time through the action of several parallel processes. MA04 describes the details of what changes over time and how, as well as the interactions between team members. First, the quote describes the time delay for learning new tasks ("Task-shifting --||→+ MA Capabilities"). Second, it describes the interaction between clinicians and MAs. This interaction is described as MA04 explains that clinicians came to trust MA04 more than the others on the team because of MA04's superior capabilities ("MA Capabilities → Clinician - MA Relationship (Trust)") and that this trust translated into MA04 being given more advanced tasks in task-shifting ("Clinician - MA Relationship Trust →+ Task-shifting"). Taken together, these links close the loop for the "Trust loop". However, uniquely to MA04, the statement also describes the functioning and perception of this feedback loop. MA04 perceives the action of this feedback loop as a self-reinforcing process whereby a person's past actions come to benefit them in the present. The action of this loop is perceived over and over again as changes occur and are observed gradually over time. This is an accurate description of how feedback works, specifically, self-reinforcing positive feedback.

By inferring one link ("MA satisfaction --||→+ MA retention"), two additional feedback loops ("Team Learning loop" and "Sharing the Load loop") can be closed by this statement. This permits MA04-10 to describe other factors which acted to MA04's benefit described in the quote, including a satisfying working relationship with the other MAs on the team and a team which has an effective capacity which is greater than the sum of its members.

In quote MA02-12, MA02 describes PCMH implementation in terms of a closely-interacting set of processes which evolve over time. MA02 did not describe perceiving the feedback loop explicitly. MA02 does not describe MA turnover problems in this statement. However, inferring it (i.e., the variable "MA Retention") and the links "MA satisfaction --||→+ MA retention →+ Capacity" allows the same three feedback loops mentioned by MA04 to be closed.

In quote CL01-25, CL01 describes how the team's achievements in implementing PCMH came as a result of the way that tasks were shared with MAs. CL01 describes how feedback is used to monitor the progress in variables as changes occur in the system and then to take action. This is an accurate description of the function of a limiting, negative or balancing feedback loop.

Similar to most participants who achieved some success in implementing PCMH, CL01 fails to mention how the team overcame the unintended consequences of task-shifting and incentives which prevented others from the same success (Insufficient Time loops 1 and 2). This omission results in two un-closed loops. Inferring the required variables and links ("Visit time spent training MAs" and "Volume of Patient Visits") permits us to observe what is required to overcome these barriers to implementation. See Table 3.9 below.

Table 3.9 Participant Perceptions of a Feedback Loop

Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable** →+ **Model Variable** (→ = causal link, --||→+ = with delay, +/- = link polarity) Interpretation	
MA04-10) "When I first started, I was a little bit intimidated because I hadn't been in the clinical setting before. And so some of the procedures and stuff were a little bit overwhelming at times. And so, I would try to adapt to whatever procedure was going on … the providers are a little bit different and so I would try to go into [it] a little slow, but also ask questions, that way I'd feel a little more comfortable about it. And the next time it would progress itself to the next level, where now I know that some of the providers will find me to help with [a] procedure because I know exactly what they need and exactly what they need to have set up. So, I think that made a huge difference for myself… "I feel like there is a lot of pressure on our team to do things a lot faster…. anytime anybody talks about our team, 'it's always so busy, it's always so busy.' … I don't think if you placed someone from another team that's not used to it into our team, it would not go well. The meshing would not work for that... "Sometimes the patients … are a lot of time, they overwhelm … the providers, which in turn overwhelms us because we get backed up. And then there is a lot more pressure on us… "I think that we kind of work off of each other's strengths, which helps the team out, and it is really good." (253/523)	• next time it would progress itself to the next level • that made a huge difference for myself	**Feedback Loop**
	• When I first started, I was a little bit intimidated • I would try to go into [it] a little slow • next time it would progress itself to the next level where now I know	**Time Delay**
	• whatever procedure was going on • to help with [a] procedure • The meshing	**TS**
	• little bit overwhelming at times • I'd feel a little more comfortable about it • next time it would progress itself to the next level • because I know exactly • that's not used to it	**MA capabilities**
	• the providers are a little bit different • now I know that some of the providers will find me • because I know exactly what they need • The meshing would not work • helps the team	**Clinician - MA relationship (Trust)**
	• I would try to adapt to whatever procedure was going on • I would try to go into [it] a little slow, but also ask	**Visit time spent training MAs**
	TS --||→+ MA capabilities →+ Clinician - MA relationship (Trust) →+ TS→+ Visit time spent training MAs	
	Trust Loop: MA04 describes how personal learning reinforces itself ("*it would progress itself to the next level*") through interaction with the clinicians on the team for direct training and for task-shifting. MAs gain the trust of MDs on the team in a self-reinforcing process, where, by working through the training and asking questions about new tasks, they build up the capability to perform new procedures. More tasks are then delegated to them, initiating more training. Learning quickly then builds MDs' trust for further teamwork.	

Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable →+ Model Variable** (→ = causal link, --||→+ = with delay, +/- = link polarity) Interpretation	
	• little bit intimidated • little bit overwhelming at times • I'd feel a little more comfortable • it is really good	**MA satisfaction**
	• Overwhelm… the providers • In turn overwhelms us • We get backed up • work off of each other's strengths • pressure on our team to do things a lot faster • so busy … so busy • a lot more pressure on us • helps the team	**Capacity**
	• I don't think if you placed someone from another team that's not used to it into our team	**MA retention**
	TS--||→+ MA capabilities →+ MA satisfaction --||→+ **MA retention→+ Capacity→+ TS**	
	Team Learning Loop: MA04 cites a self-initiated change ("*that made a huge difference for myself*") in personal job satisfaction resulting from personal learning (Trust loop) as well as learning to be part of a high-performing team. MA04 uses a hypothetical scenario which demonstrates how Team Learning works. First, a new person joins the team. It takes time to learn the team's "meshing", (i.e., their task-shifting routines), so any MA would have lower capability than the current team members, automatically lowering the average level of "MA capabilities" on the team. A new MA in this situation would feel intimidated/overwhelmed, reducing the team's average level of job satisfaction. This puts the team at greater risk of retention issues (not mentioned) which (together with reduced capability) negatively impacts the team's effective capacity to perform their work, causing the team's routines to break down.	

<table>
<tr><td>Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="3"></td><td>• The patients… are a lot of time, they overwhelm… the provider
• pressure on our team to do things a lot faster
• so busy … so busy
• a lot more pressure on us</td><td>Volume of patient visits per hour</td></tr>
<tr><td colspan="2">Volume of patient visits per hour →- Capacity --||→+ MA satisfaction --||→+ MA retention→+ Capacity</td></tr>
<tr><td colspan="2">Sharing the Load Loop: Despite the high level of pressure placed on the team, MA04 experiences a high level of job satisfaction. This is because of team members “work off each other’s strengths” (Sharing the Load loop). This high-performing team is capable of performing their work at a much higher levels of productivity (as referenced in the Team Learning example).

Keeping team members consistent over time gives personal and team learning the time needed to mature. The action of these processes contributes to a high level of capacity. Increases in the level of capacity have a gradual, positive impact on job satisfaction as work pressure becomes evenly shared across team members. This team spirit improves team members’ work attendance and retention (link not mentioned), which are prerequisites to having a consistent team in place.</td></tr>
</table>

<table>
<tr><td>Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)</td><td colspan="3">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="6">MA02-12) “At first we were … worried about working together as a team, when we have been working as an individual with a doctor. But uh, and it took a little bit for everyone to kind of get in gear and get along together as a team, and know exactly what our roles were and what we were supposed to do. And it took a little while to do that, and it seems to be working fine now.”(77/90)</td><td>• At first
• it took a little bit</td><td>• it took a little while
• now</td><td>Time Delay</td></tr>
<tr><td>• working together
• get in gear and get along
• our roles</td><td>• supposed to do
• it seems to be working fine now</td><td>TS</td></tr>
<tr><td>• get in gear</td><td>• know exactly</td><td>MA capabilities</td></tr>
<tr><td colspan="2">• worried about working together
• get in gear and get along together as a team and know exactly what our roles were</td><td>Clinician - MA relationship (Trust)</td></tr>
<tr><td colspan="3">Clinician - MA relationship (Trust) →+ TS --||→+ MA capabilities →+ Clinician - MA relationship (Trust)</td></tr>
<tr><td colspan="3">Trust Loop: MA02 describes how the changes in teamwork they took time to develop. The first step was a policy to add capacity and combine teams, from 1:1 to 2:2. This new context initiates a self-reinforcing process which permits the new team to “get in gear”.

At the first, team members are worried about how the new working relationships will work out, indicating clinicians trust in MAs is low. As MAs take on new tasks in TS, over time this it causes them to develop new capabilities. But, this takes time to show. As they show this gradual learning to clinicians, the team members learn to work together more and more closely, indicating a high level of trust.</td></tr>
</table>

<table>
<tr><td>Participant-Quote number) “Quote … <u>phrases referring to variables</u>” (word count/total words in causal statement)</td><td colspan="3">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="7"></td><td>• as a team</td><td>• we have been working as an individual with a doctor</td><td>Capacity</td></tr>
<tr><td>• get in gear and get along together as a team, and know exactly what our roles were</td><td>• worried about working
• it seems to be working fine</td><td>MA satisfaction</td></tr>
<tr><td colspan="2">0</td><td>MA retention</td></tr>
<tr><td colspan="3">MA capabilities →+ MA satisfaction --||→+ MA retention →+ Capacity →+ TS --||→+ MA capabilities</td></tr>
<tr><td colspan="3">Team Learning loop: MA02 describes how establishing a more trusting relationship with clinicians and performing at a more advanced level (Trust loop), actually increases the effective capacity of the team.

As the MAs’ capabilities grow, they experience increased job satisfaction. Over time, this improves (not mentioned here) their retention and work attendance rates, which increase the effective capacity of the team. This more reliable/capable team permits more TS. With time, these more advanced roles add to capabilities such that MAs “know exactly” how to perform the tasks of their new roles.</td></tr>
<tr><td colspan="3">Capacity --||→+ MA satisfaction --||→+ MA retention →+ Capacity</td></tr>
<tr><td colspan="3">Sharing the Load loop: MA02 describes how, despite the time it takes to learn to get along together as a team, once it is done things work better.

With the Team Learning loop, the team’s effective capacity grows. With time, this effectively larger team causes individual MAs to experience greater job satisfaction as the workload is more evenly shared and they “get along together as a team”. This increased satisfaction is shown as MAs worry less about how the team will impact their work and are confident that it is working out well. This further strengthens (unmentioned) retention, which further builds capacity.</td></tr>
</table>

<table>
<tr><td>Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="8">CL01-25) “It has become a more organized process, more straightforward from [the scheduling to the visit].
“It has become more efficient… The involvement that we have given the MAs has really helped [us] so that we can spend more time [with patients]. Really, it could be the same amount of minutes [but] I think that it is more effective the way it is right now…
“I have observed that it’s a lot better: we communicate more often, we talk about the patients and what’s going on more often. [It] definitely, ends up being a better experience for the patient.” (98/134)</td><td>• it has become a more organized process
• it has become more efficient</td><td>Time Delay</td></tr>
<tr><td>• The involvement that we have given the MAs has really helped us spend more time [with patients]
• I have observed that it’s a lot better</td><td>Feedback Loop</td></tr>
<tr><td>• The involvement that we have given the MAs has really helped us so that we can spend more time [with patients]
• the way it is right now</td><td>TS</td></tr>
<tr><td>• has really helped us</td><td>MA capabilities</td></tr>
<tr><td>• we communicate more often, we talk about the patients and what’s going on more often</td><td>Clinician - MA relationship (Trust)</td></tr>
<tr><td colspan="2">Clinician - MA relationship (Trust) →+ TS --||→+ MA capabilities →+ Clinician - MA relationship (Trust)</td></tr>
<tr><td colspan="2">Trust loop: CL01 describes how Trust reinforces itself (“the involvement we have given the MAs has really helped us”) and how clinicians watch for improvements (“I have observed that it’s a lot better”) as they delegating tasks and then observe MAs’ ability to contribute to patient care (Trust loop).

CL01 describes a self-reinforcing process whereby clinicians on a team give increasing levels of involvement to MAs. Over time, the MAs show their improved ability to help clinicians, for example by communicating cogently and often about patient care. As clinicians observe these performance improvements, they are even more willing to shift tasks.</td></tr>
</table>

Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable** →+ **Model Variable** (→ = causal link, --||→+ = with delay, +/- = link polarity) Interpretation	
	0	**Visit time spent training MAs**
	• spend more time with patients • it could be the same amount of minutes	**Having sufficient time allotted for the visit**
	• a more organized process more straightforward from [the scheduling to the visit] • more effective • more efficient	**Visits on schedule**
	TS →+ Visit time spent training MAs →- **Having sufficient time allotted for the visit →+ Visits on schedule →+ TS**	
	Insufficient Time 1 loop: CL01 further elaborates on the kinds of improvements that clinicians watch for (“*I have observed that it’s a lot better*”) with growing task-shifting (Trust loop). These impacts include a better operation of daily work (“*more organized … more effective … more efficient*”), which comes after some unintended consequences of task-shifting are sorted out. CL01 describes how the increased involvement given to the MAs contributed to developing a “more effective” use of visit time. This improved use of visit time must have come after an interim period in which some visit time was used for on-the-job training of MAs. As training is completed clinicians become able to spend more time with patients during the visits. Improved use of visit time reduces wasted time in the clinicians’ daily schedule. As clinicians perceive their daily work as being more organized, effective and efficient, they are even more willing to shift tasks.	

<table>
<tr><td>Participant-Quote number) “Quote … <u>phrases referring to variables</u>” (word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="5"></td><td>• The involvement that we have given the MAs has really helped us</td><td>Capacity</td></tr>
<tr><td>• so that we can spend more time with patients
• it could be the same amount of minutes but I think that it is more effective
• we talk about the patients and what’s going on</td><td>Visits with coordinated, comprehensive care</td></tr>
<tr><td>0</td><td>Volume of patient visits</td></tr>
<tr><td colspan="2">Capacity --||→+ Having sufficient time allotted for the visit →+ Visits with coordinated, comprehensive care →- Volume of Patient visits →- Capacity</td></tr>
<tr><td colspan="2">Insufficient Time 2 loop: CL01 elaborates still further on the kinds of improvements that clinicians watch for (“I have observed that it’s a lot better”) with growing task-shifting (Trust loop). These impacts include improved patient care (“more time with patients … it is more effective”) which also comes after some unintended consequences of the clinic incentive policy are sorted out.

The greater involvement of MAs (via the action of the Trust loop and overcoming the Insufficient time 1 loop) gradually contributed to the team experiencing an increase its effective capacity (“the involvement… has really helped us”). With this, the team was able to make a “more effective” use of visit time, such that clinicians spend more time with patients. This added clinician-patient time and MAs’ new ability to collect and convey useful information about patient care for clinicians, in turn provides patients a better overall handling of their care. While it is not mentioned, providing patients with comprehensive care reduces the total volume of patients which can be seen in the short term, as MA training takes up time in the visit which means patients may have to use two visits to accomplish what should be done in one. In the long term, clinicians could choose to use this more efficient system for seeing more patients, rather than being more comprehensive. All of these factors, in turn, put a strain on teams’ capacity, which CL01’s team appears to have overcome to a significant degree.</td></tr>
</table>

An important aspect of MA04-10 for SD-S is the concept of loop dominance. MA04 tells a story with several stages of progression. This sequence can be interpreted here as the loops taking turns being *dominant*, meaning each is the primary cause of change in the system during that period. The first one refers to an early stage of personal learning. MA04 then refers to a second stage in which the team is made to wait for new team members (a hypothesized new member) to catch up. This is the Team Learning loop. In the final stage, a high-performing team emerges which enables individual team members to function comfortably under pressure, the Sharing the Load loop is dominant.

As will be discussed in the following section, system dynamics posits mental models to be perceptions of structure which are *analogous* to the real structure of an external system[118, 119]. It is to be expected, then, that mental models may contain *analogies* where the relations between physical objects in a well-understood domain are used to describe the causal structure of the less well-understood elements of a perceived system. This type of analogy is less complex than a mental model and is therefore termed a *mental image* in SDM[119] (to avoid confusion with the mental models studied in the tradition of Gentner[215]). MA02 and MA04 independently used the analogy of the interaction of gears to describe the mechanisms of the three loops ("Trust", "Team Learning" and then "Sharing the Load"). MA04-10 described mature team relationships and task-shifting as a "*meshing*" (the linkage of cogs in interacting gears which permits smooth turning). MA02-12 described the effort of building up the mature team relationships for task-shifting as a process of team members working to "*get in gear*".

Gears are used for converting leverage, or torque to create a *mechanical advantage*, a cause and effect relationship which generates an effect which is not proportional to the cause. This is the definition of non-linearity in SDM, which is "where result is not proportionate to cause"[107] (p. 4). In SDM, nonlinearity is assumed to be one of the "essential characteristics of … social processes"[107] (p. 50). In SDM models, key system non-linearities should come about through the action of feedback loops and/or time delays[178]. The system elements MA02 and MA04 are referring to in their metaphors include several feedback loops and time delays.

Considering that quote MA02-12 follows quote MA02-11, "get in gear" could be a reference to the shifting gears of a standard-transmission automobile. As described above, quote MA04-10 describes stages, which are interpreted as being shifts in dominance among the different feedback loops. The first one, the "Trust loop", achieves the initial momentum needed to break the inertia of the system. The second, "Team Learning loop" increases speed toward the goal. Then, the third, “Sharing the Load loop” completes the transition, sustaining effort moving forward in a high-performing team. The transitions in loop dominance, like shifting gears in an automobile in the build-up to cruising speed, are “tipping points” where the forces dominating or driving the dynamics of the system change from stability to inertia and finally to sustaining momentum.

This sequence of feedback loop dominance tells a story about how PCT occurs. There is also an emotional aspect to this story. The many aspects of care which must change with PCT are aggregated in “Task-shifting” (i.e., the “*things*” in quote MA02-11), which keeps track of changes in who is responsible for different sorts of “tasks”. Similarly, “MA capabilities” tracks the MAs’ average level of mastery across all of these tasks. Referring to MA02-13 (presented as a stand-alone quote above), it appears that an important source of emotional strain during the first phase of PCT results from the time delay in developing “MA capabilities”. MA04-10 uses a hypothetical story to describe what may be an autobiographical experience of new team members experiencing emotional strain from seeing how they slow the team down.

In Table 3.9 above, the same sequence of loops is used, as it fits the descriptions. This is not assumed to *necessarily* tell a story of the sequence by which the loops might take turns being dominant in all cases. Also, just because MA04-10 and MA02-12 are interpreted to perceive these three loops, and tell stories indicating a sequence of loop dominance, I do not interpret them to have mental models *with* loop dominance. Similarly, although these interpretations show an understanding of feedback loops, which are an important cause for system nonlinearities, I do not interpret them as having mental models *with* nonlinearity. The existence of these elements in mental models is an outstanding research question in SDM[180].

3.5.6.4 COGNITIVE RIGOROUSLY INTERPRETED QUOTATIONS FOR SYSTEM DYNAMICS SATURATION

The tables below present Cognitive RIQs (*Rigorously-Interpreted Quotations – for Cognition*). These are based on the system dynamics-related items coded in CM-S. The first bolded column includes the quotations. Underlined text indicates instances where quotations are interpreted to describe an aspect of mental models which supports the usefulness of SDM as the modeling approach in this research. In the following three columns, statements are interpreted. In the second bolded column, quotations are marked with an “x” when they express the existence of the types of mental models with which research in SDM is concerned – defined as *Mental Models of Dynamic Systems*[118, 119]. There are seven aspects in Doyle’s definition. Each of these seven aspects corresponds to an aspect under this bolded column (each aspect has its own un-bolded column). In the third column, quotations are marked with an “x” when they express perceptions that changing the content of mental models is part of participants’ perceived solution to the problem. For each quotation, below the mark(s) in each column, there is an interpretation which ties the quotation to the aspect(s) treated in that column.

Quotations in all four tables contain aspects substantiating the claim that participants are “aware of having an SDM-type mental model”. Table 3.10 presents quotations also substantiating the claim that participants see that “mental models matter”. This is shown in two ways: 1) that participants see mental models as existing in the problem and 2) that they see that changing mental models is necessary to solve the problem. The yellow boxes marked with an “x” highlight the progression in quotations, from ones that focus on the existence of mental models to ones that focus on their change, and then to quotations that incorporate both aspects.

Table 3.10 Participant Perceptions of Mental Models and of Their Importance

Participant-Quote number) "Quote … phrases referring to system dynamics-related items" (word count/total words in conceptual statement)	**Aware of having an SDM-type mental model**							**Mental models matter**	
	Relatively enduring	Accessible	Limited	Internal	Conceptual	Representation of an external system (historical, existing or projected)	Whose structure is analogous to the perceived structure of that system	Existence of mental models in the problem	Changes to mental models matter for the problem
NM01-47) "I don't see the whole picture here, I only see my part, so I trust my managers and my administrators that they are going to see the whole picture. So when I give suggestions, if it is something that may not work then I know that it may be [because of] something that I am not seeing." (57/57)	x	x	x	x	x	x	x	x	
	NM01 describes having an understanding of PCMH which is largely in accordance with what SDM terms a "mental model".							Policy-making in PCT relies on the use of accurate mental models.	
CM01-48) "Yes, I would say that you are able to see the outcome. Would you be able to measure it in quantifiable terms? On some areas, probably, maybe not in all of them, because a lot of it is qualitative versus quantitative. Of course, you are going to have that with everything. The observable pieces, I don't know, I just use the examples that I have from experience, you see your wait times dropping, those are measurable, those are outcomes, that your providers are going home at the end of the night instead of charting at home trying to catch up. All those pieces are done, when patients have more time, I guess yeah, it should be observable in concept. Have we done a great job exactly getting that information? I don't know. I don't know, I would have to think about that more. I think that you could definitely (do that). I think that you should be able to see and feel the difference, in one that is run one way versus another." (173/176)				x	x		x	x	
	CM01's mental model of PCMH defines success using both intangible and observable variables.							Monitoring PCMH is still conceptual. Stakeholders must have accurate mental models in order to perceive success.	

Participant-Quote number) "Quote … phrases referring to system dynamics-related items" (word count/total words in conceptual statement)	**Aware of having an SDM-type mental model**							**Mental models matter**	
	Relatively enduring	Accessible	Limited	Internal	Conceptual	Representation of an external system (historical, existing or projected)	Whose structure is analogous to the perceived structure of that system	Existence of mental models in the problem	Changes to mental models matter for the problem
CM01-49) "How do you know that you have achieved [PCMH?] And what is it going to take to get those key … [PCMH] pieces? I don't know exactly how you would be able to exactly tell. I think that it's almost that you have to commit to the concept, you have to commit to the, not necessarily tiny little components, but maybe that's what it takes for some people. For me … personally, it's a conception of [PMCH]. You will know it when your patients are happy, you don't have wait times … you get calls back efficiently, you have your quality indicators where they need to be. So … when those [key pieces] are taken care of, then things are flowing … That's the concept of [PCMH], the services revolve around the patient not around each individual provider or team. They are working together for the patient…"It's really the concept but you have to break it down, also, to say that these are the key components or the must haves to get you there. If you are just doing those actions does that really mean that you have [PCMH], or are you [just] doing those actions? What is the end result? What is it about? It is about the patient and about their experience, and the quality of the care that they are receiving…" (225/342)				x	x	x	x	x	
	For CM01, success in PCMH involves adopting a certain conceptual understanding including future expectations for several variables.							Because measurement of success in PCMH is uncertain, achieving PCT requires changes to expectations first.	
CM01-50) "Yes and no. [PCMH is] easy to understand, but there is a lot more detail to it … you can simply say, 'planned care'. But what does that mean? And defining what that means and have everybody buy in to what that definition really truly is and (what) all those little key pieces are [is hard]." (54/53)			x			x	x		x
	For CM01, changing behavior requires developing agreement on the details of PCMH's structure.							It is difficult to build a shared understanding of policies.	

Participant-Quote number) “Quote … phrases referring to system dynamics-related items” (word count/total words in conceptual statement)	**Aware of having an SDM-type mental model**							**Mental models matter**	
	Relatively enduring	Accessible	Limited	Internal	Conceptual	Representation of an external system (historical, existing or projected)	Whose structure is analogous to the perceived structure of that system	Existence of mental models in the problem	Changes to mental models matter for the problem
CM01-51) “Well, my level of commitment is that I want to do what is best for the patient. For commitment, I would like to understand it more. I don’t feel like I have a complete understanding of it. We … have talked about the Patient-Centered Medical Home. And for some reason, the little details … are not really coming to the front of mind right now, because I am not really well-versed in it, or as much as I would like to be, or to understand how Healthcare reform is a part of that as well. Because there are so many things happening … is it moving to healthcare reform or is this not a piece of healthcare reform. There's a lot of things going on with healthcare right now … How [PCMH, health reform and Accountable Care Organizations] all relates, I am not real sure. But I do have commitment to do what is right for the patient and what is right for healthcare.” (163/248)	x	x	x	x	x	x	x	x	x
	CM01 describes an awareness of having a mental model from which details can easily come to the "front" of the mind, and is limited with respect to PCMH. CM01 desires to have a mental model of PCMH which has more detail, is more accessible, is better able to handle complex interactions and dynamics and has a broader system boundary.							CM01’s commitment to PCMH depends on the extent to which it *relates* to ‘what is best/right’.	
CM01-52) “The main piece in my mind about [PCMH is] it’s not just [x], but it is the starting from the ground up and saying ‘this is the way that things flow and work’, ‘this is the way the clinic is designed’, ‘this is the way that our [electronic medical record] falls into this’, ‘this is how pre-visit planning falls into this’. It is a whole lot easier to put something in place as a designed practice, than to try and change people that have been doing this for 20 years in an old model. Trying to get them into that concept.” (100/100)	x				x	x	x	x	x
	CM01 describes an awareness of key stakeholders having mental models. These mental models are made up of many interrelated concepts and are difficult to change.							Achieving PCT involves modifying both policy and cognitive structures.	

Table 3.11, Table 3.12 and Table 3.13 follow the same structure as Table 3.10, with an additional fourth column[49]. In this column, quotations are marked with an "x" when they express that the needed mental model changes will require engaging the target group on a sub-conscious emotional level. This is shown in two ways: 1) that *changes to the system* involve emotion and 2) that *changes to mental models* also involve emotion. Again, the yellow boxes marked with an "x" highlight the progression in quotations.

Table 3.11 focuses on quotations expressing the emotional aspect of changes to the system, first alone and then in conjunction with mental models as existing in the problem.

Table 3.12 focuses on quotations expressing the emotional aspect of changes to mental models in combination with the recognition that changing mental models is necessary to solve the problem, and then in conjunction with both aspects indicating that "mental models matter".

Table 3.13 focuses on more complex quotations that express many of these aspects together.

Taken together, these tables present quotations supporting assumptions which are crucial to the usefulness, applicability, and the likelihood of acceptability of SDM research.

[49] Table 3.10 did not have this column as those quotations did not contain this aspect.

Table 3.11 Participant perceptions of mental models, of their importance and of the emotions tied to system change

Participant-Quote number) "Quote … phrases referring to system dynamics-related items" (word count/total words in conceptual statement)	**Aware of having an SDM-type mental model**							**Mental models matter**		**Emotional engagement**	
	Relatively enduring	Accessible	Limited	Internal	Conceptual	Representation of an external system (historical, existing or projected)	Whose structure is analogous to the perceived structure of that system	Existence of mental models in the problem	Changes to mental models matter for the problem	Changes to the system involve emotion	Changes to mental models involve emotion
CM01-53) "It's great to [meet with HSDOs] in different parts of the country to see where we are and we used to be there, even though we are here now. We are not happy with where we are at, we are still are doing much better than other places and I remember when we were like that, but we are here, and even though it's still frustrating, and we haven't achieved it yet. We have made progress towards that goal." (79/217)							x			x	
	CM01 recognizes PCMH implementation involves a long time delay which is hard to observe.									Slow changes challenge expectations and cause negative emotions.	
CM01-54) "It is a challenge <<laughter>>. I love a challenge. The most exciting thing is we are really looking forward to [changing x] to kick this off. To kick this back into gear, as far as working towards the [PCMH] model… I would say [x] would be a frustration … the barrier is being removed and we are excited about the possibilities." (61/119)							x			x	
	CM01 describes PCT using an analogy ("kick ... gear") of an automobile which needs to shift down in order to build momentum.									CM01 attaches strong emotions to the potential of policy to promote PCT.	

Participant-Quote number) “Quote … phrases referring to system dynamics-related items” (word count/total words in conceptual statement)	**Aware of having an SDM-type mental model**							**Mental models matter**		**Emotional engagement**	
	Relatively enduring	Accessible	Limited	Internal	Conceptual	Representation of an external system (historical, existing or projected)	Whose structure is analogous to the perceived structure of that system	Existence of mental models in the problem	Changes to mental models matter for the problem	Changes to the system involve emotion	Changes to mental models involve emotion
CM01-55) “There is a lot more that we can try; and change isn’t always a bad thing… [it allows to ask] “Ok, now that [x] barrier is removed, how can we make [y] work better? What do we need to do to facilitate all the issues that we have had with [y]?”[and to say] "Well, this resolves some of that by having more [x] … [and] you need [x] right now, which will make [z] better." That is the exciting thing about [PCMH], when we are able to fully achieve what we want to do, that when we get there, things will flow, our patients will be happy…” (105/173)					x	x	x	x		x	
	CM01’s mental model of PCMH has a causal structure made up of variables and causal links with polarity which interact in a highly interrelated system.							A prospective mental model provides optimism and is used to prioritize decisions.		CM01's mental model includes hopeful expectations about imminent changes to the system.	
CM01-56) “Some of the changes were harder than others, and some we are still struggling with. I mean the structure, [PCMH] hasn’t really changed the structure, the whole structure of the clinic ... We have gotten to the point where we want to keep moving forward [but x] is not allowing us to have a very good design, with how we want to utilize [y]. … The idea is [changing x] would increase in efficiency, and that is a significant structural change to the [primary care] process so with that ability to be able to do that and we are really hopeful and planning on making some big improvements in efficiency in that way.” (111/171)					x	x	x	x		x	
	PCMH implementation involves many changes, some more difficult than others. The content and difficulty of changes evolves over time. The later, more difficult ones involve "significant structural changes”.							CM01 uses a mental model to make policies which involve “significant structural changes”		Emotional struggles with system change evolve over the course of implementation.	

Table 3.12 Participant perceptions of mental models, of the importance of changing them to change the system and of the emotions tied to such change

Participant-Quote number) "Quote … phrases referring to phrases referring to system dynamics-related items" (word count/total words in conceptual statement)	**Aware of having an SDM-type mental model**							**Mental models matter**		**Emotional engagement**	
	Relatively enduring	Accessible	Limited	Internal	Conceptual	Representation of an external system (historical, existing or projected)	Whose structure is analogous to the perceived structure of that system	Existence of mental models in the problem	Changes to mental models matter for the problem	Changes to the system involve emotion	Changes to mental models involve emotion
CM01-57) "[It's] simple, and it makes sense. Completely makes sense: '[PCMH]' well, yeah, let's do it. But getting it to the point where everybody is on the same page and everybody understands exactly what pieces are the must-haves is the tough piece. Because everybody has different perceptions of what the must-haves are, or the pieces that they don't want to do because they just don't want to, or the pieces that they need to do, in order to make it work.(80/81)	x	x	x	x	x	x	x		x		x
	For CM01, changing mental models requires building understanding of how concepts fit together, which is hard because of the attributes of mental models. CM01 observes significant differences across individuals' mental models.							It is hard for behavior to change without first changing mental models.		This change involves engaging people on their needs *and* wants.	
CM01-58) "There was a [clinician] … that couldn't get his [x skill] down. He couldn't get the concept, and he just got frustrated that he wouldn't be able to continue on his guarantee, making the money that he needed to be, not for lack of desire or that he thought that it wasn't a good idea, but it just wasn't for him. He had a really hard time with [x and y skills], so he couldn't see the numbers that he felt that needed to succeed. He ended up leaving. "We did have a [clinician] that did fight against it quite a bit and he was terminated. Just for, not just that, but for many other issues that, just his lack of cooperation."(122/130)	x				x				x		x
	CM01 sees the process of clinicians adopting PCMH practices as one of modifying concepts in their mental model. CM01 sees this process as harder for some clinicians than others.							Not all stakeholders are amenable to making the changes which are required by PCT.		Emotional strain from changing mental models can be for highly-sensitive, tangible issues like a gap between desired and actual salary.	

<table>
<tr><th rowspan="2">Participant-Quote number) “Quote … phrases referring to phrases referring to system dynamics-related items” (word count/total words in conceptual statement)</th><th colspan="7">Aware of having an SDM-type mental model</th><th colspan="2">Mental models matter</th><th colspan="2">Emotional engagement</th></tr>
<tr><th>Relatively enduring</th><th>Accessible</th><th>Limited</th><th>Internal</th><th>Conceptual</th><th>Representation of an external system (historical, existing or projected)</th><th>Whose structure is analogous to the perceived structure of that system</th><th>Existence of mental models in the problem</th><th>Changes to mental models matter for the problem</th><th>Changes to the system involve emotion</th><th>Changes to mental models involve emotion</th></tr>
<tr><td rowspan="2">CM01-59) “So what pieces are those must haves? And we have tried to define those in the past … When providers say, ‘I don’t care what you tell me to do, I am not going to have that MA in the room, and if you make me have one in the room, then that’s fine, they can sit there in the corner, while I do my work.’ I mean, seriously they have said that. So, it’s like ‘you can’t make me do that. That’s not the way I practice. I don’t want to do that. This is my job to chart and I don’t want someone typing for me’, even though that may make them more efficient …” (117/226)</td><td>x</td><td></td><td></td><td></td><td></td><td></td><td></td><td>x</td><td></td><td></td><td>x</td></tr>
<tr><td colspan="7">Clinicians’ mental models are not changed easily. For some, even a persuasive argument about system improvement is not acceptable.</td><td colspan="2">Clinicians associate their mental model with their very identity.</td><td colspan="2">Clinicians are described to have displayed strongly emotional, irrational responses to efforts to change their behavior.</td></tr>
<tr><td rowspan="2">CM01-60) “For providers … [and for] staff, that you have to coordinate and get them all on the same page. It just gets overwhelming … we have to convert that, and we have to change that in their minds … We have to continually resolve concerns and frustrations that they go through on a daily basis to keep the ball moving… The communication is a lot more difficult when you have larger teams of providers, larger teams of MAs, getting everyone on the same page.” (85/262)</td><td>x</td><td></td><td></td><td>x</td><td></td><td>x</td><td></td><td>x</td><td>x</td><td></td><td>x</td></tr>
<tr><td colspan="7">Clinicians’ and staff members' mental models are incomplete. Ideally, these mental models would include a shared understanding of PCMH.</td><td colspan="2">CM01 sees constantly building shared understanding as part of PCMH.</td><td colspan="2">Building shared understanding causes emotional strain, and is hard with large groups.</td></tr>
</table>

Table 3.13 Participant perceptions of mental models, of their importance, of the importance of changing them to change the system and of the emotions tied to these issues

	Aware of having an SDM-type mental model							Mental models matter		Emotional engagement	
Participant-Quote number) "Quote … phrases referring to system dynamics-related items" (word count/total words in conceptual statement)	Relatively enduring	Accessible	Limited	Internal	Conceptual	Representation of an external system (historical, existing or projected)	Whose structure is analogous to the perceived structure of that system	Existence of mental models in the problem	Changes to mental models matter for the problem	Changes to the system involve emotion	Changes to mental models involve emotion
CM01-61) "They like things stable, no change. They don't like change. 'I get into a routine and that is the way I am. I have been practicing medicine for 30+ years'". (30/30)	x								x	x	x
	Clinicians' mental models are not easily changed, they are relatively enduring.							Changing clinicians' mental models is a prerequisite for PCT.		Clinicians associate their mental model with their identity.	
CM01-62) "But this is no longer that type of practice. It is really just a process …. It is definitely a negotiation, a negotiation to find better ways to deliver care, [and to persuade MDs] that maybe change isn't so bad. It might be painful a little bit, but look what could be in the end if we could get there." (60/76)						x			x	x	x
	PCMH implementation involves the use of a mental model with causality, and patterns of expected behavior over time.							CM01 uses a "*reference mode*" to negotiate for mental model change.		This reference mode is an emotionally-compelling story.	
CM01-63 "[When I] started [implementing PCMH], I had great visions, what it would be, could be, was going to be. Getting in to it, you kind of just get beat down sometimes with it. You think, "is that really possible?" (39/46)						x		x	x	x	x
	At first, CM01's mental model of PCMH included optimistic expectations for the future. These eroded over time as the reality of implementation continually refuted original expectations.							A prospective mental model of PCT helps CM01 make sense of challenges.		Challenges to CM01's mental model cause strong negative emotions.	
CM01-64) "We do try to work with them. We try to help them along the road. 'Look, this is how this is going to help you, this is going to be more efficient' and that is how we get them on board. But the key is, getting provider buy-in, at the first of it. And then getting them excited about it, because if you can't get them to that point then they just want to keep things the way it is." (80/80)	x	x	x	x	x	x	x	x	x	x	x
	PCT stakeholders all have mental models which include a causal structure of PCMH, are susceptible to change.							The process of mental model change is at the core of PCT.		Convincing with emotions is a prerequisite to lasting mental model change.	

3.5.7 STAKEHOLDER DIALOGUE SUITABILITY

Stakeholders were engaged in dialogue throughout this research, from the scoping study to the policy analysis (as described in the methods chapter).

A wide range of stakeholder types were engaged, although only some of the stakeholder types were able to re-engage in dialogue at various points throughout the research. Some stakeholders were engaged only once due to contextual constraints. Some stakeholder categories were never engaged in dialogue (e.g., MAs). This omission was not intentional. Future researchers could include the engagement of relevant stakeholder categories, especially those previously omitted and/or less frequently engaged.

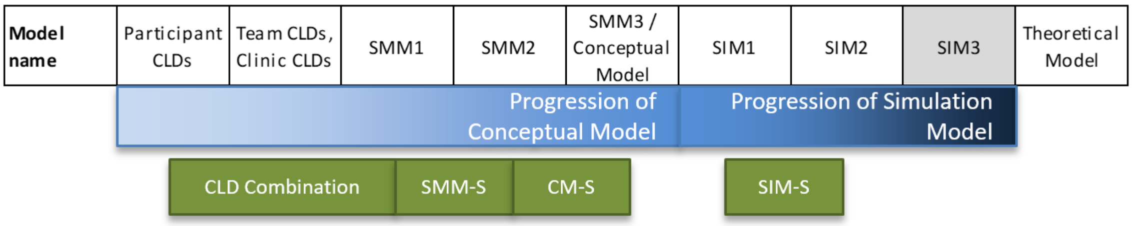
The methods for their engagement also varied. I utilized pre-existing meetings (i.e., HSDO management meetings, conferences, and mentoring meetings). I also initiated meetings specifically for stakeholder dialogue. Throughout these meetings, I made an effort to engage in open dialogue with stakeholders, to explore areas where they disconfirmed results with which they were presented (i.e., the tensions, the SMM, equations within the simulation model, model runs, decision functions).

Based on the reflection described above, I made the determination that the stakeholder dialogue was suitable, specifically that: the research results were assessed for congruence with stakeholders' experiences/knowledge in all the different ways that they can be explored in SDM.

3.6 PHASE 4 – POLICY ANALYSIS & THEORETICAL FINDINGS

This section presents results from running the final version of the Simulation Model (SIM3) under different policy scenarios (Section 3.6.1). These are scenarios which teams may experience in attempting to transition from traditional primary care to transformed primary care in PCT. Wherever possible, the scenarios are generated by making one policy change at a time. Scenarios are experiments in the model which enable us to predict the impact of different policy changes on the time paths of all variables in the system. This section presents these impacts *only for relevant variables*. Following this, the Theoretical Model is presented (Section 3.6.2). This diagram is the result of all of the model development work to date and depicts the dynamic insights of the research.

3.6.1 POLICY ANALYSIS



This section presents results obtained by using the Simulation Model in policy analysis. This analysis considers changes to policies (e.g., how much to pay clinicians, how to hire MAs) and alternative preferences (e.g., a team with more reticent clinicians, more reticent MAs, patients who are more sensitive to service delays). Simulation results are organized as follows:

Section 3.6.1.1 presents a **no task-shifting scenario**, where there is no task-shifting kickstart and the model remains in equilibrium throughout the simulation time period.

Section 3.6.1.2 presents a **Successful PCT scenario** (i.e., the base-case with the task-shifting policy).

Section 3.6.1.3 presents scenarios implementing various **clinic management policies**.

Section 3.6.1.4 presents scenarios with various **alternative preferences**.

Section 3.6.1.5 presents scenarios implementing **management policies with alternative preferences**.

3.6.1.1 NO TASK-SHIFTING SCENARIO

The Simulation Model starts out in equilibrium, an institutional *homeostasis*[216]. By reducing the kickstart amount to zero, task-shifting remains dormant in the model and the model remains in the initial equilibrium for the duration. The level of actual adherence to clinical guidelines remains at 50% overall (12.5 % from completing technical tasks (red) and 12.5% from completing non-technical tasks (green) and 25% from completing the MA-only tasks). Clinicians perform 50,000 tasks per month (25,000 technical and 25,000 non-technical tasks; there are no training tasks being performed). Clinical staff members perform 50,000 tasks per month (these are MA-only tasks, there are no advanced tasks or training tasks being performed). The MAs have 50 of the 100 possible capabilities (that would have under full task-shifting). They remain sufficiently satisfied (75% satisfied) such that they do not leave (MA personnel remains at 2 people throughout the duration, with no hiring or turnover). These results are shown in Figure 3.75 below.

As there was no kickstart amount of task-shifting, clinician willingness to shift tasks remains at zero. The four variables influencing clinician willingness to shift tasks also remain unchanged through the duration of the simulation. The current level of willingness would allow for more willingness (graph value of one). The MD's perceived salary is in line with the desired salary so this, alone, would allow task-shifting (graph value of one). The MD's perception of patient satisfaction would also allow task-shifting (graph value of one). However, the perceived level of MA capabilities (50/100) reinforces the clinicians' perspective that task-shifting should not be tried (graph value of zero). These results are shown in Figure 3.76.

Finally, from the clinic manager's perspective, the clinic's bottom line remains healthy without transformation. The MD costs are constant since they are paid a fixed salary and benefits, the MA costs are also constant as they do not gain capabilities (the requirement for increasing MA salary and thus MA cost), and revenue collected each month is greater than the monthly expenses. See Figure 3.77 below for accompanying graphs.

Figure 3.75 Equilibrium Run Showing Adherence, Tasks, MA Capabilities, Perceived MA Satisfaction, MA Personnel

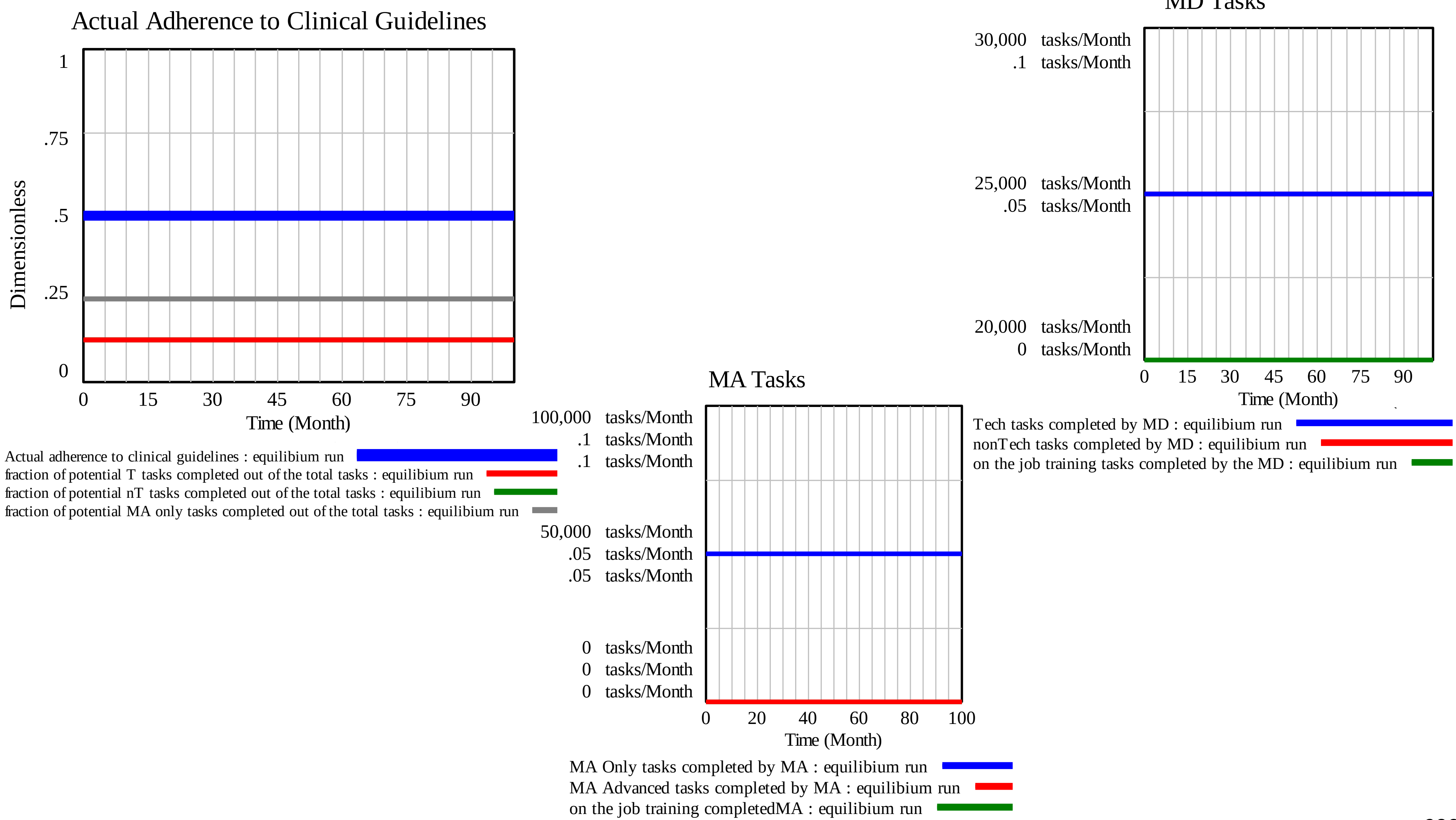


Figure 3.75 Equilibrium Run Showing Adherence, Tasks, MA Capabilities, Perceived MA Satisfaction, MA Personnel (continued)

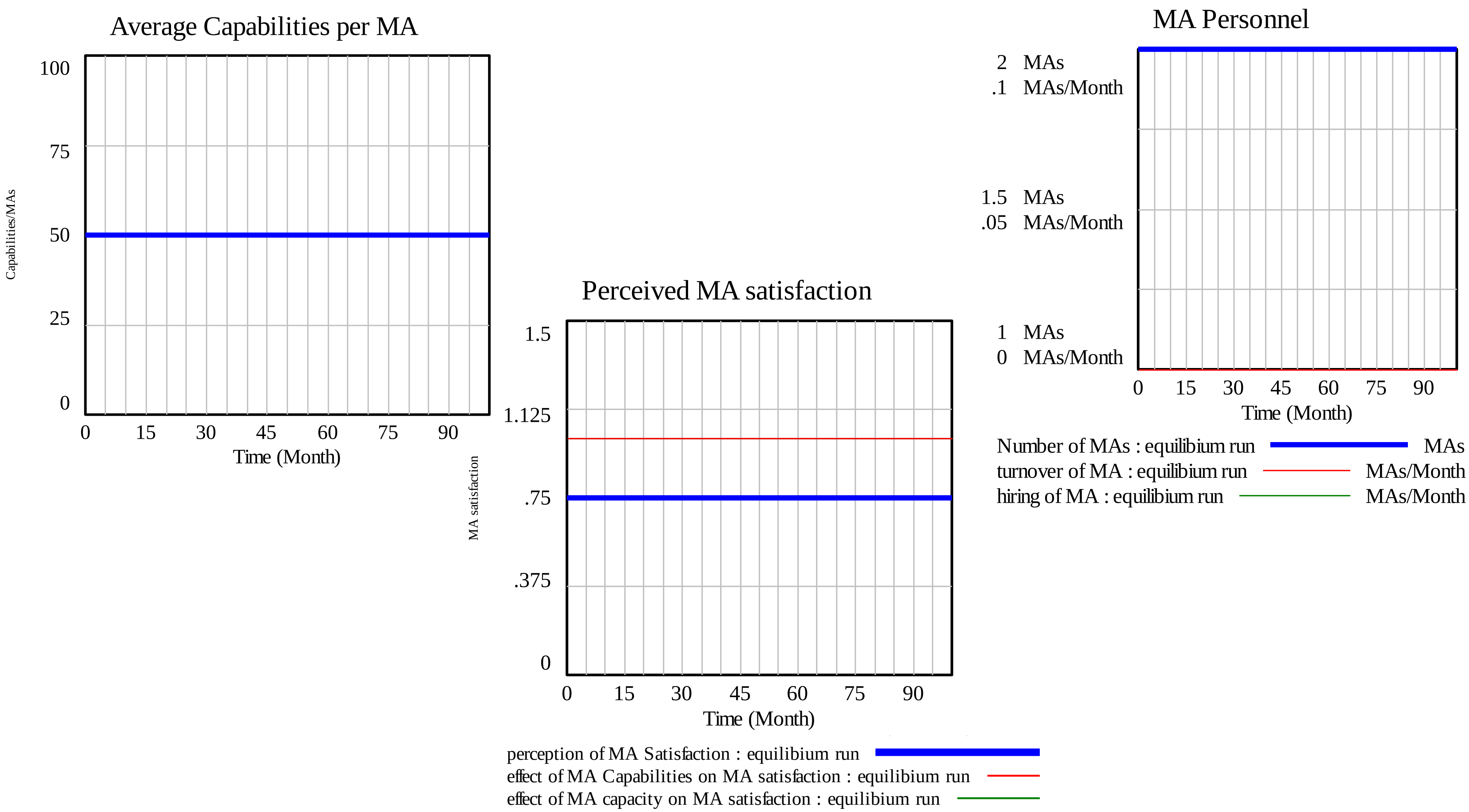


Figure 3.76 Equilibrium Run Showing Clinician Willingness

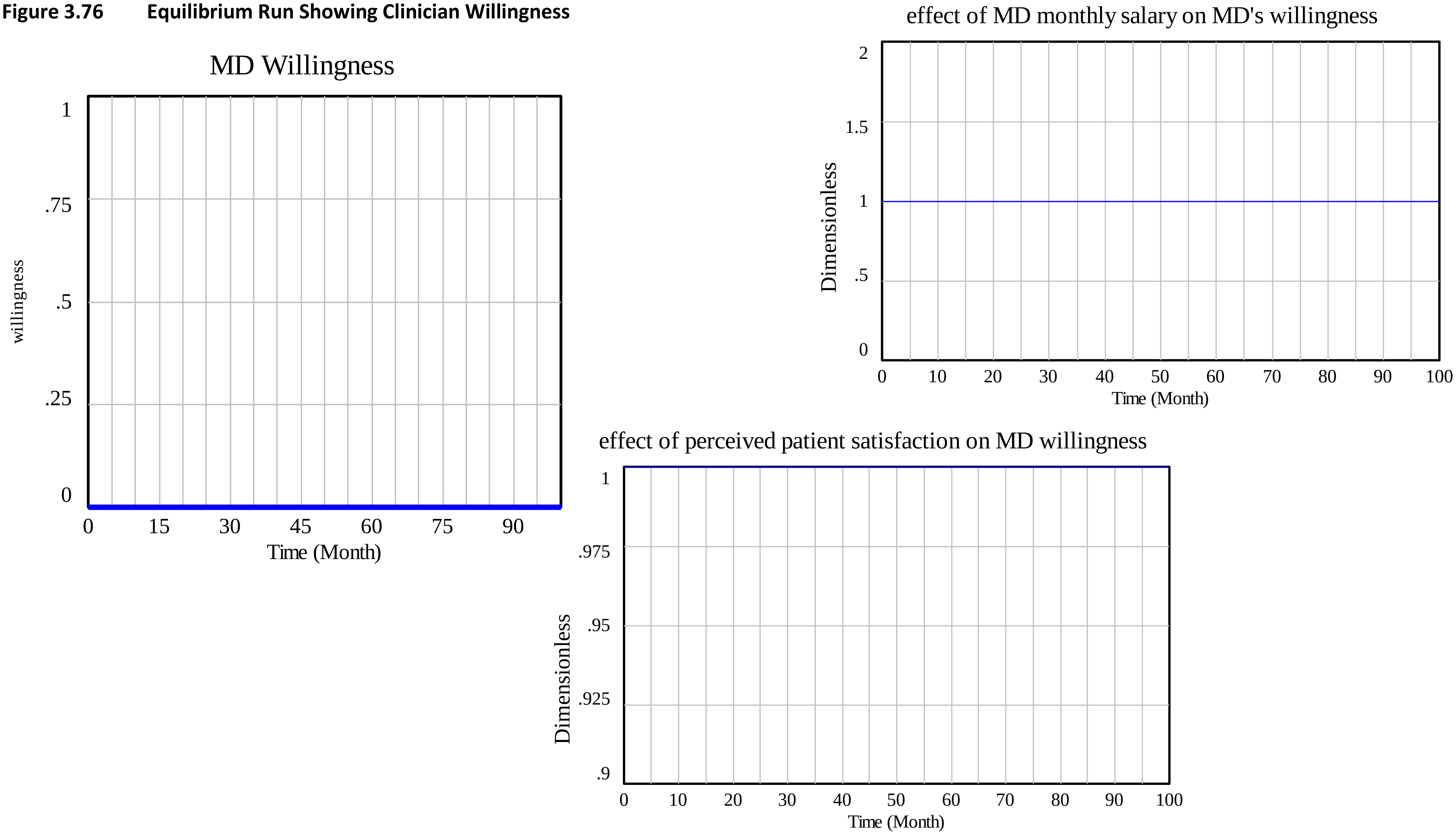
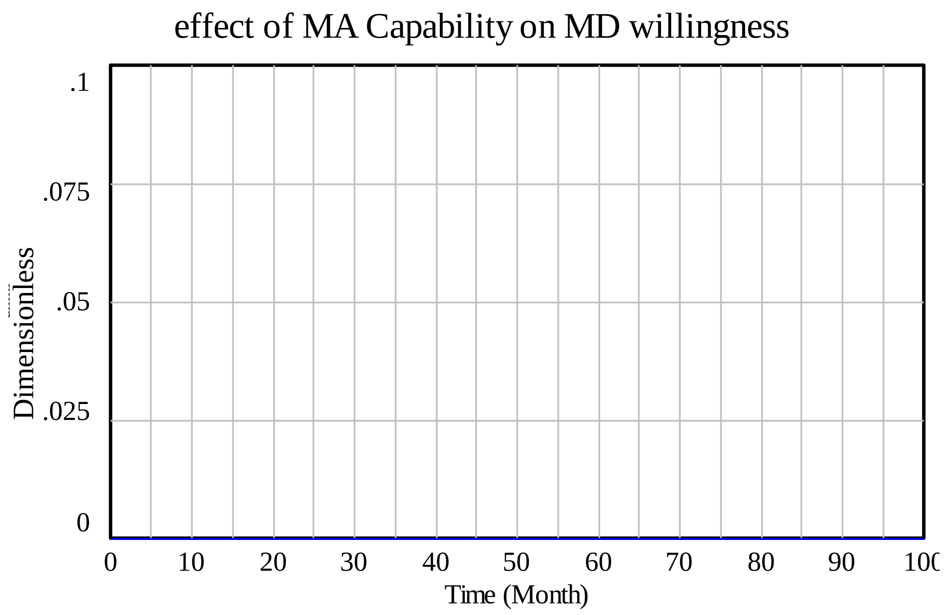


Figure 3.76 Equilibrium Run Showing Clinician Willingness (continued)



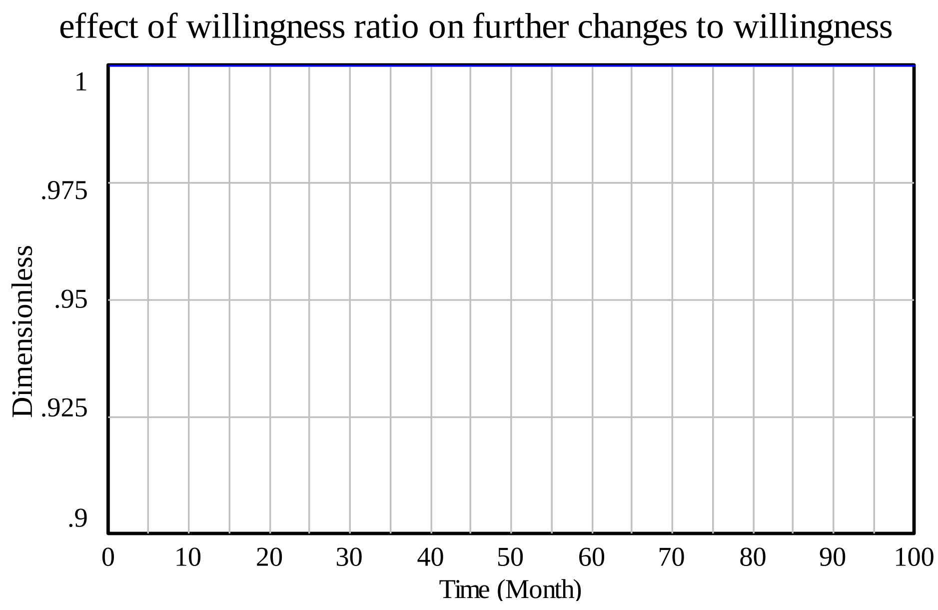
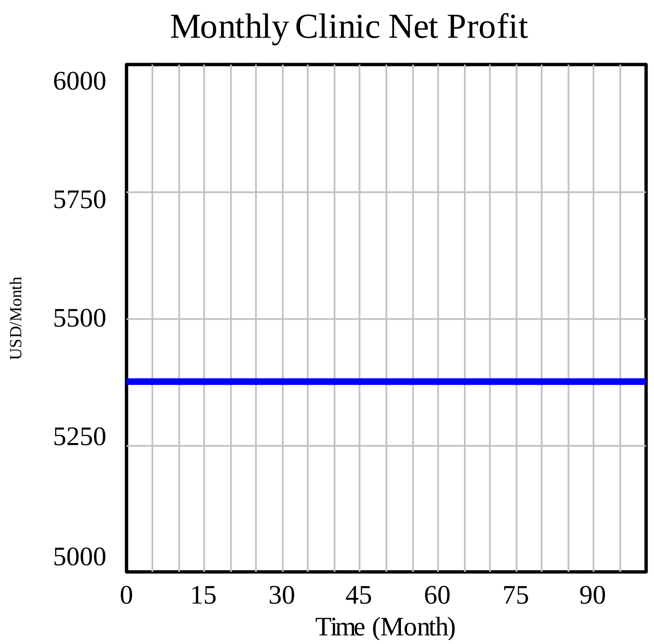
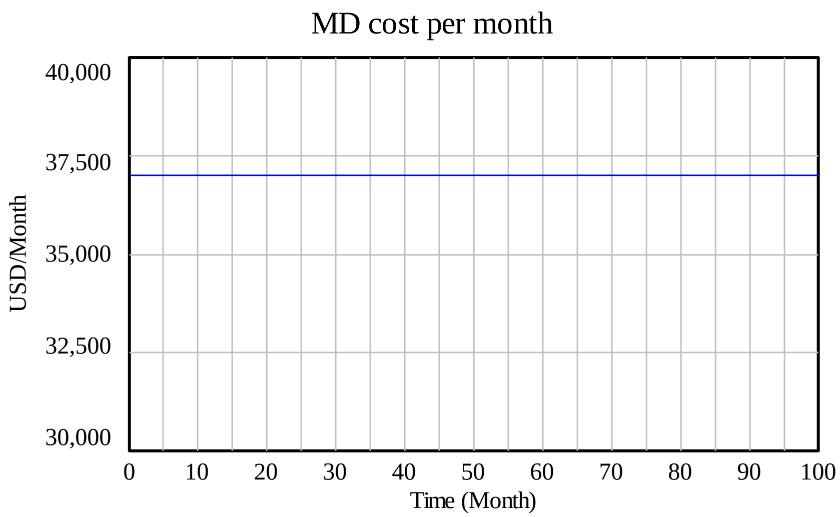
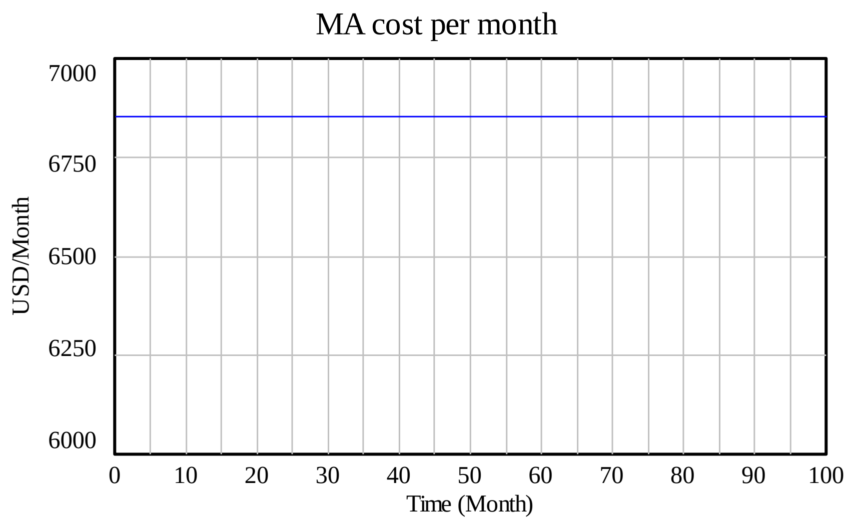


Figure 3.77 Equilibrium Run Showing Financial Variables







3.6.1.2 SUCCESSFUL PRIMARY CARE TRANSFORMATION SCENARIO

In this scenario, *Successful PCT*, the simulation also starts out in equilibrium. In this state, things operate smoothly, but there is something missing: the full value of primary care is not being realized for the panel of 4000 patients. All of this could change with the introduction of a new program: Primary Care Transformation (PCT). The team's potential for success in this program is good, it already includes clinicians (also called MDs) working together and sharing a small group of dedicated clinical staff members – with the biggest time commitment from the full-time medical assistants (MAs). The base-case also involves paying the clinicians a fixed salary over the course of the transformation and it assumes certainty about the amount of revenue received for work completed (i.e., no pink noise, see Section 3.4.2.5 above). However, the forces underlying the status quo are strong – and they are mostly entirely unknown because no one has bothered to disturb them, until now (in month 10).

The stocks in the model (e.g., tasks, MAs, Patient Satisfaction, MA Satisfaction etc.) are controlled by decisions made by the represented agents in the model (patients, managers, clinicians and clinical staff members). These decisions, or policy functions, were described in Section 3.4.2. Some of them are *policies* in the sense of making a yes/no choice to a question of what to do in a certain case. The majority of them are *preferences*, where decisions are made in a tentative manner based on a preference.

This scenario assumes an "average" value of the preferences in the model (generally, not too willing/eager to change but also not too resistant/hesitant to change).

At month 10, a small kickstart amount of task-shifting is introduced. In this scenario, PCT works as hoped and the end result is that full adherence to clinical guidelines is attained with a team consisting of six MAs and two MDs producing 1600 encounters per month for their panel of 4000 patients. This is accomplished over a period of approximately five years as MDs increasingly focus on the tasks that only they can accomplish and shift the tasks that can be shifted to the MAs. As tasks are shifted (and MDs and MAs spend the time needed to figure out how to implement these tasks on their team), MAs become more capable. Capacity strains on MA satisfaction are eventually overcome by the effect of increased MA capabilities on their satisfaction, and by the increased capacity due to hiring more MAs and retaining them as they become more capable. During these five years, MA hiring is able to keep ahead of MA turnover such that the capacity strain is less than it otherwise would be. See Figure 3.78 below for accompanying graphs.

This transformation was made possible by a small initial kickstart of task-shifting. This initial task-shifting occurs due to a management requirement that MDs "try it out a little". After the initial kickstart, the increase in task-shifting is entirely in the hands of the clinicians on the team. This willingness is informed by four feedback processes involving clinicians' perception of: MA capabilities, their monthly salary, the satisfaction of their patients and their current willingness. In the base-case, the clinicians see the MAs grow in their capabilities over time (increasing their willingness, all else being equal). They see that their salary is unaffected (not increasing or decreasing their willingness, all else being equal). They see that their patients' satisfaction dips down during the capacity trained period (temporarily decreasing their willingness, all else being equal). Finally, as they grow increasingly willing, their change in willingness slows. Together, these inputs slow the growth in willingness at times, but ultimately, full willingness (and full task-shifting) is achieved. See Figure 3.79 below for accompanying graphs.

From the clinic manager's perspective, the transformation proceeds satisfactorily while this team's transformation positively impact's the clinic's bottom line. Facility costs for this team are assumed to remain the same across this time period. The MD costs are constant since they are paid a fixed salary. The MA costs increase as their individual capabilities increase and the number of MAs also increases. See Figure 3.80 below for accompanying graphs.

Figure 3.78 Task-shifting Run Showing Adherence, Tasks, MA Capabilities, Perceived MA Satisfaction, MA Personnel

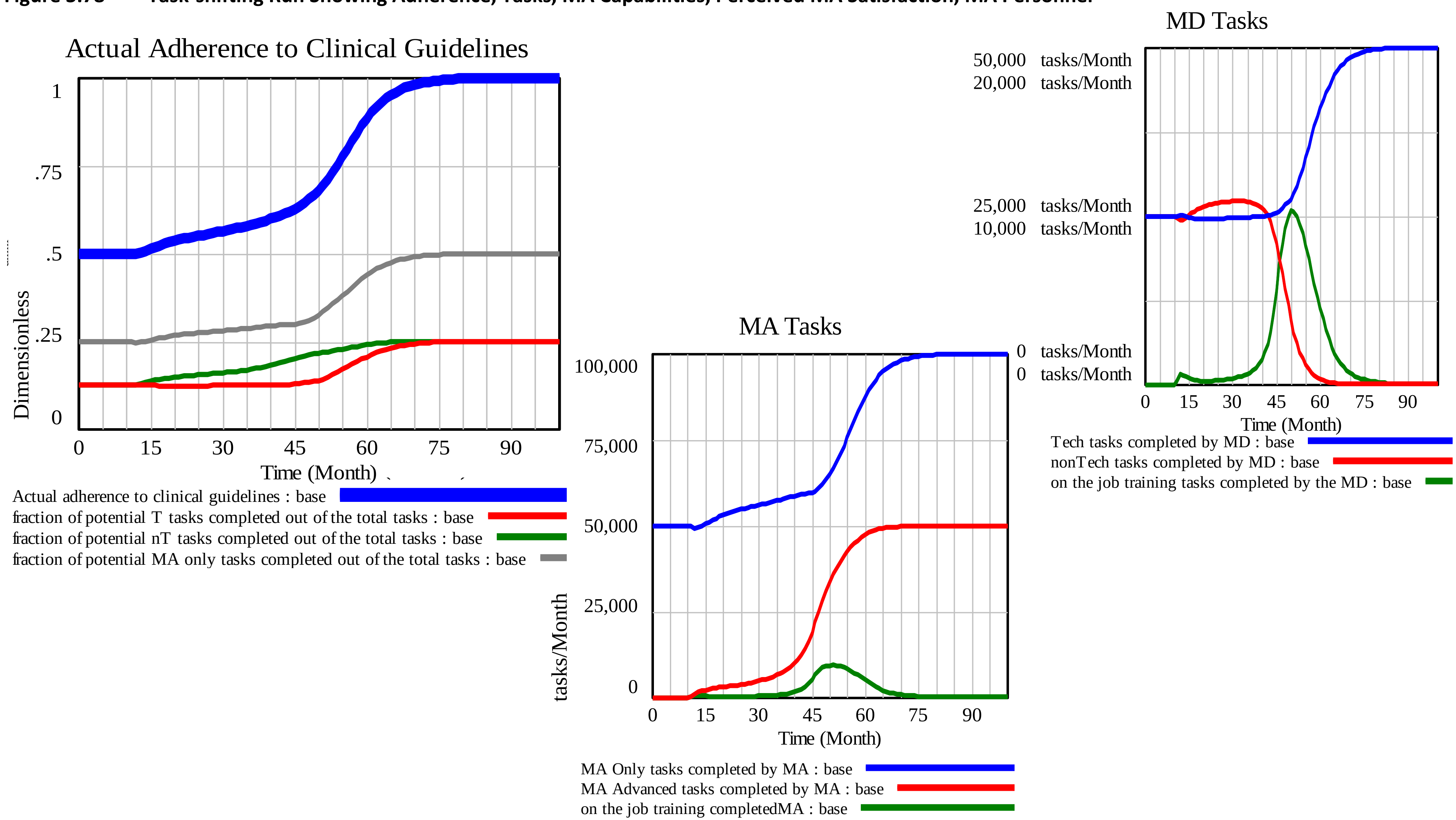
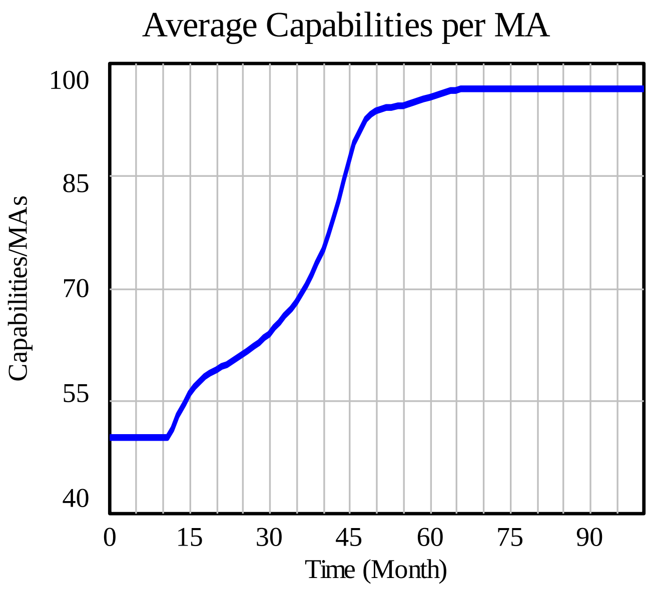
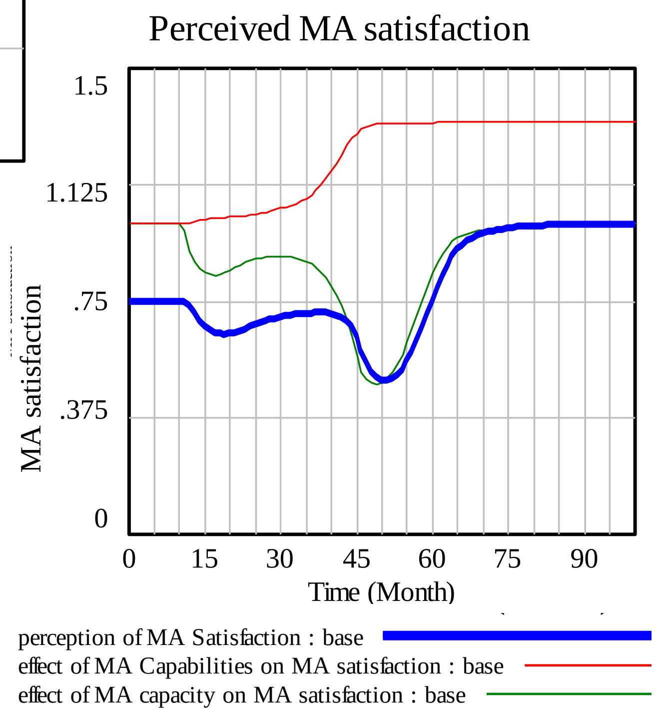


Figure 3.78 Task-shifting Run Showing Adherence, Tasks, MA Capabilities, Perceived MA Satisfaction, MA Personnel (continued)





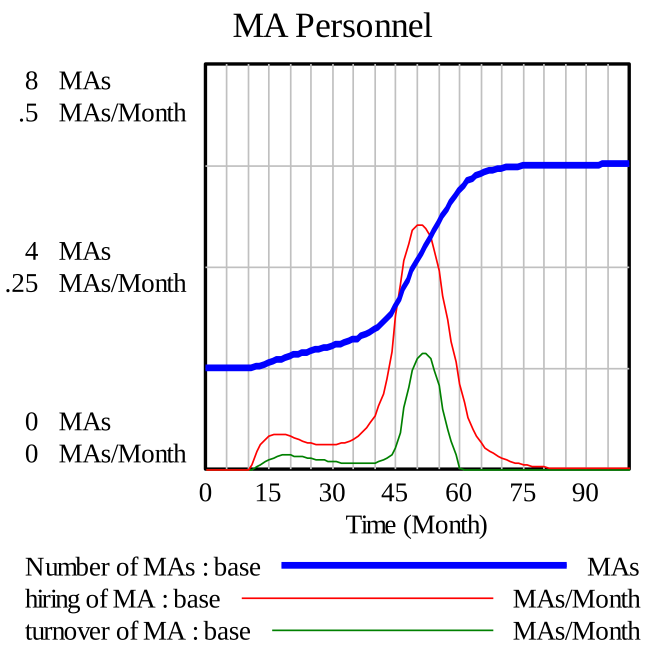
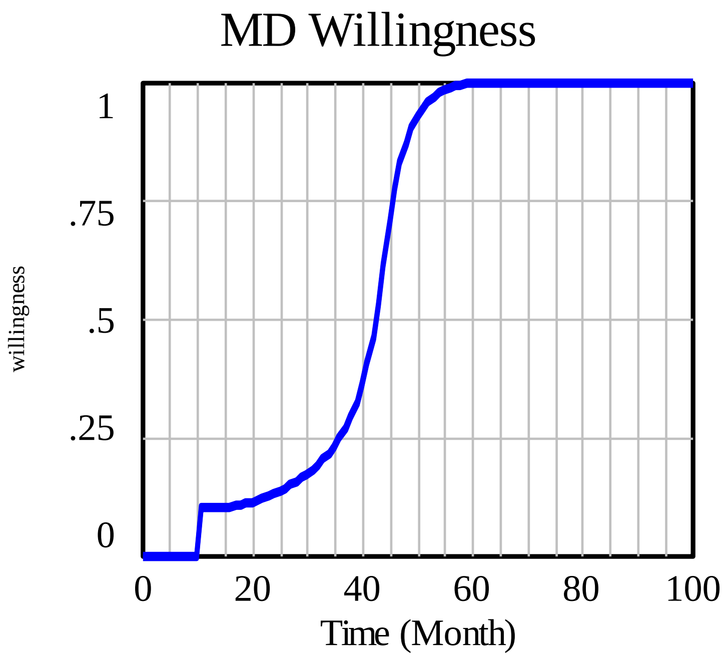
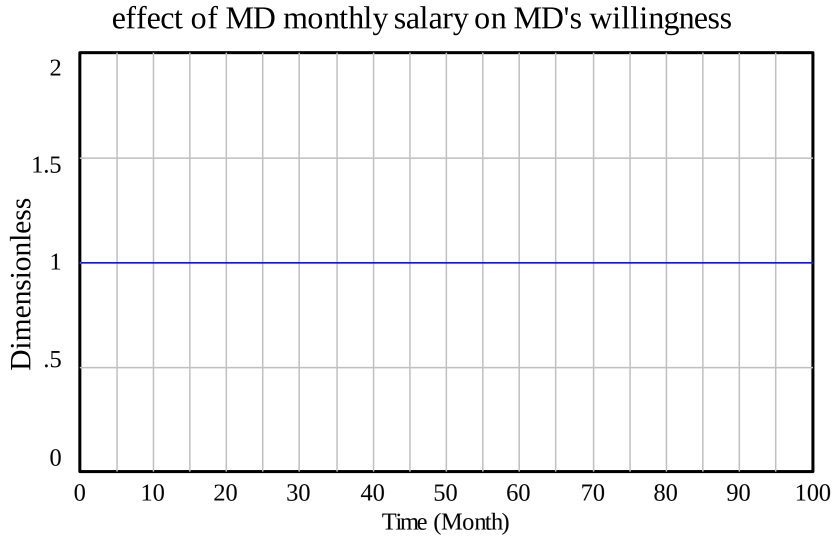


Figure 3.79 Task-shifting Run Showing Clinician Willingness





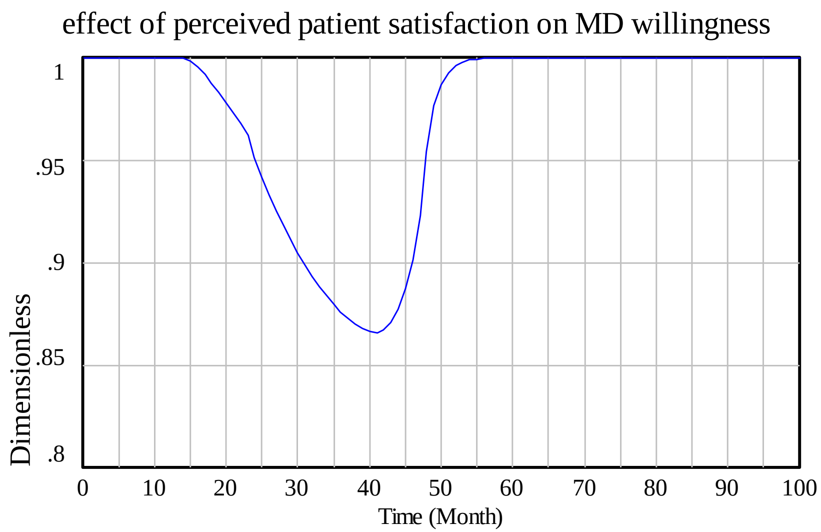


Figure 3.79 Task-shifting Run Showing Clinician Willingness (continued)

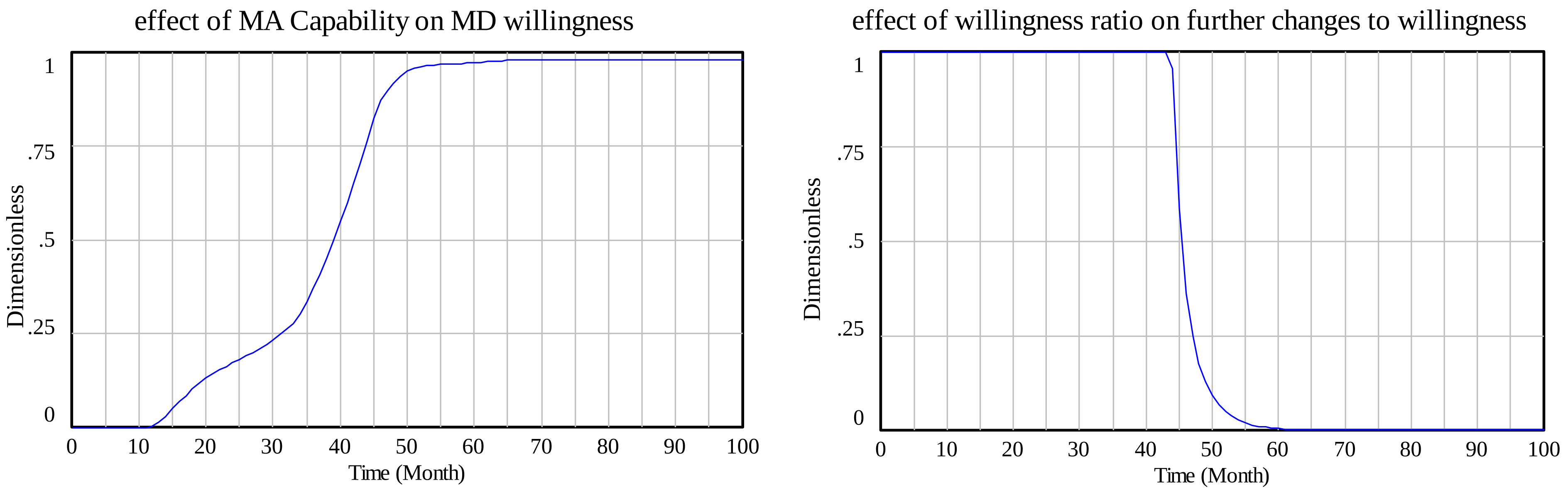
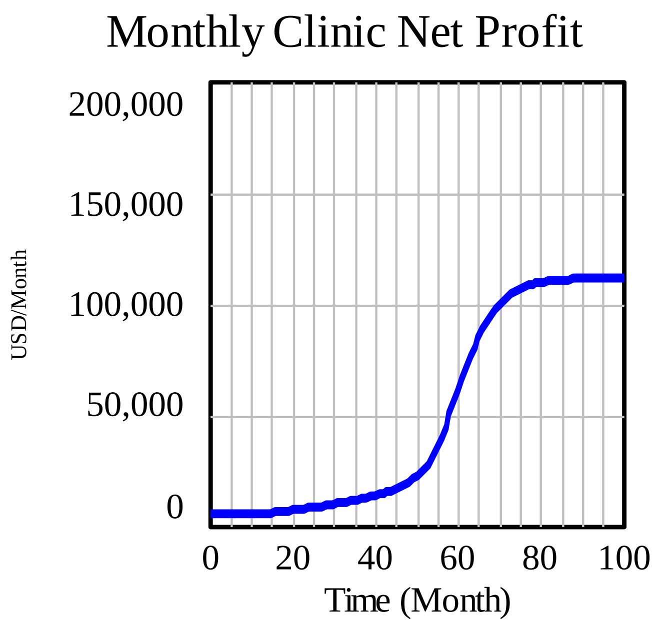
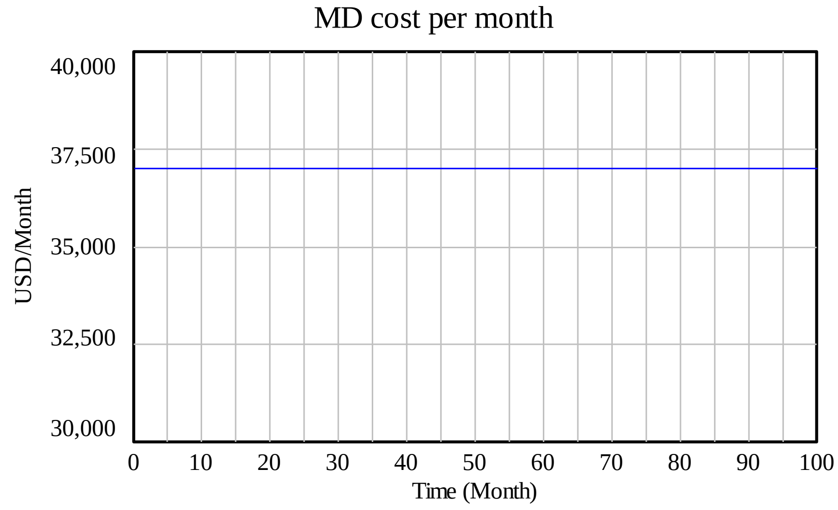
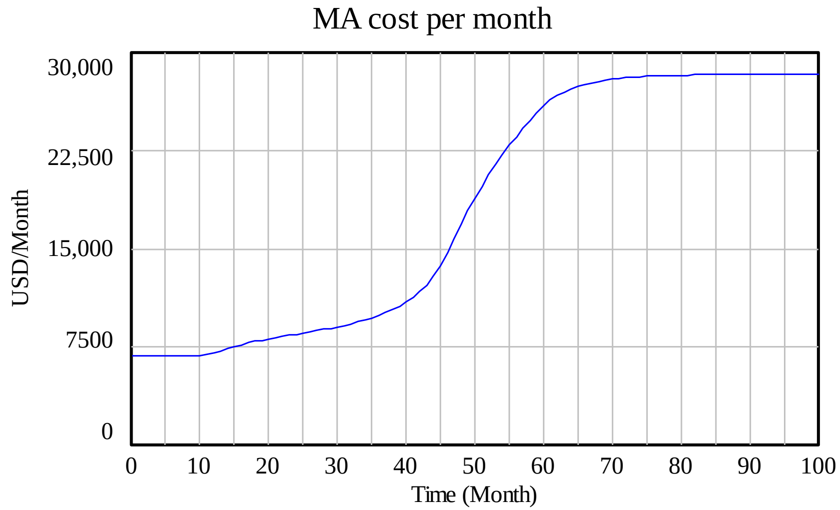


Figure 3.80 Task-shifting Run Showing Financial Variables



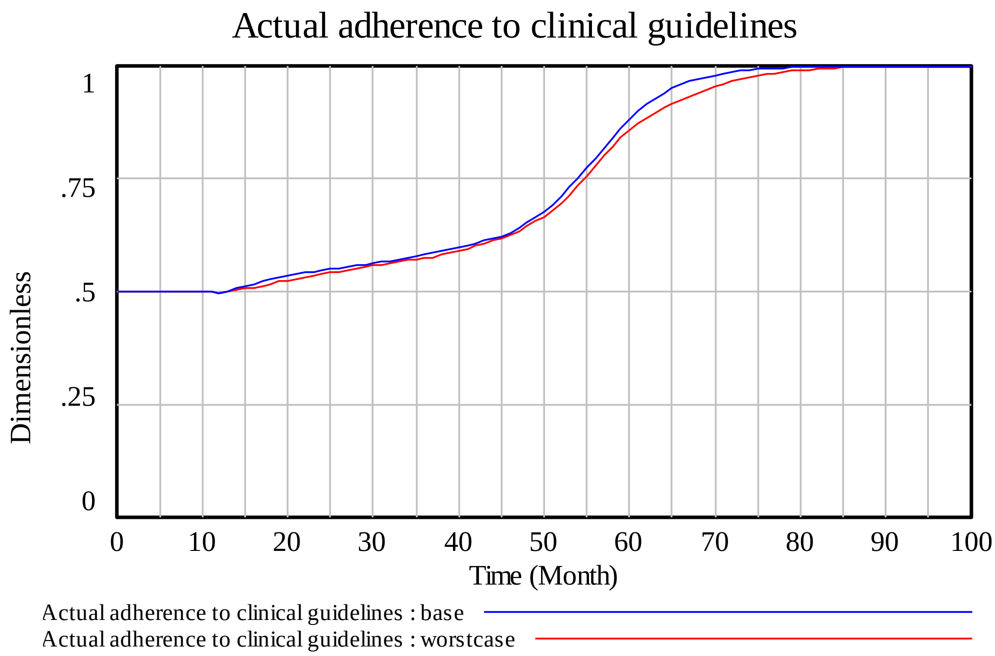




3.6.1.3 POLICIES IMPLEMENTED IN THE BASE-CASE

This section considers what happens when different policies – the yes/no decisions described above – interact with the task-shifting policy. Environmental conditions (i.e., preferences) are held constant. The worst-case policy scenario involves uncertainty in payment (pink noise on), a reticent hiring policy (only hire when the team's contribution to the clinic profit margin is above zero), along with clinicians being paid based on the number of encounters. In this scenario, there is a slight delay, but full adherence is still reached (Figure 3.81). Qualitatively, the behavior is the same as in the base-case. In such instances, I refer to the behavior as being the same.

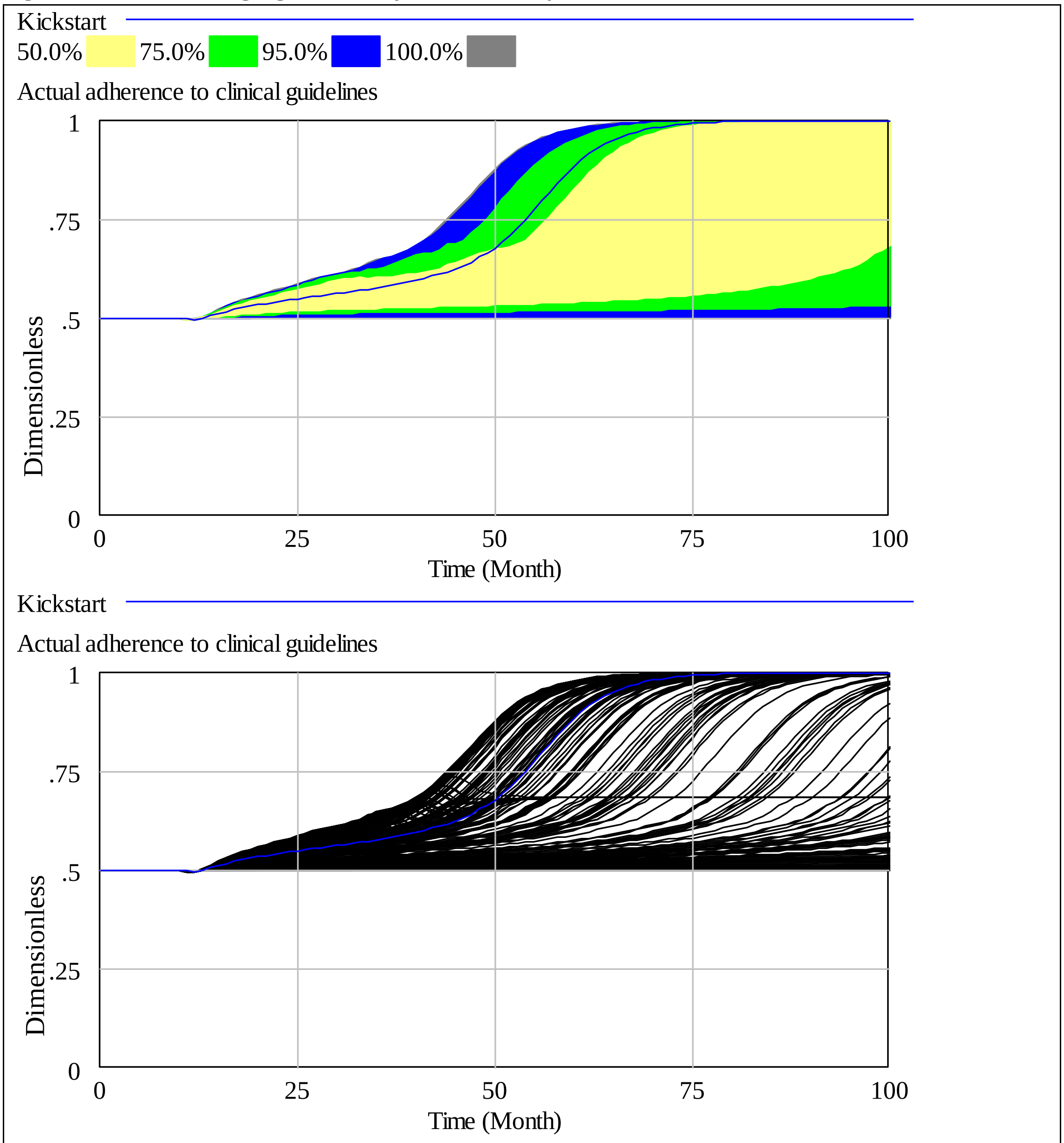
Figure 3.81 Base-case Runs with Normal & Worst-case Policies



Varying MD payment policies (encounter base salary, fixed salary, versus wRVU based salary[50]) did not change the level of actual adherence to clinical guidelines that was achieved. Switching on or off the hiring of MAs regardless of clinic net profit did not change simulation results. Uncertainty in revenue obtained (pink noise) also did not change simulation results. All these cases reproduce the original base-case results (ultimately achieving full adherence to clinical guidelines). Changing the policy regarding kickstart amount does matter (Figure 3.82 below).

[50] An alternative to paying clinicians for the number of patient visits (or encounters), essentially paying by volume, is paying by what happens in the visit. This is done using *work relative value units* (wRVUs). These are a task-based incentive where each patient care task (service) delivered is assigned a certain number of *relative value units* such that all tasks can be assigned to the same scale. This scale is then used to calculate how much to pay a clinician for the services provided, as follows: the number of patient care tasks (services) delivered in terms of wRVUs is multiplied by the compensation per wRVU.

Figure 3.82 Changing the Policy Kickstart Impacts the Base-case Run



The first graph uses colors to indicate the space occupied by 50% (yellow), 75% (green), 95% (blue), and up to 100% (grey) of runs. The second graph shows the same results, this time each line is a separate run.

In all model runs, the MAs are paid based on the tasks that they perform. The salaried option for paying MAs is not used as MAs performing non-technical tasks represent MAs with more advanced training (e.g., licensure), who would be paid more than MAs without this training.

3.6.1.4 ALTERNATIVE PREFERENCES IN THE BASE-CASE POLICY

The previous section created alternative scenarios based on changes to policy. This section deals with alternative scenarios based on assuming alternative preferences. Specifically, I conducted an analysis of the impact of alternative preferences on the outcome of interest "actual adherence to clinical guidelines" under the base-case policies (hire regardless of clinic net profit, no revenue uncertainty, and salaried clinicians). The alternative preferences are a policy analysis structure, they are represented using alternative versions of the table functions for operating policies (i.e., not those for productivity). Visuals of both the original and the alternative preferences are found in Appendix G.

In this analysis, I was interested in testing whether alternative preferences result in a different outcome for "actual adherence to clinical guidelines". I made this determination using a semi-qualitative approach. After each run, I visually inspected a graph which compared the base-case task-shifting scenario to the present one. I then used visual inspection to answer the question: does this alternative preference make a difference on the outcome of interest? When the qualitative behavior pattern was the ***same*** as that of the base-case successful trajectory, that variable was marked "**No**". This was done even in cases where there was a slight difference in the actual values across scenarios – usually because of a slight delay in reaching 100% on the outcome of interest. Delays were not considered as a change unless they exceeded *five months*. Thus, a qualitative behavior pattern is considered ***different*** than the successful trajectory when exhibiting any of the following behavior patterns: **worse than initial, suboptimal equilibrium, or same successful trajectory with a significant delay**. In these cases, I marked "**Yes**". In Table 3.14 and Table 3.15 below, I present results from this analysis.

Table 3.14 indicates that using alternative preferences for all but three of these variables does sometimes result in a different outcome for "actual adherence to clinical guidelines". The first column identifies the variable for which alternative preferences are applied (changes to the table function for that variable). The second column provides a brief description of the variable. The third column presents whether alternative preferences for the variable in question *impacted* the outcome of interest under the policies that operate in the base-case.

The "N/A" is specific to the variable "effect of MD monthly salary on MD's willingness" since the impact of alternative preferences was not observable in these scenarios because the "desired MD monthly salary" (a constant) was made to equal their actual salary in initial equilibrium under all three payment options. Once task-shifting begins, the salary only increases in the base-case.

Table 3.15 below lists the specific result of each alternative preference applied with the base-case policies. For some preferences I developed three alternatives and for others I developed more than that. Each "x" corresponds to a specific alternative preference (i.e., "xxx" means three alternatives).

Table 3.14 Impact of Preferences on Base-Case Policies

Variable Name	Variable Description (preference description)	Change?
effect of MA Backlog on inflow of nonTech tasks	The MDs observe how well the MAs are able to keep up with the tasks that they have been shifted. As the backlog of shifted tasks becomes significantly higher than the backlog of MA-only tasks, clinicians become concerned, pulling back on the number of shiftable (non-technical) tasks that clinicians attempt to have the team address.	Yes
effect of MA workload ratio on MD's desired MA staffing level	The MDs make hiring requests when the MAs are overworked. "Workload ratio per MA per MA-only task" measures the ratio between the current and normal workload for MA only tasks. When this variable is above 1, the MAs are overworked.	Yes
effect of MA Capability on MD willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is their perception of the MAs' ongoing level of capabilities. The more capable the MAs, the more willing the MD.	Yes
effect of perceived patient satisfaction on MD willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is the ratio between patients' ongoing and initial satisfaction, as perceived by the clinician. When at or above the initial patient satisfaction, the clinicians are more willing	Yes
effect of willingness ratio on further changes to willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is the ratio between their current level of willingness relative and the max level of willingness the clinicians have determined that they have.	Yes
effect of MD monthly salary on MD's willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is the ratio between their desired and actual monthly salary. When at or above their desired salary, the clinicians are more willing.	N/A
effect of proportion of Tech tasks on inflow of Tech tasks	The MD decides how many technical tasks he expects to get done during the visit based on his current workload ratio (proportion of Tech tasks out of the total tasks for the MD). The MD decides not to allow himself to book more technical tasks until he has more capacity to do them. When he sees that he has more capacity (via the decision input), he expects to get more tech tasks done and thus allows them in.	Yes
effect of time to complete backlog Tech tasks on shedding	In the initial equilibrium, the normal time to complete Tech tasks is 3 months. The decision input compares the normal time to complete these tasks to the actual time to complete these tasks. If the ratio is above 1 (actual time to complete is greater than the normal time to complete these tasks), there may be some shedding of these tasks. If the ratio is less than 1 then there is no shedding of these tasks.	No
effect of MD time to complete nT tasks ratio on shedding of those tasks	In the initial equilibrium, the normal time to complete nTech tasks is 3 months. The decision input compares the normal time to complete these tasks to the actual time to complete these tasks. If the ratio is above 1 (actual time to complete is greater than the normal time to complete these tasks), there may be some shedding of these tasks. If the ratio is less than 1 then there is no shedding of these tasks.	No
effect of MA Capabilities on MA satisfaction	An aspect influencing MA satisfaction is their level of capabilities. As they become more capable, they become more satisfied. They enjoy learning and using their capabilities to do higher level tasks.	Yes
MA willingness to stay in MA job	MAs choose to leave depending on their relative level of satisfaction (current perceived level to initial/expected level), but the exact tolerance varies across MAs.	Yes
effect of MA capacity on MA satisfaction	An aspect influencing MA satisfaction is their level of capacity. As they have more capacity, they become more satisfied. MAs expect to be able to do their job. As they shed more MA only tasks (i.e., their basic duties), they become less satisfied.	Yes
effect of MA capab ratio on change in MA capab	Learning generally begets more learning, but the exact shape varies across MAs.	Yes
effect of time to complete backlog of Tech tasks on Patient Satisfaction	Patients come to clinicians to get treated, to receive services that only clinicians can provide. Patients are satisfied as long as they receive services that they need in a timely manner. Therefore, the model uses the ratio of current time to complete to normal time to complete technical tasks. When that ratio is greater than 1, patients become less satisfied. When it is equal to or less than 1 patients are satisfied.	Yes
Keeping all else constant: a "**no**" here indicates that changes to the variable in question did *not* result in a different outcome, while a "**yes**" here indicates that they did.		

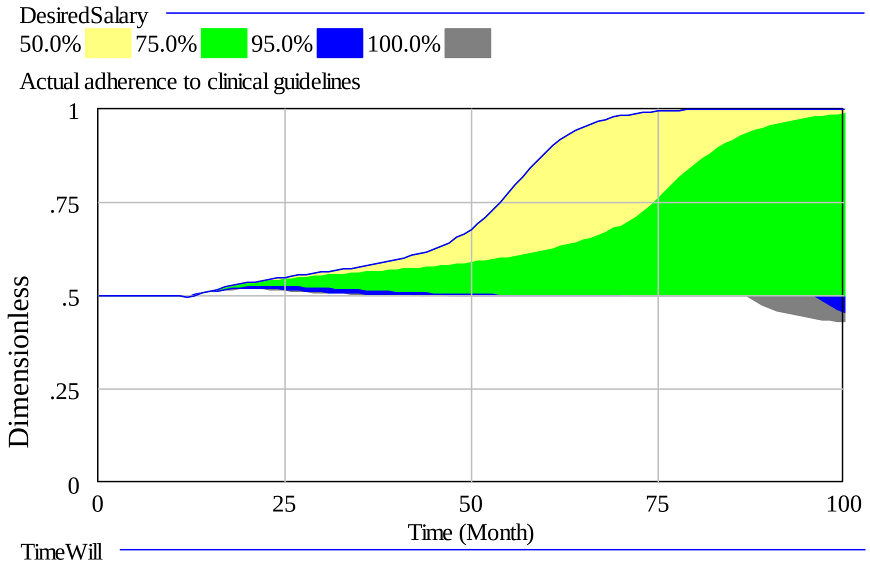
Table 3.15 The Association of Behavior Modes with Alternative Preferences

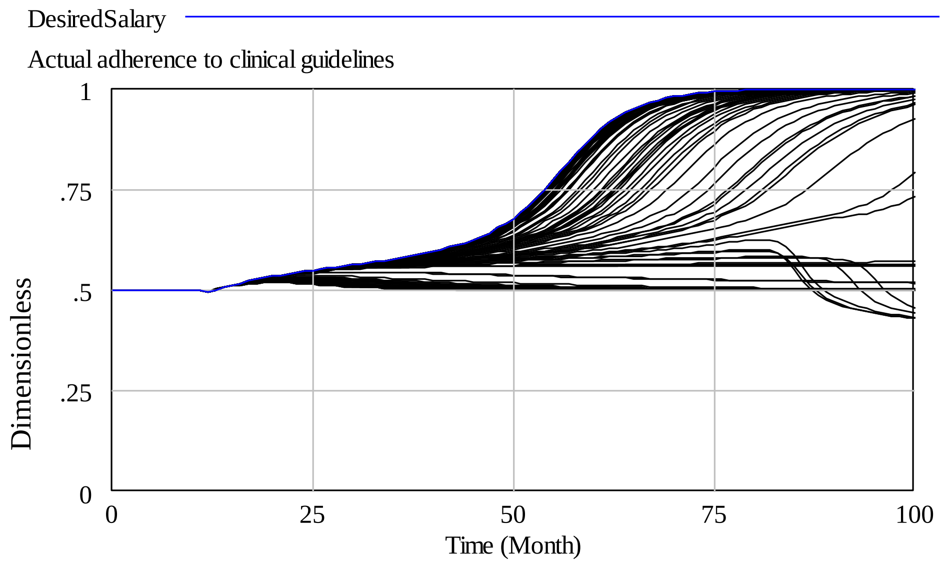
	Base-case versus				
Variable Name	**Worse than initial equilibrium**	**Back to initial equilibrium**	**Suboptimal equilibrium**	**Same Delayed**	**Same**
effect of MA Backlog on inflow of nonTech tasks	x		xx		
effect of MA workload ratio on MD's desired MA staffing level	x	x			x
effect of MA Capability on MD willingness	x		x	x	x
effect of perceived patient satisfaction on MD willingness		xx		xx	
effect of willingness ratio on further changes to willingness			x	x	xx
effect of proportion of Tech tasks on inflow of Tech tasks	xx				xx
effect of time to complete backlog Tech tasks on shedding					xxx
effect of MD time to complete nT tasks ratio on shedding of those tasks					xxx
effect of MA Capabilities on MA satisfaction			x		xxx
MA willingness to stay in MA job	x				x
effect of MA capacity on MA satisfaction			xx		xx
effect of MA capab ratio on change in MA capab			x	x	xx
effect of MD monthly salary on MD's willingness					xxx
effect of time to complete backlog of Tech tasks on Patient Satisfaction		x	xx		

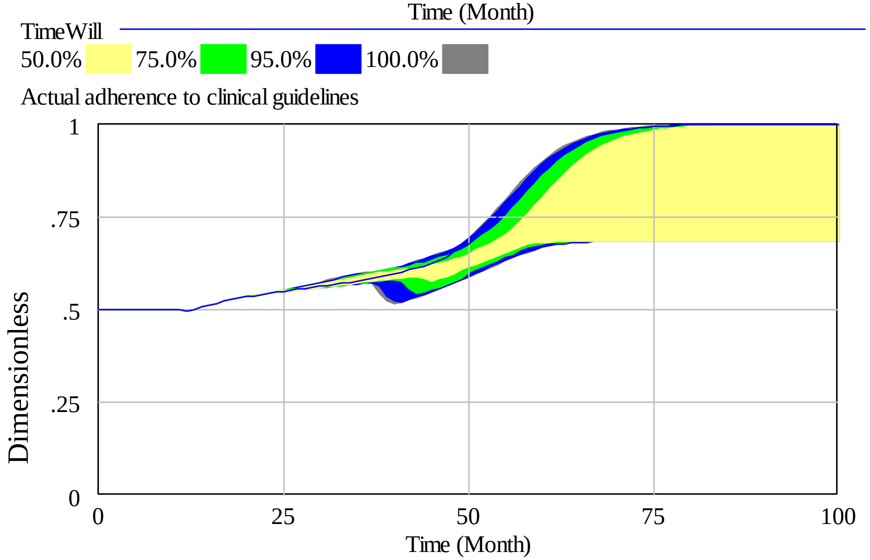
I also explored the impact of alternative preferences for variables represented by constants. Figure 3.83 below present sensitivity runs demonstrating the impacts of these alternative preferences on “actual adherence to clinical guidelines” under the base-case policies. The graphs on the left show the density of runs across the values space (i.e., the yellow region of color indicates that 50% of the runs reside inside that area of the graph). The graphs on the right show the trace of each run, more clearly visualizing the different behavior patterns observed.

In brief, all three variables have a range within which they result in the successful trajectory for the main outcome of interest, just with a smaller/greater delay; however, there is also another range for each variable where the successful trajectory is not attained. For “Desired MD monthly salary”, the latter range involves reaching an “actual adherence to clinical guidelines” that is lower than the initial. For “MD time to develop willingness” and for “training tasks needed to gain capability”, the latter range involves reaching a suboptimal outcome.

Figure 3.83 Impact of Preferences which are Constants on Base-case Policies







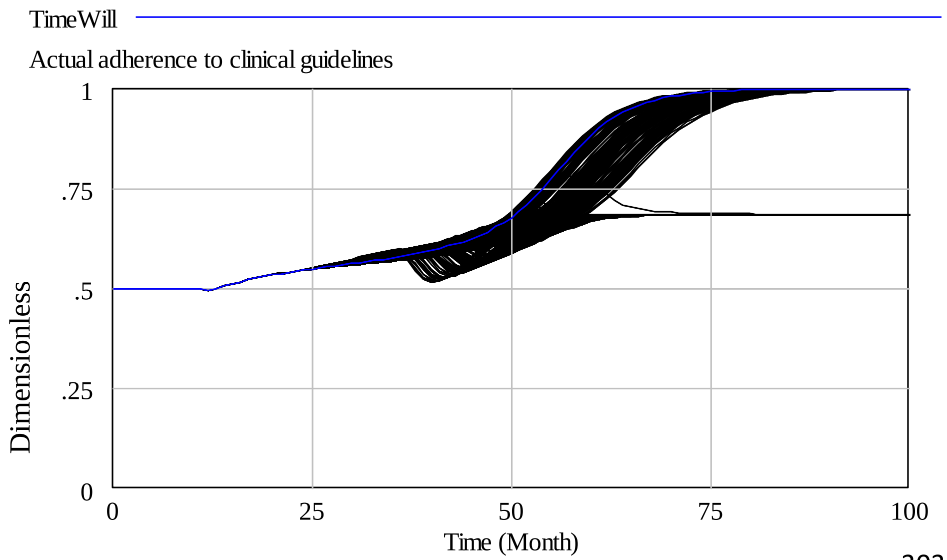
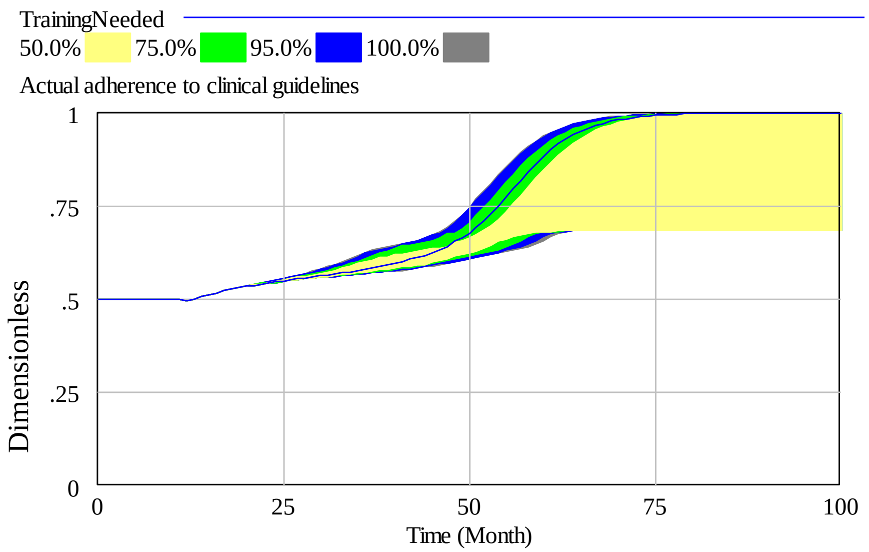
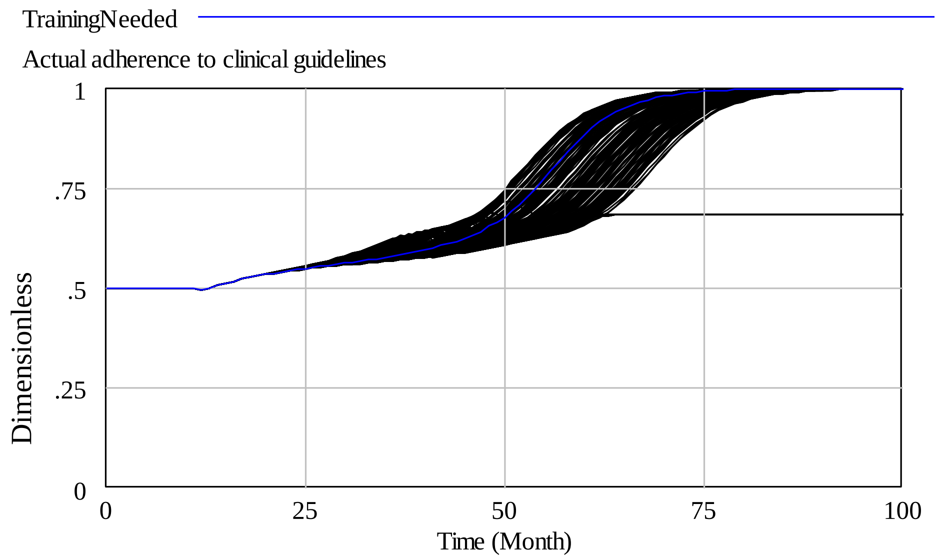


Figure 3.83 Impact of Preferences which are Constants on Base-case Policies (continued)





3.6.1.5 POLICIES IMPLEMENTED WITH ALTERNATIVE PREFERENCES

In this section, I present results from sensitivity analysis of alternative preferences on the outcome of interest “actual adherence to clinical guidelines” under different policy options. Results are divided into two sections based on the variable type: results for constants and results for table functions.

The same semi-qualitative method for determining the impact of assuming an alternative preference (see Section 3.6.1.4) was applied here.

3.6.1.5.1 CONSTANTS

I begin by presenting the impact of alternative preferences for variables represented by constants. Table 3.16 below presents a summary of the impacts of these alternative preferences on “actual adherence to clinical guidelines” (the main outcome of interest) under the different policy options. The first column identifies the scenario category – the person to whom the environmental conditions (i.e., preferences) apply. The second column identifies the variable for which alternative preferences are applied (changes to the value for that variable). The third column provides a brief description of the variable. The remaining three columns present whether changing the policy in question impacted the outcome of interest given different environmental conditions; other words, whether step-wise payment (payer reimbursement), hiring MAs based on need alone or salary payment policies would have a different impact given different values for the variable in the second column).

Keeping all else constant, a “no” here indicates that changes to the policy in question did *not* result in a different outcome for “actual adherence to clinical guidelines”, under any alternative preferences for the variable in that row. Keeping all else constant, a “yes” here indicates that changes to the policy in question *did* result in a different outcome, under certain alternative(s) for the variable in that row. This was done by visual inspection as described previously. Again, delays were not considered as a change unless they exceeded *five months*.

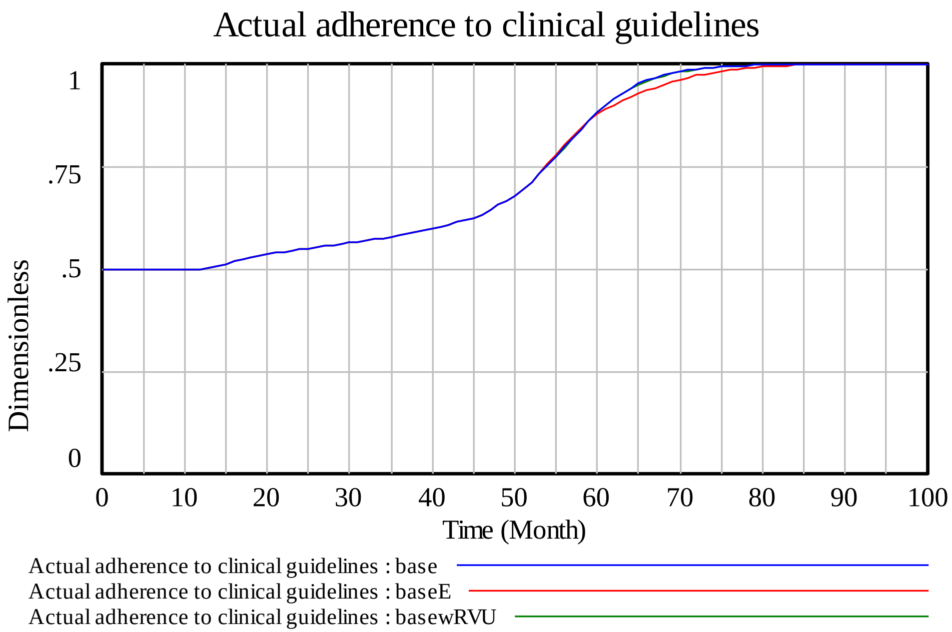
Table 3.16 Impact of Alternative Preferences – Constants

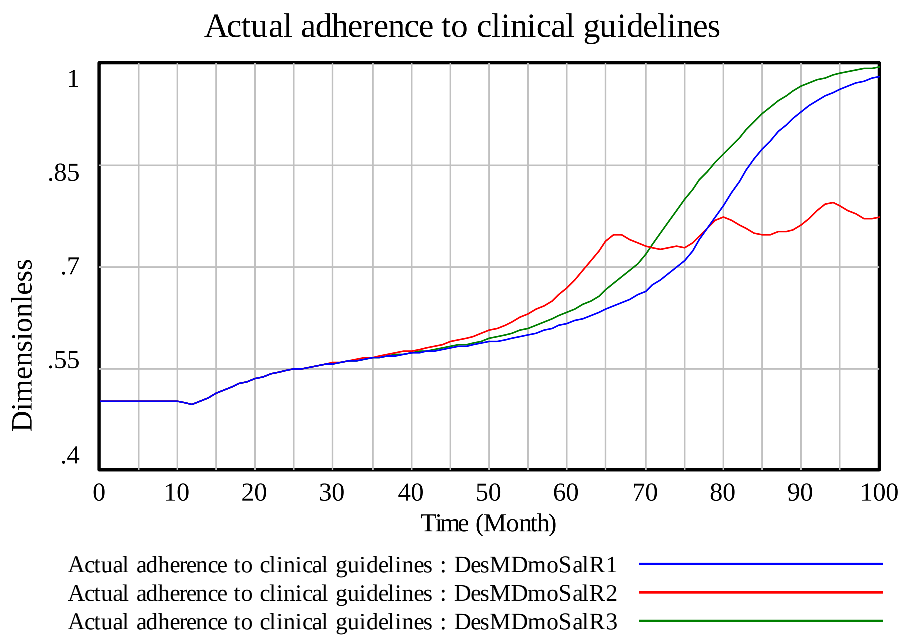
Scenario Category	Variable Name	Variable Description	Does each policy make a difference?		
			Step-wise payment	Hiring MAs Based on Need Alone	Clinic's Incentive Policy
Clinicians	Desired MD monthly salary	The monthly income clinicians see as acceptable. Receiving above that amount is fine. Receiving below that amount is concerning – how concerning depends on "effect of MD monthly salary on MD's willingness".	No	No	Yes
Clinicians	MD time to develop willingness	The adjustment time for changes in clinician willingness (i.e., the time it takes for the change to manifest).	No	No	No
Team	Training tasks needed to gain capability	The number of tasks that the MA must complete in order to gain 1 unit of capability.	No	No	No

In brief, the only set of policies to have an impact on the outcome of interest under alternative values for variables in Table 3.16 is "Clinic's Incentive Policy", when used in response to certain alternative values for "Desired MD monthly Salary". Changing policies did not change the outcome observed for all other environmental conditions evaluated for this table.

Figure 3.84 below presents a typical example of how payment policies impact alternative preferences for "Desired MD monthly salary". With "Desired MD monthly salary" at the base value of $12,330 (the graph on the left), all three salary-payment options result in the same outcome (a successful trajectory). With "Desired MD monthly salary" at $15,000 (the graph on the right), the three salary-payment options result in worse outcomes than the base-case. Each policy has different result (runs labeled DesMDMoSalR1, 2 and 3) in terms of the outcome of interest; where wRVU based incentives achieve a higher level sooner than payment of a fixed salary. Both result in a higher level of adherence than the suboptimal result attained with encounter-based incentives.

Figure 3.84 Impact of Clinic's Incentive Policy on Alternative Preferences for "Desired MD monthly salary"





3.6.1.5.2 TABLE FUNCTIONS

I focus on alternative preferences for variables represented by table functions. Table 3.17 below presents a summary of the impacts of these alternative preferences on “actual adherence to clinical guidelines” (the main outcome of interest) under the different policy options. The first column identifies the scenario category – the person to whom the environmental conditions apply. The second column identifies the variable for which alternative preferences are applied (changes to the table function for that variable). Third column provides a brief description of the variable. The remaining four columns present whether the policy in question impacted the outcome of interest given different environmental conditions.

Keeping all else constant, a “no” here indicates that changes to the policy in question did *not* result in a different outcome. Keeping all else constant, a “yes” here indicates that changes to the policy in question *did* result in a different outcome. Keeping all else constant, a “no*” here indicates that changes to the policy in question did *not* result in a different outcome *except for when revenue uncertainty also existed*. This was done by visual inspection as described previously. Again, delays were not considered as a change unless they exceeded *five months*.

As in Table 3.14 above, the “N/A” is specific to the variable “effect of MD monthly salary on MD's willingness”. The only way one can see the effect of alternative preferences for this variable is by running a scenario where salary decreases for a period. Then, alternative environmental condition expressed by this table function will be seen to differentially impact the outcome of interest.

Table 3.17 Impact of Alternative Preferences – Table Functions

Scenario Category	Variable Name	Variable Description	Does each policy make a difference?			
			Step-wise payment	Environmental Conditions	Hiring MAs Based on Need Alone	Clinic's Incentive Policy
Clinicians	effect of MA Backlog on inflow of nonTech tasks	The MDs observe how well the MAs are able to keep up with the tasks that they have been shifted. As the backlog of shifted tasks becomes significantly higher than the backlog of MA-only tasks, clinicians become concerned, pulling back on the number of shiftable (non-technical) tasks that clinicians attempt to have the team address.	No	Yes	No	No
	effect of MA workload ratio on MD's desired MA staffing level	The MDs make hiring requests when the MAs are overworked. "Workload ratio per MA per MA-only task" measures the ratio between the current and normal workload for MA only tasks. When this variable is above 1, the MAs are overworked.	No	Yes	No	No
	effect of MA Capability on MD willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is their perception of the MAs' ongoing level of capabilities. The more capable the MAs, the more willing the MD.	No	Yes	No	No
	effect of perceived patient satisfaction on MD willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is the ratio between patients' ongoing and initial satisfaction, as perceived by the clinician. When at or above the initial patient satisfaction, the clinicians are more willing	No	Yes	No	No
	effect of willingness ratio on further changes to willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is the ratio between their current level of willingness and the max level of willingness the clinicians have determined that they have.	No	Yes	No	No
	effect of MD monthly salary on MD's willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is the ratio between their desired and actual monthly salary. When at or above their desired salary, the clinicians are more willing.	No	N/A	No	Yes
	effect of proportion of Tech tasks on inflow of Tech tasks	The MD decides how many technical tasks he expects to get done during the visit based on his current workload ratio (proportion of Tech tasks out of the total tasks for the MD). The MD decides not to allow himself to book more technical tasks until he has more capacity to do them. When he sees that he has more capacity (via the decision input), he expects to get more tech tasks done and thus allows them in.	No	Yes	No	Yes

Scenario Category	Variable Name	Variable Description	Does each policy make a difference?			
			Step-wise payment	Environmental Conditions	Hiring MAs Based on Need Alone	Clinic's Incentive Policy
Clinicians (continued)	effect of time to complete backlog Tech tasks on shedding	In the initial equilibrium, the normal time to complete Tech tasks is 3 months. The decision input compares the normal time to complete these tasks to the actual time to complete these tasks. If the ratio is above 1 (actual time to complete is greater than the normal time to complete these tasks), there may be some shedding of these tasks. If the ratio is less than 1 then there is no shedding of these tasks.	Yes	No*	No*	No
Clinicians (continued)	effect of MD time to complete nT tasks ratio on shedding of those tasks	In the initial equilibrium, the normal time to complete nTech tasks is 3 months. The decision input compares the normal time to complete these tasks to the actual time to complete these tasks. If the ratio is above 1 (actual time to complete is greater than the normal time to complete these tasks), there may be some shedding of these tasks. If the ratio is less than 1 then there is no shedding of these tasks.	Yes	No*	No*	No
Clinical Staff Members	effect of MA Capabilities on MA satisfaction	An aspect influencing MA satisfaction is their level of capabilities. As they become more capable, they become more satisfied. They enjoy learning and using their capabilities to do higher level tasks.	No	Yes	No	No
Clinical Staff Members	MA willingness to stay in MA job	MAs choose to leave depending on their relative level of satisfaction (current perceived level to initial/expected level), but the exact tolerance varies across MAs.	No	Yes	No	No
Clinical Staff Members	effect of MA capacity on MA satisfaction	An aspect influencing MA satisfaction is their level of capacity. As they have more capacity, they become more satisfied. MAs expect to be able to do their job. As they shed more MA only tasks (i.e., their basic duties), they become less satisfied.	Yes	Yes	No*	No
Clinical Staff Members	effect of MA capab ratio on change in MA capab	Learning generally begets more learning, but the exact shape varies across MAs.	Yes	Yes	No*	No*
Patients	effect of time to complete backlog of Tech tasks on Patient Satisfaction	Patients come to the doctor to get treated, to receive services that only the doctor can provide. Patients are satisfied as long as they are able to receive the services that they need in a timely manner. Therefore, the model uses the ratio of current time to complete to normal time to complete technical tasks. When that ratio is greater than 1, patients become less satisfied. When it is equal to or less than 1 patients are satisfied.	Yes	Yes	No*	Yes

Keeping all else constant, a "**no**" here indicates that changes to the policy in question did not result in a different outcome. Keeping all else constant, a "**yes**" here indicates that changes to the policy in question did result in a different outcome. Keeping all else constant, a "**no***" here indicates that changes to the policy in question did not result in a different outcome except for when revenue uncertainty also existed.

The following paragraphs present a summary of these results along with further detail on the type of impact had.

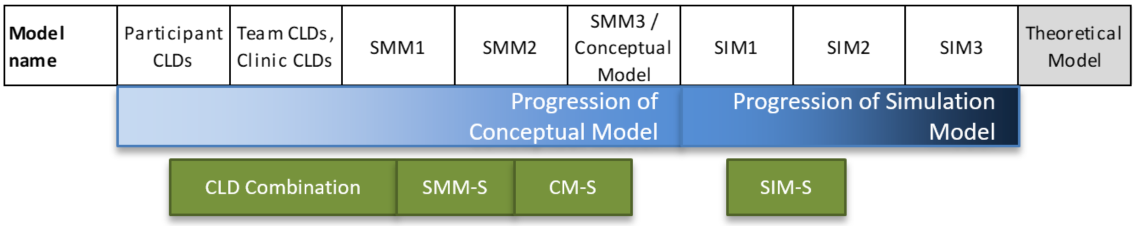
Alternative preferences were observed to impact the outcome of interest for all variables in Table 3.17; however, for two variables this impact was only observed with the existence of revenue uncertainty (as indicated by "no*").

Revenue uncertainty was seen to differentially impact the outcome of interest for five of the 14 variables in this Table 3.17 (under a subset of alternative preferences for these variables). In each case, the impact was to produce a worse outcome. For "effect of MA capacity on MA satisfaction", in environmental conditions that produced an effect, the effect was to bring the main outcome of interest from suboptimal equilibrium to a worse than the initial equilibrium value. For the other four variables, in environmental conditions that produced an effect, the effect was to bring the main outcome of interest from the successful optimal equilibrium outcome to reaching only a suboptimal equilibrium.

Hiring MAs based on need alone (i.e., "hire regardless of clinic net profit") was seen to influence the outcome for five variables (and only in the presence of revenue uncertainty), under a subset of alternative preferences for these variables. The impact was always positive such that hiring MAs based on need alone (rather than hiring them only when the team's contribution to the clinic net profit was above zero) improved the outcome. For four of the five variables, the improvement was from achieving only suboptimal equilibrium to achieving the desired successful outcome. For "effect of MA capacity on MA satisfaction", the improvement was from worse than initial to achieving a suboptimal equilibrium.

Clinic's Incentive Policy (i.e., fixed salary, encounter-based, or wRVU-based salary for clinicians) were seen to influence the outcome of four variables (one of which was only observed with the existence of revenue uncertainty). For "effect of proportion of Tech tasks on inflow of Tech tasks" and "effect of MD monthly salary on MD's willingness", switching from fixed salary to encounter-based salary resulted in a suboptimal result (either reaching a suboptimal equilibrium or trending upwards but still lower than the 100% within the simulation's time horizon). For these same variables, switching from encounter to wRVU-based salary improved the outcome, moving from a suboptimal result to the successful trajectory. This was also the case for "effect of MA capab ratio on change in MA capab", under scenarios involving revenue uncertainty (and specific environmental conditions). However, for "effect of time to complete backlog of Tech tasks on Patient Satisfaction", fixed salary and wRVU-based salary both resulted in suboptimal equilibria under a subset of environmental conditions whereas (under those same environmental conditions) encounter-based salary resulted in the successful trajectory for the outcome of interest.

3.6.2THEORETICAL FINDINGS



The main learnings so far have come from the process of modeling and validation. My theoretical findings emerge from the process of developing the Theoretical Model. This model combines the Policy Structure Diagrams which were developed to summarize the workings of the Simulation Model (Section 3.4.2) into a one-page hybrid System Policy-Structure Diagram which shows how all of the policy structures, or decision functions, in the model interact.

This process requires simplification. However, this is not done by removing components but by aggregating compatible components together to show the most important causal relationships. This process permits reflection about what the model is really saying about the causal structure of the system. Because of the "inherent causal nature of theory"[217], such reflections can be considered theoretical findings. The Theoretical Model conveys my theory of PCT based upon my dissertation work.

In this section, I begin by presenting the Theoretical Model (Section 3.6.2.1). Then, I present the insight I have gained from this process into the core structure linking the primary care tenets (Section 3.6.2.2). Next, I present my theory of the decision processes surrounding this basic structure. These processes shed light on the tensions (Section 3.6.2.3). Next, I present my theory of the changes to this core system structure in PCT (Section 3.6.2.4). Finally, I present my theory of the tensions experienced during PCT (Section 3.6.2.5).

3.6.2.1 THE THEORETICAL MODEL

The Theoretical Model is presented in Figure 3.85 and Figure 3.86 below. This is my theory of primary care transformation – it is my most comprehensive *dynamic hypothesis* and the result of my grounded dynamic analysis. *It is a visualization of the Simulation Model, bringing together the policy structure diagrams presented in the model walk-through* (Section 3.4.2).

Figure 3.85 presents the key. Some aspects will be familiar (e.g., the use of variables and links with polarity as well as feedback loops). New aspects include pointing out where people's preferences and payer as well as management policies interact with PCT structure. Also, the primary outcome of interest is highlighted in red. Links have also been colored to correspond with the colors used to identify the same loop in the Conceptual Model. When a link is part of more than one feedback loop, one color was chosen such that all loops would still be somewhat visible.

I will now provide an example walk-through of the Theoretical Model. The health system's expectation is for primary care to provide fully comprehensive care to patients. Primary care has yet to deliver on this promised value. The actual comprehensiveness is shown with a large red circle. What can be done? The factors which this theory holds constant are shown in plain text on the left: the number of patients, of clinically-recommended tasks per patient and the number of MDs. None of these things can be changed to reach the target of *Comprehensiveness Delivered (Actual Adherence)*. To reach the target goal, the *Services Delivered* needs to increase.

Clinicians keep their expectations closely-tied to reality: *Desired Comprehensiveness* is based on the amount of work the care team currently has before them. This is a function of the three *Backlog of Tasks* stocks – the three white boxes in the middle part of the diagram. When the backlogs grow, this tells clinicians that the team has more work to do than they can complete on time. So, clinicians conclude that less should be expected of the team in the future, reducing their goal. This causes clinicians to take on less work for their team in the future. When the backlogs are within acceptable levels, this tells clinicians that the team is keeping up and that the current workload is appropriate.

Apart from this, clinicians make several other decisions which impact PCT (blue circles with white labels). These are described here from top to bottom. Clinicians advocate for hiring more MAs, based on how well their MAs are doing at keeping up with their main responsibilities in *MD Workforce Planning*. They modulate how willing they are to shift tasks to MAs based on their perception of how capable these team members are in *Perceived MA Capability Ratio*. They hold tight to their inherited identity as the sole providers of primary care services in *MD Caution*. They make difficult choices about how much to cut corners in *Task Shedding (2)*. They fight to keep their patients happy in *Perceived Reputation*. Because they are not paid for training, task shifting can lower clinicians' salary. So, they hold back on how sharply to shift tasks to MAs in *MD Salary Ratio.*

In my theory, clinic managers influence PCT through three decisions described from top to bottom of the diagram (maroon bubbles with white text). Managers decide what information to take into account in the salary they offer to clinicians in *Clinic Incentive Policy*. They decide whether or not to hire MAs in *Clinic Hiring Policy*. They choose how much task-shifting to require and how long in *Policy kickstart*.

Clinical staff members (MAs) influence PCT through five decisions (teal bubbles with white text) on the right of the diagram. They decide how satisfied they are with their job in *Turnover of MAs*. MAs modulate how sensitive they will be to capacity strain in *MA Capacity Tolerance*. They modulate how sensitive they will be to having the opportunity to learn new things in *MA Capability Appreciation*. They learn as much as they can in *Learning Curve*. They control how engaged they are in gaining new capabilities in *Speed of Uptake*.

Finally, patients and payers also influence PCT. Patients consider how able their clinician is to provide them with needed services in a timely manner in *Patient Satisfaction*. Payers design policies for reimbursement for services in *Payer Policy*.

These choices are all made in relative isolation. Patients, payers, MAs, clinic managers and clinicians all decide on their own what to do under changing circumstances. These choices are made based on information that is readily available and that they can be confident in because it reflects current conditions of things they can easily observe.

The color of arrows and orientation of feedback loops on the Theoretical Model diagram retain some of the visual elements of the Conceptual Model. Blue and gray arrows on the right of the Theoretical Model diagram label causal relationships in the blue and gray-labeled loops on the right of the Conceptual Model (see Figure 3.4): Team Learning, Sharing the Load and Trust. Dark and light green arrows in the middle and upper left label causal relationships on the upper left of the Conceptual Model: Sufficient/Insufficient Human/Financial Resources. Dark and light orange arrows near the bottom of the diagram label the causal relationships on the lower left of the Conceptual Model: Sufficient/Insufficient Time. Pink arrows on the outside left and bottom label an expanded loop *Incentive for Task-Shifting* where the goal/gap of clinician salary now directly influences the *MD Willingness to Shift Tasks* – the box in the lower right.

Figure 3.85 Key for Theoretical Model

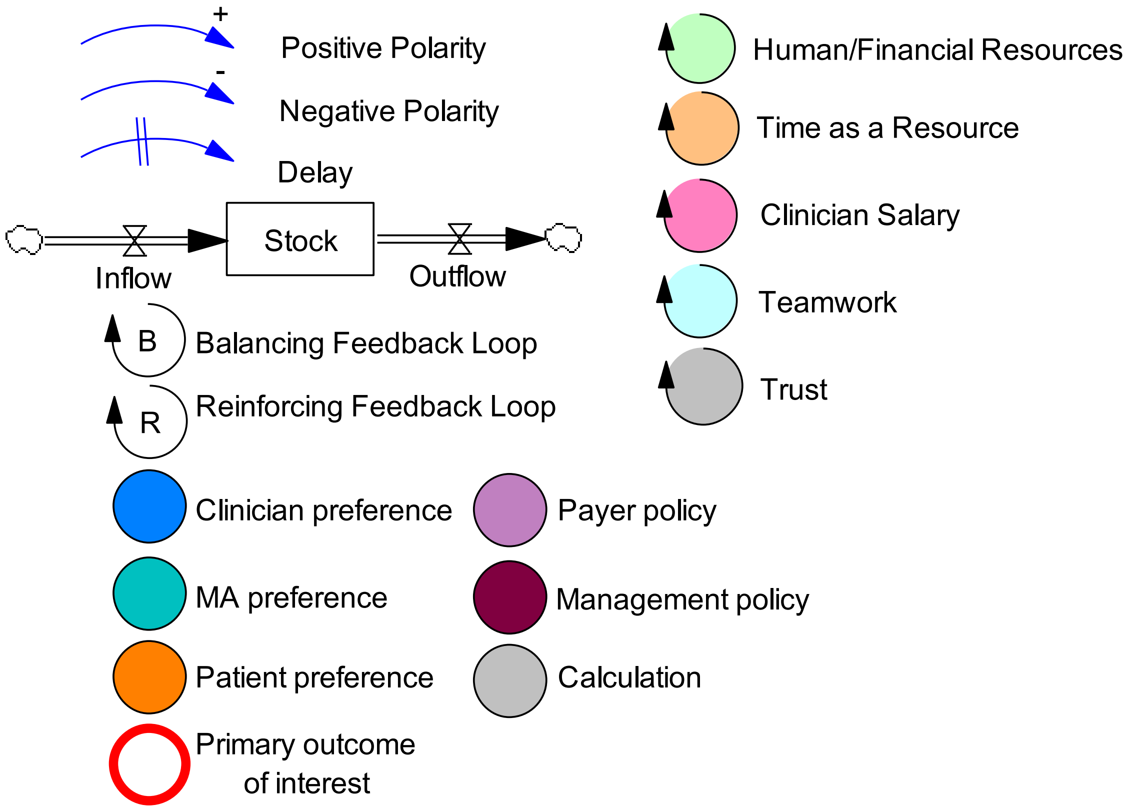


Figure 3.86 PCT Theoretical Model

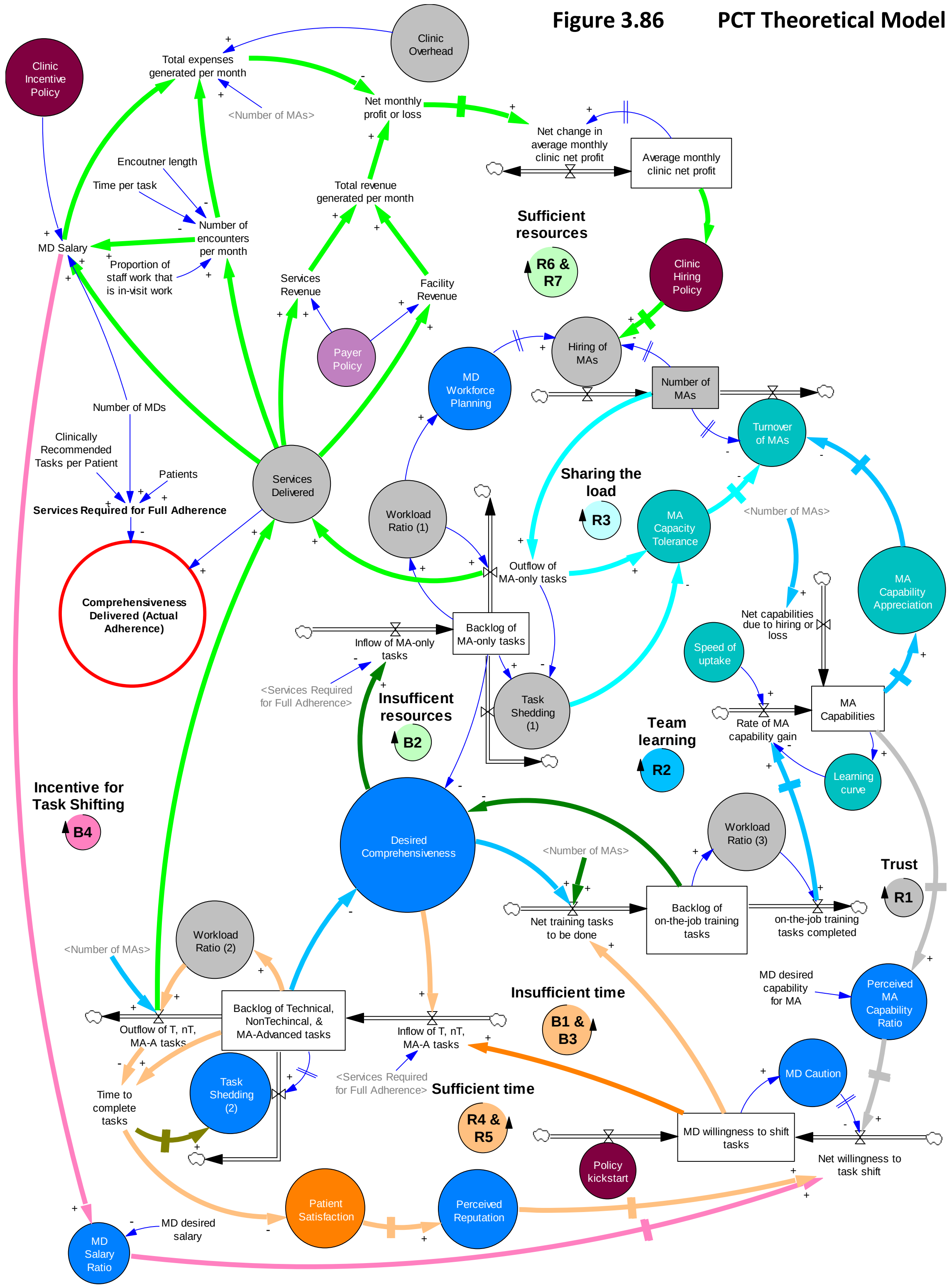


Figure 3.86 PCT Theoretical Model (continued - zoom for top half)

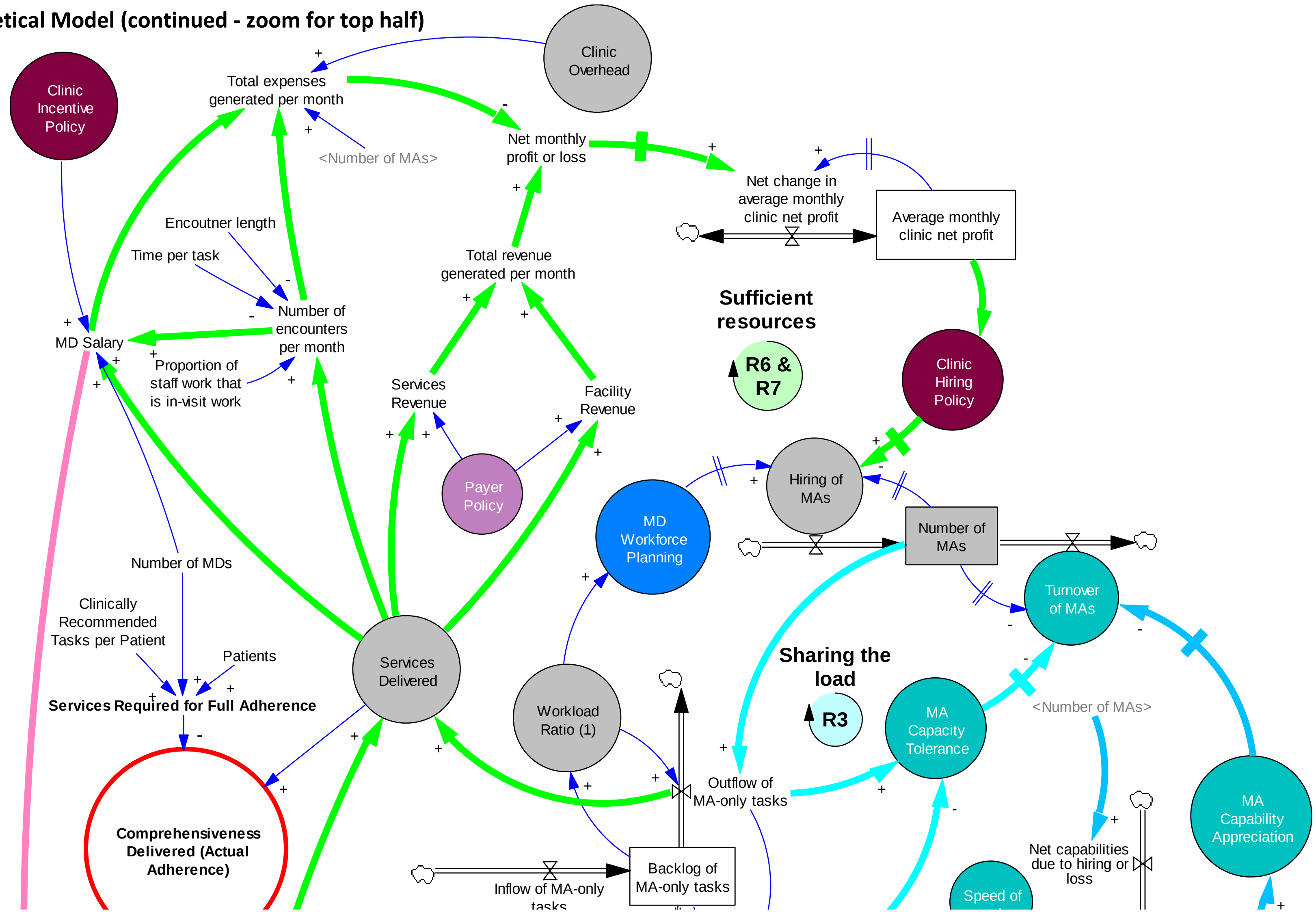
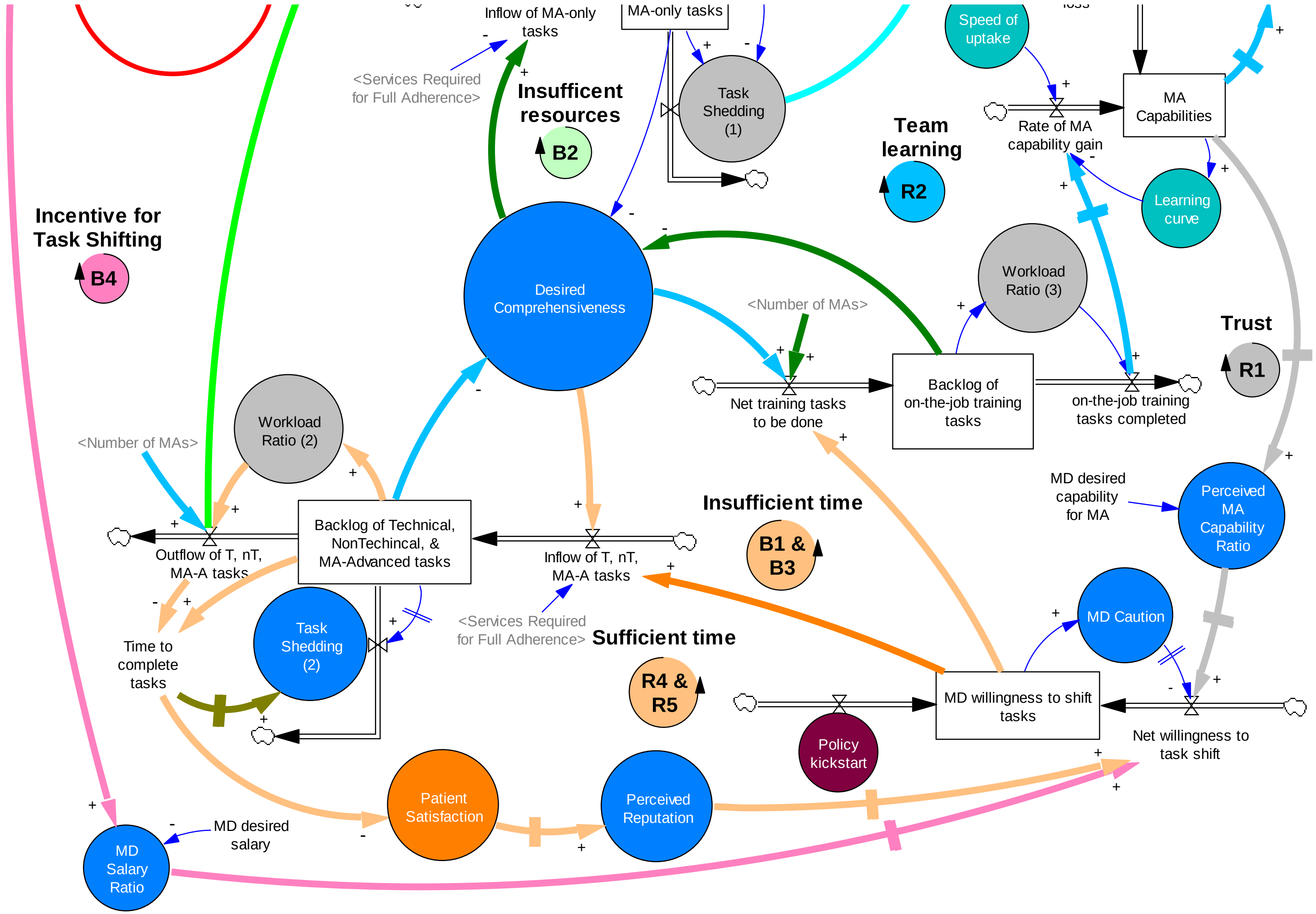


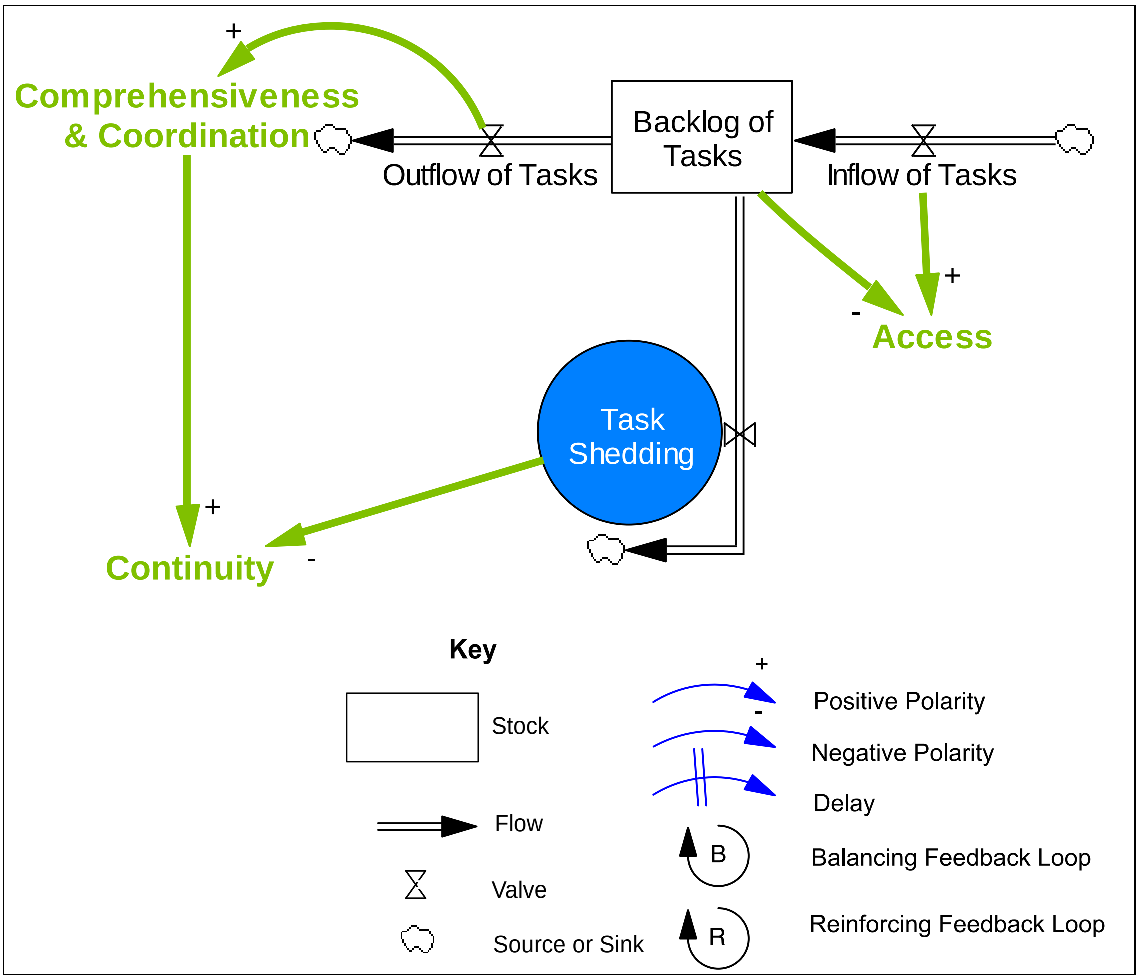
Figure 3.86 PCT Theoretical Model (continued - zoom for bottom half)



Aggregating the patient care tasks stocks from the Theoretical Model into one *Backlog of Tasks* permits reflection on how delivering patient care tasks occurs. This reflection represents my findings for the theory of primary care.

Specifically, I found that the tenets of primary care interact by means of four basic coupled elements: one stock with one inflow and two outflows, as shown in Figure 3.87. The *Backlog of Tasks* stock and *inflow of tasks* together define **access**. The *outflow of tasks* shows completed patient care tasks, which defines the **comprehensiveness** and **coordination** provided. The combination of Comprehensiveness and coordination together with *task shedding* define **continuity**.

Figure 3.87 The Core of Primary Care – Service Operations and the Tenets



Patient care services can be thought of as one stock of patient care tasks which demands the attention of care team members. The stock is like a bathtub with a spout that brings water in (inflow of tasks), one drain (outflow of tasks) and one bucket (task shedding). The drain is the right way for water to leave the tub and the bucket is for use in emergencies.

Care teams seek to maintain the stock[51][218, 219] *Backlog of tasks* at a stable level of **access** to care for their patients. Going above means patients' access to care is constrained and teams are stressed; and going below means team members are idle. For this reason, the clinician watches to bring in only as much work (*Inflow of Tasks*) as they can complete (*Outflow of Tasks*) and lapses in care (*Task Shedding*, i.e., tasks never completed) are kept at a minimum. Green links and variables show how this structure relates to the four tenets of primary care.

Access, therefore, is a function of how much work a team brings in and the amount of backlog in the system – the higher the backlog of services that the care team is expected to complete, the longer patients have to wait to receive those services. When those services are completed (*Outflow of Tasks*), they "count" toward the care team's adherence to clinical guidelines. Some of these services are delivered in an encounter and others are delivered outside an encounter (e.g., coordinating care with other service-providers). Thus, completed tasks count toward **comprehensiveness *and* coordination**. As they are completed by the patient's care team, patients experience **continuity.** When those services are shed (i.e., the team decides they do not have the time to deliver them after all), the service disappears from the backlog but it does not accrue to the team's adherence to clinical guidelines ("task shedding" outflow). Patients experience this as a lapse in care which harms their relationship with the care team – decreasing the crucial element of **continuity**. This can be experienced as a too-short visit or in other ways.

Work is generated as clinicians identify patient care tasks to be completed by the care team. This is the second element contributing to patient **access**.

US clinicians are estimated to average around 50% adherence to clinical guidelines **(comprehensiveness and coordination)** for their patient panel[76]. These clinicians operate at the level of comprehensiveness possible in the current system. It is estimated that a 21-hour-work-day[220] would be needed for a clinician to accomplish 100% adherence to clinical guidelines for the average panel of 2000 patients – an impossible feat! This comprehensiveness gap causes tension within the primary care system, the subject of the next section.

---

[51] This agrees with the definition of primary care provided by Stange, that it is a "buffer" (p.S34). See Oliva for an explanation of how a stock acts as a buffer.

3.6.2.3 PRIMARY CARE TENSIONS

In the Theoretical Model, I do not label the small feedback loops surrounding the *Backlog of Tasks* stocks. This section describes these loops, which are important for any primary care work, whether or not the clinic is actively engaged in transformation. These feedbacks act to resist changes to the level of the *Backlog of Tasks* stock, causing tensions when system changes (like PCT) are implemented. During periods of stability and modest change, these tensions serve to keep the system in balance – and may go unnoticed.

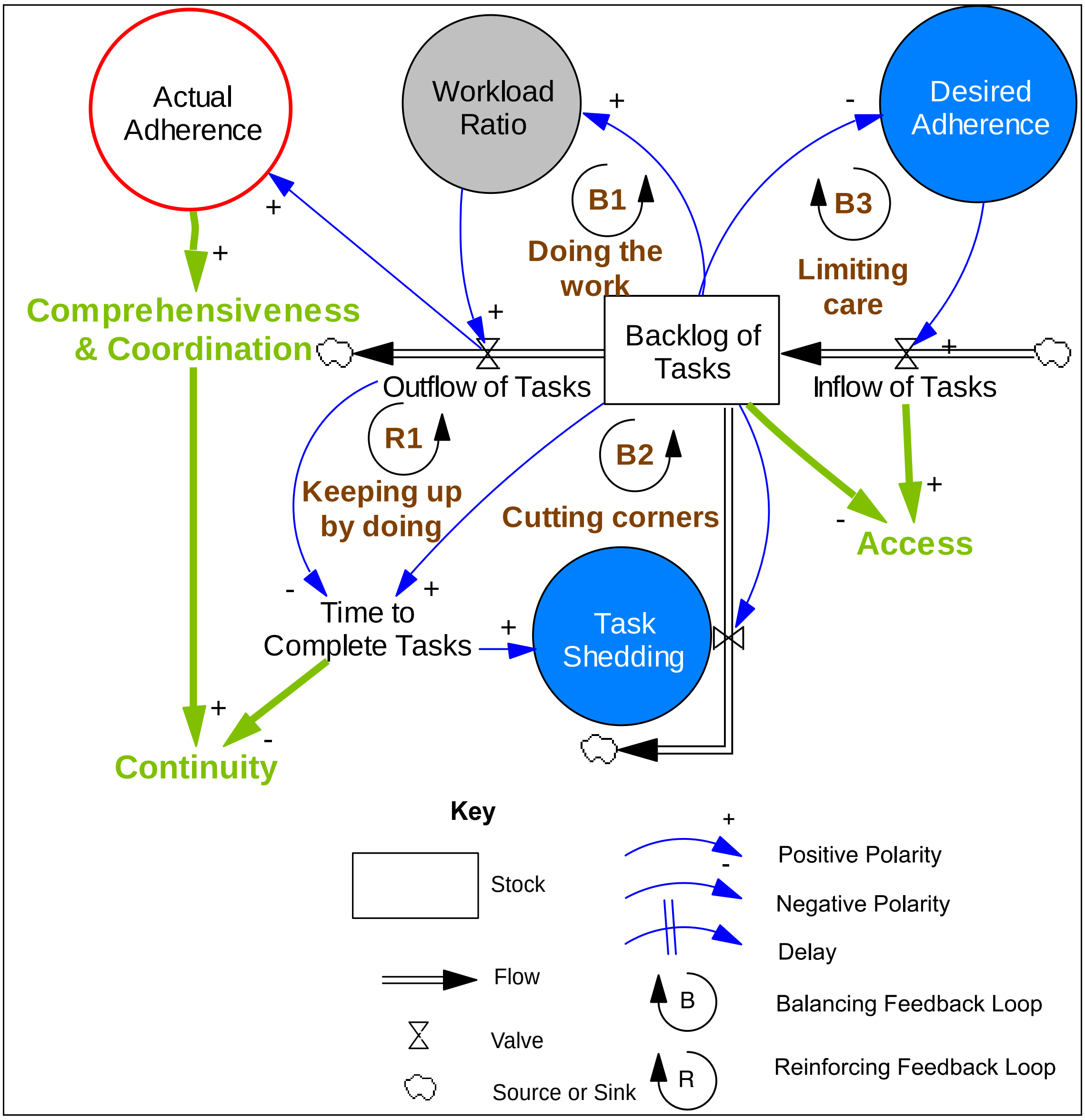
Figure 3.88 below adds four feedback loops to the core structure above. Blue arrows from the stock are information flows that inform clinician choices regarding the use of the bucket (task shedding) and the spout (inflow of tasks). As this information informs choices, the clinician closes the feedback loops that allow them to monitor the stock and act accordingly to keep the system under control. Green links and variables show how this structure relates to the four tenets of primary care. Blue circles represent clinicians' conscious decisions, gray circles represent subconscious decisions, and the red-outlined circle represents the system goal.

In **Loop B1: *Doing the work***, team members compare the current backlog of work to do and adjust their productivity to keep up as best they can. At very high levels of backlog, this loop can turn positive as a *haste makes waste* dynamic takes over and puts the team further behind, but in the ideal, it is a balancing loop that limits the growth of the backlog[52][113] (p. 577-578,582)[198]. In Loop **R1: *Keeping up by doing*,** the team keeps the backlog under control by completing tasks. When this fails, clinicians can exercise the option to increase **Loop B2: *Cutting corners*** which reduces the work pressure on the team allowing them to keep up with the work (e.g., neglecting promised follow-up). **Loop B3: *Limiting care*** is one aspect of the many factors clinicians take into consideration in assessing their ability to deliver comprehensive care. In this loop, clinicians work to prevent getting behind by limiting the care they offer to patients (e.g., doing less in the visit or offering fewer follow-ups).

[52] The link from *workload ratio* to *outflow of tasks* is shown here as positive relationship. Nevertheless, my model relies on the *Yerkes-Dodson Law*, as interpreted Andersen & Richardson. This allows for this link to be positive or negative, depending on the workload ratio (the current backlog of tasks divided by the standard backlog of tasks).

With the bathtub, the operating goal is to keep the amount of water in the tub at a constant level (i.e., watching the stock). Clinicians exercise full control of the spout (inflow of tasks) and the bucket (task shedding), but they have only a limited say on the capacity of the drain (outflow of tasks – the productivity and staffing of their care team). If water starts to accumulate above that level, clinicians may resort to using the bucket, if it goes below that level clinicians may increase the inflow at the spout. **Balancing these flows is at the core of the tensions.** To meet the bathtub goal, the water level goal must be reached and the spout (inflow) set equal to the capacity of the drain (outflow). This requires closely-monitoring the water level and adjusting the flow rates (spout and bucket). **This process of monitoring a stock, and making changes to flows as its level changes, is the essence of feedback.** However, perceiving the differences in the rates of flow it **is a very difficult task** to accomplish. Clinicians seek to balance their team's workflow in much the same way.

Figure 3.88 Feedbacks in the Core of Primary Care – Service Operations and the Tenets



3.6.2.4 PRIMARY CARE TRANSFORMATION

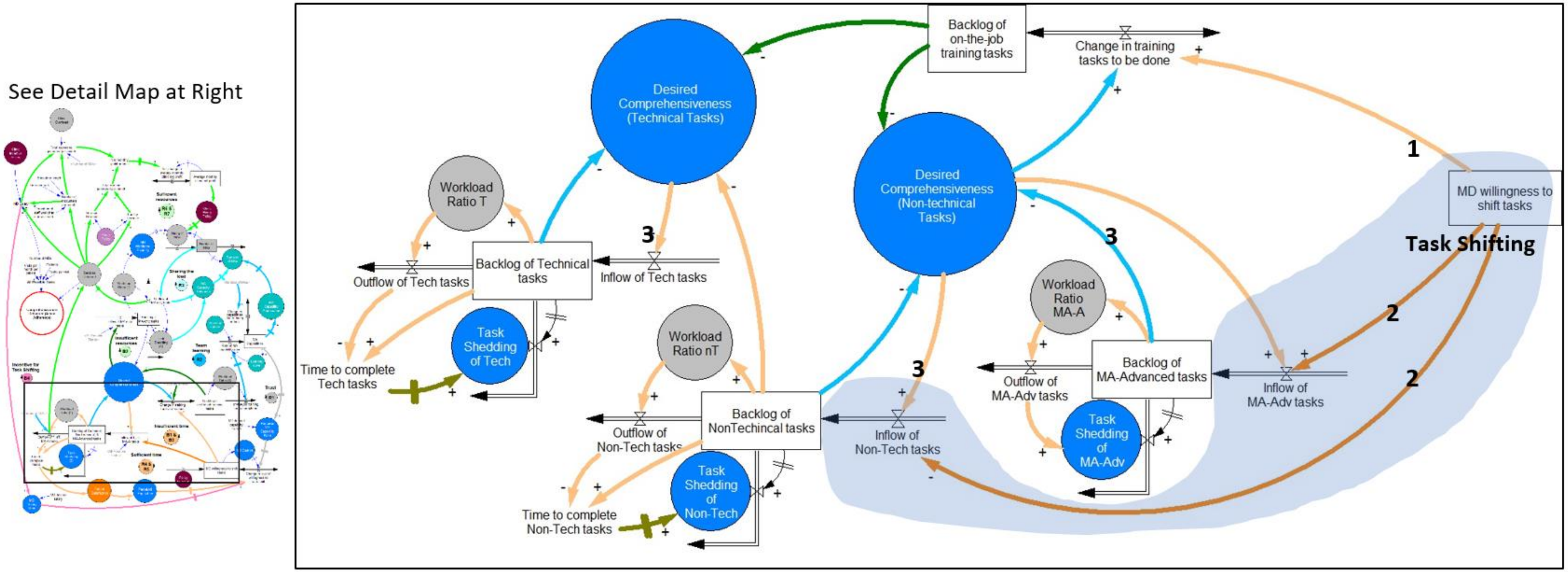
In this section, I disaggregated the stock of *Backlog of Tasks* used above into four separate stocks, each representing a distinct subset of patient care tasks. Task shifting introduces two more stocks: the training tasks associated with task shifting and the willingness of clinicians to shift tasks.

Figure 3.89 is a detail map associated with the Theoretical Model which illustrates how task shifting works from the clinician perspective. It does so by disaggregating the "*Backlog of Tasks*" stock[53] and showing how willingness and training interact with these stocks. As above, black lines show stocks and flows (the bathtubs, spouts, drains and buckets) and circles show policies: gray for subconscious productivity, blue for clinician decisions.

This figure uses numbered arrows to show that increasing willingness to shift tasks (the blue shaded region) does three things: (1) **adds** training tasks needed for the entire team to get up to speed on the new way of working, (2) **shifts** patient care tasks and (3) **adds** patient care tasks. These impacts are not complementary to one another, especially at first. For example, given the context of teams operating at full capacity, there is very limited time to provide training and due to the delays in the system (not visible in this detail map) these changes can lead to implementation failure or stagnation when the duration of training is over- or under-estimated.

[53] The *Backlog of Tasks* stock described above is an aggregate representation of the four types of patient care tasks considered in my model. MA-Only tasks are omitted from Figure 3.89 and Figure 3.90 for simplicity of display and because they are more indirectly related to task-shifting.

Figure 3.89 Task-Shifting & Task Backlogs in Primary Care Transformation



The dark blue region on the diagram highlights the mechanism for Task-shifting. It is the two flows: *Inflow of Non-Tech tasks* and *Inflow of MA-Advanced tasks*; and the stock *MD willingness to shift tasks*. When the level of the stock changes, it triggers changes in the two rates of flow. If the stock increases, then the *Inflow of Non-Tech tasks* goes down meaning that the clinician is assigning themselves a reduced number of Non-Technical tasks. The *Inflow of MA-Advanced* tasks goes up meaning the clinician is now assigning these tasks to the MAs.

I have found that the successful trajectory for PCT appears to involve four phases. Using the Simulation Model, I present one scenario[54] (“*Successful PCT*”) involving these four phases to successful transformation, starting with the initial team (two clinicians and two clinical staff members, taking care of the average panel of patients). Figure 3.90 below shows the task trajectory[55] for this scenario. Blue lines are tasks performed exclusively by clinicians (Technical tasks) and clinical staff members (MA-only tasks). Red lines are tasks that can *plausibly* be done by any member of the team – the tasks being *shifted*[56]. Green lines are training tasks (i.e., instruction, supervision and team problem solving). The four phases are described here:

- **In the first five months (Phase 1, months 10-15),** task shifting (and thus PCT) begins. Clinicians shift a small portion of tasks to clinical staff members. The team spends time training and staff show ability to perform newly-shifted tasks. Task shifting increases only slightly during this period since they are working through the tasks already shifted, on top of their regular duties.
- **The next 30 months (Phase 2, months 15-45)** may look like relative stagnation for clinicians but successfully completing this phase is crucial to reaching the later phases. Training continues, little by little, resulting in the clinical staff members on the team accumulating an average capability of three-quarters that needed to reach full comprehensiveness. Thus, task-shifting continues, in fact it picks up, reaching the inflection point for task-shifting nontechnical tasks in month 45. This shows up as inflection points in task completion shortly thereafter.
- Over the past two phases, the *clinician desired comprehensiveness has slowly increased* as the care team has kept up with tasks of the practice. **The next 5 months (Phase 3, months 45-50)** consist of a rapid increase in the comprehensiveness that the practice is able to deliver indicating to clinicians that they can confidently expect more (desired) comprehensiveness and thus they *turn up the spout more quickly*.
- **The remaining 50 months (Phase 4, months 50-100)** finish out the transformation. In month 50, the inflection point for desired comprehensiveness was reached – and the spout-turning slows. Training continues but it takes up less and less of care team time. Before this point, the increased comprehensives was only due to task-shifting. Now, clinicians can finally start doing more technical tasks.

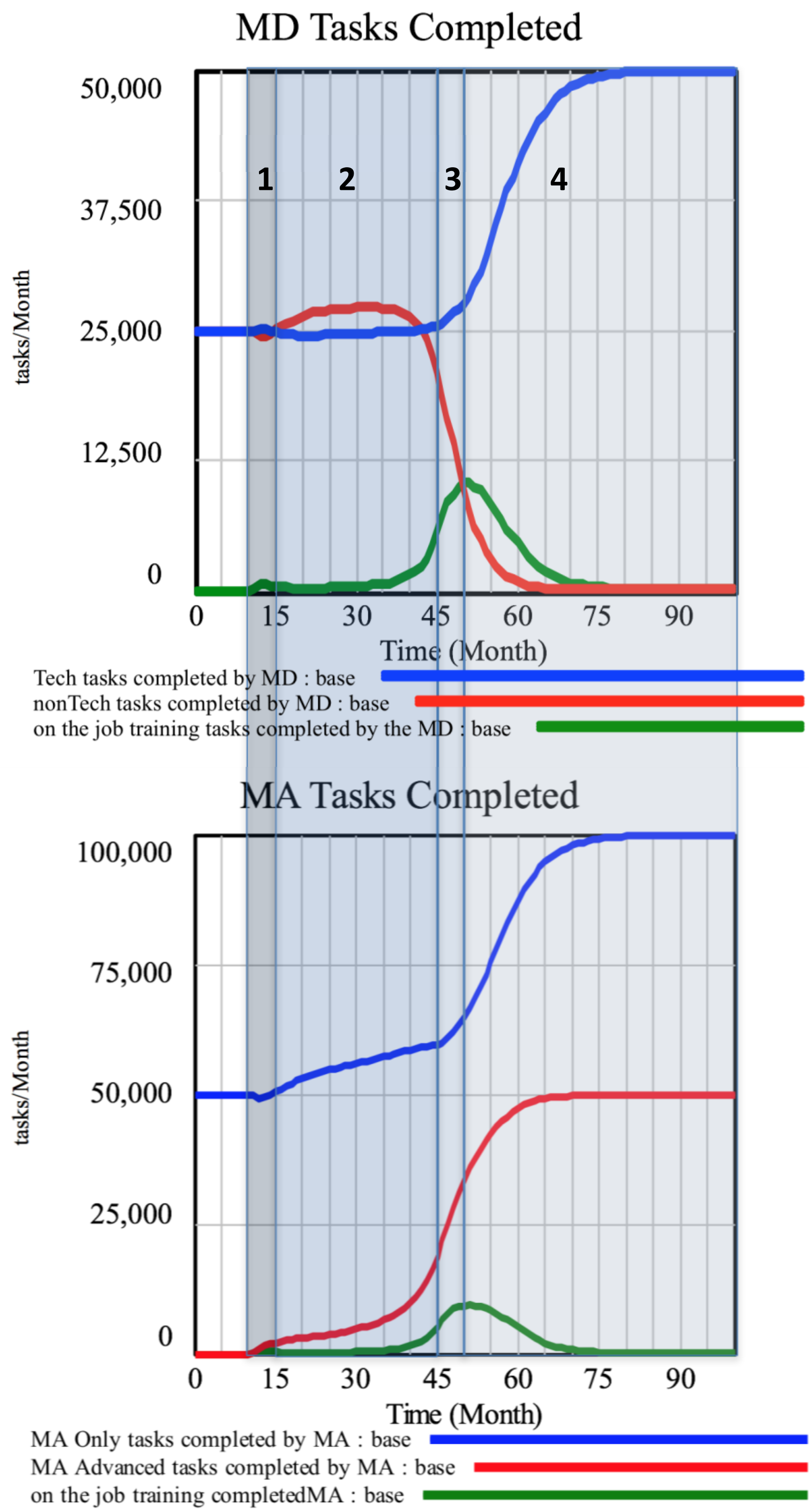
---

[54] This scenario tells a story of the successful transformation of primary care from current norms to realizing its full potential and value for patients. In policy analysis, I refer to it as the “Task-Shifting Base-Case” scenario (see Section 3.6.1.2 for a description).

[55] Also see Section 3.6.1.2 for additional variable graphs for the base task-shifting scenario.

[56] These are re-named from *Non-technical tasks* when they are performed by clinicians to *MA-Advanced tasks* when they are performed by clinical staff members.

Figure 3.90 One Successful Trajectory of Primary Care Transformation



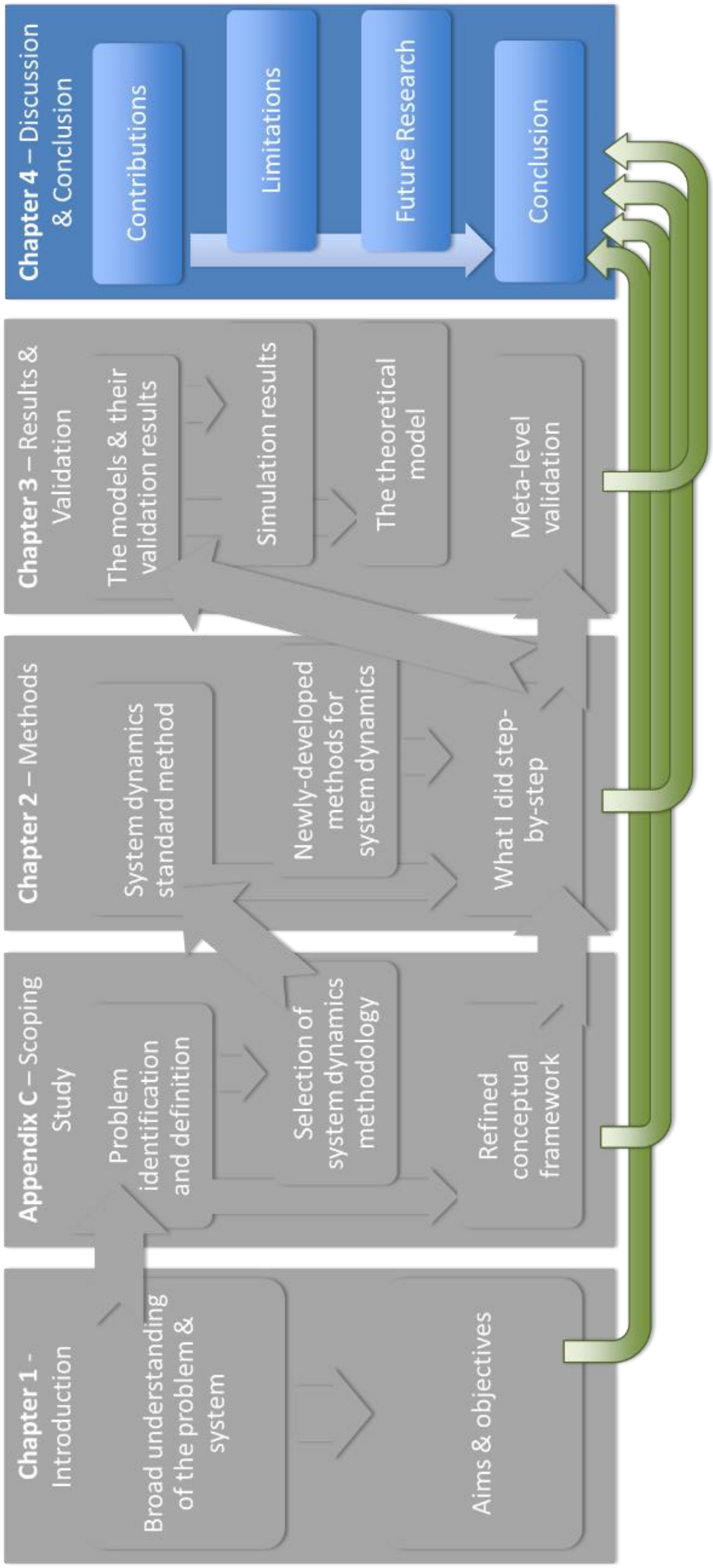
3.6.2.5 PRIMARY CARE TRANSFORMATION TENSIONS

The Scoping Study (Appendix C) identified tensions in PCT. The Conceptual Model specifically identified tensions between short-term and long-term loops. The Theoretical Model presents the operational structure of these tensions. Policy analysis using the Simulation Model presents behaviors associated with the tensions.

This research finds that the tensions are forces that prevent teams from building up the high-performing team which can produce the resources needed to sustain PCT.

Overcoming the tensions begins with **Trust**. The task-shifting kickstart policy introduces a small amount of trust where there was none before. This is the trust that MAs will be able to perform tasks they have never performed before. It is the stock of *Willingness to Shift Tasks*. As tasks are shifted, teams engage in **Team Learning** – they spend time in on-the-job training tasks and clinical staff members begin to develop new capabilities. These capabilities reinforce clinicians' trust and encourage clinical staff members to stay on the job. With time, a stable team of clinical staff members develops. This leads to a **Sharing the Load** effect where each clinical staff member feels the burden is shared fairly and this reinforces their decision to stay on the job. These virtuous cycles of Trust, Team Learning and Sharing the Load provide the momentum needed to allow the long-term loops of **Resources** (human, financial and time) to overcome the vicious short-term loops that present a constant drag on progress.

CHAPTER 4: DISCUSSION & CONCLUSION



4.1. INTRODUCTION

Primary care clinicians face a daunting task: to care for everyone's health without neglecting a multitude of interrelated, potentially-significant, complex factors in coordination with all relevant facets of patient lives. What's more, many primary care systems are not helping as they could. Health systems can better support primary care clinicians in achieving their important mission with a grounded, dynamic theory of primary care in hand.

Primary care scholars have called for integrative, transdisciplinary, participatory, multimethod research approaches to improve understanding of the role of this integrative, transdisciplinary patient-centered profession in the fragmented US health system[2, 221]. I pursued the task of scoping the problem and selecting appropriate methods. Nevertheless, making these methods *adequate* required me to be creative.

In this chapter, I reflect on the results (section 4.2.1) as well as on the journey that took me there (sections 4.2.2-5). In section 4.3, I describe the limitations and, in 4.4, I discuss potential areas for future research. Section 4.5 concludes the chapter with an extended summary of what I learned for my theoretical aim (regarding the transformation of primary care) as well as my methodological aim (regarding paradigm issues, emotional engagement and empirical validity).

4.2. MY CONTRIBUTIONS

Table 4.1 below lists the contributions of this thesis, grouped by contribution area (grey). The first column provides general descriptions of each contribution. The second column lists the section of the chapter where that contribution is discussed. The third column summarizes each contribution. The remaining columns to the right mark the aim(s) and objective(s) to which each contribution corresponds.

In this chapter, **<u>bold with underline</u>** is used to indicate when these contributions are claimed in text.

Table 4.1 Contributions of This Thesis

General Description	Section Number	Detailed Description of Contribution — Aims: Objectives	Theory: Primary care transformation	Methods: Paradigm Issues	Methods: Emotional Engagement	Methods: Empirical Validity
4.2.1 CONTRIBUTIONS TO THEORY OF PRIMARY CARE TRANSFORMATION						
More precise definition of key concepts	4.2.1.1	Stock and flow structure of service delivery in primary care, and its link to the tenets of primary care	x			
		Operational core of the tenets and the tensions in primary care service delivery	x			
		Core structure of primary care transformation with task-shifting as key ingredient	x			
		Structural explanation for the various developmental pathways (and phases) observed by Quigley *et al.* Tuepker *et al.,* Van Cleave *et al.,* Cronholm *et al.,* and The National Demonstration Project	x			
	4.2.1.2	Tensions in primary care transformation include those in service delivery as well as others	x			
		A theory of the *structural* tensions underlying the interpersonal tensions which have been observed to rise with primary care transformation	x			
		Perception delays and preferences are points of leverage in primary care transformation	x			
		Tensions exist between short term and long term loops, thus successfully managing them requires patiently balancing resources with workload	x			
More clarity on mental models	4.2.1.3	The role of clinician identity change in primary care transformation	x			
		Mental models definition for primary care transformation research	x			
		Individuals implementing primary care transformation understand it differently from one another	x			
		The core of shared understanding in primary care transformation is task shifting	x			
		The role and impact of different clinician mental models in primary care transformation	x			
		Precise constructs for improving understanding of mental model change (be it Ruddy *et al.*'s *premise transformation*, Cronholm *et al.*'s *mental models*, Berwick's *mental model shifts* or the National Demonstration Project's *identity change*)	x			
More clarity on the time delay for PCT	4.2.1.4	A simulation enabling understanding of the timeline for establishing realistic expectations of what primary care transformation can achieve in a variety of plausible circumstances	x			

		Aims	Theory	Methods		
General Description	**Section Number**	**Detailed Description of Contribution** / **Objectives**	**Primary care transformation**	**Paradigm Issues**	**Emotional Engagement**	**Empirical Validity**
4.2.1 CONTRIBUTIONS TO THEORY OF PRIMARY CARE TRANSFORMATION (continued)						
More clarity on adaptive reserve	4.2.1.5	Insights from my model for elements of primary care transformation that The National Demonstration Project links to *adaptive reserve*	x			
		Definition for *adaptive reserve*	x			
		Equation for *adaptive reserve*	x			
Policy insights	4.2.1.6	The task-shifting policy is the highest point of leverage	x			
		There is no policy that induces success regardless of preferences	x			
		The effects of certain preferences can be counter-acted by policies and/or other preferences	x			
4.2.2 CONTRIBUTIONS TO SYSTEM DYNAMICS – GENERALLY						
More precise definition of key concepts	4.2.2.1	Proposed types of information for sorting mental data (and methods for identifying & using them)		x		
	4.2.2.2	Conceptual, formulational and data validity		x	x	x
New methods	4.2.2.3	Methods providing a step-wise, grounded approach to developing and validating system dynamics models		x	x	x
Empirical evidence	4.2.2.4	Evidence for the differential perception of the elements of structure		x	x	x
4.2.3 CONTRIBUTIONS TO SYSTEM DYNAMICS THEORY – PARADIGM ISSUES						
Synthesis of philosophical work	4.2.3.1	Synthesis of philosophical works of Schumacher, Wilber and Stange on the *Four Quadrants of Knowledge*		x		
		Synthesis of the above with philosophical works of Schultz, Hatch, Gioia and Pitre on approaching multiple paradigms		x		
Synthesis of philosophical work	4.2.3.2	Use of system dynamics theory on causality and validation as the *second-order theoretical concept* (i.e., the mechanism) for crossing the *transition zones* between paradigms		x		

General Description	Section Number	Detailed Description of Contribution Aims Objectives	Theory Primary care transformation	Methods Paradigm Issues	Emotional Engagement	Empirical Validity
4.2.3 CONTRIBUTIONS TO SYSTEM DYNAMICS THEORY – PARADIGM ISSUES (continued)						
Theoretical reflection on approaches to multiple paradigms using system dynamics	4.2.3.2	Assumptions of system dynamics theory under each of the *Four Quadrants of Knowledge* and their research implications		x		x
		Assumptions of system dynamics models within each of the six *transition zones* of the *Four Quadrants of Knowledge*		x	x	x
		Identification of system dynamics best practices as forming a paradigm-crossing strategy and explicitly mapping that strategy to the *Four Quadrants of Knowledge*		x		x
		Three alternatives that researchers using system dynamics have for dealing with multiple paradigms		x		
New methods	4.2.3.3	The *interplay* paradigm-crossing strategy (i.e., integrative **methodology**)		x	x	x
Case contribution	4.2.3.4	The case presented here uses system dynamics as an integrative paradigm in an *interplay* paradigm-crossing strategy		x		x
Overall	4.2.3.1-4	Strengthen system dynamic's paradigmatic position as an integrative paradigm		x		
4.2.4 CONTRIBUTION TO MIXED METHODS						
New methods	4.2.4.1	Use of mixed methods for a scoping study			x	x
	4.2.4.2	Methods for developing and validating mental models				x
		Methods for mental model *measurement*				x
More precise definition of key concepts	4.2.4.3	Synthesis of system dynamics definitions and proposed mechanisms for emotional engagement			x	
New methods & Case contribution		Methods for and a case of developing stories with numbers that attempt to engage emotionally with stakeholders			x	
New visualizations		Newly-developed visualizations are designed to engage emotionally with stakeholders			x	
More precise definition of key concepts	4.2.4.4	Terminology for integrative work using *stories to accompany numbers* within each of the six transition zones of the *Four Quadrants of Knowledge*		x		
New methods		Integrative methodological approach which maintains paradigm and methods integrity		x		
Case contribution		A case for generating *stories to accompany numbers* using system science methods		x		x

General Description	Section Number	Detailed Description of Contribution	Aims	Theory	Methods		
			Objectives	Primary care transformation	Paradigm Issues	Emotional Engagement	Empirical Validity
4.2.5 CONTRIBUTION TO VALIDITY IN SOCIAL SCIENCE							
Synthesis of philosophical work	4.2.5.1	Summary of the philosophical work of Trigg on *expectations for integrative minor content theories*			x		x
New methods and theory		New theory and methods for research aiming to develop and validate *integrative minor content theories*					x
New methods	4.2.5.2	Newly-developed methods employ qualitative data to develop and test both qualitative and quantitative models for SDM			x		x

4.2.1. CONTRIBUTIONS TO THEORY OF PRIMARY CARE TRANSFORMATION

The aim of this section is to show the usefulness of my theory by comparing and contrasting my Theoretical Model with other theories to show its usefulness for generating insight. The Theoretical Model (Section 3.6.2) summarizes my contribution to the theory of primary care – it is my most comprehensive *dynamic hypothesis* and the result of my grounded dynamic analysis. Theorists in social science[217] use box and arrow diagrams to present theoretical models. In my Theoretical Model and in the diagrams presented in this chapter, I use these symbols[57] as well as additional ones from system dynamics (flows, link polarity and delays).

I begin by discussing my insights on the system structure of primary care and of primary care transformation (PCT). Section 4.2.1.1 describes the stock-and-flow structure for service delivery in primary care and its interaction with the four primary care *tenets*, the operational structure of the tensions involved in *any* primary care setting and the core structure of primary care transformation. The section concludes with a model-based explanation for the various developmental pathways observed in past studies of PCT.

Next, I consider insights from the full Theoretical Model. Section 4.2.1.2 describes the additional tensions involved in a *transforming* primary care setting including a detailed example for one specific tension. Next, I present insights regarding the link between causal structure and interpersonal tensions. I also discuss the importance of perception delays, preferences, and a patient, balanced, long-term approach in achieving successful PCT.

Following this, I present some of the conceptual insights my research has to offer for understanding PCT more generally. Section 4.2.1.3 presents my contributions to clarifying several aspects of the important concept of *mental models in PCT*. Section 4.2.1.4 presents my contribution to understanding the impact of time delays on PCT timelines. Section 4.2.1.5 demonstrates the insights available from my model regarding the concept of *adaptive reserve* that has been proposed by others studying PCT.

Finally, in section 4.2.1.6, I present the policy insights I have gained from this dissertation work; namely that management's approach to task-shifting offers the highest point of leverage, that policies alone cannot guarantee success (because of preferences) and that the effects of certain preferences can be counter-acted.

[57] I introduced these symbols in a gradual way (using the metaphor of a bathtub) alongside my generalizations about primary care transformation in Section 3.6.2 above. I do so in a technical way when I describe Figure 2.4.

4.2.1.1. PRIMARY CARE TENETS, TENSIONS AND TRANSFORMATION

Receiving adequate primary care involves providing patients with all the clinically-indicated **services**, delivered and/or coordinated by their clinical team (made up of clinicians, medical assistants[58] and other clinical staff members), and providing them in a timely manner. Delivering these services involves the four primary care tenets of access, coordination, comprehensiveness and continuity (see Table 1.1). **This work shows the causal structure of the tenets as being closely coupled in the primary care system. The simple stock and flow structure in Figure 3.87 in Section 3.6.2.2 helps to see this *coupling* clearly.**

This close coupling makes optimizing the system difficult. In such situations, high-leverage policies can be hard to find. Reflecting on recent health reform efforts, former US Secretary of Health and Human Services Sylvia Burwell identifies access as the piece of the system that is easy to see, measure and improve:

> "…when one looks at the progress of the Affordable Care Act against the three measures [access, affordability, and quality], I think that our best progress was against that one [access]. *Because it could be targeted and focused*. The affordability and quality… are sometimes harder because *the levers are more diffuse*, and … we do *struggle with measuring* quality, which is *related* to affordability because you don't just want affordability."[222]

When attempting to improve primary care, the tenet which is most easily seen is *access*, because it includes the stock of *Backlog of Tasks* and stock variables are more readily perceived than flows. The understanding in Figure 3.87 is useful because it represents these relationships using a *precise* definition which begins to build understanding of the many intangible concepts that relate to primary care and primary care transformation.

This definition can be easily converted into an integral equation, a system of differential equations or into a familiar qualitative metaphor like the bathtub (assuming no splashing or evaporation). Either way, once the stock/flow visualization is understood, **it is clear what is happening**: "the quantity in the stock is always the accumulation of the inflows to the stock less its outflows"[113] (p. 207). Also, this diagram **represents the links in a way that is easy to modify and improve** upon in later work. Finally, this understanding **provides the basis for exploring** how other system components influence the primary care tenets and the core (the following sections).

[58] For simplicity, variable names use "MA" to refer to all types of clinical staff members.

That said, an element of mystery still remains. First, we humans have a difficult time predicting what will happen in even the simplest stock/flow structures. This is likely because, as Sterman observes, "no one can measure the instantaneous value of *any* flow"(emphasis added)[113] (p. 207). Second, while simple in appearance, this kind of system is *dynamically* complex, it resists optimization of its parts in isolation.

These findings have much in common qualitatively with the influential work of Murray and colleagues on *Advanced Access*. This intervention design considers these *flows,* their links to access and continuity as well as tools for measuring them[223] and the "operational specifics" of both the care and the change processes for implementing them[224] (p. 324). While unintended consequences (e.g., being *over-paneled*) are recognized[223], its underlying math[225] excludes the role of stocks and feedback mechanisms by using optimization[113] (p. 11, 208)[106]. This aspect and the measurement challenges cited above make it difficult to avoid these consequences in practice. Indeed, research on PCT has documented unintended consequences. Studies by Magill, Rittenhouse and Driscoll observed that Advanced Access led to unintended consequences in continuity[49, 226, 227]. In the bathtub metaphor, this would be turning up the spout too high and then having to bail the water out with buckets.

The structure in Figure 3.88 (Section 3.6.2.3) represents the operational core of the tenets and of the tensions in primary care service delivery. These are the feedback loops enfolding the core stock and flow structure of primary care. There are three balancing loops: **(B1) *Doing the work*, (B2) *Cutting corners*** and **(B3) *Limiting care.*** There is also one reinforcing loop, **(R1) *Keeping up by doing*.**

These tensions, are consistent with integrative work studying service operations by Senge & Oliva[228] and Oliva & Sterman[219, 229, 230] which found that "under work pressure, service personnel struggle to keep a balance between the flows of incoming and outgoing orders while maintaining reasonable working hours and sustaining service quality"[219] (p. 27). This contribution gets me most of the way to achieving my theoretical aim. The remaining tensions are felt as consequences of PCT.

Understanding PCT requires first expanding the stock and flow structure described in the previous two sections. I am now able to describe the basic structure of PCT: the shifting of certain tasks within the care team. In a multi-disciplinary care team, clinicians separate out tasks into ones to be done by them and ones to be done by others on the team. In essence, the structure described above is replicated four times as four types of task.

The first two types are: Technical clinician-only tasks and clinical staff member-only tasks. The next two types are non-technical tasks that can *plausibly* be done by any member of the team. PCT involves *shifting*[59] these tasks from clinicians to other clinical staff members (called "MA-Advanced tasks").

Increasing willingness to shift tasks has three impacts on the system: (1) adds **training** tasks needed for the entire team to get up to speed on the new way of working, (2) **shifts** patient care tasks and (3) **adds patient care tasks** (see Figure 3.89 for a detail map of task-shifting). These impacts are perceived in four phases of transformation. **My four phases are as follows (see Figure 3.90 in Section 3.6.2.4 above)**:

- **In Phase 1,** task-shifting increases only slightly during this period since they are working through the tasks already shifted, on top of their regular duties.
- **In Phase 2,** task-shifting continues, in fact it picks up, reaching an inflection point.
- (Over the past two phases, the clinician desired comprehensiveness has slowly increased as the care team has kept up with tasks of the practice.)
- **In Phase 3,** there is a rapid increase in the comprehensiveness that the practice is able to deliver, thus clinicians turn up the spout (inflow of patient care tasks) more quickly.
- **In Phase 4,** the team finishes out the transformation: the inflection point for desired comprehensiveness was reached, clinicians turn up the spout more slowly until the new equilibrium point is reached.

These findings have important similarities and contributions relative to other research describing PCT over time.

Quigley and colleagues studied PCT inside 14 locations of an urban, federally-qualified health center – a very different context from most of the sites studied in this research. They interviewed 56 participants and found a "seemingly common pattern in approaching and implementing [PCT]" across sites, with four phases[231] (p. 13). Phase 1 involves "building PCMH infrastructure" *(ibid.)*. Teams are created, additional staff members are sometimes added (both clinicians and clinical staff members), and regular meetings are instated, involving "[patient] data review communication and learning" *(ibid.)*. Phase 2 involves assigning medical assistants to doctors and "working to promote a PCMH culture focused on meeting patient needs and providing whole person care" *(ibid.)*. Phase 3 involves increasing access (via urgent

[59] Different terms are used, including: task-shifting, task-delegation, task-sharing, and mobilizing support staff. In all cases, after this step, the clinical staff member is authorized to perform that task, until it is retracted.

care, extended hours and technology). Phase 4 involves reworking the patient visit, more frequent team meetings and customer service training.

Below, I present the four phases shown by Quigley *et al.* and describe how their concepts (in bold) are addressed in my model or how they are not addressed therein[231].

Table 4.2 How My Model Addresses Key Concepts from Quigley *et al.*

Quigley Phase	Concepts which my model addresses	Not addressed in my model
1	**Team-based problem solving** is the stock of on-the-job training tasks (at the top of Figure 3.89 above). My concept **for adding staff members** involves adding capacity to teams gradually, a fraction of a *full-time equivalent* staff member at a time – similar to the roles played by professions Quigley lists.	**Adding clinicians**
2	**Promoting a PCMH culture** involves increases in the stock of *willingness to shift tasks* (at the right of Figure 3.89 above). My concept for **PCMH culture** also considers changes to *desired comprehensiveness* (blue bubbles, Figure 3.88 and Figure 3.89).	**Empaneling patients** and **assigning MAs to clinicians** were addressed but not delayed in my model
3	**Increasing access** is, using the bathtub metaphor, turning up the spout - offering more care to the existing panel of patients (inflow of tasks in Figure 3.87, Figure 3.88 and Figure 3.89).	**Extending working hours** and **adding urgent care**
4	**Reworking visits** involves, for clinicians, doubling the number of technical tasks provided and, for clinical staff members, taking on all of the remaining tasks involved with full comprehensiveness.	**Customer service training**
Note: Bolded text are concepts from Quigley *et al.*[231].		

Similarly to my *Successful PCT* scenario, Quigley *et al.*[231] report a *progression* with early phases involving the strengthening of the care team through both team-based problem-solving and staffing resources. Building on this infrastructure, the later phases deliver more comprehensive patient care.

Tuepker *et al.* studied PCT inside 15 primary care clinics within the US Veterans Administration Integrated Service Network[232]. They engaged 241 participants in focus groups and/or interviews. The clinics are described as in an *early phase* of implementation. For one participant, this involved: first, 6-8 months of struggle followed by 8-10 months of improved experience while continuing to develop teams, culminating with visible *progress* in the most recent month (17). While not explicit, a sequential progression is evident in the way that themes are described.

Van Cleave and colleagues performed 48 interviews and 60 medical record reviews at 12 pediatric practices in *late-stage* PCT across the US, specifically focusing on care coordination. Without defining specific phases, they describe an evolution from a reactive stance toward patient care to a proactive, more comprehensive one. This *evolution* was facilitated by a process of team-based problem-solving[233].

Cronholm *et al.*[124] interviewed 118 individuals at 17 practices, all engaged in the *early phases* of a state-led multi-payer-supported PCT effort. The authors described the early phases as requiring cultural changes, such as introducing "team-based processes" (p.1198). These changes required changes to practice members' "mental models" (p.1200). They cite that "the most tension in shifting the required mental models was displayed between clinician and MA participants ... [in] moving away from clinician-centric care" (p. 1200). A facilitator of this change was clinicians taking into account their current level of comprehensiveness. The concepts of *tensions* and *mental models* in PCT are considered in next two sections.

The National Demonstration Project[234, 235] studied 36 practices across the US for two years where the intervention was facilitated transformation (with no facilitation as the control). They found that transformation "represents a fundamental reimagination and redesign of practice replacing old patterns and processes with new ones" where there are "multiple components…[which] are highly interdependent" and where each change "ripples throughout the practice, affecting all other *work processes and individual roles*" [234] (p. 255-6). Rather than a successful sequence of phases, they describe practices as having varying "developmental pathways" that, they found, depended on their initial characteristics and *adaptive reserve* [235]. Concepts[60] from this important study are considered in later sections.

The structure yielding the Successful PCT scenario can also produce a wide range of other plausible system stories. This theory adequately explains what happened in the *successful* PCT experiences described by the five studies above. It also offers precise answers for why success happened, while making some acceptable simplifying assumptions. It can also consider what might have happened under different circumstances. **My theory improves understanding of transformation by providing a structural explanation for the various developmental pathways (and phases) observed by Quigley *et al.*[231], Tuepker *et al.*[232], Van Cleave[233], Cronholm *et al.*[124] and The National Demonstration Project[234, 235].**

Regardless of the terminology employed, the circumstances of a given case, or the trajectory of a PCT experience, in my theory, PCT always involves team-members working together to deliver more comprehensive care. Although not always explicitly stated, and when stated, referred to in different ways (e.g., implementing new functions, expanded role for staff

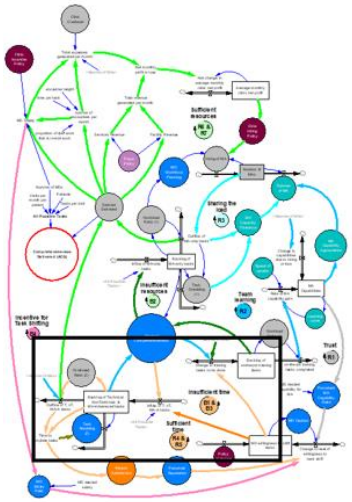
[60] Their concepts are listed here, with my terms in italics: interdependence and ripple effects (*tensions*), changing work processes and individual roles (*mental models*), timeline of expectations (*time delays*) and adaptive reserve.

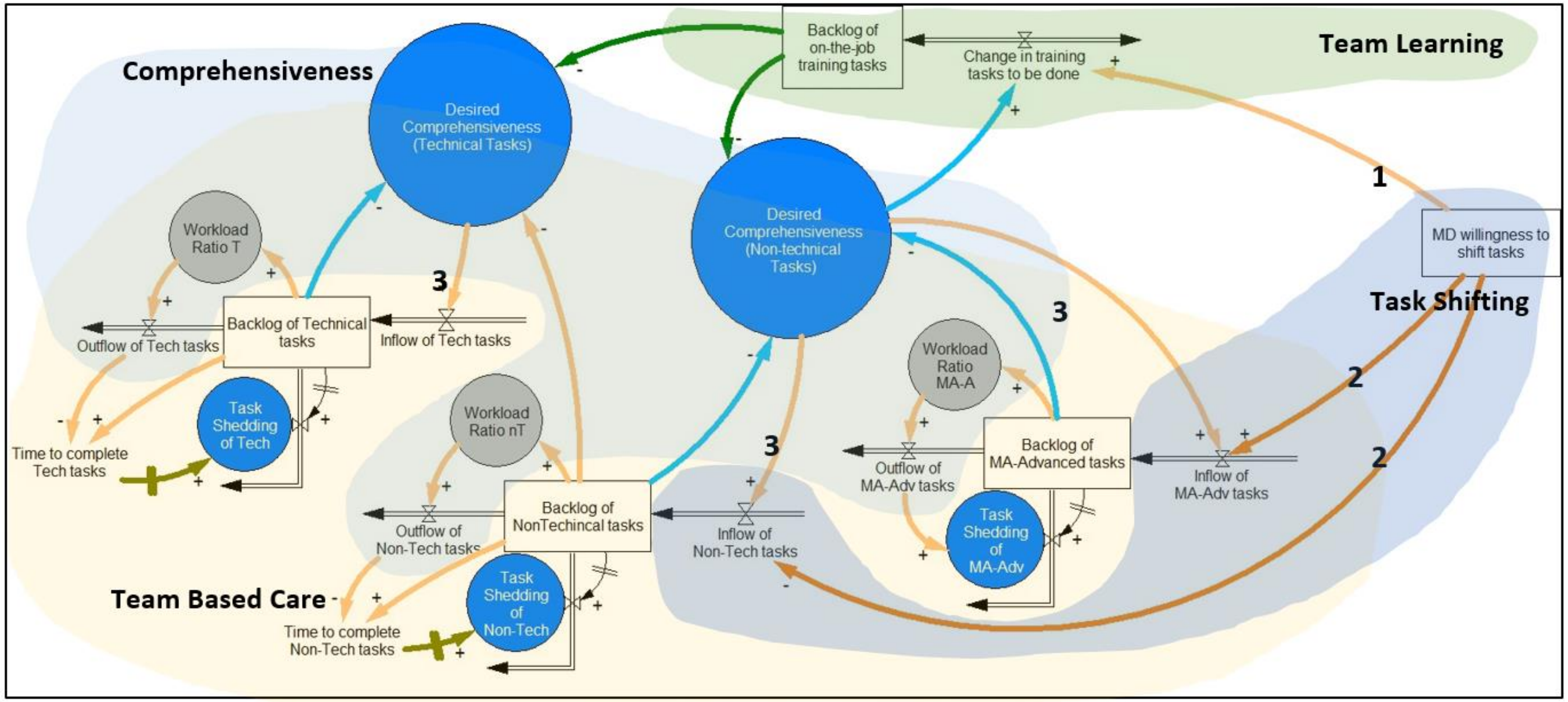
members, task-sharing), **task shifting is the driver of this change**. In all three studies presented above, researcher called task-shifting out as indispensable.

My theory of PCT offers a more explicit understanding of task shifting and its role in PCT. Figure 4.1 below shows how the progression discussed in these studies (qualitative shifts in emphasis over time) is occurring inside the primary care system. Team-based primary care (yellow region) is altered when *well-functioning* teams begin to *develop*. This process begins when *willingness to task shift* goes positive and initiates team learning (green region), task shifting (purple region), and increased comprehensiveness (blue region) (labeled 1, 2, 3, respectively). While all three aspects are active across phases, PCT participants will tell a story about one aspect being important at first and another one being important later on. The Simulation model matches this description over time and may explain how participants' experiences are formed.

Figure 4.1 Task Shifting & Task Backlogs With Qualitative Concepts

See Detail Map at Right





The dark blue region on the diagram highlights the mechanism for Task-shifting as previously decried in Figure 3.89 This figure here adds three other regions. The green region highlights the mechanism for Team Learning (completing on-the-job training tasks). The yellow swatch highlights the mechanism for Team Based Care (completing tasks as a team given that tasks are shifted among team members). The light blue region highlights the mechanism for Comprehensiveness (desiring to do and actually completing patient care tasks).

4.2.1.2. PRIMARY CARE TRANSFORMATION TENSIONS

The four phases I described above for a Successful PCT scenario can be thought of as phase shifts which are caused by changes in the dominance of the feedback loops operating in a complex system. These shifts in loop dominance occur as tipping points are passed, as this scenario passes through its various phases.

Participants might describe such an experience with PCT as the National Demonstration Project authors did: "Practice development is often slow and barely detectable. But it also occurs in sudden bursts"[236] (p. S73). **It is these difficulties, sensing the impact of current efforts and predicting when significant changes will occur, that make tensions visible. Tensions were always present, but without transformation they are dormant.** These include the primary care tensions that make it difficult to improve comprehensiveness as well as the PCT tensions which compound this by making it difficult to secure the resources that would make such improvement sustainable.

At month 10, in all scenarios, the tensions begin to be felt as the team wrestles with how to implement task-shifting within their team. In suboptimal and failed scenarios, this sensing/predicting challenge makes them felt all the more acutely. Suboptimal and failed scenarios in the Simulation Model helped to elucidate this finding.

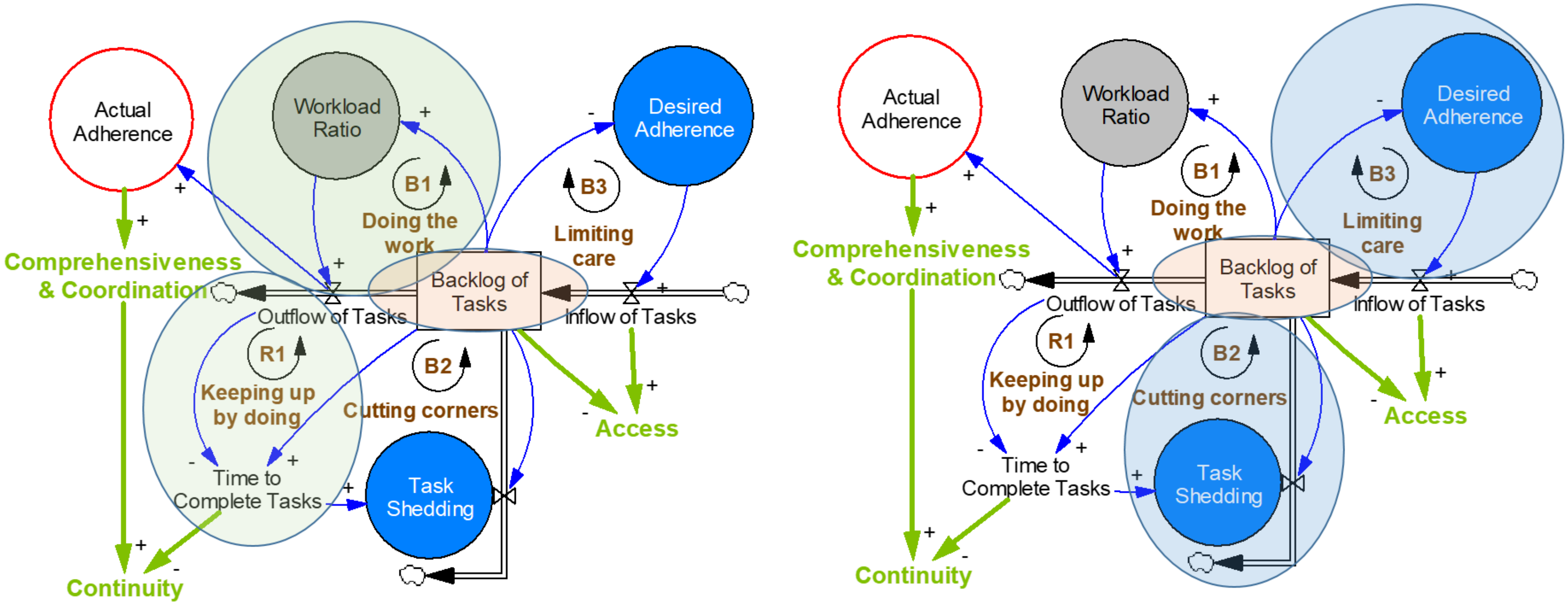
I will now provide an example of how this works for a primary care tension identified in the scoping study as "*Fostering a well-functioning system* versus *Getting through the day*" using Figure 4.2. On the right side, I show what happens as teams get overwhelmed by this tension. At first, PCT exacerbates this tension as the team experiences a gradually-growing *Backlog of Tasks* – shown with an orange overlaying region over the stock.

If the team opts for the short-term gain of getting through the day, then the feedback loops with blue overlaying regions become dominant (right side of the figure). This is done by decreasing the *Backlog of Tasks* by Limiting Care and Cutting Corners. This can lead to the team never achieving the long-term gain of transformation.

On the left side, I show what happens for teams which overcome this tension. They decrease the *Backlog of Tasks* by Doing the Work and Keeping Up by Doing. The short-term loops of *Getting through the day* become dominating the long-term loops of *Fostering a well-functioning system*. This is the path to achieving greater comprehensiveness.

Understanding why teams end up on one side or the other is an important question and a significant motivation for this research.

Figure 4.2 Fostering a Well-functioning System versus Getting Through the Day



PCT provides teams with additional resources with the goal of helping them end up on the left side of the figure. However, developing these resources also involves overcoming a new set of tensions.

The fundamental solution to the problem portrayed in Figure 4.2 above is to increase the *Outflow of Tasks*, which is the team's capacity to decrease the *Backlog of Tasks* by doing the work (or the drain in the bathtub metaphor from the previous sections). Increasing the stock of MA Capabilities (shown on the middle right side of the Theoretical Model, Figure 3.85) is one way of doing this. I consider here what determines how changes in this stock will begin to ripple through the system via the Trust and Team Learning feedback loops.

In Table 4.3 below, I describe the preferences and perception delays which relate to the MA Capabilities resource (right-hand column). These delays and preferences can attenuate or exacerbate the effects of PCT tensions. The first column identifies the loop for each one. The second column presents them in a causal chain and the third column identifies the element type.

In the Team Learning loop, MAs recognize increases in their capabilities and preserve them for the team by staying on the team (as turnover would result in the loss of the capabilities gained by the more capable individual leaving). In the Trust loop, clinicians recognize increases in MA Capabilities and allow more task-shifting to continue growing these capabilities. The reverse is also true in both cases.

Table 4.3 Preferences & Perception Delays of Changes in MA Capabilities

Loop	Causal Chain	Type	Description
Team Learning	MA Capabilities --||→+	Delay	MAs' perception of the change in their capabilities is delayed. This delay prevents them from reacting correctly to changes in their capabilities.
	MA Capability Appreciation	Preference	Are MAs on the team pleased by the change in their capabilities? MAs described as being "like a medical student" generate more satisfaction from increases in their capabilities. The opposite is also true for MAs that prefer their job description to remain unchanged.
	--||→- *Turnover of MAs*	Delay	MA's response to changes in their capabilities is delayed. This is a *second* delay that impacts them changing their willingness to stay.
Trust	MA Capabilities --||→+	Delay	Clinicians' perception of the current level of MA Capabilities is delayed. This delay prevents them from reacting correctly to changes in MA Capabilities.
	Perceived MA Capability Ratio	Preference	The relative eagerness or hesitation of clinicians to engage in further task-shifting based upon their perception of the change already achieved in MA Capabilities. Being too eager can overload the team with new tasks to perform, yet being too hesitant stunts the team's ability to transform.
	--||→+ Net willingness to task shift	Delay	Clinicians' response to changes in MA Capabilities is delayed. This is a *second* delay that impacts them changing their willingness to shift tasks.

These preferences and perception delays impact how the team experiences the tension and thus also their decisions. In both loops, team members' decisions affect flow rates to stocks which directly feed back onto the resource of MA Capabilities and both can lead to either a virtuous or a vicious cycle depending on the preferences and perception delays. Favorable conditions (for preferences and delays) give team members more of the resources they need (i.e., MA Capabilities) to overcome the tensions and increase the *Outflow of Tasks* – the fundamental solution. Otherwise, conditions may prevail which add pressure to the team to opt for the short-term loops in order to keep up (thus going to the right side of Figure 4.2 above). In these cases, tensions are felt more acutely and negative emotions increase (see Section 3.5.6).

My finding that tensions in the primary care system have important consequences for PCT is consistent with the findings of others studying PCT who observed *tension and stress* along with PCT. For example, Cronholm and colleagues observed that the mental model shifts and practice culture changes required in early phases caused "much tension and stress"[124] (p. 1200).

The National Demonstration Project team found interpersonal tension to be one visible sign of problems in the underlying structure of primary care (the structural tensions I am interested in):

> "The intense pace and magnitude of change soon revealed and exacerbated deeper dysfunction within *the relationship infrastructure of practices* including tension among physicians and among practice staff, ineffective communication patterns, and avoidance of potential conflict and difficult conversations that produced stalemate."[234] (p. 256-257, emphasis added)

My theory describes how these interpersonal tensions arise – because of tensions in the system structure. These system structures are there in traditional care (the model's initial equilibrium) but they are not felt by participants as a long-standing balance has been struck. Model runs begin in equilibrium with the core functions of the system operating as on an average US primary care team. The bulk of the system structure – especially time delays – is invisible during this stage.

It is revealed after the disturbance of the task-shifting kickstart policy which sets off the change process. When teams see the tensions, they must decide how to address them. The fundamental solution will require team cohesion (i.e., perceiving capability and capacity, and acting accordingly) and team problem-solving (i.e., on the job training). It is at this point that specific cases of function or dysfunction are made visible within that system structure. What makes the difference? **Suboptimal and failed scenarios point to two challenges that are to be overcome for the team to achieve Successful PCT: 1) shorter perception delays and 2) more middle-of-the-road preferences.** How teams approach these challenges causes them to travel on different evolutionary paths and (in the meantime) to feel interpersonal tension as the team wrestles to find their way through transformation.

The system's causal structure is the same in all scenarios. Thus, the Theoretical Model (Section 3.6.2) is capable of shedding light on the structural and interpersonal tensions in PCT. The purpose of my research was to learn about the feedback structures underlying the tensions which arise in PCT (see Appendix C and Section 3.5.6.3.1). These structures have been explored at the level of participants' shared understanding in the Conceptual Model (see Section 3.3.3). I formalized this understanding in a Simulation Model which permits controlled experiments (Section 3.4.2). Developing the Theoretical Model permitted me to reflect upon the implications of these findings for primary care and PCT.

Scholars conducting research on service operations across industries have found similar tensions. Along with the mental models which are used to respond to them, these tensions can lead organizations to "reducing the level of service they deliver, often locking entire industries into a vicious cycle of eroding service standards"[230] (p. 894). Reversing these cycles requires improving our understanding of the structures and practices that cause them.

I observed tensions between short and long term loops (like the one described above) where improved comprehensiveness through task-shifting is facilitated by patiently balancing the resources (time, financial, human) with the steadily increasing workload. While the example above focused on MA Capabilities, the Theoretical Model describes many resources that must be coordinated appropriately. In a different context (the insurance industry), Senge describes strikingly similar findings for successful transformation of a service organization: a strategy that balances aggressive hiring and steady elevation of quality standards is required to improve both quality and cost over the long run.

Specifically, Senge states:

> "Successful strategies depend on no single factor, but upon *coordination*. You hire and train people at a steady rate, you reduce turnover, you let backlog slip somewhat at first, and you strengthen the quality target steadily. You pay close attention to the intangibles of quality and [service personnel] effectiveness. This strategy takes some patience, but after five [simulated] 'years' of it, you're shocked by how lucrative your [scenario] has become." [237](emphasis in original, p. 334)

As described in this section and the previous one, it is plausible to assume that, as similar phases were described across multiple studies, and for many of the same reasons, a common structure exists across contexts and throughout the phases of PCT. As such, I posit that, for the purposes of understanding PCT and improving primary care, my theory proposes a structure that is both a "relatively enduring"[118] (p. 414) feature of reality and one with the power to "constrain and influence the choices"[98] (p. 208) of PCT participants in any context. I conclude that it is the way primary care stakeholders respond to the tensions that determines if the outcome is a successful transformation like the scenarios described above, or another one (i.e., suboptimal, failed or worse than before).

4.2.1.3. MENTAL MODELS IN PRIMARY CARE TRANSFORMATION

The goal of PCT is to help *ordinary primary care teams deliver extraordinary comprehensive primary care* by developing a high performing team; shifting tasks such that all team members' capabilities are fully utilized and preserving enough capacity to do so without overburdening the team during the transition.

PCT researchers point out that PCT is a dramatic change for primary care clinicians. On the one hand, it helps them live up to their identity as "someone who treats more than medical needs" by caring for their patients in a holistic way like parents do for their children[238] (p. 246). On the other hand, it involves changing that identity by requiring task-shifting[124, 235], which means the clinician is not the exclusive care-giver; possibly the equivalent of adding a team of nannies into the relationship. It is a dramatic change, but the patient-clinician relationship remains, now with the added relationships that clinical staff members have with patients and clinicians. **My model considers PCT as a clinician identity change: from one where patient care tasks are segregated such that only clinicians form relationships with patients, to one that allows clinical staff members to also care for their patients and form relationships with them**.

Research on cognition uses wide definitions of what a mental model is and contains[119], going beyond the professional identities considered in research on PCT[235]. **One contribution that I make is to widen PCT researchers' perspective on mental models (specifically, bringing the definition from system dynamics).**

The understanding of PCT described in the previous sections comes from my use of SDM. I began by eliciting participants' mental models from interview transcripts. Combining these into a shared mental model permitted me to gain an understanding of the complex interactions in the primary care system, as well as the way that they synergistically produce different PCT experiences. Validating this shared understanding with a validation-set of transcripts allowed me to develop a sufficiently comprehensive conceptual model and to make detailed measurements[239] of *these* participants' mental models. **All of the mental models I studied demonstrate that individuals implementing PCT understand it differently from one another (as visualized in SMM-S and CM-S diagrams).**

From the beginning of PCT in the US, researchers have feared that full transformation of primary care would falter. Authors of the flagship National Demonstration Project study evaluating PCT feared that: "Beyond the skeleton of the Joint Principles [of the Medical Home], is there even a **shared understanding** of what the heart and soul of a PCMH really is?"[234]

(p.258) Having such a shared vision, or shared mental model, is considered to be an important aspect of *successful* PCT[124, 235].

My qualitative diagrams show that **the core of participants' shared understanding of PCT is task shifting.** Links surrounding this variable had the most mentions and the most explicit mentions. These links tell a progressively more mysterious story of tensions involving feedbacks and time delays. Section 4.2.1 portrays the dynamic implications of this story, starting with the core of primary care.

In their final iteration, my Simulation and Theoretical models (i.e., representations of this shared understanding), involve two pieces: what system dynamicist Richmond calls the *infrastructure* and the *nervous system*[112]. The *infrastructure* shows the core structure of the service delivery system and task shifting. The *nervous system* shows the feedbacks that make it so difficult to balance the tenets and transform.

These models are designed to improve stakeholders' (e.g., participants, policy-makers) understanding of PCT. My interview data indicate strongly that understanding the system structure will aid in implementation by improving clinician buy-in, decision-making and team learning. With respect to the latter two, there are several policies under the control of clinicians that are very important for system behavior. Clinic managers described the task of investigating these policies and persuading clinicians to modify them (i.e., mental model change) as being emotionally-fraught, difficult and very important for PCT (see Section 3.5.6.4). This work **describes what clinician mental model changes are needed, what they involve and what would happen in their absence by precisely defining these *clinician* policies (i.e., preferences) and modeling their implications for system behavior.** This has the potential to make mental model change easier by helping clinicians (and others) to better understand what to hold on to and what to let go of in PCT[218, 240].

Over twenty years ago, the influential American health services scholar Don Berwick argued that improving healthcare worldwide requires six mental model shifts[241]. Table 4.4 below presents Berwick's six shifts (left-most columns) and the precise constructs that my theory provides (right column). In the grey box below each shift, I provide added commentary.

Table 4.4 Berwick mental model shifts

Berwick Mental Model Shifts[241] (p. 435)* — **From assuming that…**	**To…**	**Precise constructs from my theory of Primary Care Transformation****
"future performance levels will be approximately the same as current levels"	"believing in the pervasive **possibility** of breakthrough"	*Desired comprehensiveness*, Willingness to shift tasks
In the steady state, the actual comprehensiveness equals the desired level: 50% adherence to clinical guidelines. All of the dynamics in the model begin with a one-month positive pulse to the Willingness to task shift stock at month 10. This small belief in the possibility of achieving change kick-starts the positive *and negative* cycles in the model. Improved comprehensiveness comes at once, but the forces permitting steep increases take several years to develop.		
"measurement induces improvement"	"emphasizing **leadership of change** as the key to improvement"	Clinic Incentive Policy, Clinic Hiring Policy, *MD Workforce Planning,* Willingness to shift tasks
Important factors in achieving successful PCT include leadership taken by clinic management in the way they use the policy tools at their disposal (hiring policy and incentive pay). Clinician's main role in the improvement of healthcare is as leaders of care teams where they demonstrate leadership by their willingness to shift tasks and by advocating for additional clinical staff when care teams need them to avoid burnout.		
"professional and organizational boundaries must be carefully preserved"	"reducing those **boundaries"**	*Perceived MA Capability Ratio*, *MD Salary Ratio, MD Caution, Perceived Reputation,* Willingness to shift tasks
Many forces promote boundaries between clinicians and clinical staff members. Individual clinicians bring this to bear on the teams in the way they perceive the capabilities of their clinical staff members (as well as their perception of patient satisfaction, clinician salary and current level of task-shifting). How they internalize them influences their willingness to shift tasks. Ultimately, PCT is about their willingness to reduce these boundaries.		
"patients are passive and caregivers are active"	"working from strong notions of **equal partnership"**	*Desired comprehensiveness*, *Perceived Reputation, Task Shedding (2)*
In the steady state, patients and caregivers tolerate 50% adherence to clinical guidelines. As this desired comprehensiveness increases, patients and providers necessarily have more meaningful and frequent interactions, forming a more equal partnership for patient health. Task shedding erodes these relationships. Clinicians pay attention to patient relationships, in the model, by monitoring their Perceived reputation.		
"traditional forms of space and equipment are well designed"	"valuing fundamentally **new designs"**	Non-Technical tasks, On-the-job training tasks
While equipment and space are not explicitly considered, the re-design of primary care is explicit. For clinical staff members to take over and perform Non-Technical tasks appropriately new information technology will be needed, and teams will need space to perform the frequent on-the-job training (team meetings, huddles, problem-solving). Providing for on-the-job training with both space *and time*, enables team learning to occur (e.g., through small-scale plan-do-study-act tests like those described by Berwick[242]).		
"medical care operates in an environment of scarcity"	"noticing and employing what it has in **abundance"**	Clinic Incentive Policy, Clinic Hiring Policy, *Perceived MA Capability Ratio, MD Salary Ratio, MD Caution,* Number of MAs, MA Capabilities
There are many ways that agents monitor resources and take actions in the model. Scarcity mindsets involve being reticent to take action, even when needed resources are sufficiently abundant to justify it. Clinics have options for paying clinicians (incentives) and providing resources to care teams (hiring) which can deliver abundant resources or reinforce a sense of scarcity. Clinicians in my model have several policies for how much of a resource they would need before employing it, including the way they perceive MA Capabilities, how sensitive they are to perceived changes in their take-home pay, and how willing they are to share tasks (MD Caution). Finally, medical assistants are a capable resource which is relatively abundant compared to clinicians.		
*- emphasis added **-Policies are plain text, preferences are in Italics, stocks are underlined, and flows are italics underlined. Variable names are those from the Theoretical Model.		

Nevertheless, despite this early call to action, the work of Cronholm and colleagues[124, 195] shows that mental models are both an important and an *under-appreciated* aspect of PCT.

Concerned that PCT may be seen solely as a technical change to be implemented, primary care scholars emphasize that transformation is personal, and therefore mental models matter. Mirroring considerations from my research (see System Dynamics - Saturation), these scholars see mental models as an enduring feature of reality, worthy of scientific study and meriting attention of policy because they hold back needed changes. Ruddy *et al.* write that:

> "Our habits of premise, the frames of reference that explain how we understand ourselves and our worlds and our interrelatedness with those worlds and others in them [(i.e., our mental models)], are likely to be the biggest hindrances to sustainable, largescale practice transformation. Current policy definitions of and incentives for transformation lack *appropriate acknowledgment of the importance of personal and professional premise transformation*, and the adaptive change necessary for effective engagement with and leadership of mandated technical process changes. When we undergo these premise transformations, changing our frames of reference about who and where we are, we tend to feel a deep sense of confusion and disorientation, even fear"[240] (emphasis added, p. 625).

My theory provides precise constructs for improving understanding of mental model change (be it Ruddy *et al.*'s premise transformation[240], Cronholm *et al.*'s mental models[124], Berwick's mental model shifts[61] [241] or the National Demonstration Project's identity change[235]). Specifically, it identifies where they exist within the structure of PCT and how they impact a scenario's trajectory.

The importance of mental models has been recognized in process improvement more broadly. Organizational scholars and system dynamicists Repenning and Sterman argue that: "the paradox posed by useful innovations that so often go unused" (p. 292) can only be overcome by developing theories which integrate the cognitive structure of the "beliefs and behaviors of those working" *(ibid.) with* the "physical structure of the organization and its processes" *(ibid.)*; affirming that the sources of implementation problems lie neither in the innovation nor in the social context "but, rather, are rooted in the ongoing interactions among the physical, economic, social, and psychological structures [(i.e., mental models)] in which implementation takes place" *(ibid.)*[205].

To succeed as an innovation, PCT needs to work within the existing system; and, to do so, it needs to be part of the dominant mental model, the *theory in use*[169]. PCT, like quality

[61] The *mental model shift* considered by Cronholm and National Demonstration Project are one of the six types identified by Berwick. My comments to "professional and…" below relate to their definitions as well.

improvement in general[243], is a problem in need of an integrative, interdisciplinary theory to make sense of the technical-structural changes as well as the personal changes required. Crabtree *et al.* note of PCT's technical-structural changes, "not only do these structures and processes need to be in place, but they also need to be fully integrated into the day-to-day delivery of care"[63] (p. S83). Achieving the hoped-for mental model shifts facilitates *integrating* PCT into the day-to-day workings of primary care.

The task-shifting policy is essentially persuading the clinician to begin transformation by trusting clinical staff members with some of their patients' care. It introduces a new instance of the *Backlog of Tasks* stock (Figure 3.88). This is a redesign of the primary care system, a system *transformation* in itself. This technical change requires a significant change to clinicians' identity: trusting their team to take on some of their patient care work is a change in the clinician's operating policy, or mental model. System dynamicists (and psychologists, such as Powers[244]) consider this kind of mental model change to be an essential feature of *learning*[113] (p. 16). In this sense, my model simulates the challenges clinicians face in **learning** the right lessons from the task-shifting policy (the one month of compulsory although minimal task-shifting) – challenges due to dynamic complexity[245].

4.2.1.4. TIMELINE OF EXPECTATIONS – THE OVERALL DELAY

Another fear stated by the National Demonstration Project was: "What if the timeline of expectations is too short?"[234] (p. 258). Indeed, time delays has turned out to be a key point of contention in evaluating PCT. While it is recognized that PCT involves multiple phases[231-233, 246] and can take several years to play out[194, 234] (p. 257), evaluators have failed to account for the full delay[195, 246].

From the first time I tried to tell stories with the initial shared mental model, it was apparent that there was a struggle between what I saw as short-term and long-term feedback loops. The underlying factor determining which loop-types dominated was stakeholders' ability to wait through the delays – did they get stuck in the short term loops or could they get beyond them.

With the Simulation Model, I have gained an even greater appreciation of these delays. Delays matter and tensions are felt only when the system is knocked off balance. PCT involves rapid change, involving multiple interacting stocks and significant time delays of differing lengths. Because of this, we cannot quickly get the current state of reality to match our expectations, to bringing the system into balance. This causes strong feelings of tension and stress.

Delays are readily perceived but participants have a difficult time assessing their duration and the implications of their interaction; making the problem hard to understand and address[213, 214]. As a consequence, participants adopt "new policies and actions long before the results of old policies and actions can be properly assessed"[196] (p. 144). Specifically, stakeholders are pressured by the system to blame PCT and seek refuge in the traditional model of care.

My model names the delays and how they interact within the system structure: showing how physician burden can be maintained at a manageable level through the growing-pains of PCT. Each individual delay may be up to only a few months long but, together, they result in PCT taking years to complete as their effects interact in the model. **The model also enables understanding of the timeline for establishing realistic expectations of what PCT can achieve in a variety of plausible circumstances.**

4.2.1.5. ADAPTIVE RESERVE

Based on available empirical evidence[77-79], Altschuler *et al.*[220] found that, with task-shifting, comprehensive care could be delivered in the US using the existing clinician workforce. While they usefully outline feasible future states for transformed care teams, a more complex understanding is needed to map out *the journey* for getting there.

In their theory development work for PCT, the National Demonstration Project provided the concept of *adaptive reserve* as the guide for this journey. Table 4.5 below presents descriptions of adaptive reserve[234, 236, 247] from *National Demonstration Project* research team papers (second column). These are divided into *what it is not*, how it relates to system *behavior* or phases of transformation, and *measures* (using the three measures of adaptive reserve proposed in Jaen *et al.*[247]). Alongside this, I present related **insights from my Simulation Model** (third column).

Table 4.5 Insight from my model into adaptive reserve

National Demonstration Project descriptions		Insight from my model
What it is not[234]	"[Not the] core structure … capabilities to manage basic finances and clinical and practice operations during times of stability and modest change" (p. 256)	In my model, the core structure is the decision processes surrounding delivery of patient care (see Figure 3.87) and finances. These are common to any primary care practice – regardless of PCT implementation. In equilibrium, adaptive reserve is present, and not being exhausted.
	"In many practices, change began as an initial flurry of physician-led, just-do-it, top-down actions. Although initially successful in some practices, this approach proved ineffective in the long-term" (p.256)	I initiate PCT in the model using a *just-do-it* policy of *task-shifting kickstart*. Depending on the decision processes used in the system, this initial change can lead to all the trajectories. What proves ineffective in the model is the attempt to sustain this initial level of implementation when the system cannot support it.
Behavior [234, 236]	"[adaptive reserve greatly affects] a practice's ability to keep pace with rapid development and change"[234] (p. 256)	Were a practice to use *adaptive reserve* in decision-making, it has the potential to affect its ability to keep pace with transformation. My data did not say this happens, so it is not part of the feedback structure of my model. Instead, clinicians consider current workload only, not the adaptive reserve.
	"The frenetic pace and magnitude of the NDP quickly outran the practices' capability for change and required them to develop their capability for organizational learning and development. We labeled this capability the adaptive reserve…"[235] (p. S54)	An important feature of organizational learning is the ability to make needed changes to mental models and goals[113] (p. 16). In the model, clinicians continually observe and act in response to their team's performance. Their mental models change when they initiate and continue task-shifting. Their *Desired Comprehensiveness* (goal) changes as they react to the results of task-shifting.
	"Transformation toward a PCMH appears to require a strategic developmental approach that starts with assuring a strong structural core, and then implements smaller changes that help to build the adaptive reserve. Only then can larger, more complex changes begin."[234] (p. 257)	Policies leading to successful transformation are characterized by gradual positive approaches to task-shifting, increasing comprehensiveness and hiring new staff. Clinicians are given sufficient time to develop the needed trust in their clinical staff members as incentives encourage on-the-job training. Deficits in all three types of tasks (technical, nontechnical/MA-Advanced and MA-Only) are addressed as all involved seek to retain clinical staff members by paying attention to their capabilities development *and* workload. A growing team workload is balanced by hiring clinical staff members.
Measures [247] (p. 1)	"People in this practice operate as a real team."	The Simulation Model begins with a team of 2 clinicians working together, with 2 clinical staff members, sharing a common patient panel. The process of making this into more of a *real team* is task shifting, which begins with training and only progresses when high levels of trust exist within and across professions.
	"When we experience a problem in the practice we make a serious effort to figure out what's really going on."	At the team level, on-the-job training tasks are created any time task shifting occurs. These tasks require *serious effort* because they take time, whether used for protected group reflection or other time used for problem-solving, such as re-working the team's errors.
	"Leadership in this practice creates an environment where things can be accomplished."	Management determines the type of environment for this transformation through hiring and incentive policy. Having sufficient capacity (via sufficient staffing) allows the team to have the time for completing on-the-job training tasks while at the same time having a manageable workload (so capabilities can be gained and task-shifting can succeed). Similarly, having sufficient salary (via appropriate incentives) allows clinicians to continue being willing to shift tasks.

Above, I show how my theory treats the aspects of adaptive reserve that they described. Inspired by their work, I took an interest in this concept. I wanted to include it in my model in some way but struggled to see how it fit. A nonscientific internet search shows that the term is also used by scholars in physical therapy[62].

Below, I provide the definition presented by Robert Orr, a scholar in physical therapy: "adaptive reserve is our body's ability to *adapt in a positive manner* to a training stimulus … when completely rested [it] … is at its fullest … easier (recovery) workouts can be performed when [it] … is slightly depleted."[248] (p. 3, emphasis added). Trainers plan out the "volume and intensity of training loads" (p. 1) over time in an athlete's *periodized training* program. In conducting this "load planning" (p.3), trainers should consider the impact of the program on an athlete's *adaptive reserve* over time, their performance goal and the timeline of expectations. With this information, and "an understanding and careful manipulation of both volume and intensity," a *periodized training* program can protect adaptive reserve as it builds toward the goal via the application of "effective progression and overload" [248] (p. 1). This framing expresses the *physics* of adaptive reserve suitably to use it as a metaphor for my model of PCT.

I define *adaptive reserve* as an individual and/or team's ability to adapt in a positive manner to the added effort required by primary care transformation. PCT involves developing a stronger, more capable, high-performing team – one that has *adapted* itself through each progressive increase in task-shifting in a *positive manner*. In my model, the *periodized training* program involves the clinicians' preferences for how to move things along (the influences on task-shifting willingness). Through these preferences, the clinicians determine the inflow of patient care tasks (volume) and training tasks (intensity) – together they constitute the *load planning* in the *periodized training* program. The team's goal of full adherence to clinical guidelines is like the *performance goal* of the athlete.

In the previous section, the overall timeline of expectations for PCT depends on the interaction of many delays and it is not possible for clinicians in the real world to accurately estimate this beforehand. What they can see is the work they have to do (the *Backlog of Tasks* stocks). Careful manipulation of the volume and intensity that would make up a *periodized program* for PCT would require an understanding of the impact of the program on the team's *adaptive reserve* over time.

[62] This is the profession known as physiotherapy outside the US.

So, how would clinicians be able to see the *adaptive reserve* of their team? They can see the stocks of *Backlog of Tasks* (patient care + training tasks) which determines the *current workload*. They also know what their *normal workload* is. They know how busy they are compared to normal, which is their *workload ratio*. The only missing piece is what I call "*the maximum happy workload ratio*". This is the *workload ratio* at which the team reaches its peak productivity – beyond which they become less productive. Thus, **a simple equation for adaptive reserve is:**

$$\text{Adaptive reserve} = the\ maximum\ happy\ workload\ ratio - \frac{current\ workload}{normal\ workload}$$

Adaptive reserve should not get below zero. There are powerful feedbacks at play that prevent getting back to normal afterward – fatigue, burnout etc. These are important themes in the study of PCT – themes that I found in my qualitative interviews and also reported by others. But these are also themes that have been reported with other big changes, changes that involve increased volume and training such as implementation of health information technology (e.g., an electronic medical records system)[124, 195].

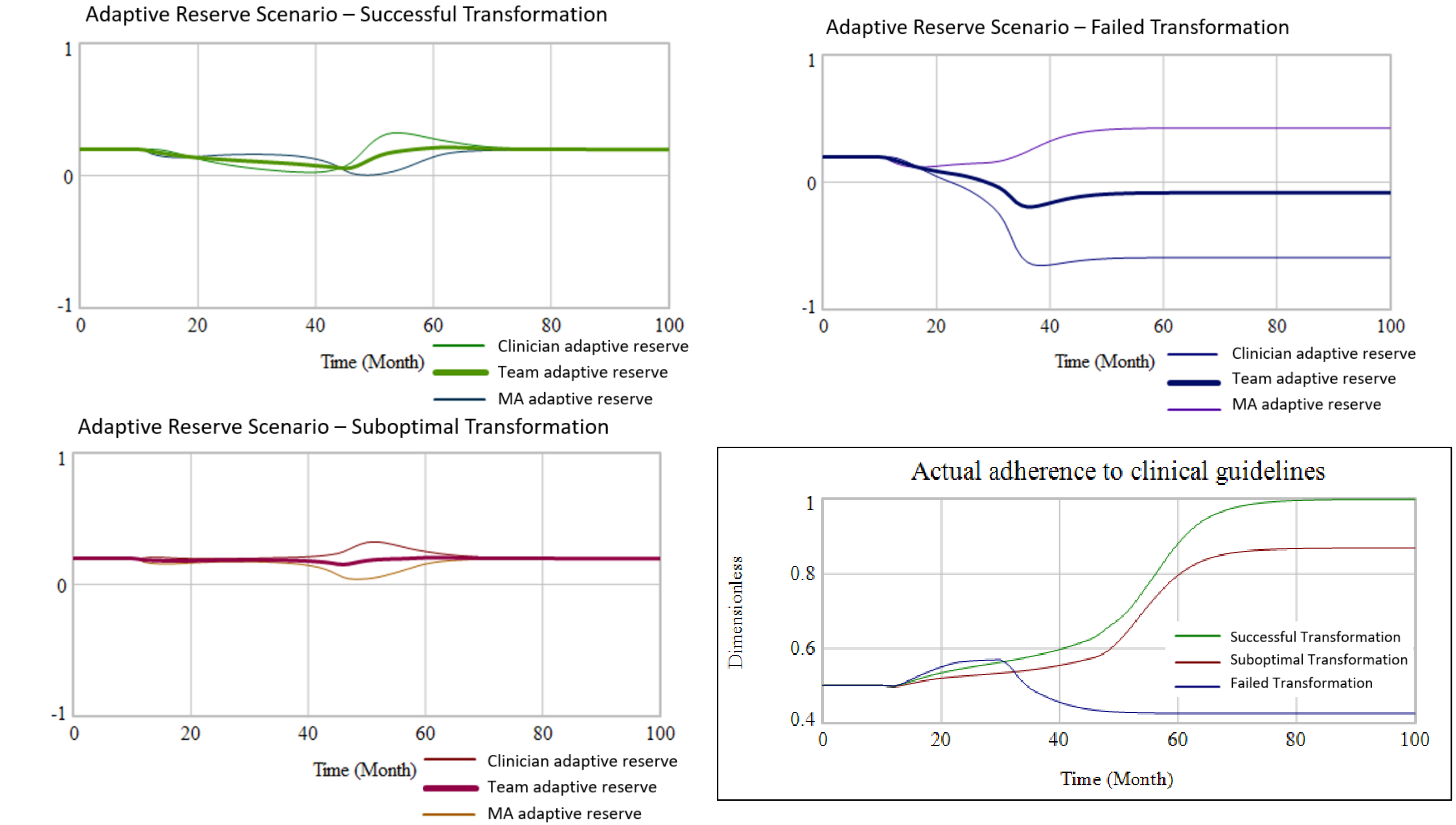
To consider what happens to adaptive reserve in the context of transformation, I present the three scenarios used for reference mode reproduction: Successful PCT, Suboptimal PCT and Failed PCT. Figure 4.3 below reproduces the reference modes (black bordered graph) alongside adaptive reserve for clinicians, MAs and teams. The variables calculate the adaptive reserve with the equation above and do not change the model's feedback structure. In the Successful PCT scenario[63], the team members take turns depleting their adaptive reserve, with the whole team recovering by the end of the transformation. In the Suboptimal PCT scenario, clinicians are hesitant to increase the inflow of tasks. In this scenario, the adaptive reserve for MAs gets depleted while that for clinicians increases as clinicians hesitate to take on more tasks while shifted tasks are successfully being done by MAs. In the Failed PCT scenario, clinicians are overly eager to increase the inflow of tasks. In this scenario, the adaptive reserve for clinicians gets depleted while that for MAs increases as clinicians give up on task-shifting and try to keep up with the increased inflow of tasks by themselves.

[63] Successful PCT is referred to as *Task-Shifting Base-Case* in Section 3.6.1.2 and in Section 3.5.5.1.1, and is the same scenario used in Figure 3.90.

During transformation, the goal is *not* to preserve adaptive reserve, but rather to *use it* – it takes work to build muscle. This is evident in the figure above. The team achieving Suboptimal Transformation uses less adaptive reserve – and clinicians hardly use any. Successful Transformation required more work on the part of MAs and clinicians alike. That said, adaptive reserve must be protected. Failed Transformation occurs when adaptive reserve is not protected from complete depletion[64].

[64] I am not predicting that PCT will drive some clinicians into a permanent failed state. The model's power to predict the results of policies diminishes as soon as the failure mode begin (as described in Section 3.4.1.3).

Figure 4.3 Three Scenarios: Actual Adherence to Clinical Guidelines, Individual and Team Adaptive Reserve



4.2.1.6. POLICY INSIGHTS

The model's ability to tell many stories in the same structure permits it to offer answers specific to one's context – one's policies and preferences. Results for specific policy and preference combinations are provided in Section 3.6.1.5. I begin this section with an overview of those results. I present these as part of my contributions to the theory of primary care transformation since they are part of its structure.

In a typical context (two clinicians and two clinical staff members on a team, each paid a fixed salary, with management willing to hire based on need), a small initial kickstart pulse of task-shifting can produce successful transformation. Clinicians need not be over-burdened by PCT. The kickstart makes it so task-shifting is maintained at 10% for 1 month after which it dips back down to a sustainable level (below 10% but above 0%) which is enough to build the momentum through a gradual positive approach.

The most impactful leverage point is the policy regarding the task-shifting kickstart amount. It impacts transformation regardless of the other policies and preferences in place. The lower the kickstart amount, the slower the change. The higher the kickstart the faster the change to a point, after which a suboptimal transformation results (see Section 3.6.1.3).

With the kickstart policy set to its base-case value, the next-most impactful leverage points are preferences. **There are some preferences that induce success or failure regardless of the other policies being implemented.** For example, with base-case preferences, an otherwise worst-case policy environment (payment uncertainty, a reticent hiring policy, and encounter-based payment) only slightly delays transformation.

Another way of saying this is that **there is no policy that induces success regardless of preferences.** There are several preferences that can derail transformation – regardless of supportive policies. Failure modes include: worse than initial, suboptimal equilibrium, and same successful trajectory with a significant delay for the outcome of interest "actual adherence to clinical guidelines".

That said, there are also some preferences whose problematic effects can be counter-acted by certain policies and/or other preferences. For example, one failure mode comes from having clinical staff members who are very sensitive to capacity constraints (i.e., they easily quit from being overworked) – perhaps because a competitor is recruiting MAs from the clinic and offers a less stressful work environment. A clinic policy of hiring clinical staff members based on need, rather than based on current finances, addresses the capacity constraints as they arise. Otherwise, the seemingly-judicious hiring restraint policy has the unintended consequence of preventing hiring when it was needed most and putting PCT in jeopardy. Another way to solve this problem is to have clinicians who are more sensitive to their team's workload – thus more strongly advocating for hiring. This prevents failure by hiring before the clinic finances dip down to where a hiring freeze would go into effect.

Therefore, stakeholder-specific recommendations involve both policies and preferences.

- **Clinical staff members** – Their absenteeism and turnover require teams to duplicate training effort and slow down transformation. Clinical staff members should communicate their preferences before choosing to skip work or leave the job altogether; for example, how much do they value growing in their role and having their workload under control. They can also contribute to PCT by actively engaging in the on-the-job problems solving – finding the best way for their team to transform.
- **Clinicians** – Their resistance to task-shifting slows down transformation. That said, being overly eager does not give the team enough time to problem-solve as more tasks are added. In deciding how much to task-shift, how much to expand services offered, and how much staffing to request, clinicians should pay attention to their team's preferences, to their evolving workload and capabilities, and to patient satisfaction. Also, they should communicate their salary concerns to management so that supportive incentive policies can be implemented if not already in place.
- **Clinic managers/Owners** – Their incentive and hiring policies can create a supportive environment for care teams wrestling with transforming themselves. That said, the initial approach that they take to task-shifting can make or break transformation entirely. When teams are burdened already, a pulse is less burdensome than a step; the pulse requires that clinicians shift tasks at a low level for a month after which managers/owners support them as clinicians choose how much to task-shift. Assuming teams have middle-of-the-road preferences, the system structure will get them to goal eventually; the pulse gives them the time they need to build up the resources to get there. Turnover and absenteeism among clinical staff members as well as insufficient progress in task-shifting are warning signs and deserve managers'/owners' attention. So, after setting up the initial task-shifting approach, their job is to figure out team member preferences and respond to them accordingly.
- **Health Service Delivery Organizations (HSDOs)** – Everyone has a role to play in the creation of a shared vision: teams need to develop a shared understanding of what they are doing and managers need to communicate clearly their goals and how they expect their plans to work toward achieving the end goal. This shared understanding results in

more middle-of-the-road preferences. Also, managers should ensure that net profits captured during transformation remain with the clinics so that they can be reinvested in hiring additional clinical staff members when needed. Finally, it is standard practice for management to require staff to read management books. My interviews indicate that the 5th Discipline[169] would be useful for teams undergoing transformation.

- **Payers** – Reimbursement policies can create revenue uncertainty for clinics. When managers are uncertain that the clinic will be appropriately compensated for expanding services, transformation can stall prematurely (e.g., if they hire only when clinics are in the black). Payers should consider alternative payment policies that are more supportive of transformation – as payment policies involving step changes in reimbursement (the current standard) are insensitive to the gradual process of transformation.
- **Healthcare education**
    - **Clinical staff member education** – PCT requires a capable workforce of clinical staff members to participate in task-shifting. This involves creating curricula for established professions (e.g., pharmacists) and new career paths (e.g., for care managers and *advanced* medical assistants) as well.
    - **Medical education** (clinician graduate school and residency) – PCT requires that clinicians be capable of guiding the learning of their team. Each team will have a specific context and way of doing things. The team needs to undergo the on-the-job training (problem solving[65][169] with each increase in task-shifting) to find a workflow that works for them in their context. This capability likely has a course component as well as a practical component, thus I suggest it be considered for both school and residency education.
    - **Healthcare administration** (masters of business administration, healthcare administration) – PCT also requires that managers are capable of guiding a complex organizational transformation. This involves more than process improvement and optimization – as shown in this dissertation. Organizational behavior courses need to be better integrated with quantitative courses – teach mixed methods, teach students that *all models are wrong some are useful*[113] (p. 890)[10]. In addition to making decisions about money and hiring, PCT requires managers to build personal mastery, to encourage learning and dialogue and to develop and communicate shared vision on teams. Coursework on mental model change and systems thinking[169] would thus be useful.

---

[65] Based on interviews analysis, this would require fostering systems thinking on the part of team members.

4.2.2. CONTRIBUTIONS TO SYSTEM DYNAMICS – GENERALLY

This section describes my contributions to the methods and theory of system dynamics. (My theoretical contributions to paradigm issues in system dynamics can be found in Section 4.2.3).

I begin with methods-related contributions. Section 4.2.2.1 presents my contributions to methods for sorting mental databases. Section 4.2.2.2 presents my contributions to the validity types underlying model validation methods. Section 4.2.2.3 presents my newly-developed methods for working with mental databases in developing and validating system dynamics models.

Finally, in Section 4.2.2.4, I present some evidence about the relative perceptibility of the elements of system structure used in system dynamics.

4.2.2.1. INFORMATION TYPES

In System Dynamics theory, mental data is the most empirically-valid conceptual basis for system dynamics models[100, 107, 138, 249]. Sterman holds that mental data "[includes: 1] descriptions of decision processes, [2] internal politics, [3] attributions about the motives and characters of others, and [4] theories to explain events, but these different *types of information* are mixed together" (numbering and emphasis added)[113] (p. 157). **I contribute four types of information for sorting mental data -- and methods for finding and using each type** (see Appendix E Section E.2.1)**:**

- *Causal information* (Sterman's types 1 and 4),
- *Conceptual* information
- *System dynamics-related items* and
- *Extraneous information* (Sterman's types 2 and 3)[66],

In reflecting on the theoretical implications of my research process, I found that the information types, especially the *extraneous* type, could potentially be more useful than I had thought. My process (see Appendix E) forced me to submit my problem statement to data and decide whether the problem statement continued to hold water or whether the extraneous information is relevant and therefore the problem statement should be revised to reflect that understanding.

[66] Extraneous to model conceptualization and not included in the conceptual model.

In reviewing the content of my extraneous information, I noticed that it tended to be instances of participants presenting their experience in ways that achieve a specific agenda, rather than earnest descriptions of their experiences. I was, in essence, coding out the non-purposive text data – the politics and personal attributions.

In reflecting on the theoretical implications of my research process, I found others developing typologies for qualitative data. For example, Van Maanen[250], defined information types from an ethnographic perspective and emphasized that categorizing correctly is crucial to developing high-quality theory. Kim & Andersen (see Section 2.4.2 *purposive text analysis*) support this view for SDM, stating that “the modeler must be able to assume, with a reasonable degree of confidence, that the mental models of the decision makers or stakeholders are revealed in the discussion”[117] (p. 313).

Thus, all four newly-proposed information types are relevant, regardless of the methodology, data sources, etc. and drawing the distinction among them is crucial to the research process and the product of the work. Table 4.6 below presents these newly-developed information types, as they relate to others contributing thoughts in this area. The theoretical descriptions of information types described by Sterman, Van Maanen, Kim and Andersen are compatible with the ones I produced.

Table 4.6 Information Types – Existing & Newly-proposed Ones

<table>
<tr><th colspan="3">Existing information types</th><th rowspan="2">Newly-proposed information types and definitions</th></tr>
<tr><th>Kim[117] (p. 312-313)</th><th>Van Maanen[250] (p. 542)</th><th>Sterman[113] (p. 157)</th></tr>
<tr><td rowspan="3">Purposive text: “the participants in the discussion have a sophisticated knowledge of the system … capture the focused discussion on the system and the problem at hand … should reflect a frank and unfeigned conversation of the decision-making group”</td><td rowspan="3">Operational: "documents the running stream of spontaneous conversations and activities … pertain to the everyday problematics of informants going about their affairs … behavior per se"</td><td>“descriptions of decision processes”</td><td rowspan="2">Causal*: Identifies cause and effect between at least two variables</td></tr>
<tr><td>“theories to explain events”</td></tr>
<tr><td></td><td>Conceptual*: Provides detailed definitions of meanings attached to concepts or assumptions (e.g., system boundary, time step, time horizon)</td></tr>
<tr><td rowspan="2">Grand-standing: “taking rhetorical positions that they do not believe in strongly”</td><td rowspan="2">Presentational: "those appearances that informants strive to maintain (or enhance) in the eyes of the fieldworker, outsiders and strangers in general, work colleagues, close and intimate associates, and to varying degrees, themselves."</td><td>“internal politics”</td><td rowspan="2">Extraneous: Not relevant to the problem statement</td></tr>
<tr><td>“attributions about the motives and characters of others”</td></tr>
<tr><td colspan="4">*- Note: these can also contain system dynamics-related items, a fourth newly-proposed information type. System dynamics-related items can be found in causal information (showing perception of complex elements of structure: delays, feedbacks, stocks, flows and nonlinearities) and conceptual information (showing that mental models and emotions are important to problem owners) (see Appendix E for more detail).</td></tr>
</table>

4.2.2.2. VALIDITY SUBTYPES

System dynamicists have considered validity at several levels of a research project[105, 166]. I considered the validity types for models to be most relevant to this dissertation. These include: Conceptual Validity, Formulational Validity subtype 1 and 2, Experimental Validity Types 1 to 3 and Data Validity[105]; **I propose subtypes for some (4 for Conceptual and 3 for Data Validity) and additional aspects of relevance to others (for Formulational Validity subtype 1 and 2) (see Appendix D Section D.2).**

In brief, these new types assist in clarifying the aspects of conceptualization and data which are being validated by a given validation method, or model test. These are consistent with theory of and expectations for model validation in SDM[107, 161, 249] and build on previous work by Lane[105], following guidance from Barlas[123] for developing validation methods.

4.2.2.3. NEW METHODS FOR WORKING WITH MENTAL DATABASES

Mental databases are most often used in SDM for conceptualization. Considering what can be attained with improved conceptualization methods, Forrester cautions that "rules … that assure description of a successful model" are elusive across applied and scientific domains[120] (p. 12). **The methods presented here provide a grounded approach to developing and validating the dynamic hypothesis as one progresses** and in so doing they *improve the chances of developing a useful* model.

The system dynamics standard methodology comprises a wide variety of methods and is aptly described as a *toolbox*[161], not just a single tool. Following Lane's advice to document my process and think theoretically about it[145], I present a table that lists additional tools for the SDM toolbox (Table 4.7) followed by my reflection on how these methods relate to expectations for methods development in system dynamics. These tools are designed for: study design, model development, model validation, and meta-level validation. The columns in the table identify each method by name, its use and its purpose.

These methods were created in accordance with the guidance of system dynamicists with regard to developing new methods. For model development, Morecroft recommends that "new structuring aids would complement, not contradict, the existing principles for formulation"[108] (p. 287). Foremost among these for me was Forrester's rule that model conceptualization methods should be "guided toward the [later steps of] equation writing and simulation" and avoid giving the impression that conceptualization alone can be used to diagnose or treat problems[120] (p. 13). While using CLDs for conceptualization is considered a best practice by some, there is still disagreement among experts on this matter[121]. This is due to their limited ability to organize information about decision processes[155] and the potential they have to mislead users about stock-flow relationships[251] along with other limitations[156, 252]. Given this, I use them strictly where they are best suited, to “portray system linkages”[155] (p. 22). Despite these limitations, I found CLDs to be a very useful tool, because their limited use of elements of structure[156] makes them accessible to a novice with limited access to formal system dynamics training, like myself. They worked as expected by enabling me to produce “quick wins at the start of [the] study” [156] and by helping me to conceptualize the system structure which I needed for formulating my simulation model.

For model validation, Barlas recommends that new methods balance being “formal, rigorous validation procedures suitable for system dynamics models” with an awareness “of the limitations inherent in such formal procedures”[123] (p. 204)[67][105]. Following this guidance permitted me to design and apply my new model validation methods and to contribute to system dynamics theory of validation (see previous section).

My methods are designed to be used in concert with SDM standard methods and to be seamlessly integrated into the work of interdisciplinary research teams. **In so doing, my methods can help researchers to more closely appreciate the situation and influence participants while minimizing coercion by recognizing and addressing social realities**[114] (p. 6)[105]**.**

[67] It is for purposes of gaining this awareness that Barlas recommends reading Lane (1995).

Table 4.7 New Methods which Enhance the Use of Mental Databases in System Dynamics

General use	Method	Specific use	Purpose
Research Design	*Cross validation strategy for SDM saturation*	Creates a *model development set*, *model validation set*, and saturation reserve, (using purposive sampling, over-sampling and partitioning of interviews)	To permit more rigorous use of mental data for model development and validation
Model Development	*CLD Combination*	Provides a structured way of combining individuals' CLDs into a single CLD with the potential of being a Shared Mental Model	To resolve differences in perspective using a simple set of rules
Model Validation	*SMM-S (Shared Mental Model Saturation)*	Tests the extent to which a model combining multiple perspectives (from individuals or different contexts, i.e., the output of CLD Combination) represents shared understanding (i.e., is a Shared Mental Model) and has adequate boundaries for its purpose	To confirm a CLD to be a Shared Mental Model
	CM-S (Conceptual Model Saturation)	Tests the extent to which a model of shared understanding can be generalized outside its original context and thereby be considered a Conceptual Model	To confirm a Shared Mental Model to be a Conceptual Model
	SIM-S (Simulation Model Saturation)	Tests the extent to which a simulation model represents the structure and behavior of scenarios in mental data	To confirm a simulation model to be *ready* for model analysis and policy analysis
Meta-level Validation	*SD-S (System Dynamics Saturation)*	Tests the extent to which SDM is an appropriate way of studying the situation	To enable others to better understand, evaluate, reproduce and extend studies
	Data Suitability	Reflects on and reports on crucial aspects of any research project	
	Methods Suitability		
	Stakeholder Dialogue Suitability		
NOTE: Others have explored similar territory. For example, Luna-Reyes *et al.* used interviews to validate the relationships in a CLD[253] and Von Raesfeld used rigorous coding and qualitative data analysis to explain model structure (as with the Causal RIQ) and to communicate modeling conclusions using stories, while simultaneously raising the validity/generalizability[254].			

I will now describe how these formal methods enhance the way system dynamicists can use mental databases in their research. CLD Combination integrates diverse perspectives and leads one to believe that the resulting model represents shared understanding[68]. Next, aspects of this assumption are tested: SMM-S tests the ways in which the resulting model represents the shared cognition of the groups who produced it. CM-S tests this for a group in a similar context – affirming the shared understanding and demonstrating the potential for generalizability beyond the original context. Finally, the assumption that the Simulation Model represents diverse perspectives, in the form of dynamic hypotheses, is tested in SIM-S. Results

[68] In other words, that the resulting cognitive structure could be obtained from any individual when specifically and appropriately probed. This does not mean that all participants explicitly identify the same structural elements but that, as practitioners with deep experience in PCT, they have access to the entire model even though they only drew on parts of it during the semi-structured interview.

from all the methods can be used to communicate with stakeholders to build shared understanding in the group regarding the model content and policy recommendations.

How did these novel methods improve my research? In many ways, the methods made the research possible. When I began this research, I had formed a firm belief in the importance of my *dynamic hypothesis* (i.e., the *tensions* described in Appendix C and Section 3.5.6.3.1) from my experience observing clinics, journaling and conducting interviews, but only a vague idea about its structure or how it influenced important system behaviors. CLD Combination provided me with a way to *simulate* the Sensemaking process of negotiating an intersubjective meaning out of equal individual meanings[255]. SMM-S then allowed me to test if this process was likely to have reached a conclusive end (i.e., reached saturation) in a Shared Mental Model or not. CM-S then permitted me to return to the raw interview transcript data to test the extent to which this model was comprehensive of a separate set of individuals' understandings of system structure. Next, SIM-S permitted me to consider the match between the behavior of my simulation model of PCT with individuals' mental models of the structure-behavior link, or their dynamic hypotheses. Finally, the meta-level validation methods permitted me the opportunity to document this process, to explore its potential acceptability and the potential acceptability of my findings.

Model building is an iterative process and requires methods which can withstand the demands of "repeated recycling" back to prior steps to answer questions that arise later in the process[120] (p. 13). My validation tests repeatedly bring the modeler back to *raw data* on how stakeholders describe the system. When a test is failed, the methods show the modeler what action to take so as to continue progressing.

The methods I developed were created for SDM, inside the field's theory, and are a creative way of using qualitative data in an SDM study for conceptualization and validation. This creative approach is in line with best practice recommendations in the field for model conceptualization[121].

System dynamicists paint a bleak picture for novices' ability to adopt SDM. Peterson, Eberlein and Richmond describe novices in the early phases of modeling as *leaping blind over an abyss*[111, 112]. System Dynamics founder Jay W. Forrester goes further, describing how, despite "multitudes of papers" using SDM, they "reveal little about how the theories [(i.e., models)] came into being" (p.12) because conceptualization "has rested heavily on … 'canned' models … apprenticeship … [and] trial and error" (p. 13)[120]. It is a problem of both methods *and practice*, where SDM novices (including stakeholders) see the qualitative model conceptualization phase of research as "*baffling*"[120] (p. 13 emphasis added).

The textbook advice is to:

> "Get a preliminary model working as soon as possible. Add detail only as necessary. Develop a working simulation model as soon as possible. Don't try to develop a *comprehensive* conceptual model prior to the development of a simulation model. Conceptual models are only hypotheses and must be tested. Formalization and simulation *often uncover flaws* in conceptual maps and lead to improved understanding. The results of simulation experiments inform conceptual understanding and help build confidence in the results. *Early results* provide immediate value to clients and justify continued investment of their time."[113] (p. 81, emphasis added).

I followed this advice and started building simulation models soon after the scoping study was completed. I wrestled both with how to add structure only as necessary and how to reconcile differences across participants. I found myself having false starts on the simulation side, so I kept refining my conceptual understanding, digging deeper into the qualitative data. When I returned to simulation, I struggled to set simulation model boundaries until I had built up enough confidence in my conceptual model.

Qualitative work continued to progress as I built the Simulation Model and, when I got to building the willingness sector (Section 3.4.2.6), the Conceptual Model was key to understanding all the pieces that needed to influence that willingness in the simulation.

An important boundary issue I had struggled with was the lack of a decision-making role for patients in the Conceptual Model. I had a variable in the Conceptual Model for the clinician-patient relationship ("Clinician develops relationship with patient"), but CM-S did not confirm or disconfirm this variable or the link. This irked me because my understanding of operational thinking[179] required that patients be *in* the simulation model – not to mention the fact that I was modeling the *Patient-Centered* Medical Home.

As I progressed through the *standard method* simulation, I uncovered an issue with the expected dynamics: when paid a fixed salary, clinicians did not react to being behind schedule because their salary was not impacted. This indicated that a feedback was missing as being behind would harm clinicians' relationships with patients.

My previous qualitative work helped me to improve *my understanding* and figure out the missing link. I turned to the Conceptual Model and noticed that the link between "Capacity" and "Clinician develops relationship with patient" tells a similar story to the chain of links between "Capacity" and "Visits on schedule" but that the latter also closes a feedback loop. I used that understanding to inform the structure for incorporating the clinician-patient relationship into the simulation model tests ("Behavior Boundary Adequacy/Structure Sensitivity" see Section 3.5.5.1.9).

Care teams receive regular reports from patient satisfaction surveys, which include an estimate of their wait times. Clinicians' may use these estimates as way of gaging schedule pressure, which they may have a hard time seeing over long time intervals. These surveys were mentioned in interviews, but never involving causal relationships. That is why they were not explicitly in the conceptual or simulation models until now.

Addressing these model formulation issues required me to have confidence in the structures I was introducing while also allowing me to be creative. I based my intuition on the Conceptual Model and figured out how to operationalize it using the grammar and existing puzzle pieces in the Simulation Model. The *imprecision* of the Conceptual Model complemented the *precision* of the Simulation Model – it was my rope bridge across the *abyss*.

Regarding the textbook advice: of the qualitative models I developed, the final Theoretical Model is the only one I would consider to be *comprehensive* and it would not have been available to me without simulation. In that sense, this is sound advice (for getting to "z"). My first models I had any confidence in were CLDs of individual participant's transcripts. I was just as confident that they were grounded in data as I was that they were flawed – each identified only pieces of the system (only a portion of its causal links and variables).

Where I depart from the advice above is in the *flaw-resolving impact* of formalization and simulation at this early stage (doing formalization and simulation from the very beginning, at "a"). I found the greatest flaw-resolving impact from rigorously-interpreting the mental data. Bringing the participant CLDs together – *after pruning* to a tractable size – provided a more complete picture of the causal mechanisms behind observed tensions in the system. Rigorously testing this picture against validation-set data, and engaging in stakeholder dialogue around it, improved it even more and helped me to feel confident that I had captured a hypothesis worth testing quantitatively – a *conceptual model*. From here, I heartily support the next piece of advice: model validation tests worked just as described.

Finally, regarding the concluding advice on producing *early results*, Graham & Els[203] argue that this depends and that quality is also very important, as follows. Early results are a *double-edged sword*. In some contexts they initiate virtuous cycles, and vicious ones in others. At the same time, they argue that the *quality* of work by novices can make or break the field and that quality rests heavily on obtaining a "prior understanding of feedback dynamics" before simulating (p. 3).

My methods are accessible to novices and represent a formalized process for obtaining this understanding in higher-quality qualitative models and grounded simulation models. It is my hope that they make "becoming truly skilled" in SDM easier in the future for novices like myself to accomplish[203] (p. 6). At the same time, my methods also provide opportunities for quick wins along the way while also setting expectations for quality that can be met (by transparently indicating that models and insights generated are for the modeling and validation step that produced them).

4.2.2.4. COGNITIVE LIMITATIONS

It has already been established that cognitive limitations are an important part of system dynamics theory. These limitations are described by Lane as follows:

> "People—we—have limited cognitive capacity. We tend to think that for each effect there is a single cause. We assume that causal chains are short. We do not think in terms of feedback effects. We find it hard to accept, even see, information we do not expect. When we try to explain things, we are satisfied the moment we feel we have found a convincing cause for something and we stop searching for more complex explanations. We try to confirm our own explanations rather than challenge them."[256] (p. 634)

This dissertation **provides additional clarity regarding the differential perception of the elements of structure.** Groesser & Schaffernicht provide the following comprehensive list of the elements of structure of mental models of dynamic systems: "variables (stocks, flows, intermediate variables), causal links between the variables, link polarities, delays in the causal relationships, linear and nonlinear relations between variables, feedback loops, and the polarity of feedback loops" and loop dominance[180] (p. 63).

I coded the validation set interviews for *system dynamics-related items*. These are cases where I interpret a participant to explicitly describe issues of particular relevance to system dynamics (see section 4.2.2.1 above). I distinguish between statements where a participant describes such an item, but does not mention it explicitly (what I call *indirect perception*) and ones where the item is described explicitly (*direct perception*).

Direct perception of time delays was somewhat uncommon in the model-validation-set interviews: found in roughly 6% of causal statements. These statements often tied delays to negative emotions, especially for clinicians.

Direct **perception of feedback** was even rarer with roughly 0.6% of causal statements. These statements were worded awkwardly, and participants showed an otherwise meager degree of systems thinking. Interpreting these quotes uncovered a sequence in the phases ascribed to the statements, offering some evidence for perception of **loop dominance.**

Indirect perception was much more common for feedback than for delays.

System dynamics theory of causality correctly assumes **perception of stocks.** Upon review of my simulation model, I modeled a majority of variables in the Conceptual Model as stocks. Nevertheless, it was my experience that participants did *not distinguish* variables as being stocks – as distinct from flows or intermediate variables – even when they obviously were (e.g., *Hiring new MAs*, and *MA Retention* are the flows impacting the stock of MAs, *Capacity*).

Finally, **perception of non-linearity** was also extremely rare, primarily expressed as gear metaphors alongside perceptions of information feedback.

I draw two conclusions from this evidence. First, the **system dynamics-related elements of system structure** are *not* readily perceived. Second, albeit not *readily*, **delays** are significantly *easier* to perceive than feedback. This result should not come as a surprise, given the literature on cognitive limitations and the elements of system structure[133, 134] and on systems thinking[181, 182]. Thus, one contribution is to provide documentation of limited cognition.

As I reflect on these findings in relation to relevant current debates in the literature, I also have four recommendations.

1. **In response to the suggestion by Groesser and Schaffernicht[180] that the elements of structure be named in the definition of mental models of dynamic systems, I argue that *they should not be named in the definition*.** These elements are important in reality and are a core part of our theory, but are not closely aligned with the way people talk. Because of their relative paucity in mental data, adding these elements explicitly to Dolye & Ford's definition of mental models[118] would be detrimental because it could lead novice SDM users to the conclusion that SDM-type mental models are not *ubiquitous* and taken for granted, but only belong to people who can directly perceive the full suite of the elements of structure (i.e., nobody). Because Doyle & Ford's definition focuses on the way people think, it was very helpful in pinpointing the things that I should look for in answering the question: does my target group see SDM as a useful way of addressing the issue?
2. **In response to the suggestion by Groesser and Schaffernicht[180] that the design rules for SDM *diagrams* be revised to include significant non-linear relationships, I propose that this would be most clear with Policy Structure Diagrams[113, 114].** I

used a System Policy-Structure Diagram to show where all of the decision functions[69] are located in my Theoretical Model, including those with significant nonlinear relationships. Modifying it to distinguish them would be a relatively minor change.

3. **Because of their complementary features, researchers using CLDs for system conceptualization will benefit from using System Policy-Structure Diagrams (e.g., my Theoretical Model) to summarize their Simulation Models.** There are non-trivial trade-offs between diagrams with more and less complex elements of structure: showing conserved stock-flow accumulation makes it difficult to show interacting feedback loops and vice versa[156]. Considering my findings, it appears that CLDs utilize the elements which are *most readily perceived*, an added rationale for their use in conceptualization[121, 203]. Using a CLD for the Conceptual Model, the research story starts out with an understandable presentation of *the feedback dynamics* which reflects both the data and *the modeler's* understanding at that time. On the other hand, I found that stock and flow diagrams (i.e., System Structure Diagrams and Policy Structure Diagrams) make visible elements that *are less directly perceptible* and were more useful for exposition of the Simulation Model. Using a hybrid System Policy-Structure Diagram (i.e., the Theoretical Model) to represent the Simulation Model concludes the research story by showing how decisions interact with the system infrastructure. This process helped me to see the improvement that simulation helped me to make in my understanding and that is why I believe my *whole* research *story* is best told showing both model types as book-ends.
4. **When one's goal in eliciting mental models is to empirically establish the complex elements of structure, formal instruments must be used. Otherwise, semi-structured interviews should suffice.** This is because the more readily-perceived elements of structure, those in CLDs, can be *readily empirically-elicited* from mental data like my interview transcripts while the more complex elements can be *empirically-elicited only with* formal instruments (e.g., for nonlinear functions[257] or via specific probes in semi-structured interviews[107] (p.58, 103, 128) and metaphors in facilitated workshops[112, 258] (for stocks and flows)) and otherwise are best *inferred* using operational thinking[157, 179].

As SDM research uses the methods proposed here, better evidence could emerge about the extent to which these elements of structure are perceived and what it means for SDM.

[69] The term decision function is used interchangeably with the terms policy function and operating policies in system dynamics and in this dissertation.

4.2.3. CONTRIBUTIONS TO SYSTEM DYNAMICS THEORY – PARADIGM ISSUES

In Chapter 1, I identified three outstanding methodological issues for MMHSR. I also found that these issues apply to system dynamics, although to different degrees. Here, I consider how the research process I developed relates to the outstanding system dynamics issue of *paradigm issues*. Such reflections are not common. As such, this section includes several contributions to system dynamics theory – for the *niche* of reflecting on the theoretical implications of one's research process.

My contributions to paradigm issues rely on the work of several philosophers. Section 4.2.3.1 presents my synthesis of the philosophical works of Schumacher, Wilber and Stange on the Four Quadrants of Knowledge – the framework I use to discuss paradigm issues. This framework is expanded by synthesizing it with the philosophical works of Schultz, Hatch, Gioia and Pitre on research approaching multiple paradigms.

Section 4.2.3.2 presents how system dynamics theory informed my multiple paradigm research approach[70]. Following this, I present frameworks based on the above synthesis which show the assumptions used in system dynamics theory across the Four Quadrants of Knowledge as well as their implications for research and for models. Next, I identify system dynamics best practices as forming a paradigm-crossing strategy which I explicitly map to the Four Quadrants of Knowledge framework. This section concludes by presenting three alternatives that researchers using system dynamics have for dealing with multiple paradigms.

With this background in place, I have a sufficiently deep basis for a discussion of my contribution to paradigm issues in system dynamics. Section 4.2.3.3 locates the methodological contribution of this dissertation in terms of paradigm issues: an interplay paradigm-crossing strategy, in other words an *integrative methodology*. Section 4.2.3.4 presents my work on PCT as a case study of research using this methodology.

Taken as a whole, these contributions build understanding of the system dynamics field's potential as an integrative methodology and paradigm.

[70] Note: I used the system dynamics theory on causality and validation as the second-order theoretical concept (i.e., the mechanism) for crossing the transition zones between paradigms.

4.2.3.1. PARADIGM INCOMMENSURABILITY

The modern concept of the *paradigm* can be traced back to the work of philosopher Thomas Kuhn[259]. There are two aspects to his concept: a scientific worldview and a community of practitioners[113, 260, 261]. To be fully-fledged, a paradigm must have a theory, methods and a body of empirical cases which are shared and used by a "self-consistent communit[y] of like-minded scientists"[113] (p.849)[260].

Paradigm incommensurability exists because the communities holding a worldview make it hard for their members to associate with other communities[262], arguing that the assumptions about reality embedded in any two worldviews are mutually incompatible. In other words, incommensurability arises as a research community strives to maintain coherence by excluding certain ideas; for example, by using certain terms and definitions in what Deetz calls "discursive practices of unity and separation"[263] (p. 193). That said, *paradigm commensurability* is possible to the extent that paradigms favor inter-paradigmatic dialogue (i.e., allowing association with ideas from or researchers using other paradigms) as a means to solving problems in the real world (see Deetz[263]).

Kuhn's goal was understanding paradigm shifts across the history of science, so he did not offer a way to *compare and contrast* current paradigms. Two additional philosophers add clarity here. Schumacher proposed a philosophical map with four fields of human knowledge[264] into which all paradigms can be placed. Wilber organizes these fields into the *Four Quadrants of Knowledge* framework[265].

This framework permits considering similarities and differences in assumptions across paradigms; and in so doing it maps out where methodological pitfalls associated with paradigm incommensurability as described by Sale *et al.* (see Sale *et al.*[99]) may occur.

Across a series of articles, Stange and colleagues use these insights to argue for a new integrative paradigm for research in primary care to better understand its integrative function in health systems[2, 266-268]. This sparked for me the idea to also use this framework for a new integrative paradigm for research as I reflected on my case (primary care transformation), my newly-developed methods, and my potential theoretical contributions.

Figure 4.4 below **presents my synthesis of the work of** Schumacher, Wilber and Stange on the *Four Quadrants of Knowledge*. This framework is useful as it provides a map with which to engage as one navigates around the various ways of knowing and, in so doing, confronts paradigm incommensurability in the process.

Figure 4.4 The *Four Quadrants of Knowledge* Framework

	Inner Reality	Outer Reality
Individual	**Quadrant 1** the intention (self-awareness)	**Quadrant 4** a behavior in reality (matter, experiments)
Collective	**Quadrant 2** the culture (collective consciousness, context)	**Quadrant 3** the social (living systems, concrete social structures)

Note: Definitions are summarized in Edwards[269] (p. 272). They are based on the work of Wilber[265] (p.127), and Schumacher[264] (p. 24, 28). They also contain words from Stange *et al.*[2] (p. 289).

Separately, Burrell & Morgan (organizational theorists) developed a four quadrant framework[71] to classify sociological worldviews and theorists[262]. Although this framework is more well-known, I prefer the *Four Quadrants of Knowledge* framework because it considers philosophy of science more broadly and deeply[72][104] (p. 522)[102, 270-272], achieving the same end of providing a way to talk about paradigm issues without the same[264] "almost hegemonic"[263] (p. 191)[73] means and thus permitting me to describe my contributions at a high level of abstraction (one above the milieu of sociological paradigms). Furthermore, as it was used to propose an integrative paradigm for primary care, couching my contributions in this framework makes them more accessible and relevant to this audience.

Before considering how one might intentionally develop/use an integrative paradigm, one must consider how to conduct research involving multiple paradigms. Organizational theorists Shultz and Hatch identified three mutually-exclusive alternatives for doing this[273]. They are:

---

[71] Burrell & Morgan also consider inner and outer reality for their columns. However, they distinguish across rows using a view of society (*radical change/ regulation) dimension*.

[72] Lane (1994-1999) uses Burrell & Morgan's framework, but advises integrative research in SDM to adopt a *Four Quadrants of Knowledge*-style framework with Burrell & Morgan's subjective/objective dimension (inner/outer reality) and an additional agent/structure dimension (individual/collective) (p. 522 in Lane (1999)). In later works (2001-2008), Lane uses a figure with mechanisms involving integrative social theories that can map onto such a framework, but is not designed to do so.

[73] In his paper, Deetz argues that the dimensions of a framework (e.g., inner/outer reality and individual/collective in Figure 4.4 above) have the potential to be useful to the extent that they generate dialogue. Because Burrell & Morgan's framework uses the terms of functionalism (e.g., *radical*) to describe the other paradigms, it therefore privileges functionalism. At the same time, its purpose is paradigm classification. For Deetz, these contextual elements stifle dialogue generally and make it hard for researchers operating beyond functionalism to represent their work clearly and correctly.

- *Incommensurability* – using only one paradigm (ignore other paradigms)
- *Pragmatic/integrationist* – using multiple paradigms and not prioritizing theoretical issues
- *Paradigm-crossing/integrative*– intentionally and explicitly using multiple paradigms and addressing theoretical issues in the process

Table 4.8 below presents the strategies for paradigm-crossing, where the strategy employed is determined by (1) the assumption regarding the permeability of paradigm boundaries and (2) the researcher's interest in looking at similarities and/or difference between paradigms. Unlike the other strategies that either focus on similarities or differences, *interplay* (top right) relies on the "maintenance of tension between contrasts and connections"[273] (p. 534).

Table 4.8 Paradigm Crossing Strategies

		Interested in similarities or differences, or both?		
		Similarities	**Differences**	**Both**
Boundaries are permeable?	**Yes**	Bridging		Interplay
	No	Sequential	Parallel	

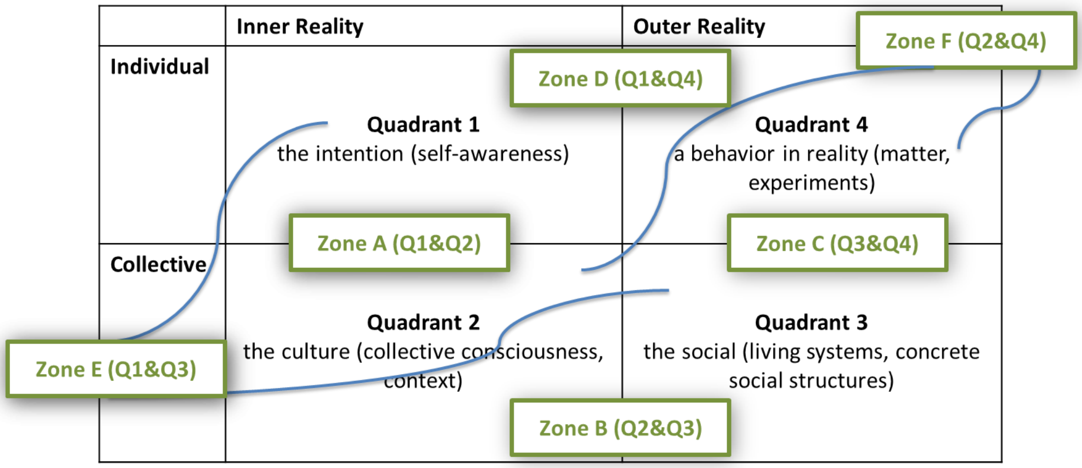
NOTE: This table is based on the textual description provided by Shultz & Hatch.

For research assuming permeable paradigm boundaries, Shultz and Hatch refer to organizational theorists Gioia and Pitre's concept of the transition zone[274][74]. This is a conceptual area where "paradigms become indistinguishable to the researcher" as similarities between paradigms are emphasized[273] (p. 534).

Figure 4.5 below **presents my synthesis of the work of** Figure 4.4 above as well as Schultz, Hatch, Gioia and Pitre. It is useful because it presents the transition zones (green boxes) as applied to the *Four Quadrants of Knowledge* framework. Blue curvy lines are used to point out transition zones located across from each other on opposite corners of the framework.

[74] Gioia & Pitre describe the theory behind four transition zones in terms of the Burrell & Morgan Framework. I adapt this for the *Four Quadrants of Knowledge* framework.

Figure 4.5 Transition Zones and More Basic 4 Quadrants of Knowledge Framework



Each transition zone requires identifying how the quadrants that each is linking are similar and explaining how the link is made. Gioia & Pitre refer to the linking mechanism as a *second-order concept;* something to "[make] related or analogous concepts become more evident" inside the transition zone[274] (p. 592). The theory chosen provides a lens that focuses the researcher on similarities across the paradigms being crossed.

I have here described the theoretical context that I relied on for considering how my work relates to paradigm issues. This is a contribution to theory as these various schools of thought were not previously brought together**.** By integrating these concepts, I have gained insight into a theoretical grounding for the SD paradigm as an integrative paradigm (discussed in next section).

4.2.3.2. PARADIGM-CROSSING IN SYSTEM DYNAMICS

Paradigm-crossing in system dynamics requires clarifying the integrative aspects of system dynamics theory. This clarity can conceivably come from synthesizing system dynamics theory with existing integrative social theories or building within existing system dynamics theory. Lane has warned that the first path is "a difficult task" for two reasons: first, this literature suffers from "abstruseness of style and imprecision of terms" and second, it is biased "towards the subjective aspects of social theory"[102] (p. 300). Indeed, previous efforts to delve into this subject[275] have been poorly-received (see Pruyt[276]) for both of these reasons.

Nevertheless, organizational theorists point to such integrative theories as offering the *second-order concepts* necessary for paradigm-crossing[273, 274]. Considering its potential as an integrative theory, as laid out by Lane[101, 102], **I use system dynamics theory on causality and validation as the *second-order theoretical concept*** (i.e., the mechanism[75][274]) for bridging the transition zones between paradigms (Table 4.9 below). These two components of system dynamics theory involve its primary purpose (adding precision to causal-descriptive theories through modeling) and its unique contribution (validation of the structure-behavior link).

Doing so has allowed me **to express some of the assumptions of system dynamics research within each of the *Four Quadrants of Knowledge*** (Figure 4.6 below) as well as to explore **what the transition zones would look like** (Figure 4.7 below); thereby testing the waters for SDM's potential as an integrative methodology.

When I first considered how system dynamics practice might map to the *Four Quadrants of Knowledge*, it was unclear what to put in Quadrant 4 and Zone C. The previous mapping of system dynamics practices by Lane did not locate any pre-existing work to the upper-right quadrant[76] of his framework[104]. For Stange *et al.*, research in Quadrant 4 "seeks to isolate a phenomenon from its context, so that it can be understood in its purest, most singular form" (p. 289) and is done best with methods exploring issues "over time"[2] (p.290).

This corresponds well to Forrester's concept for simulation (below):

> "*Controlled laboratory experiments* on [social] situations are now possible with computers to do the work required by mathematical *models that simulate the system* being studied. Unlike real life, *all conditions but one can be held constant* and a particular time-history repeated *to see the effect of the one condition that was changed*. Circumstances can be studied that might *seldom* be *encountered in the real world*. *Daring changes that might seem too risky* to try with an actual company can be investigated."[107] (p.43, emphasis added)

However, for Zone C, Stange's definition was less helpful. For Gioia and Pitre, the right-hand transition zone involves the application of "*activist values"* to transform approaches which serve systems into ones that serve people (i.e., radical change)[274] (p. 594). This could fit with Forrester's view that SDM "appeals to *activists*"[120] (p. 3) and the last two sentences in the quote above, but it fails to explain what happens in this zone.

[75] As above, something to "[make] related or analogous concepts become more evident" inside the transition zone (p. 592).
[76] Note: he was using the Burrell & Morgan framework.

With running-the-simulation now in Quadrant 4 (per Stange and Forrester just above), I conclude that Zone C involves establishing a link between stakeholders' dynamic hypotheses linking structure and behavior and the stories with numbers represented by the Simulation Model.

Table 4.9 System Dynamics Theory on Causality and Validation

The system dynamics theory on **causality** postulates that:

• "All systems [have] the same fundamental [elements of] structure of [stocks] and flows *structured* into feedback loops that cause all changes through time" (p. 10, emphasis added)[120]
  ○ People's mental models of dynamic systems contain the conceptual elements of this causal structure[118, 119]
    ▪ Which include: "variables (stocks, flows, intermediate variables), causal links between the variables, link polarities, delays in the causal relationships, linear and nonlinear relations between variables, feedback loops, and the polarity of feedback loops", loop dominance[180] (p. 63), time horizons and system boundaries[113] (p. 16)
    ▪ Which can be elicited from mental databases[116, 117];
    ▪ Nevertheless, cognitive limitations[109, 118, 119, 180, 256, 277] must be taken into account in developing simulation models (e.g., by using triangulation of perspectives[102, 113] (p. 187), *searching questions*[107] (p.58, 103, 128) and operational thinking[157, 179])
  ○ The ultimate purpose of SDM is to improve mental models[256, 278-280], thereby increasing the rate at which improvements are made to causal structure to improve system behavior[102, 120], or *learning*[107] (p.43-46)[113] (p. 14-19).

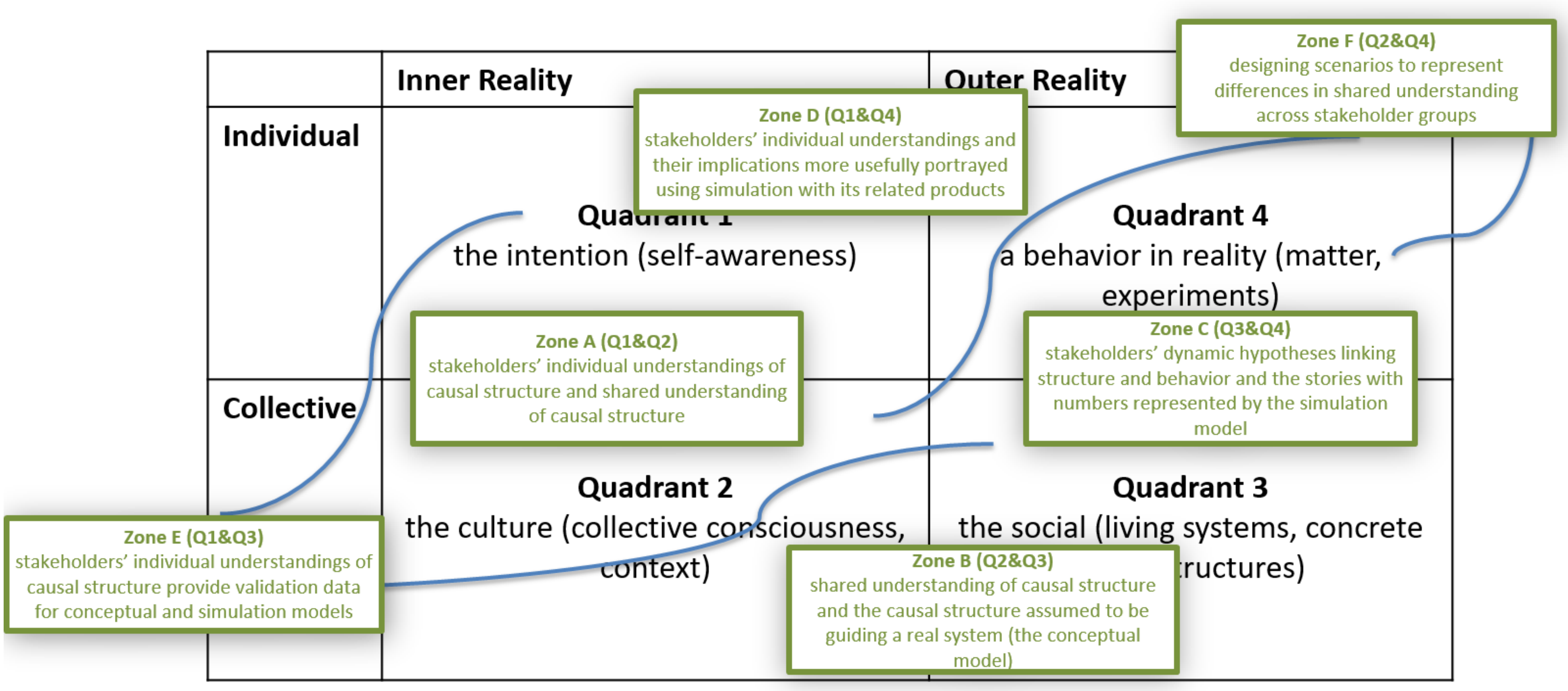
The system dynamics theory on **validation** postulates that:

• While "there is no way to prove [model] validity …. One can achieve … a degree of confidence in a model"[120] (p. 4), that it is both useful[10, 100, 123] and "adequate for the purpose under consideration"[120] (p. 4)
  ○ This can be achieved by using a framework of
    ▪ validation tests[105, 113, 123] (p. 869-889)
    ▪ and validity types[105]
    ▪ which is characterized by the "quite distinctive" principle[100] (p. 455) of *right behavior for the right reason*[107] (p. 117) implying a qualitative basis for model structure which results in
      • an emphasis on qualitative modeling/validation methods[107] which are empirically-rigorous[116, 117, 138, 139, 175] and
      • the principle that validation is a qualitative, judgmental process[107] (p. 129)[100, 160, 161]
    ▪ "by means of which the confidence in [the usefulness of] a model is gradually enhanced … [as] a matter of degree, never an absolute property" until reaching a threshold set by relevant stakeholders[161] (p. 176).

Figure 4.6 Assumptions of System Dynamics Research within Each of the *Four Quadrants of Knowledge*

	Inner Reality	**Outer Reality**
Individual	**Quadrant 1** the intention (self-awareness) – individuals perceive reality and process it through their individual understanding of it – so we conduct research in this quadrant **to understand their individual mental models**.	**Quadrant 4** a behavior in reality (matter, experiments) – behavior that can be seen empirically – so we conduct research in this quadrant **to use our understanding of *the physics* of the system to understand how an individual disturbance to system structure results in a particular response** (how a system behaves).
Collective	**Quadrant 2** the culture (collective consciousness, context) – groups perceive reality and process it through their context – so we conduct research in this quadrant **to understand shared understanding**.	**Quadrant 3** the social (living systems, concrete social structures) – structure of reality, having an exterior form that can be seen empirically – so we conduct research in this quadrant **to understand that structure.**

Figure 4.7 **Assumptions of System Dynamics Models within Each of the Six Transition Zones of the *Four Quadrants of Knowledge***



Forrester wrote the *Industrial Dynamics* book with the ambition of improving social science. In Chapter 1, I referenced Hoover & Donovan's definition of social science as that branch of *science* with the responsibility to answer "'*what can be done to* ***improve the human condition*** *– and what matters are* ***beyond our ability*** *to change?*'" (p. viii, emphasis added)[97] in arguing that the choice of methods effectively sets the boundary of what is in *our ability to change*. In Industrial Dynamics, Forrester's ambition was to expand this boundary[107] (p. 6) by means of a new "total viewpoint and discipline" (p. 115) whose value is demonstrated in a set of path-breaking empirical cases[107]. The book has been called a "construct paradigm"[281][p. 1037], likely because it contains the seeds for the three paradigm criteria referenced earlier: theory, methods, and empirical cases. It employs a nascent paradigm-crossing strategy to develop mathematical models from "verbal description, experience, field observations and such [numeric] data as are available"[107] (p. 9).

While most SDM practitioners ground their work in a social theory that can best be associated with functionalism[104], other paradigms have also been used[100, 104]. The prevailing view among system dynamics practitioners for dealing with multiple paradigms is pragmatic[104][77][282] or pluralist[116]; however, some argue that SDM research projects actually have the contours of an integrative paradigm in practice[101].

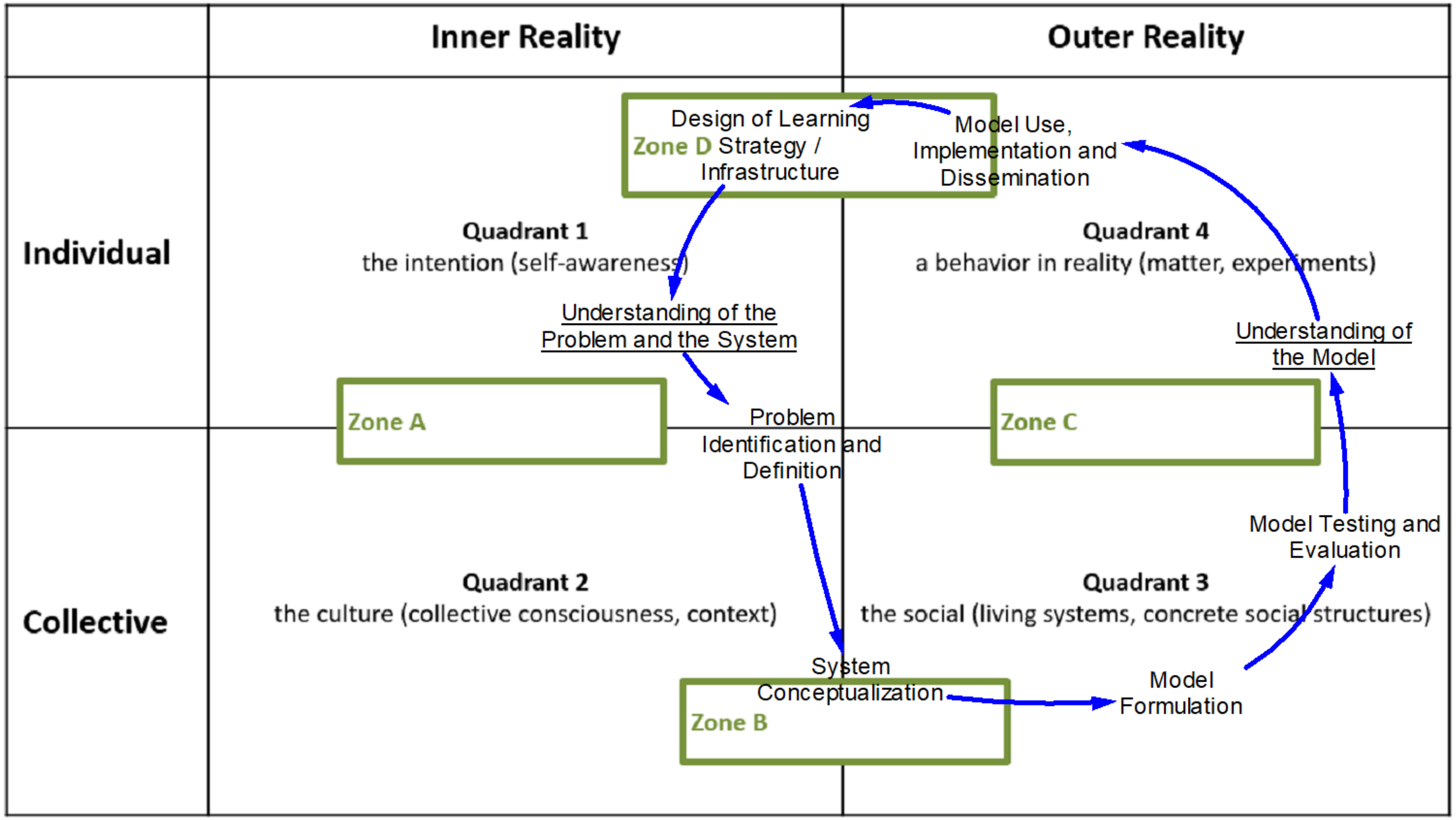
With the quadrants and zones made clear in the context of system dynamics, I am ready to consider system dynamics practice in terms of theory. To do so, I use the work of Martinez-Moyano & Richardson[121] which documents *best practices* for the field. They convened system dynamics masters to develop the theory of practice in use in system dynamics. While there was not unanimity among the masters on all points**, best practices were documented for achieving what is, in my view, a paradigm-crossing strategy**.

[77] This is distinct from Pragmatism (see Barton for more on this philosophy).

In Figure 4.8 below, I show my interpretation of **how a *best practices* system dynamics research project would map** to the *Four Quadrants of Knowledge framework*. Stages are unformatted and outputs are underlined. Thick arrows show the process flow of purposeful learning and thin arrows show the informal learning of researchers and stakeholders along the way. *Problem identification and definition* as well as *System Conceptualization* are intentionally placed along borders in the figure[78]. Which side of the border they lie on depends on the problem being studied. Problems with a very well-established history of rigorous system dynamics work (like many business contexts(see models in Warren[283]), may start in Quadrant 3.

[78] For the first, the best practice considers involving problem owners (plural) in a process focused on describing the problematic system behavior and looking at available data. For the second, the best practice is to engage with mental models of problem owners to understand their dynamic hypotheses and identify the stocks (nouns) that describe the system. As described, the first stage starts in Q1 and, with the second stage, ends up in Q3.

Figure 4.8 Best Practices in System Dynamics Modeling as they Cross Paradigms



NOTE: This figure is based off of Figure 2 in Martinez-Moyano & Richardson[121]; placement on the *Four Quadrants of Knowledge* and in relation to *transition zones* are my contributions). Zones E and F are not shown. Depending on how the researcher designs the methods for "Problem Identification and Definition", they may pass through Zone E or go from Quadrant 1 to 2 to 3, without crossing a transition zone. Zone F may be used in some "Model Formulation" approaches.

A paradigm-crossing strategy may be occurring in practice, with modelers performing the conceptual leaps around the Quadrants in iterative cycles, in their minds and with their stakeholders. However, *how* this occurs is rarely documented (with the lone exception of Yearworth & White[116]) or considered theoretically[145], perhaps producing what Lane sees as a gap between SDM's stated functionalist paradigmatic location and its use of an integrative methodology in practice[100, 104, 281].

My general thoughts regarding paradigms and system dynamics are these: **whether or not researchers are attempting to use system dynamics to break through paradigm incommensurability, they need to consciously engage with paradigm theory (as I have done) – to reflect on their approach to using multiple paradigms and to document how and why they made the choices they did.**

The best practices are clearly multi-paradigm (as shown in Figure 4.8 above). Thus any use of system dynamics should be multi-paradigm research. I argue that **SDM researchers choose one of these three alternatives below for dealing with multiple paradigms**, whether consciously or not. **I explicate each alternative and argue for what could be done to bring more theory to each approach**. Those having an affinity to:

- **Incommensurability** could instead consider the concept of "meta-paradigm crossing" (p. 595) where the "viewpoint beyond that of an individual paradigm" *(ibid.)* is, nonetheless, "*rooted in* a specific paradigm"[274] (p. 596, emphasis added)
    - This approach permits practitioners' mostly-functionalist research to use a bridging strategy of paradigm crossing.
- **A pragmatic/integrationist approach** could consider the paradigms which Pruyt argues are appropriate for various specific approaches to SDM[276].
    - This approach permits practitioners to continue being flexible regarding chosen methods; not prioritizing theoretical issues.
- **A paradigm-crossing/integrative approach** could consider adding to theory, methods and cases such that system dynamics earns its place as a "formal approach to dissolving dualisms"[104] (p. 521) with unique contributions to make to developing an integrative paradigm.
    - **This approach permits practitioners to** intentionally and explicitly using multiple paradigms and address theoretical issues in the process.

Fully Integrative paradigm-crossing strategies should pass through all four transition zones and enter, stand in, and then exit all *Four Quadrants of Knowledge*. When this is accomplished in multiple cases and documented at a theoretical level (a generalizable description of how this is done), only then can system dynamics address these dualisms and establish itself as an integrative paradigm.

Without this, claims of an integrative paradigm[284] and hopes of a paradigm shift based on the strength of one or another SDM process (see Richmond[112, 258]) will continue to be dismissed as under-theorized[104], while calls for developing the integrative-paradigm potential of system dynamics[102, 285] continue mostly[275] unheeded.

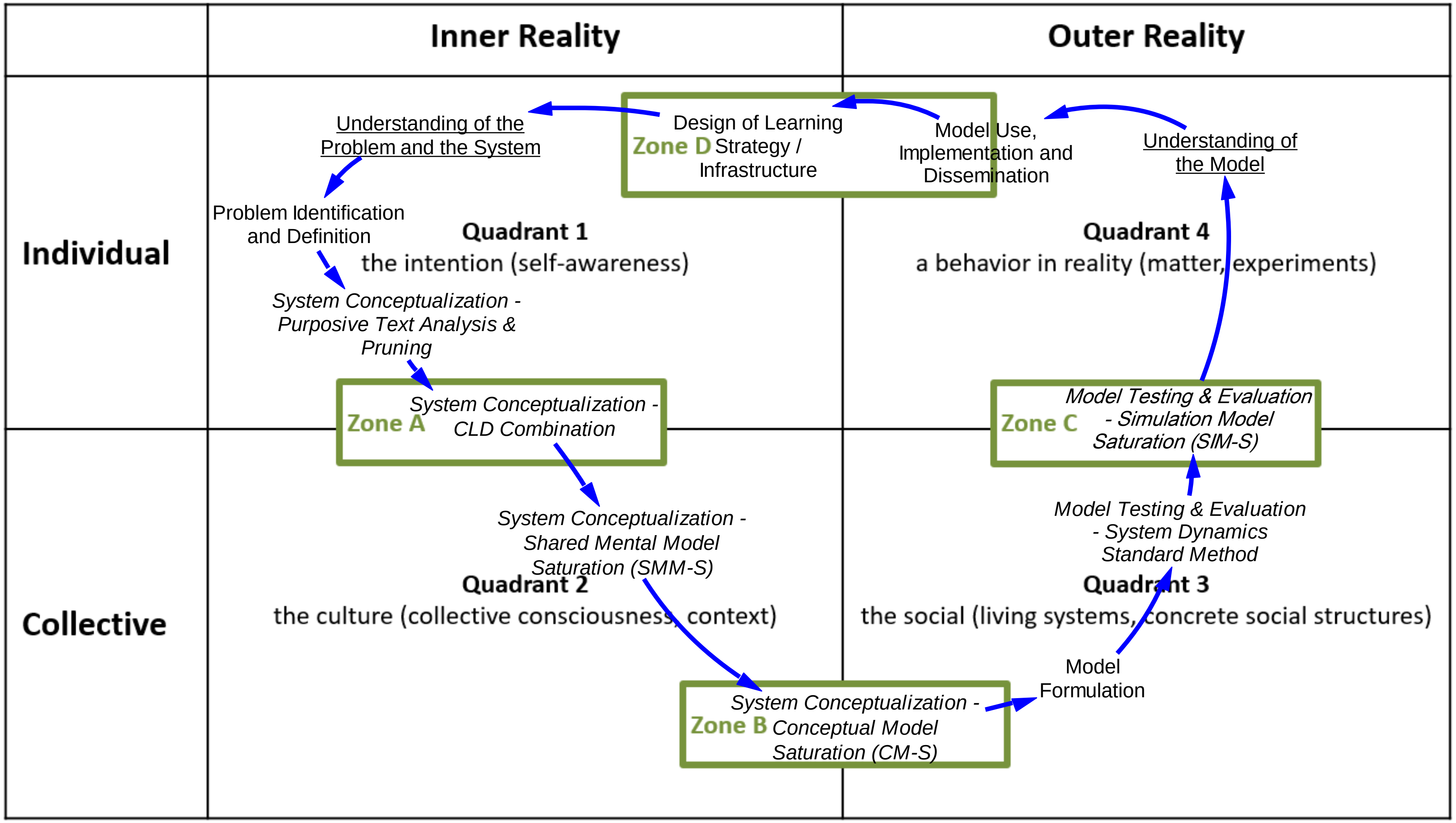
In the next section, I present the **use of the interplay paradigm-crossing strategy (with SDM theory on causation and validation as a second-order concept)** to enter, stand in, and then cross through each quadrant in the *Four Quadrants of Knowledge* framework. In so doing, I explicitly document and reflect on that crossing. Furthermore (above I presented the theory), next I reflect on the methods and case[79] that this dissertation presents to **strengthen system dynamics' position as an integrative paradigm.** To my knowledge, this is the first time that this has been done using SDM.

4.2.3.3. SYSTEM DYNAMICS AS AN INTEGRATIVE METHODOLOGY

Chapter 1 also referenced the use of system dynamics *in practice* as having an integrative paradigm, but lacking the necessary formal *integrative methodology*. While it may lack some of the necessary components of an integrative social theory[101, 102], system dynamics theory, as I have here demonstrated, provided me with the basis for developing methods to conduct research using an integrative methodology. These methods allowed me not only to stand inside each paradigm of the *Four Quadrants of Knowledge* as I built a better understanding of the problem being investigated but also to cross paradigms (and in so doing, to consider the similarities and differences among them). Figure 4.9 below **presents this integrative methodology** by modifying two of the six stages of the best practices in SDM(see Figure 4.8)[121] (System Conceptualization is split into four stages and Model Testing & Evaluation into two stages). The newly-developed methods are described further in the next section.

[79] As above, a fully-fledged paradigm has four pieces: theory, methods, cases and a community of practitioners. A reflection for the final piece is outside the scope of this dissertation.

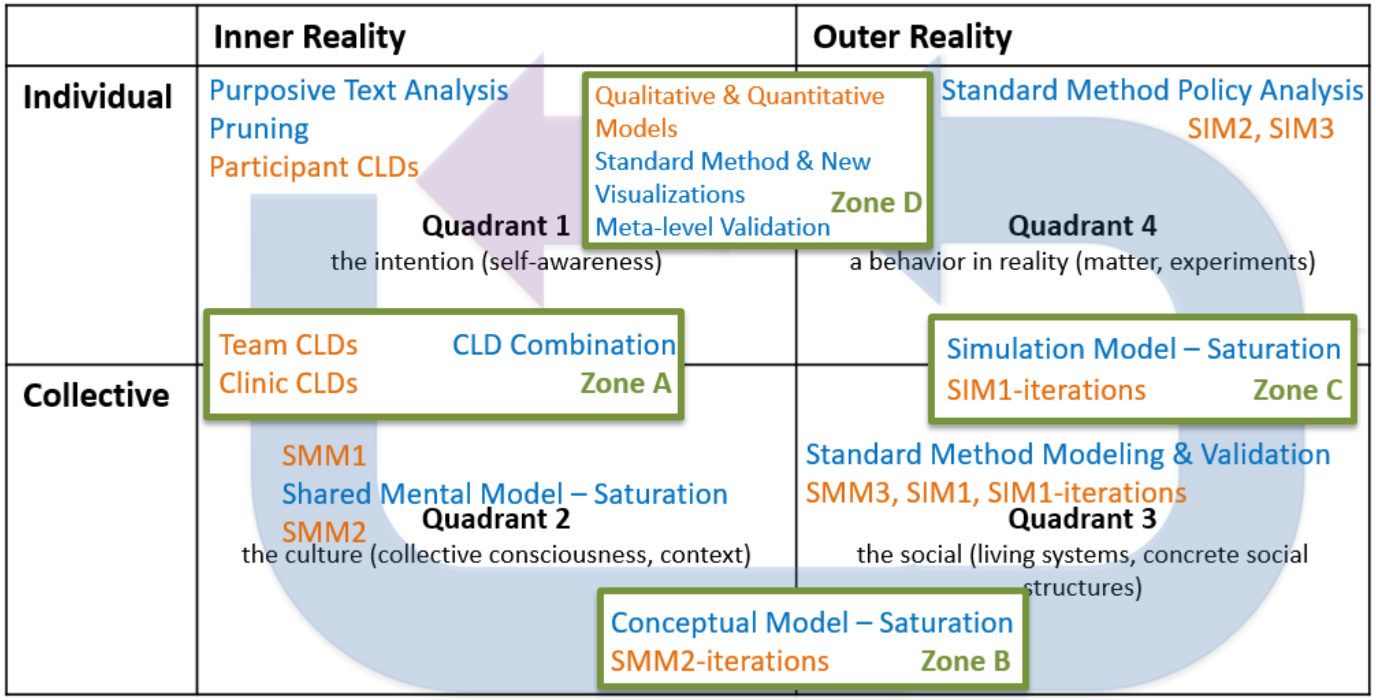
Figure 4.9 A Proposed Integrative Methodology for System Dynamics Modeling



NOTE: For simplicity, Zone E (crossing Quadrant 1 and 3) and Zone F (crossing Quadrant 2 and 4) are not displayed. For how Zone E is crossed, see Table 4.10. Zone F is not crossed.

Figure 4.10 presents how this strategy travels counter-clockwise around the *Four Quadrants of Knowledge framework*, passing through the four *transition zones*[274]. This path is prescribed by the use of SDM (see previous section). Interplay does not inherently have a sequential approach, rather it considers connections and contrasts from each standpoint with other standpoints. This strategy enters, stands in and then exits each paradigm (quadrant). Similarly, it enters and stands in each transition zone. However, before exiting either one, one also looks to other standpoints and takes stock of what is similar and different as one travels across the various paradigms. Visualizations are used for both similarities and differences.

Figure 4.10 Using Interplay in System Dynamics Modeling as a Paradigm Crossing Strategy



Model name	Participant CLDs	Team CLDs, Clinic CLDs	SMM1	SMM2	SMM3 / Conceptual Model	SIM1	SIM2	SIM3	Theoretical Model

Progression of Conceptual Model | Progression of Simulation Model

Note: Methods are in blue and model names are in orange. Arrows represent my research process. For a detailed description of how similarities and differences across quadrants and zones are treated, see Table 4.10. For simplicity, Zone E (crossing Quadrant 1 and 3) and Zone F (crossing Quadrant 2 and 4) are not displayed.

Having my problem statement and methods selection in hand, I begin in Quadrant 1 with mental data at the individual level and use Purposive Text Analysis to extract variables and relationships from the causal information in that data, producing individual CLDs. I then use Pruning to extract only the links which express an individual's perception of feedbacks and delays.

I then enter Zone A, the transition zone between *stakeholders' individual understandings of causal structure and shared understanding of causal structure*. In this space, I apply CLD Combination to aggregate individuals' mental models using their actual organizational groups. I treat each CLD equally and compare, contrast and then combine the models to create models which represent my estimate of each group's shared understanding in Team and Clinic CLDs. I exit the transition zone when I aggregate across clinics producing Shared Mental Model 1 (SMM1). SMM1 represents the shared understanding, grounded in individual and sub-group understandings.

In Quadrant 2, I validate SMM1 using SMM-S. This test checks the boundaries of the model and shows areas which are more and less understood/shared, producing SMM2.

I then enter Zone B, the transition zone between *shared understanding of causal structure and the causal structure assumed to be guiding a real system (the Conceptual Model).* In this space, I apply CM-S to double-check the structure of SMM2 with conceptual information and causal information contained in the model-validation-set interviews, mental data which was not used to develop SMM1. At the same time, this test checks the boundaries of the model and shows areas which are more and less understood/shared. Model revisions are made as called for in this method. I exit the transition zone when I conclude the CM-S Test has been passed, producing SMM3. Now called the Conceptual Model, SMM3 is a coherent, orderly framework of a problem made up of the variables and relationships within a system boundary[144].

Zone E, the transition zone between *stakeholders' individual understandings provide validation data for conceptual and simulation models,* also leads me to Quadrant 3. In this space, I analyze a subset of the *model validation set* interviews to create *Rigorously-Interpreted Quotations* (RIQs) and present them along with SMM3, as validation data.

In Quadrant 3, I use SMM3 as a blueprint for developing the simulation model (SIM1) using standard SDM methods for model development and validation. For model development, a key aspect was operational thinking[179]. In this quadrant, SDM validation looks both forward to simulation experiments and communication with stakeholders and backward to participants' mental models. Looking backward, I explore the relevance and consistency of the simulation model with inner reality. Also, I describe the simulation model using various diagram to communicate the contents of the various sectors in a way that is accessible to stakeholders. Looking forward, I explore its suitability to the purpose of building understanding of the problem and exploring plausible system improvements.

I then enter Zone C, the transition zone between *stakeholders' dynamic hypotheses linking structure and behavior and the stories with numbers represented by the simulation model*. In this space, I re-employ the causal information found in RIQs to verify that the structure and behavior found in the individual-level mental data can be well-represented by the simulation model (SIM-S). Revisions are made as called for in this method. I exit the transition zone when I conclude the SIM-S Test has been passed, producing SIM2.

In Quadrant 4[80], I create SIM3 by adding policy analysis structures to SIM2 (e.g.., on/off switches for different policy scenarios). I then simulate the effect of policy changes on the system, controlling for all other aspects (using standard SDM methods for policy analysis). These runs test the effects of policy in the context of the *physics* of the system[112, 157, 179]. Looking forward, I explore the role of diverse individual preferences under plausible system conditions.

I then enter Zone D, the transition zone involving *the emergence of individual understandings from exploring the system model and related products*. In this space , I employ the Simulation Model (SIM3), validation results and visualizations to communicate the full extent of the understanding obtained through this research for the purpose of improving readers' mental models and , it is hoped, their policy choices.

---

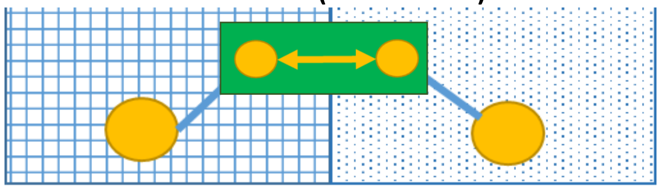
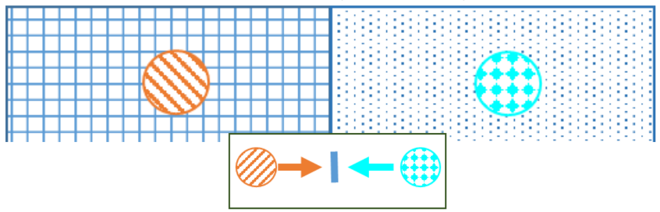
[80] Technically, there is another zone, Zone F between Quadrant 2 and Quadrant 4. I did not step foot in this zone.

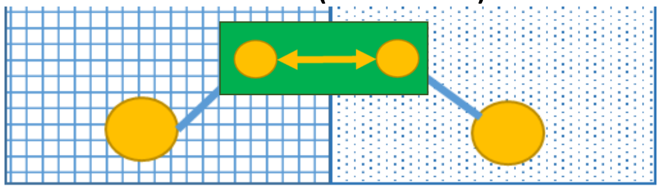
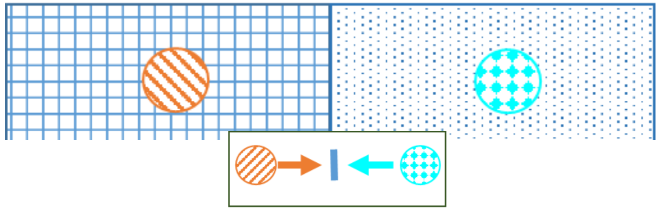
Table 4.10 below walks through this interplay paradigm crossing strategy. The first column identifies where I am standing in the *Four Quadrants of Knowledge framework*. The second column presents the model(s) in that location, as well as the main assumptions about the models[81] and the data and/or methods that modify (shown with an asterisk) those models to produce the resulting model. When in a quadrant, this is done using one cell. When in a transition zone, two cells are used (to show being in and then exiting the zone). The third column describes the lens that is used to view the differences in ways of knowing in the present location, either looking backward ("previous" is underlined) or forward ("next" is underlined). The final column identifies how the difference is visualized; when backward, it is concerned with relevance to and consistency with appreciate of the situation, when forward, it is concerned with the suitability to purpose. Zone F is grayed out as it was not visited. There is no text for differences when in Quadrant 1 as this is where I start my journey.

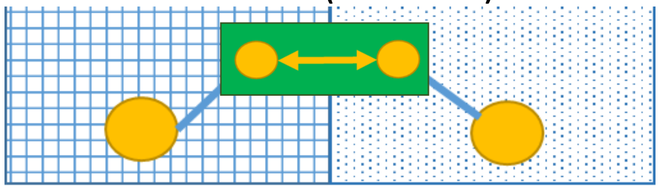
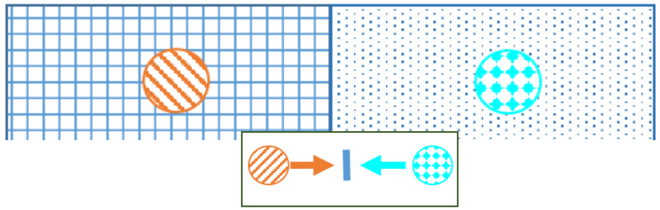
At this point, we are closing the loop in the theory building process[138] (see Section 2.2.2) as we navigate the four quadrants and build a *minor content theory*[82][103, 145] specific to the problem being studied.

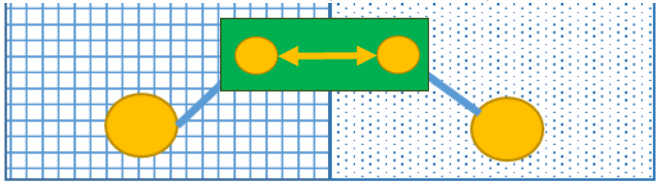
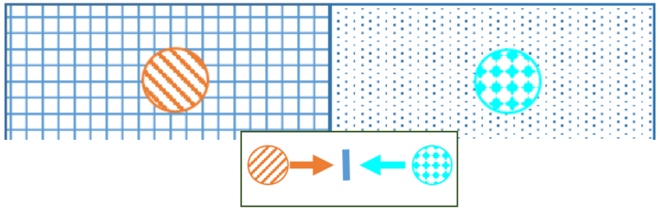
[81] SUM is used in Zone A to denote the assumption that the sum total of the mental data in one set of models is equivalent to that in the more aggregated models.
[82] A minor content theory is a dynamic hypothesis specific to the problem being studied – it identifies the causal relationships thought to bring about system behavior (see Section 2.2.2).

Table 4.10 SDM Interplay Paradigm Crossing Strategy

	Similarities (Connections) 		Differences (Contrasts) 	
Location	Model names = **The main assumptions** about models * *data and methods*		**What lens do we use to view the differences in ways of knowing relative to the previous location (type of visual)**	**Visualized aspect of differences (Called out via figures, tables or text)**
	In the transition zone	Exiting transition zone		
Quadrant 1	Individual Mental Map = A stakeholders' **individual understanding** of causal structure (all variables and links), of the system, for a problem Individual Mental Map * *pruning =* Individual CLD = **Dynamically-meaningful pieces** of maps (only feedbacks and delays)		This is where I started my journey.	
Zone A	SUM(Individual CLDs) = SUM(Team CLDs) = SUM(Clinic CLDs) =	Shared Mental Model #1 = sum of **perspectives** on causal **structure**	Variation among • Individual CLDs (table) • Team CLDs (table)	• Limited understanding of the system's causal structure, based on experience alone • Extent of shared understanding of causal structure
Quadrant 2	Shared Mental Model #1 vs Clinic CLDs = Shared Mental Model #2 = **shared understanding** of causal structure		Variation for Clinic CLDs in the level of • Complexity (table) • Agreement (CLD) • Understanding (CLD)	Extent of shared understanding of causal structure in the *assumed* Shared Mental Model

	Similarities (Connections) 		**Differences (Contrasts)** 	
Location	<u>Model names</u> = **The main** assumptions about models * *data and methods*		**What lens do we use to view the differences in ways of knowing relative to the <u>previous</u> location (type of visual)**	**Visualized aspect of differences (Called out via figures, tables or text)**
	In the transition zone	Exiting transition zone		
Zone B	<u>Shared mental model #2</u> vs *raw diverse perspectives* = <u>Shared mental model #2</u> <u>iterations</u> =	<u>Shared mental model #3</u> = sum of **shared perspectives** on causal **structure**	Model revisions • Summary (table) • Justification (RIQs – Rigorously-Interpreted Quotations) Variation in level of agreement • Across individuals (CLD) • Across professions (CLD)	• Model improvements from a new set of diverse perspectives on causal structure • Extent of shared understanding of causal structure across new diverse perspectives
Zone E	<u>Shared mental model #3</u> vs *diverse causal perspectives* ** purposive text analysis*	*Validation data*	• Variation in the meanings attached to variables and links (RIQs – Rigorously-Interpreted Quotations)	• Extent of differences in articulating the causal structure across diverse perspectives
Quadrant 3	<u>Shared mental model #3</u> = <u>Conceptual Model</u> = the **causal structure** hypothesized to be guiding the real system (<u>Conceptual Model</u> * *operational thinking*) * *model validation tests* = <u>Simulation Model #1</u> = the **causal structure** hypothesized to be guiding real system **behavior**		What is beyond understanding in the simulation model • Tested for relevance and consistency (validity test results) • Described in accessible terms (diagrams)	• Extent to which the simulation model is relevant to and consistent with stakeholders' perspectives • Content of the system's causal structure (respecting and superseding cognitive limitations)

	Similarities (Connections) 		Differences (Contrasts) 	
Location	Model names = **The main** assumptions about models * *data and methods*		**What lens do we use to view the differences in ways of knowing relative to the next location (type of visual)**	**Visualized aspect of differences (Called out via figures, tables or text)**
	In the transition zone	Exiting transition zone		
Quadrant 3	Shared mental model #3 = Conceptual Model = the **causal structure** hypothesized to be guiding the real system (Conceptual Model * *operational thinking*) * *model validation tests* = Simulation Model #1 = the **causal structure** hypothesized to be guiding real system **behavior**		Capability of simulation model to run realistic experiments • Tested for suitability for purpose (validity test results)	Extent to which the simulation model is suitable for running realistic experiments
Zone C	Simulation Model #1 vs *diverse causal perspectives (dynamic hypotheses)* = Simulation Model #1 iterations =	Simulation Model #2 = sum of **diverse perspectives** on the causal structure underlying **behavior**	System story reproductions • Categorized by degree of difference with models • *Right behavior, right reasons* tested across multiple variables (run charts)	• Extent of necessary detail in the models with respect to participants' detailed stories • Suitability of the model's story-telling
Zone F (Q2 -> Q4)				

	Similarities (Connections) 		Differences (Contrasts) 	
Location	<u>Model names</u> = **The main** assumptions about models * *data and methods* In the transition zone	Exiting transition zone	**What lens do we use to view the differences in ways of knowing relative to the <u>next</u> location (type of visual)**	**Visualized aspect of differences (Called out via figures, tables or text)**
Quadrant 4	<u>Simulation Model #2</u> + policy analysis variables = <u>Simulation Model #3</u> = the **causal structure** hypothesized to be guiding the real system **behavior**, equipped with tools to change it <u>Theoretical Model</u> = sum(Policy Structure Diagrams)		Variation in preferences • Described (table) • Implemented (alternative environmental conditions) • Located in the causal structure (diagrams)	Extent to which diverse preferences can be explored
Zone D	Stakeholder understanding * Simulation Model #3 * SUM (*Validity*) * SUM (*Visualizations*) =	Improvement in stakeholder understanding obtained through the research	Accessibility of research results • Summarized (models) • Evaluated (model and meta-level validation) • Visualized (visualizations)	Implications of limited understanding

NOTE: Interplay relies on the "maintenance of tension between contrasts and connections"[273] (p. 534). The drawings in the first row present the general concept: yellow circles represent connections and circles of other colors represent contrasts across two paradigms (represented by the blue boxes with different patterns). When one is working within one paradigm, one considers contrasts with the other paradigm. When one is in a transition zone (green box – second set of columns in this table), one is using the connection between paradigms to travel from the one paradigm to the other; differences are also reported (white box – third set of columns in this table). Thus, connections and contrasts are considered as one works to better understand the problem being investigated. The same process is used to travel between a paradigm and a transition zone.

4.2.3.4. PRIMARY CARE TRANSORMATION AS A CASE OF PARADIGM INTERPLAY

An essential feature of a scientific paradigm is a body of work – made up of empirical cases – illustrating its methods and the advantages they bring[113, 260, 286] (p.849). **By providing an empirical case study that uses system dynamics as an integrative paradigm in an *interplay* paradigm crossing strategy, this dissertation provides evidence for an *integrative* system dynamics paradigm.** This strategy relies on newly-developed qualitative methods, which work in concert with standard SDM methods.

The case involves understanding the complex dynamics involved in primary care transformation described in Section 4.2. The result is an improved understanding of this transformation process as well as of the interaction between the main theoretical tenets of primary care systems. The case involves one HSDO that contains heterogeneous practice styles – each care team has latitude in how they practice and each clinic has adapted to its local market environment. This has offered the opportunity to consider contrasts as well as similarities across contexts and diverse perspectives of individuals, thus starting to explore generalizability.

The resulting *minor content theory*[103, 145] is made up of causal relationships between variables complete with units of measure and measurements representing an average care team undergoing primary care transformation – starting at traditional primary care with the potential to achieve full adherence to clinical guidelines – in the US. The theory includes many claims that can be tested further in this context and in other contexts as well. The theory can also be used as a blueprint for the development of a detailed calibrated model.

4.2.4. CONTRIBUTION TO MIXED METHODS

As part of my methodological aim, I created methods when needed. In doing so, I also developed some contributions which apply to MMHSR, including to two[83] of the three outstanding methodological issues described in Chapter 1 (emotional engagement and paradigm issues).

Section 4.2.4.1 presents my contribution of a mixed methods approach for conducting a scoping study. Section 4.2.4.2 presents my contribution of new methods for developing and validating mental models as well as methods for mental model *measurement*.

Objective #2 concerns addressing issues with respect to emotional engagement. I begin Section 4.2.4.3 by using system dynamics definitions to attain a more precise definition of some key concepts related to emotional engagement. Next, I present my contributions (methods and visualizations) to developing a *story with numbers* that attempts to engage emotionally with stakeholders; with this dissertation as a case demonstrating their application.

Finally, I present a more precise definition of key concepts relating to paradigm issues in MMHSR. Section 4.2.4.4 begins by proposing some terms (as suggested by mixed methods scholars Teddlie and Tashakkori[287]) for mixed methods research aiming to conduct integrative work using stories to accompany numbers within each of the six transition zones of the Four Quadrants of Knowledge. MMHSR scholars have called for integrative methodological research. However, they warned that such work should respect the integrity of methods with respect to paradigms[96]. Therefore, after presenting the terminology, I present new methods for such an approach which maintain paradigm and methods integrity. I conclude by presenting my contribution to the development of the mixed methods paradigm: this dissertation as a case demonstrating the generation of stories to accompany numbers using system science methods.

[83] My work on the third objective has contributions not only to MMHSR but also to social science more broadly. Therefore, it is considered in Section 4.2.5 below.

4.2.4.1. USE OF THE SCOPING STUDY

Using mixed methods in a scoping study[84] (and documenting methods results and conclusions/actions taken) was useful for several reasons:

1) To clarify the problem statement
2) To inform the conceptual framework
3) To identify methods for further investigation
4) To build confidence in the proposed work (on the part of funders, research team members and problem stakeholders)
5) To get to know the data available (their suitability for further use)

Looking back, I now see that this step was crucial to my dissertation. However, in today's fast-paced policy, managerial and research environments, it is a step that is tempting to skip. I feel it important to highlight since there is a lot of inertia in work today to spell out the study design before data collection. Doing so locks the research team in to assumptions that may be inappropriate for the problem studied at hand. Past research work may not necessarily point to assumptions that are appropriate to the current work.

Miller *et al.* agree. In a special issue of Health Services Research on the use of MMHSR, they highlight the need for matching "each phenomenon to the most appropriate method" (p.2130) before attempting to "fit each piece into the larger mosaic" *(ibid.)* of an integrative methodology[96].

My contribution is **the use of a mixed methods study for a scoping study** – one that (first gains clarity on the research question and relevant assumptions and then) matches the problem to the appropriate method(s) before fitting them into the larger methodological mosaic (as Miller *et al.* have requested[96]). Mixed methods, with its embedded, multi-disciplinary research teams has a large potential for carrying out this kind of work.

4.2.4.2. ACCESSIBLE MENTAL MODELS METHODS

I contribute methods for developing and validating mental models, (individual, group and shared), which are accessible to qualitative researchers without extensive system dynamics training. Furthermore, my validation **methods permit more than mental model elicitation,**

[84] The scoping study is presented as an appendix (rather than a chapter of the thesis) because it was preliminary to the SDM work that is the focus of this dissertation. This in no way diminishes its importance to my research process.

but the actual mental model *measurement* which Doyle *et al.* argue is necessary for studying improved understanding[239].

4.2.4.3. EMOTIONAL ENGAGEMENT

Miller *et al.* envision the future of MMHSR as a research process producing findings that are: "a mosaic from which an emotionally engaging and empirically valid research story is created"[96] (p. 2129).

The concept of emotional engagement has been explored in different ways in diverse social science disciplines. **My contribution here is to synthesize insights from system dynamics both on the meaning of emotional engagement and the mechanism/force for generating it.**

In their work to improve society by helping people reason better, social scientists have found it necessary not just to produce results which appeal to *logos*, but also to *ethos* and *pathos*. Renowned organizational theorist, systems thinker and system dynamicist, Peter Senge, describes why emotional engagement matters:

> "It has been my experience that when a group of people discover for themselves how their own actions are generating major problems they are experiencing (and usually attributing to forces outside their control), the experience can be genuinely powerful. As with the engineering team …, they invariably say *'Look at what we are doing to ourselves.'* Moreover, seeing how they created a systemic structure that is shaping their reality is a powerful force for creating new actions and new structures."[285](emphasis added, p. 9)

For Senge, this genuinely powerful discovery experience ***is*** the emotional engagement that, even for engineers, causes a powerful force for changing policy. Forrester, the creator of system dynamics, also identifies this phenomenon, as he describes why such a powerful force is needed; because, in the opposite direction, there are countervailing forces hampering policy change:

> "Almost always, the reasons [for problems in a system] will lie in policies that people know they are following and *which they believe will lead to solutions* to the troubles. Implementation often involves reversing deeply embedded policies and strongly held emotional beliefs. It is not that people disagree with the goals, but rather how to achieve them. Even with widespread intellectual agreement with a system dynamics model and with the recommended improved policies, there may still be great discomfort with the prospect of changing from traditional actions."[120] (p. 5, emphasis added)

As these quotes make clear, system dynamics theorists agree with mixed methods scholars Miller *et al.* that research will have more impact on policy when results include stories with numbers that are emotionally-engaging[96].

In the system dynamics tradition of probing these cognitive psychological assumptions[119, 277], Stave *et al.* identified four forms this improvement can take: problem-related insights, structural insights, dynamic insights, and paradigmatic insights[278]. These insights accumulate as stakeholders place *confidence* in the process and results of a system dynamics intervention[258, 279], building up the *motivational force* described by Senge above.

From its founding, system dynamics has been more interested in improving understanding than in providing prediction[288]: "it may seem paradoxical but the results of a quantitative system dynamics study are qualitative insights"[103] (p. 17). Thus, communication (through a participatory research process and/or a visualization of research results) is key for emotional engagement that connects with one's subconscious, where the mental models which drive decision-making (and thus policy changes) can be changed.

The research process presented in this dissertation can be split into two categories: (1) scoping study and (2) system dynamics modeling study – each of which had emotional engagement.

In system dynamics, the scoping study typically has one of two manifestations: (1) the stakeholder identifies this method as the one they would like to pursue – in so doing, making the case themselves for its use[203], or (2) researchers make the case that this method is needed.

I was somewhere in the middle; on the one hand I was a researcher, part of a research team; on the other hand, that team was an *embedded* research team that also included HSDO management personnel.

My scoping study (Appendix C) tells a mixed methods research story. It presents the evidence that led me to selecting system dynamics. This was useful in engaging with the research team, the funder, and project stakeholder advisory committees in allowing this part of the work to move forward and, I hope, it will be useful to my audience more broadly as they consider whether or not to even read the SDM analysis presented in this work.

The remaining work (Chapters 2 and 3) tells a system dynamics research story. In that research story, I present a simulation model which tells stories with numbers (e.g., in Section 2.4.5.5.2 the simulation runs are compared to participants' stories). While stakeholders have not yet seen these results, they did provide me with confidence that my research was ready for the outside world and I hope it will provide equal confidence for them in my approach and findings. Specifically, the meta-level validation results present evidence for the empirical rigor of this work and that the research approach is in line with the way problem owners see the problem.

PCT researchers external to my team expressed the sense that: "once you've seen one (organization attempting PCT), you've seen one"– each experience is highly contextualized, so it is hard to learn about "the system". Based on my continued interactions with these stakeholders, I have a sense that this work is engaging them in wanting to see how the story ends. These include interactions at the National Advisory Committee for the project, various poster presentations at national and international conferences (see Appendix A).

Now regarding problem owners, my sense is limited. It would be naïve to assume that they would read this dissertation; as the saying goes "your committee and your mother are the only people who are likely to read it in its entirety". That said, I have had some interaction with them in the context of scoping study findings, via the Local Advisory Committee for the project and presentations of the team to the HSDO management. In all cases, the response was positive, wanting to know more.

Thus, both the scoping study and the system dynamics modeling study present (methods and) a case of developing stories with numbers that attempts to engage with individuals' understanding at a subconscious, emotional level.

SDM work has two primary aims: building understanding and system improvement. Because of this emphasis on stakeholders' *understanding*, SDM has always used visualizations to aid model *users* in understanding the assumptions and implications of models. These include diagrams for conceptualizing and summarizing model structure, friendly algebra for equations, graphs for reference modes showing behavior-over-time theories and expectations, charts for simulation runs, and tables for summarizing model assumptions and policy implications[113, 114].

The majority of the visualizations I employed in Chapter 3[85] come from the SDM standard method. However, I also developed a set of new visualizations specific to my methods (see Chapter 3 for detailed descriptions). In Table 4.11 below, I group these visualizations by type and identify the method they are used in, how they are used, and why they are used. These visualizations show (1) the confidence which *I have* in the model, (2) the diversity of perspectives upon which it is based, (3) the diversity of perspectives that the models capture, and (4) the types of data in participants' mental models.

[85] The majority of diagrams are policy structure diagrams and sector maps. The Theoretical Model is a hybrid diagram which encapsulates the simulation model using policy structure diagrams.

All of these things visualize the research process and its results, and thereby *aid users in judging* how much confidence to place in the modeling process and results *as well as in understanding* the models and where they can act within the system; thus, ultimately, to more easily engage with the findings and discover *insights* (made available by the research) that will improve their understanding.

By contributing methods designed to engage with qualitative data, I strongly ground the work in problem owners' experiences and I also *show* them that I did so and check my work in the process. **These newly-developed visualizations are designed to engage emotionally with stakeholders** to connect with their subconscious and to motivate change in their mental models that reflects an improved understanding (primarily of system structure[289]).

Table 4.11 Newly-developed Visualizations

Type of Visualization	Methods with new visuals	Specific use	Purpose
Charts*	SIM-S	Run charts compare the model's structure and behavior to participants' descriptions of their dynamic hypotheses	Display the degree of saturation of the simulation model with respect to individual understandings
Diagrams*	SMM-S, CM-S	Custom CLDs display the level of agreement between qualitative models and • Participants' combined mental model CLDs, by clinic (SMM-S) • Participants' causal statements (CM-S) individually and by profession	Display the degree of saturation of the qualitative models with respect to shared understandings
Saturation curves	SMM-S, CM-S	Custom graphs explore the extent to which new model elements have ceased to emerge from analysis of • combined mental model CLDs (SMM-S) • participant interviews (CM-S)	Display the degree of saturation of model elements in the data
Information Accumulation Graphs	SD-S	Custom graphs explore the content of interview transcripts by information type.	Display the extent to which individual understandings include system dynamics-related elements of system structure
Rigorously-Interpreted Quotations	CM-S, SIM-S, SD-S	Custom tables display how quotations are interpreted to • indicate how capable the qualitative model is of exposing variables and causal links in participants' mental models (CM-S) • show participants' descriptions of their dynamic hypotheses (SIM-S) • show participants' awareness of having SDM-type mental models, that these matter, that they can (and indeed must) change and that doing so involves emotional engagement (SD-S)	Display the degree of saturation of the Conceptual Model with respect to individual understanding. Display descriptions of participants' dynamic hypotheses. Display the extent to which individual understandings include and prioritize SDM-type mental models.
Tables*	SMM-S	• Variation in Clinic CLDs match to SMM1	What were some differences between the model and its inputs?
	CM-S	• Re-conceptualized elements table	Did validation change the model?
		• Systems thinking by participant	How much of the system do participants see?
	Meta-level validation	• Participant perceptions of a time delay	What indicates seeing a *delay*?
		• Data Suitability table	Where did my model come from?
		• Methods Suitability table	What trade-offs did I make?

*These are not new visualization types but I did create new uses for these. NOTE: Visualizations were also developed for the scoping study. These are not presented here.

4.2.4.4. INTERPLAY FOR MIXED METHODS – A CASE & THEORY

There are complementarities between simulation and mixed methods studies. For one, a mixed methods study is a good way to identify appropriate simulation methods, using rigorous definitions of their assumptions[290], as described in the scoping study. Also, the goal of mixed methods ("fully capture the complex interactions among system components, including interactions among multiple levels of analysis[86] and over time"[96] (p. 2125) is the goal of system dynamics. This dissertation shows how system dynamics theory and methods can contribute to filling gaps in mixed methods and a case for how this can be done.

In terms of theory, in Section 4.2.3.1, I presented the *Four Quadrants of Knowledge* with *transition zones*. While I use system dynamics as a second-order concept, others in MMHSR may find other ways to create an integrative study. One contribution I believe my research makes to such efforts broadly is **to provide terminology for integrative work using Miller *et al.*'s concept of *stories to accompany numbers* mapped onto the *Four Quadrants of Knowledge*** (their recommended framework)[96].

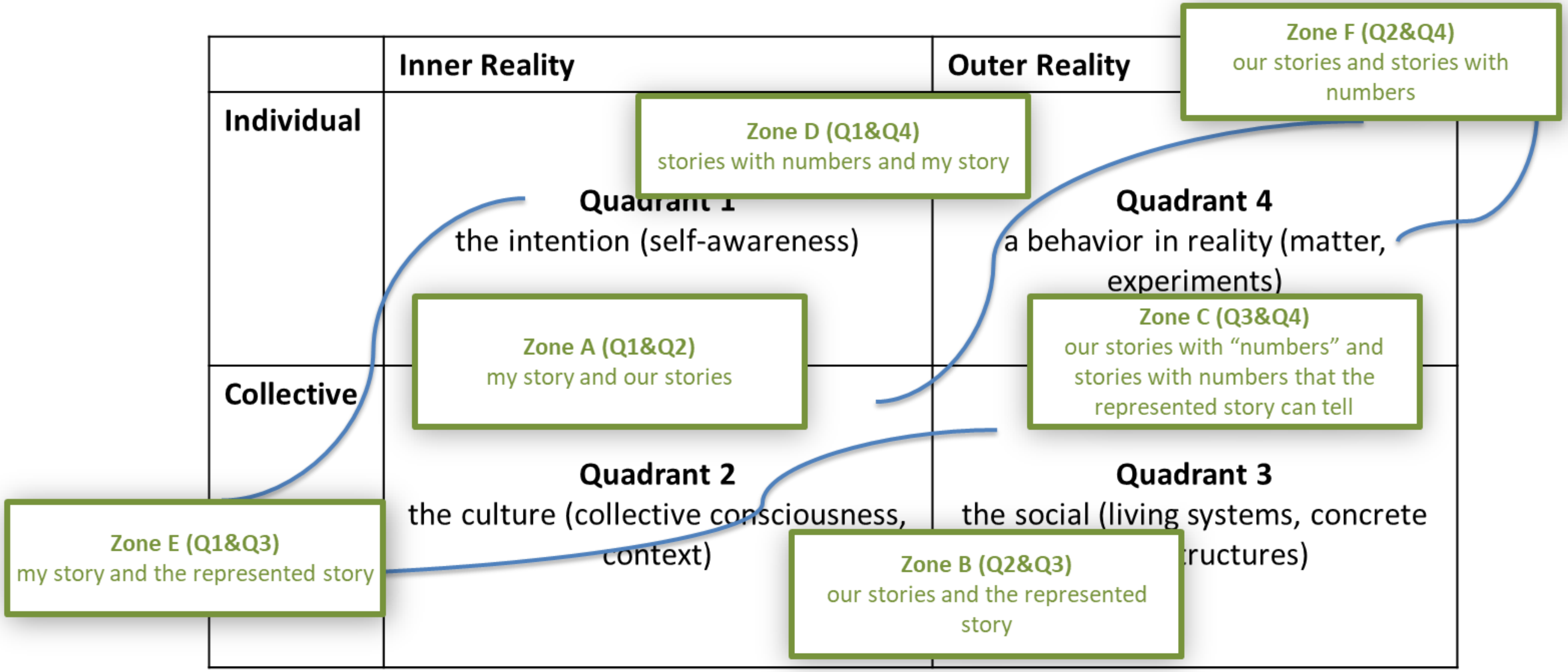
Miller *et al.* bemoan the existence in MMHSR of a "separate and unequal" (p. 2126) approach to methods, data collection and in the ways that results are published, calling for the development of "a more complete methodological mosaic"[96] (p. 2126). They point out that a "particularly effective means" (p. 2129) for doing so is creating research "stories to accompany numbers" *(ibid.)* by "integrating [the] divergent approaches"[96] *(ibid.)*. **The use of the *interplay* paradigm-crossing strategy** in system dynamics, as I have here demonstrated, provides one way of integrating the divergent approaches that capture and then not only report but weave together the stories with numbers. At the same time, **this strategy maintains paradigm and methods integrity**.

Their plain-language concept of *stories with numbers* inspired me to develop Figure 4.11 below. This figure synthesizes the concept of *stories with* numbers with the *Four Quadrants of Knowledge* and *transition zones*. In contrast to Figure 4.7 above, this version of the quadrants with transition zones is in plainer language that can apply to mixed methods system dynamics work as well as to mixed methods generally. The more of these zones that MMHSR studies can cross, the more integrative their work will be. Interplay involves documenting the

[86] An important part of a system dynamics study is setting the model boundaries which involves deciding how to treat various levels of analysis. My model incorporates several levels of analysis. While the stocks consider issues that the team sees, the model uses other types of variables to also incorporate necessary variation among individuals, clinics and the larger market or organizational context.

similarities and differences between the paradigms that methods consider in each quadrant and zone (as in Table 4.10 above).

Figure 4.11 Transition Zones for Mixed Methods



Note: To be clear, in Zones A, D and E, "my story" does not refer to the researcher's story but to the story in each participant's individual perspective.

In the research on PCT, there has been a more widespread adoption of *ideas* from systems science than applications of systems science *methods in cases* (see, Nutting 2010). This dissertation work **presents methods along with a case for generating *stories with numbers* using system science methods.** Stories are obtained in a rigorous process and then told using numeric simulation results and qualitative mental data (including quotations as well as visualizations of system structure and behavior) such that understanding is gained from the description *and* the behavior that accompanies it. In so doing, I illustrate the usefulness of a systems science approach (system dynamics) and of my newly-developed methods.

I present these thoughts with the caveat that, currently, mixed methods does not explicitly include simulation methods in descriptions of potential study designs (c.f. Fetters *et al.*[291]), leading some observers to conclude that, even when qualitative methods may play a large role in simulation, it is not part of mixed methods as currently conceptualized[292]. Instead, such studies have been described as *mixed qualitative-simulation*[292] and *multi-method*[116]. Another potential challenge could be that mixed methods is not associated with an integrative paradigm, but rather with the pragmatist, pluralist, and dialectical paradigms[87][293] (p. 26-29)[287] (p. 779)[276, 294].

I couch aspects of my work in mixed methods because that is where it began and I have continued to mix qualitative and quantitative methods throughout this research process. To learn from my work, mixed methods researchers can learn from the theory, methods and case whether or not they wish to engage in mixed methods simulation modeling.

[87] Broadly speaking, these paradigms hold that pragmatic problems are best studied using a flexible toolkit of social science methods (qualitative and quantitative), and that paradigm incommensurability is overcome by allowing the use of a variety of paradigms to support research across the phases of a research project.

4.2.5. CONTRIBUTIONS TO SOCIAL SCIENCE

In Chapter 1, I identified two paradigm issues for MMHSR. In my research I have found that these are really issues for social science more broadly. Those issues are: the individual/society problem and paradigm incommensurability. Section 4.2.5.1 deals with the individual/society issue by summarizing the philosophical work of Trigg and my contributions to the development of integrative social theories. Section 4.2.5.2 deals with paradigm incommensurability by showing how I employ qualitative data to develop and test both qualitative and quantitative models. Both sections emphasize my contributions to empirical validity in addressing these paradigm issues.

4.2.5.1. ON GENUINE INSTITUTIONAL FACTORS

Lane suggested that the individual/society debate in social science represents an important opportunity for system dynamics – one which requires new methods that are in line with the "empirically grounded and practically minded approach that is the strong heart of system dynamics"[102] (p. 306). With the objective of addressing issues of empirical validity for MMHSR, I conducted a deep investigation of the system dynamics theory of validation and proposed several contributions (validity subtypes and validity tests summarized in 4.2.2.2 and 4.2.2.3).

Because I chose to build within existing system dynamics theory, it is useful to consider what the expectations philosophers of social science have for integrative social theories (even minor content theories[88][103, 145] like my model of PCT)**.**

Roger Trigg, a philosopher of social science[89], describes these expectations in his introductory text to the philosophy of the social sciences. Aiming to develop a theory which can hold across contexts, which is not just internally coherent but also corresponds with reality, it turns out, is not a small thing in social science. It is highly problematic because **context** cannot be studied without considering **individuals**, the two are inextricably linked. So, the only way to develop integrative theory is to discover the structures or, for Trigg, the *institutions* that link **them**[98].

[88] A minor content theory is a dynamic hypothesis specific to the problem being studied – it identifies the causal relationships thought to bring about system behavior (see Section 2.2.2).
[89] I refer to Trigg because he wrote my introductory textbook to social science. His work is synthesizing many current philosophers. It represents *current thinking* for students in an accessible way. Going to the original texts of individual social theories would be outside the scope of this dissertation.

Trigg has argued that the individual/society problem is a paradigm issue for all social science. It is "the major problem facing social science" (p. 205) because it leads to research which ignores the "intertwined relationship of individual and social setting" (p. 207) which plays out over time through "the fact of unintended consequences" (p. 207)[98]. I will describe here how, ***for Trigg,*** **solving this problem requires research which crosses paradigms seamlessly from the individual to the collective and from inner reality to outer reality and back again to individuals.**

For Trigg, a theory of a context, society or institution (i.e., a system structure) must be more than "simultaneous choices" (p. 207), it "has to take into account the passage of time" *(ibid.)* because consequences persist over time[98]. The accumulation of consequences creates the structure "which affects generations to come, just as [our] … context … has been created by the … choices of those … gone before" *(ibid.)*. Structure exists outside of "its current members [and is] … more than an accidental conglomeration of individuals"; it takes the form of "cooperative relationships that extend through time [which] … constrain and influence the choices of those who belong" (p. 207-208)[98]. So, a theory of structure needs to explain the consequences of simultaneous choices as enduring features of reality which take the form of stable causal relationships which influence individuals' agency.

Developing a theory of structure should begin with understanding these relationships which "seem … to be dependent on human understanding and could not exist without it"[98] (p. 208). Not just any shared concept qualifies, for example, "human rights … do not appear to exercise any causal power over us"[98] (p. 209). A persistent structure is *emergent*, it is "the product of a collective rather than an individual construction"[98] (p. 210). This gives it attributes which are beyond the limitations of perception, "[it's] character can often outstrip everyone's understanding of it" *(ibid.)*[98]. So, theory-building needs to be based in *individual* understanding, *shared* understanding and elements of structure which can be *beyond* understanding.

Integrative theories could be presented via "[the metaphor of] rule-governed games" and validated via precise formulation: "a *test* for the presence of ***genuine institutional factors*** [in a theory] is whether or not we could **codify the rules** explicitly"[98] (p. 211, emphasis added). **So, integrative theories could be expressed as white-box simulation models which are validated for structure.**

A few of Trigg's general comments are also pertinent here. First, validation of social science theories should consider at least "some form of correspondence with reality" (p. 116, emphasis added) and the "internal coherence" (p. 117) of theory, which "unless reality is fundamentally chaotic, [will be] a precondition for understanding"[98] *(ibid.)*. Second, broad outward boundaries are important because they enable "a theory to bring together aspects of knowledge which were previously separate"[98] *(ibid.)*. Depth of inward detail is also valuable because, while "many have valued simplicity, … in our quest for the nature of reality it can be disastrous to assume that it is less complicated than it is"[98] *(ibid.)*. So, integrative theories expressed as models should be validated for behavior as well as structure. Doing so improves their accessibility to the target audience, and the choice of inner and outer boundaries is very important.

System dynamics offers social scientists a potentially-integrative methodology. How can it meet this potential? In Table 4.12 below I present one social science philosopher's expectations for developing valid integrative social theories (column one) and then map my contributions to them (columns two and three). In so doing, **I make my contribution to the individual/society debate: (1) a summary of Trigg's expectations for integrative theories and (2) my contributions to SDM's validation theory[90] and methods as they relate to these expectations.** My newly-developed methods have the potential to help social science researchers within and outside system dynamics to develop theories that meet the expectations for integrative minor content theories.

[90] System dynamics' theory of validation includes a set of four validity types which are the underlying goals of model testing. Each type is made up of two to four subtypes. My contribution was to create new subtypes and to make some of the existing ones clearer than before. The validity types (and subtypes with their shortened descriptions) considered in SDM are listed here, subtypes which I developed in this dissertation are underlined, those which I extended are italicized. Conceptual Validity (CptV1 Variables & boundaries, CptV2 Links, CptV3 Saturation, CptV4 Culture); Formulational Validity (*FV1* Language limitations, FV2 Conceptual equals simulation, *FV3* SDM guidelines); Experimental Validity (EV1 Structural design, EV2 Insights gained); Data Validity (DV1 Mental data, DV2 Written data, DV3 Numerical data). See Section 2.3.2.1 for more information on validation in SDM.

Table 4.12 Mapping My Contributions to Expectations for Developing Valid Integrative Social Theories

Trigg's Expectations for Integrative Social Theories	My Contributions To:	
	Validity Types (see Section 2.3.2.1)	Methods (see Section 2.3.2.3 and 4.2.2.3)
Use research methods which **cross paradigms seamlessly** from the individual to the collective and from inner reality to outer reality and back again to individuals.	This added clarity (CptV1-4, FV1, FV3, and DV1-3) improves modelers' ability to test their models using data representing the **inner and outer reality of individuals and collectives**.	All newly-developed methods facilitate the **crossing of paradigms**. Figure 4.12 gives an overview of how this is done. The implementation is described in detail in Table 4.10.
Explain in the theory how **intended and unintended consequences** of simultaneous choices become enduring features of reality taking the form of **stable causal relationships** which influence individuals' agency.	This added clarity (CptV2, CptV4, and FV3) improves tests' ability to **justify causal relationships** in their cultural context, to evaluate their quality, and to visualize them.	CLD Combination is a method allowing researchers to merge participants' descriptions of the **intended and unintended consequences** of their actions to identify the **enduring features** of the system in which they are acting (Figure 2.9).
Base the methods for theory-building in **individual** understanding, **shared** understanding and elements of structure which can be **beyond understanding**.	Similarly, this clarity (CptV1-4, FV3, and DV1) improves tests' ability to justify the claims that a model represents system elements of i**ndividual and shared understanding as well as those which are beyond understanding**.	CLD Combination also permits aggregation of **individual** understanding into **shared** understanding (see Sections 2.3.1.2 and 3.2). My model validation methods (Table 4.7) help to assess the comprehensiveness of a model's representation of **individual** and **shared** understanding (SMM-S and CM-S) as well as its representation of elements that exist **beyond direct perception** (SIM-S).
Validate the theory **by codifying** its **rules** explicitly (as in a white-box **simulation** model)	This added clarity (CptV4, FV1, and FV3) also improves the way **qualitative models** are evaluated for **quality** as well as the way one justifies the **cultural rules** (goals, policies, etc.) one is modeling.	I develop a method for testing the **cultural rules** described by participants against their **codification** in the simulation model (SIM-S). I also consider whether specific SDM model coding assumptions (causal links, delays, feedbacks, etc.) can be found in participants' descriptions of their context (SD-S).
Validate the theory for **correspondence** with reality and internal **coherence** so that it is useful	The refined definitions (CptV1-4, FV1, FV3, and DV1-3) help modelers to assess the **quality** of their empirical **data,** to use these data to test the **coherence** of their **models** and the models' **correspondence** to **data.**	My model validation methods (SMM-S, CM-S, and SIM-S) evaluate my models' **internal coherence** and **correspondence with reality** by taking participants' system stories (of system structure and behavior) and comparing them to stories told using the models.
Consider **boundaries** such that we do not assume that the theory is more or less than it should be (inner and outer boundaries)	Finally, the refined definitions (CptV1-4) allow for measuring the **suitability** of a model in terms of its **outer** boundary and its **inner** detail **complexity**.	SMM-S and CM-S use Saturation curves to assess the suitability of the **outer boundaries** of the qualitative models. Story-telling in SMM-S, CM-S and SIM-S checks the suitability of the **inner boundary** by assessing how well the models' elements relate to important parts of real stories).

4.2.5.2. MULTIPLE-PARADIGM EMPIRICAL VALIDITY

As described above, paradigm incommensurability is as much about scientists and the way they do their work as it is about their worldviews. According to ethnographer and organizational theorist Van Maanen, the purpose of qualitative methods in the social scientific process is to describe how a behavior emerges from a given context. Qualitative research *can involve* defining, measuring and quantifying relationships between *quantitative* variables – "having intimate familiarity with a research setting [enables one to build] … quantitative indicators of what to outsiders might seem to be rather impressionistic and ill-defined concepts"[295] (p. 524-525). Nevertheless, it is more common for qualitative researchers not to assume that this is feasible[295], opting to limit their qualitative work to the Inner Reality side of the *Four Quadrants of Knowledge*.

The product of qualitative research is an "interpretive framework" which should be "firmly accepted" and "set" before it can be converted into an "analytic formula" for use in quantitative analysis[295] (p.521). To achieve *acceptance*, methods should rigorously consider "the language chosen [by researchers] to represent a given social world" in descriptions and frameworks. By so doing, qualitative research can permit the "unscrambling [of] social processes … over time"[295] (p. 524), which are "virtually impossible to comprehend over the short run" *(ibid.),* and prevent the "empirical de-focusing and … analytic deterioration"[295] *(ibid.),* which can occur in studies when empirically-rigorous multi-paradigm approaches are not considered [295].

I developed and implemented methods which employ qualitative data to develop and test both qualitative and quantitative models for SDM. By doing so, I provide one way of *focusing* a qualitative-informed research project and thus guarding against the tendency for *analytic deterioration*.

This is accomplished by working within the cognitive limitations of stakeholders to develop more aggregated, but still limited, models and then using mental data to remove my own interpretations from the models to the extent possible. As described in the previous section, I provide a systematic way of enabling stakeholders to correctly assess the empirical validity of SDM work (by conducting, documenting and reporting that validity assessment).

4.3. LIMITATIONS

Limitations with respect to data and methods were briefly discussed in Section 3.5.2 and Section 3.5.3. In this section, I consider these and other limitations of the research process.

With respect to data collection:

- **Interviews were limited in their ability to capture change over time, risking *presentism* (using today's lens to reflect on the past).** They were conducted during a brief period with one interview per individual. I tried to guard against presentism by anchoring participants in specific points in time using a detailed timeline of PCT and non-PCT events at the HSDO (which I developed in an archival review). I used this timeline to help ground the interviews in the sequence of PCT milestones and when they took place, asking participants to think back to that time and their thoughts then.
- **Interview content is mediated by participants' propensity for self-reflection**[250]. Relying on their statements assumes that these reflect the depth of their experiences. I attempted to reduce this limitation using follow-up questions and probes to explore different aspects of a subject/experience. I developed Saturation Curves to assess whether eliciting one more participant's mental model will provide new elements (CM-S and SMM-S curves). Thus, whether or not a participant was more or less capable of self-reflection, I had a sense across participants as to whether I could stop data collection.
- **I did not formally collect (i.e., ask them to draw) table function shapes from participants**. I constructed alternative team member preferences (table functions) based on my interpretation of what participants expressed in the interviews and my understanding of what plausible extreme conditions might be. It is possible that further data collection would identify additional alternative preferences that exist and that produce results different from the ones presented here.
- **My standards of relevance**[250] **influenced the interview content and the transcript analysis.** Regarding interview content, I was participating in the discussion – thus I may have asked follow-up questions in more depth for some subjects than others. In transcript analyses, I focused on causal and conceptual information – thus results are mediated by this standard of relevance. I developed Information Accumulation Graphs to assess the extent to which my standard of relevance corresponding to what participants spent time talking about. A mismatch would indicate that I need to go back to the data and reflect further.

With respect to model building:

- **In building the Shared Mental Model, I started with the most complex clinic CLD.** This anchoring may have biased the CLD Combination process as I looked at the other clinic CLDs to see how they would *modify* this one. I attempted to address this limitation via rigorous subsequent validation tests, including (ones most closely related) SMM-S, CM-S, and SIM-S. In each test, I subjected the work-in-process models to participant experiences to see how well they were able to match what was expressed. Revisions were made to the work-in-process models as needed.
- **The care team consists of many different professions – and this representation is limited in my model.** The model ramps up from 2 medical assistants to a team involving licensed clinical staff members as well (e.g., nurses, social workers, pharmacy doctors). All are called “MAs” in my model but, to some extent, that is a misnomer as, to get to full task-shifting, some tasks can only be shifted to personnel who are licensed to do them (e.g., medicine reconciliation can be done by a pharmacy doctor but not by a medical assistant). Thus the model limits the ability to see the different types of clinical staff members that join the team and the specific tasks that they take on.
- **Important causes of PCT failure may not be included in my model.** My model focuses on endogenous causes of PCT failure from the perspective of a team in a hypothetical stand-alone clinic. Exogenous causes were considered when they directly influenced the feedback structure (e.g., clinic policies). However, further removed causes of failure were not included (e.g., financial reinvestment policies at a clinic-owning HSDO, implementation of health information technology changes, poor management practices).
- **In the literature, some argue for risk-stratification** such that more services are offered to those who are at higher risk for negative outcomes (e.g., hospital readmission). My model treats a team’s patient population in aggregate (as a panel, as conceptualized in primary care, where some are higher risk than others such that the average services per patient are those in the model). This limits the model’s ability to explore how PCT would be impacted by various methods for operationalizing risk-stratification.
- **Quantitative data relied upon for model building came primarily from peer-reviewed literature and participant interviews.** Specific limitations for each parameter from the literature are provided in the associated reference. Detailed calibration of my model was not possible in the analysis timeframe as the appropriate quantitative data were not available. I did work to obtain the operations data available from the organization

and the outcomes data available from the government, but upon close examination, the data were not able to generate the information needed for the simulation model (see Section 3.5.2 for more detail on data limitations).

With respect to model validation:

- **Interviews were within one HSDO – so diversity of experiences may have been limited.** I relied upon the sampling frame (design for both oversampling (Section 2.3) and purposive sampling across clinics (Section 2.4.2)) to capture diverse experiences, including those of clinics and teams that had been attempting PCT for varying lengths of time and using different approaches. In validation tests, I repeatedly subjected my work-in-process models to experiences of HSDO personnel in different roles and different settings as well as to stakeholders of different kinds outside the HSDO. Finally, in this chapter, I compared my findings to those of others in other settings (Section 4.2.1) to see the extent to similarities found.
- **In the literature, some argue for the use of numeric data for model validation.** Model testing with numeric data has proven to be a useful method for discovering problems in models – and insights as well. I was able to test the relevant aspect of validity of my model which such tests consider (i.e., Experimental Validity) using parameters from literature and interview sources. There are other model tests which consider Experimental Validity which I did not perform. With suitable numeric data, running these tests would improve this aspect of my model's *validity*.

With respect to results:

- **Results presented may be *unintentionally misleading*[250].** I tested this using appropriate validation tests (e.g., CM-S). I also present tables allowing the reader to compare raw mental data with my resulting model, one variable at a time and then as a causal chain, alongside my written interpretation of how the quotation supports the links in the chain (i.e., Causal RIQs Appendix F). This thorough examination tests my interpretive process by repeatedly checking the *cohesion* between subsets of model structure and *validation* data[117] (p. 313)
- **The model is not capable of exploring methods for and impacts obtained from mental model changes *during* implementation.** A rough approximation might be switching from one preference (table function) to another part way through the simulation. The model does however show how having different preferences impacts PCT differently.

- **There are other policies that could conceivably be implemented.** I focused on the policies that were being considered at the HSDO when I collected the data as these were the policies possible at the time (and ones whose implementation was well-understood). Since then, additional policies have been proposed, for example: the creation of Accountable Care Organizations, and bundled payments, as well as special payments for comprehensive care services (e.g., for Transition Care Management, CPT code 99496)[23]. These are not explicitly in the model.

4.4. FUTURE RESEARCH

Numerous ideas for future research have come out of this work, both for PCT and for methods development. Below I present some examples.

Specifically, for PCT:

- **Study the use of the model and findings by stakeholders** (e.g., what do they learn, what elements of the results help them to learn it, how does that impact their behavior and outcomes).
- **Improve the Simulation Model** by collecting data on the sensitive variables, including on shapes of table functions describing preferences.
- **Improve the Theoretical Model** by assessing loop dominance sequences and by assessing possibilities for simplification.
- **Conduct *bounded rationality* model analysis**[108, 109]to portray how the tensions arise from the interaction between the general structure of primary care service delivery and mental models, and to explore the impact of different information-processing structures (e.g., considering *adaptive reserve* (Section 4.2.1.5) in clinician decision-functions).
- **Test the model against data** (mental and/or numeric data) to evaluate its usefulness in slightly-different contexts (e.g., small, medium and large HSDOs within the US) and to expand its range in very-different ones (e.g., the US military health system, the US veterans' health system, and HSDOs in other countries). In 2015, the World Health Organization launched its *global strategy on people-centered and integrated health service*[296] (p. 48). Future research in this arena could thus use (and test) the model in specific country contexts to inform policy decisions.
- **Explore the relevance and potential use of findings to inform health systems strengthening efforts more broadly** (e.g., to the health goal of the sustainable development goals[296]).
- **Testing the model with respect to rival theories** for PCT (e.g., risk-stratified patient panels, diffusion of innovation at clinic level, advanced access).
- **Modifying the model to parse out the different professions** (and the different tasks that they are licensed to perform) on care teams as they become engaged in PCT.
- **Further exploring the role of systems thinking**[169, 181, 182] and how that impacts PCT, with its accompanying mental model changes, building on the work of Pratt *et al.*[238].

- **Consider** how best to mobilize social networks to change mental models in the context of PCT.
- **Modifying the model to explore changes in preferences** part-way through PCT.
- **Modifying the model to include additional policies** (see limitations).
- **Modifying the model to include other aspects of the health system** (e.g., urgent care, specialty visits and hospital readmissions) and making endogenous the consequences of delivering a lower level of adherence to clinical guidelines (i.e., sicker patients, seeking more care outside primary care, experiencing more hospital readmissions – morbidity and mortality).
- **Modifying the model so that clinicians and the patient panel can change over time** (e.g., new clinicians build up a panel).
- **Modifying the model to include composite variables** like adequate staffing and adaptive reserve – calculations that are only possible in the model and could potentially be used to improve decision-making.
- **Producing a model that can be used for online** simulation games, policy analysis and other purposes (e.g., education to enhance learning under dynamic complexity).

For methods and theory:

- I have shown how I did certain things and I want to find out how others are approaching the same issue (**i.e., broader literature review**).
    - Surveying the methods used in *scoping studies* broadly (in SDM, simulation and social science) to determine the nature of a problem (i.e., how are dynamic complexity and system-structure-as-cause observed) and thus the methods to be used for better understanding it.
    - Surveying the methods used inside SDM research to obtain and supply evidence that system dynamics-related elements of system structure are enduring and significant.
    - Surveying the methods used by others claiming to use *paradigm-crossing strategies*.
- **Develop explicit strategies for communicating** the various results developed here for *stories with numbers* and explore their usefulness (engagement, improved understanding, and changed behavior).
- Continuing to develop an explicit strategy for more closely working with **stakeholders** during implementation of these methods.

- **Explore generalizability of methods and theory** in a totally different research and problem context and assess their usefulness.
    - Use my **interplay paradigm-crossing strategy** (or pieces of it) in research – do so using system dynamics as I have done or using other methods.
    - **For those preferring *incommensurability* or *pragmatic/ integrationist* approach (over paradigm-crossing)**, future research would be to use the theoretical frameworks presented above to try reflecting on their work in terms of its approach to dealing with multiple paradigms.
    - Use the **validity table** presented in Chapter 2 (Table 2.9) to consider which validation methods one should use such that all validation subtypes are tested. Also use proposed method for assessing cessation.
    - **Assess cognitive aspects** of my methods with potential study designs from the research agenda proposed by Doyle[277]; linking with integrative theories to explore where mental models come from (e.g., the cognitive styles of professions raised by Rittell[297] and observed in the Shared Understanding Diagram) as proposed by Senge[285] and Lane[101, 102].
- In parallel with the previous point, assess how best to **modify the methods** presented here for **low-resource research settings** (e.g., low-funding, fast-turnaround or crisis situations) or in **different types of teams** (e.g., governmental, interdisciplinary, organizationally-embedded).
- Identify/devise specific **methods for use** where similarities between collective inner reality and individual outer reality are brought together (transition zone F).

4.5. CONCLUSION

The growing importance of primary care transformation (PCT) has led to calls for methods to be developed which place equal weight on qualitative and quantitative information, utilizing an (as yet unidentified) integrative paradigm approach[99, 221] (p. 294, 351-355), considering emotional engagement[124], and in an empirically-valid way[124]. Similarly, system dynamics researchers have recently called for more rigorous use of qualitative methods in model conceptualization and theory development[138], for a deeper and broader consideration of the ways SDM studies engage with individual and group understanding at a subconscious level[100, 289, 298], and for increased use and transparency of model validation[161, 162].

In Chapter 1, I described the need for a methodological aim in order to adequately meet the theoretical aim undertaken in this dissertation. This aim is divided into three objectives: addressing paradigm issues, engaging the subconscious emotional level on which decision-making is based, and addressing issues of empirical validity. Using a mixed methods process (Appendix C), I selected system dynamics modeling (SDM) methods and relied primarily upon interview transcripts for this dissertation). The SDM standard method provided useful methods to analyze the descriptive mental data found in these transcripts, from problem articulation through to theory development. Along the way, I designed and justified the design of new methods (see Chapter 2) using scholarly works on *system dynamics theory* and other modeling theory and health services research(ones already referenced in the dissertation and others[178, 283, 291, 299, 300]. These methods consider paradigm issues, emotional engagement and empirical validity. My results (Chapter 3) then serve as a case for using these methods and this final chapter provides a reflection on my experience.

Both mixed methods in health services research (MMHSR) and SDM were not designed with a pre-existing *social* theory in mind, leaving their sociological paradigms an open question and scholars in both fields view the task of clarifying the philosophical basis as widely neglected[101, 287]. In both cases, scholars have called for improving the way qualitative methods are used with the end of developing an integrative paradigm in both MMHSR[99, 221] (p. 294, 351-355)[96] and SDM[101, 102, 104, 105, 284]. In this work, I have contributed to theory, methods and cases and thus ultimately to a nascent *integrative paradigm using interplay to cross the Four Quadrants of Knowledge* that I hope will be developed further through future research and strengthen the work currently being done in both fields.

SDM is typically concerned with problems involving diverse, contradictory perspectives (referred to in the literature as strategic or wicked problems, messes, a swamp[106]). So, model conceptualization naturally involves developing a "shared diagnostic framing if progress is to be made"[298] (p. 80). Despite the central nature of this aspect of research, the manner in which diverse perspectives of participants are resolved in model conceptualization is not a part of the formal methodology[106].

That said, Morecroft's text on the SDM standard method recognizes these issues and describes conceptualization as an art form balancing creatively hearing out diverse perspectives with "ruthless pruning" toward a convergent "sharable" model[114] (p. 6). I heartily agree on all three points. **Regarding pruning:** I had to remove significant amounts of excess material in the form duplicative variables and dynamically-weak links (those not involving delays or feedback loops) before developing a model that I could safely assume to be a shared understanding or which could even begin to tell the stories about tensions and variation in my problem statement. **Regarding creativity**: being a novice, this process forced me to engage creatively with my data which eventually resulted in the development of new methods. **Regarding the art form:** I hope that the contributions I outline above can add more science to the craft[301], enabling researchers to better improve the human condition[97].

This hope is grounded in Sterman & Wittenberg's explanation that research can improve a paradigm's potential for long-term success by "strengthening the **processes** of theory articulation and testing" through the "development of **tools** tailored for particular contexts" which allow for "**more rigorous** theory development [and testing]" and providing **"studies** testing the … theories", [286] (p. 338).

I see three underlying problems as having prevented progress on these qualitative aspects of methods development in SDM. First, the debate around qualitative methods in system dynamics has not received the *continued* attention it merits[203]. Second, there are circumstances which constrain researchers wishing to address this. These are "the milieu in which the field is based; the field's search for respectability; difficulties in boundary setting within the field … the social constructions which govern academic careers … and the perceived lack of palatable alternatives"[302] (p. 254). Third, SDM researchers are people with "personalities, nationalities, aesthetic values, faith, and other personal and social factors"[303] (p. 45) . These constraints underlie paradigm issues across domains and, in my experience, they apply to both MMHSR and SDM.

Lane argues that SDM researchers should engage and document their process and think about it in terms of theory, to "generalize experiences… to abstract away from specific cases and to try to look at appropriate generalities"[145] (p. 566). I have documented my process and have thought about it in terms of theory. In doing so, I found that paradigm-crossing was a very useful way of organizing my thoughts.

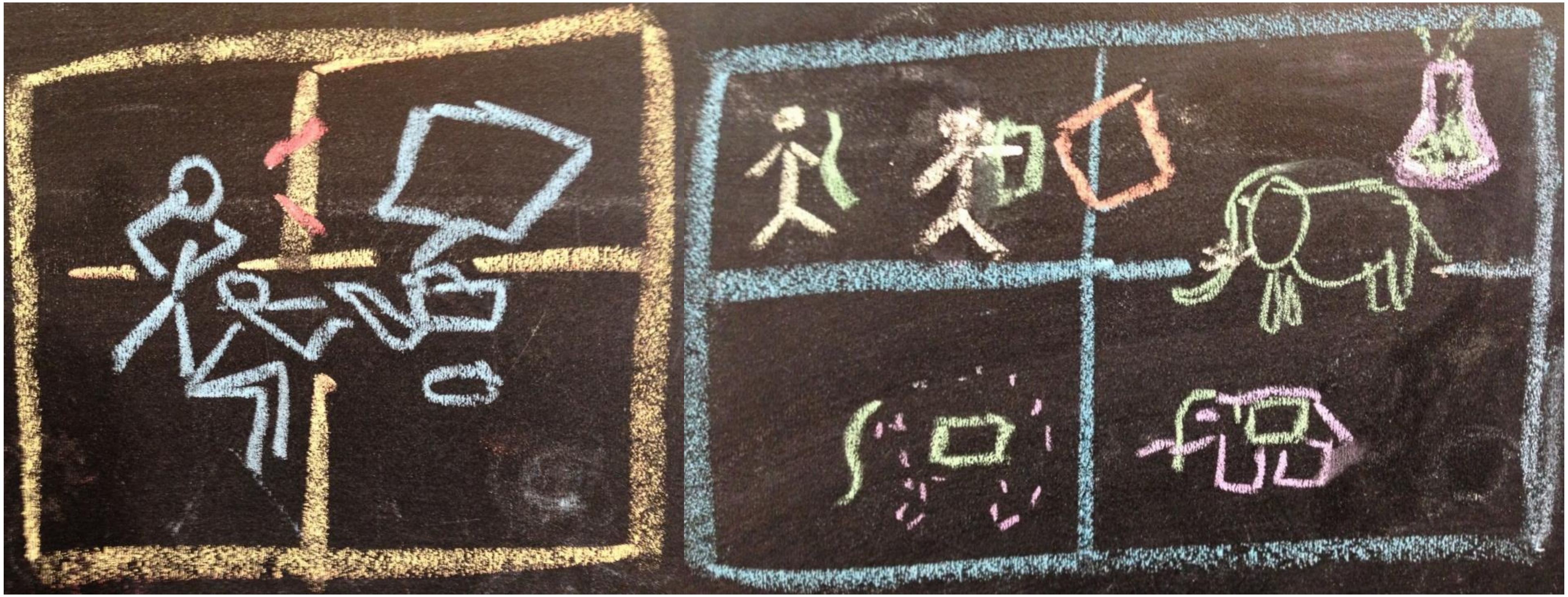
Lane states that if we choose to explore a paradigm crossing/integrative approach, we gain the *opportunity to engage* with an important debate in social theory[102]. And I have found this to be true as well, as my experience has led me to proposing system dynamics as an integrative paradigm – thus unintentionally entering the debate in social theory regarding paradigm incommensurability.

I end this dissertation with a final reflection on where I am now relative to where I began. I began with *connecting the dots*[304] (and the desire to do so more broadly than just in 2 cases, to gain understanding) and SDM standard methods (and the desire to use qualitative data for rigorous conceptualization as quantitative data fell short).

On the left of Figure 4.12 below, I show, in simplistic terms, my interpretation of how a *best practices* system dynamics research project would map to the *Four Quadrants of Knowledge framework*: stakeholders provide a modeler with their understanding of the problem, the modeler sits with the group and portrays the implications of their understanding in a way that was previously inaccessible to the stakeholder (hence their perception crosses from upper-left to upper-right quadrants). In this practice, stakeholders see outer-reality as a system by means of a computer screen showing a simulation model.

I juxtapose the *best practices* with my strategy by placing the latter on the right of Figure 4.12. To describe this process in simplistic terms, I will borrow from the allegory of the blind men and the elephant[305, 306]. There is a window (orange box) through which stakeholders touch (experience) the outer reality (the system, the green elephant on the right side). Stakeholders provide diverse perspectives of the system – they each see a part. I bring those perspectives together in the lower left quadrant and the semblance of the elephant starts to take shape we can see how the pieces that these people identified link together – with some connecting of the dots left unfinished. Then, that understanding is converted into a simulation model (purple boxy elephant) that is a problem-oriented, adequate approximation of the real system (the green elephant on the right side). Finally, that simulation model is used to enable experiments (purple beaker with bubbling green fluid) for stakeholders to learn from without harming the elephant, and thereby improving their understanding of the real system.

Figure 4.12 Simplistic Views of SDM Practice – Standard Practice & Interplay Strategy



Validation builds confidence but emotion leads to change. Confidence builds emotion to change when addressing all the validity types in a way that stakeholders can engage with and understand. Also reflecting on system dynamics' potential for changing mental models, Black reasons that "math equations are not concrete to many people … without elaborated explanations and 'telling the story', models can appear opaque and un-understandable (not concrete) to participants"[298] (p. 82). Weaving together *stories with numbers* in an integrative paradigm puts qualitative and quantitative research on equal footing, and in so doing also provides stakeholders with the tools, the concrete empirically-valid results to bring about understanding and a desire to change.

Returning to the allegory above, when I started upon this research, I often heard: "once you've seen one, you've seen one." Using an integrative methodology helped me to see that everyone was seeing the elephant but that they did not have the tools to bring their understanding together and thus thought they saw different animals. Having a model which can tell many stories allows one to see the elephant in each of the different perspectives.

Within the HSDO I studied, some participants cited problems as reasons for their resistance to implementation and others described addressing the same problems in their continued quest for PCT. The theory presented here is based on all of these experiences that were described. The Theoretical Model provides a brief summary of the structures leading to success and failure. The Simulation Model allows one to explore the dynamics in these structures, solidifying their understanding such that they can find the solutions that will meet the particular challenges found in their implementation of PCT.

Specifically, simulation allows us to run experiments with different governing realities. It allows us to run hypothetical controlled trials – in so doing, it addresses the inability to conduct randomized controlled trials in the transformation of primary care.

Our current reality is that there is a misalignment between payers and our system of care delivery. We *cannot wait* until the payment model is changed. Health care providers and administrators are operating the health system today – they need solutions that will work today – the best that can be attained under the current circumstances. The simulation model identifies leverage points – care team members' preferences as well as management policies – that are conducive to successful PCT. Preferences and polices have broadly been talked about before (*cf*., Cronholm and colleagues[124, 195]); the difference is that I am identifying the ones that matter and when/how they matter in the system of service delivery (as I provide an understanding of the system structure within which they operate). The theoretical and simulation models portray how they influence each other and how they synergistically affect the outcome given the feedbacks and delays among them.

Our current reality is also that, as the cliché goes: *healthcare is always changing*. We are doing our best to *keep up* with these changes and keep the goal of PCT in our sights – to continue working toward it and to sustain the gains we make. New structure can be added to the simulation model to explore the impact of such contextual changes and how management policy recommendations and care team member preferences may need to shift to accommodate those changes. In so doing, care teams and administrators can successfully retain their goal of PCT despite the *winds of change* (those currently felt as well as those only forecasted).

I have developed a grounded, dynamic theory of primary care transformation. In so doing, I hope to build a fuller understanding of *the connections among components* of the primary care system, and how these relate to *the complex interactions among stakeholders* (and with their environment) as they generate successful, lackluster and failed PCT implementation experiences. Left unrecognized and/or misunderstood, these structures have made PCT an elusive target.

In the US experience, few HSDOs have attained this success[307]. Also, there are many which have not yet begun this transformation in earnest—all can learn from past experience and improve. Internationally, there is a call for PCT[95, 296, 308, 309], yet achieving transformation has long proven elusive[95]. Ultimately, PCT is a journey, as healthcare is always facing new challenges, but some organizations have seen significant improvement in adherence to clinical guidelines. I believe that translating the implications of this research into practice would greatly facilitate the intended broad-sweeping transformation of primary care.

REFERENCES

1. Bourdieu, P., *Vive la crise!* Theory and society, 1988. **17**(5): p. 773-787.
2. Stange, K.C., W.L. Miller, and I. McWhinney, *Developing the knowledge base of family practice.* Family Medicine 2001. **33**(4): p. 286-297.
3. Agency for Healthcare Research & Quality. *Award Recipients: Transforming Primary Care Practice*. 2011 [cited 2011 December 30]; Available from: http://www.ahrq.gov/research/findings/factsheets/primary/transpcaw/index.html.
4. Kozakowski, S.M., K. Becher, T. Hinkle, R. Blackwelder, C. Knight Jr, A. Bentley, and P.A. Pugno, *Responses to Medical Students' Frequently Asked Questions About Family Medicine.* American family physician, 2016. **93**(3): p. Online-Online.
5. World Health Organization, *Declaration of Alma-Ata*, in *International Conference on Primary Health Care*. 1978: Alma-Ata, USSR. p. 3.
6. American Family Physician. *Immunizations (excluding Influenza)*. 2017 [cited 2017 May 5]; Available from: http://www.aafp.org/afp/topicModules/viewTopicModule.htm?topicModuleId=63.
7. Twiddy, D., *Chronic Care Management in the Real World.* Family practice management, 2014. **22**(5): p. 35-41.
8. Scholle, S.H., S.E. Asche, S. Morton, L.I. Solberg, M.A. Tirodkar, and C.R. Jaén, *Support and Strategies for Change Among Small Patient-Centered Medical Home Practices.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S6-S13.
9. MedlinePlus. *Choosing a primary care provider*. 2017; Available from: https://medlineplus.gov/ency/article/001939.htm.
10. Sterman, J.D., *All models are wrong: reflections on becoming a systems scientist.* System Dynamics Review, 2002. **18**(4): p. 501-531.
11. Sekscenski, E.S., S. Sansom, C. Bazell, M.E. Salmon, and F. Mullan, *State practice environments and the supply of physician assistants, nurse practitioners, and certified nurse-midwives.* New England Journal of Medicine, 1994. **331**(19): p. 1266-1271.
12. Ritter, A. and T. Hensen-Turton, *The primary care paradigm shift: an overview of the state-level legal framework governing nurse practitioner practice.* Health Law, 2007. **20**: p. 21.
13. Sinsky, C.A., R. Willard-Grace, A.M. Schutzbank, T.A. Sinsky, D. Margolius, and T. Bodenheimer, *In search of joy in practice: a report of 23 high-functioning primary care practices.* The Annals of Family Medicine, 2013. **11**(3): p. 272-278.
14. Bodenheimer, T. and B.Y. Laing, *The teamlet model of primary care.* The Annals of Family Medicine, 2007. **5**(5): p. 457-461.

15. Farrell, T., A. Tomoaia-Cotisel, D.L. Scammon, K. Brunisholz, J. Kim, J. Day, L.H. Gren, S. Wallace, K. Gunning, J. Tabler, and M.K. Magill, *Impact of an Integrated Transitions Management Program in Primary Care on Hospital Readmissions.* Journal of Healthcare Quality 2015. **37**(1): p. 81-92.
16. Tomoaia-Cotisel, A., T.W. Farrell, L.I. Solberg, C.A. Berry, N.S. Calman, P.F. Cronholm, K.E. Donahue, D.L. Driscoll, D. Hauser, J.W. McAllister, S. Mehta, R.J. Reid, M. Tai-Seale, C.G. Wise, M.D. Fetters, J.S. Holtrop, H.P. Rodriguez, C.P. Brunker, R. Day, E.L. McGinley, D.L. Scammon, M. Harrison, J.L. Genevro, R. Gabbay, and M.K. Magill, *Implementation of Care Management: An Analysis of Recent AHRQ Research.* Medical Care Research & Review, 2016.
17. Farrell, T., A. Tomoaia-Cotisel, D.L. Scammon, J. Day, R. Day, and M.K. Magill, *Care management: Implications for medical practice, health policy, and health services research*. 2015, Agency for Healthcare Research and Quality: Rockville, MD.
18. Tomoaia-Cotisel, A. and T. Farrell, *Care Management: Key Elements and Implications for Medical Practice, Health Policy and Health Services Research*, in *AHRQ Webinar*, J.L. Genevro, Editor. 2015, Agency for Healthcare Research and Quality.
19. Bodenheimer, T. and R. Berry-Millett, *Care management of patients with complex health care needs.* Policy, 2009. **1**: p. 6.
20. Oboler, S.K., A.V. Prochazka, R. Gonzales, S. Xu, and R.J. Anderson, *Public expectations and attitudes for annual physical examinations and testing.* Annals of internal medicine, 2002. **136**(9): p. 652-659.
21. Becker, H.I., M.R. Longacre, and D.M. Harper, *Beyond the Pap: assessing patients' priorities for the annual examination.* Journal of Women's Health, 2004. **13**(7): p. 791-799.
22. Prochazka, A.V., K. Lundahl, W. Pearson, S.K. Oboler, and R.J. Anderson, *Support of evidence-based guidelines for the annual physical examination: a survey of primary care providers.* Archives of Internal Medicine, 2005. **165**(12): p. 1347-1352.
23. American Medical Association. *CPT Network and Knowledge Base for Coding Inquiries*. 2017 [cited 2017 December 23]; Available from: https://apps.ama-assn.org/cptkb/.
24. Institute of Medicine Division of Health Manpower Resources Development, *A manpower policy for primary health care: report of a study*. 1978, National Academy of Sciences.
25. Donaldson, M.S., K.D. Yordy, K.N. Lohr, and N.A. Vanselow, *Primary care: America's health in a new era*. 1996, National Academies Press.
26. Starfield, B., L. Shi, and J. Macinko, *Contribution of primary care to health systems and health.* Milbank Q, 2005. **83**(3): p. 457-502.

27. American Academy of Family Physicians, American Academy of Pediatrics, American College of Physicians, and American Osteopathic Association. *Joint Principles of the Patient Centered Medical Home*. 2007 [cited 2008 September 1]; Available from: http://www.pcpcc.net/content/joint-principles-patient-centered-medical-home.

28. Stange, K.C., P. Nutting, W. Miller, C. Jaén, B. Crabtree, S. Flocke, and J. Gill, *Defining and Measuring the Patient-Centered Medical Home.* Journal of General Internal Medicine, 2010. **25**(6): p. 601-612.

29. Shi, L., *The impact of primary care: a focused review.* Scientifica, 2012. **2012**.

30. Rohde, J., S. Cousens, M. Chopra, V. Tangcharoensathien, R. Black, Z.A. Bhutta, and J.E. Lawn, *30 years after Alma-Ata: has primary health care worked in countries?* The Lancet, 2008. **372**(9642): p. 950-961.

31. Lewin, S., J.N. Lavis, A.D. Oxman, G. Bastías, M. Chopra, A. Ciapponi, S. Flottorp, S.G. Martí, T. Pantoja, G. Rada, N. Souza, S. Treweek, C.S. Wiysonge, and A. Haines, *Supporting the delivery of cost-effective interventions in primary health-care systems in low-income and middle-income countries: an overview of systematic reviews.* The Lancet. **372**(9642): p. 928-939.

32. Baicker, K. and A. Chandra, *Medicare spending, the physician workforce, and beneficiaries' quality of care.* Health Aff (Millwood), 2004. **23**(3): p. 291-291.

33. Chan, M., *Return to Alma-Ata.* The Lancet, 2008. **372**(9642): p. 865-866.

34. Lawn, J.E., J. Rohde, S. Rifkin, M. Were, V.K. Paul, and M. Chopra, *Alma-Ata 30 years on: revolutionary, relevant, and time to revitalise.* The Lancet, 2008. **372**(9642): p. 917-927.

35. Gunn, J.M., V.J. Palmer, L. Naccarella, R. Kokanovic, C.J. Pope, J. Lathlean, and K.C. Stange, *The promise and pitfalls of generalism in achieving the Alma-Ata vision of health for all.* Med J Aust, 2008. **189**(2): p. 110-112.

36. Gunn, J., L. Naccarella, V. Palmer, R. Kokanovic, C. Pope, and J. Lathlean, *What is the place of generalism in the 2020 primary care team.* Canberra: APHCRI, 2007.

37. Starfield B, *Primary care: balancing health needs, services and technology*. 1998, Oxford University Press: New York.

38. Shi, L., J. Macinko, B. Starfield, R. Politzer, J. Wulu, and J. Xu, *Primary care, social inequalities and all-cause, heart disease and cancer mortality in US counties: a comparison between urban and non-urban areas.* Public Health, 2005. **119**(8): p. 699-710.

39. Shi, L., J. Macinko, B. Starfield, J. Wulu, J. Regan, and R. Politzer, *The relationship between primary care, income inequality, and mortality in US States, 1980–1995.* The Journal of the American Board of Family Practice, 2003. **16**(5): p. 412-422.

40. Shi, L., J. Macinko, B. Starfield, J. Xu, J. Regan, R. Politzer, and J. Wulu, *Primary care, infant mortality, and low birth weight in the states of the USA.* Journal of Epidemiology and Community Health, 2004. **58**(5): p. 374-380.
41. Starfield, B., *New paradigms for quality in primary care.* The British Journal of General Practice, 2001. **51**(465): p. 303-309.
42. Shi, L. and B. Starfield, *The Effect of Primary Care Physician Supply and Income Inequality on Mortality Among Blacks and Whites in US Metropolitan Areas.* American Journal of Public Health, 2001. **91**(8): p. 1246-1250.
43. Friedberg, M.W., P.S. Hussey, and E.C. Schneider, *Primary care: a critical review of the evidence on quality and costs of health care.* Health Affairs, 2010. **29**(5): p. 766-772.
44. Alexander, J.A., M. Paustian, C.G. Wise, L.A. Green, M.D. Fetters, M. Mason, and D.K. El Reda, *Assessment and Measurement of Patient-Centered Medical Home Implementation: The BCBSM Experience.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S74-S81.
45. Berry, C.A., T. Mijanovich, S. Albert, C.H. Winther, M.M. Paul, M.S. Ryan, C. McCullough, and S.C. Shih, *Patient-Centered Medical Home Among Small Urban Practices Serving Low-Income and Disadvantaged Patients.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S82-S89.
46. Calman, N.S., D. Hauser, L. Weiss, E. Waltermaurer, E. Molina-Ortiz, T. Chantarat, and A. Bozack, *Becoming a Patient-Centered Medical Home: A 9-Year Transition for a Network of Federally Qualified Health Centers.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S68-S73.
47. Dohan, D., M.H. McCuistion, D.L. Frosch, D.Y. Hung, and M. Tai-Seale, *Recognition as a Patient-Centered Medical Home: Fundamental or Incidental?* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S14-S18.
48. Donahue, K.E., J.R. Halladay, A. Wise, K. Reiter, S.-Y.D. Lee, K. Ward, M. Mitchell, and B. Qaqish, *Facilitators of Transforming Primary Care: A Look Under the Hood at Practice Leadership.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S27-S33.
49. Driscoll, D.L., V. Hiratsuka, J.M. Johnston, S. Norman, K.M. Reilly, J. Shaw, J. Smith, Q.N. Szafran, and D. Dillard, *Process and Outcomes of Patient-Centered Medical Care With Alaska Native People at Southcentral Foundation.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S41-S49.
50. Gabbay, R.A., M.W. Friedberg, M. Miller-Day, P.F. Cronholm, A. Adelman, and E.C. Schneider, *A Positive Deviance Approach to Understanding Key Features to Improving Diabetes Care in the Medical Home.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S99-S107.

51. McAllister, J.W., W.C. Cooley, J. Van Cleave, A.A. Boudreau, and K. Kuhlthau, *Medical Home Transformation in Pediatric Primary Care—What Drives Change?* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S90-S98.

52. McMullen, C.K., J. Schneider, A. Firemark, J. Davis, and M. Spofford, *Cultivating Engaged Leadership Through a Learning Collaborative: Lessons From Primary Care Renewal in Oregon Safety Net Clinics.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S34-S40.

53. Reid, R.J., E.A. Johnson, C. Hsu, K. Ehrlich, K. Coleman, C. Trescott, M. Erikson, T.R. Ross, D.T. Liss, D. Cromp, and P.A. Fishman, *Spreading a Medical Home Redesign: Effects on Emergency Department Use and Hospital Admissions.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S19-S26.

54. Rittenhouse, D.R., L. Schmidt, K. Wu, and J. Wiley, *Contrasting trajectories of change in primary care clinics: lessons from New Orleans safety net.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S60-S67.

55. Solberg, L.I., A.L. Crain, J. Tillema, S.H. Scholle, P. Fontaine, and R. Whitebird, *Medical Home Transformation: A Gradual Process and a Continuum of Attainment.* The Annals of Family Medicine, 2013. **11**(Suppl 1): p. S108-S114.

56. Day, J., D.L. Scammon, J. Kim, A. Sheets-Mervis, R. Day, A. Tomoaia-Cotisel, N.J. Waitzman, and M.K. Magill, *Quality, satisfaction, and financial efficiency associated with elements of primary care practice transformation: preliminary findings.* Ann Fam Med, 2013. **11 Suppl 1**: p. S50-9.

57. Tomoaia-Cotisel, A., D.L. Scammon, N.J. Waitzman, P.F. Cronholm, J.R. Halladay, D.L. Driscoll, L.I. Solberg, C. Hsu, M. Tai-Seale, V. Hiratsuka, S.C. Shih, M.D. Fetters, C.G. Wise, J.A. Alexander, D. Hauser, C.K. McMullen, S.H. Scholle, M.A. Tirodkar, L. Schmidt, K.E. Donahue, M.L. Parchman, and K.C. Stange, *Context matters: the experience of 14 research teams in systematically reporting contextual factors important for practice change.* The Annals of Family Medicine, 2013. **11 Suppl 1**: p. S115-23.

58. Magill, M.K., R.L. Lloyd, D. Palmer, and S.A. Terry, *Successful turnaround of a university-owned, community-based, multidisciplinary practice network.* The Annals of Family Medicine, 2006. **4**(suppl 1): p. S12-S18.

59. University of Utah Health. *About*. 2017 [cited 2017 April 1]; Available from: http://healthcare.utah.edu/about/.

60. Tomoaia-Cotisel, A., S.D. Allen, D.L. Scammon, and D. Fuller. *Exploring the Impact of Perceived Versus Actual Competition on a Clinic's Sensitivity to its Stakeholders*. in

Association for Marketing & Health Care Research Annual Conference. 2013. Big Ski, MT: presented by Andrada Tomoaia-Cotisel.

61. *AHRQ Conference*. in *Moving Ahead: Leveraging Knowledge and Action to Improve Health Care Quality*. 2012. North Bethesda, MD.
62. Martsolf, G.R., R. Kandrack, E.C. Schneider, and M.W. Friedberg, *Categories of practice transformation in a statewide medical home pilot and their association with medical home recognition.* Journal of general internal medicine, 2015. **30**(6): p. 817-823.
63. Crabtree, B.F., P.A. Nutting, W.L. Miller, K.C. Stange, E.E. Stewart, and C.R. Jaén, *Summary of the National Demonstration Project and recommendations for the patient-centered medical home.* The Annals of Family Medicine, 2010. **8**(Suppl 1): p. S80-S90.
64. Solberg, L.I., *How can we remodel practices into medical homes without a blueprint or a bank account?* The Journal of ambulatory care management, 2011. **34**(1): p. 3-9.
65. Friedberg, M.W., M.B. Rosenthal, R.M. Werner, K.G. Volpp, and E.C. Schneider, *Effects of a medical home and shared savings intervention on quality and utilization of care.* JAMA internal medicine, 2015. **175**(8): p. 1362-1368.
66. Friedberg, M.W., E.C. Schneider, M.B. Rosenthal, K.G. Volpp, and R.M. Werner, *Association between participation in a multipayer medical home intervention and changes in quality, utilization, and costs of care.* JAMA, 2014. **311**(8): p. 815-825.
67. Grumbach, K., T. Bodenheimer, and P. Grundy *The Outcomes of Implementing Patient-Centered Medical Home Interventions: A Review of the Evidence on Quality, Access and Cost from Recent Prospective Evaluation Studies*. Ann Fam Med, 2009.
68. Strickland, B.B., J.R. Jones, R.M. Ghandour, M.D. Kogan, and P.W. Newacheck, *The medical home: health care access and impact for children and youth in the United States.* Pediatrics, 2011. **127**(4): p. 604-11.
69. Rosenthal, T.C., *The Medical Home: Growing Evidence To Support a New Approach to Primary Care.* Journal of the American Board of Family Medicine, 2008. **21**(5): p. 427-440.
70. Shi, L., D.C. Lee, M. Chung, H. Liang, D. Lock, and A. Sripipatana, *Patient-Centered Medical Home Recognition and Clinical Performance in U.S. Community Health Centers.* Health Serv Res, 2016.
71. Institute of Medicine, *Primary Care: America's Health in a New Era*. 1996, Institute of Medicine.
72. Bodenheimer, T., E. Chen, and H.D. Bennett, *Confronting The Growing Burden Of Chronic Disease: Can The U.S. Health Care Workforce Do The Job?* Health Affairs, 2009. **28**(1): p. 64-74.

73. Egan, B.M., Y. Zhao, and R.N. Axon, *US trends in prevalence, awareness, treatment, and control of hypertension, 1988-2008.* Jama, 2010. **303**(20): p. 2043-2050.
74. Ford, E.S., C. Li, W.S. Pearson, G. Zhao, and A.H. Mokdad, *Trends in hypercholesterolemia, treatment and control among United States adults.* International journal of cardiology, 2010. **140**(2): p. 226-235.
75. Cheung, B.M., K.L. Ong, S.S. Cherny, P.-C. Sham, A.W. Tso, and K.S. Lam, *Diabetes prevalence and therapeutic target achievement in the United States, 1999 to 2006.* The American journal of medicine, 2009. **122**(5): p. 443-453.
76. McGlynn, E.A., S.M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E.A. Kerr, *The quality of health care delivered to adults in the United States.* New England journal of medicine, 2003. **348**(26): p. 2635-2645.
77. Østbye, T., K.S.H. Yarnall, K.M. Krause, K.I. Pollak, M. Gradison, and J.L. Michener, *Is There Time for Management of Patients With Chronic Diseases in Primary Care?* The Annals of Family Medicine, 2005. **3**(3): p. 209-214.
78. Yarnall, K.S., T. Ostbye, K. Krause, K. Pollak, M. Gradison, and J. Michener, *Family physicians as team leaders:"time" to share the care.* Preventing Chronic Disease, 2009. **6**(2): p. 1-6.
79. Yarnall, K.S., K.I. Pollak, T. Østbye, K.M. Krause, and J.L. Michener, *Primary care: is there enough time for prevention?* American journal of public health, 2003. **93**(4): p. 635-641.
80. John Billings, Nina Parikh, and Tod Mijanovich, *Emergency Department Use: The New York Story*. 2000, The Commonwealth Fund: New York.
81. Bradley, C.J., S.O. Gandhi, D. Neumark, S. Garland, and S.M. Retchin, *Lessons For Coverage Expansion: A Virginia Primary Care Program For The Uninsured Reduced Utilization And Cut Costs.* Health Affairs, 2012. **31**(2): p. 350-359.
82. Smulowitz, P.B., R. Lipton, J.F. Wharam, L. Adelman, S.G. Weiner, L. Burke, C.W. Baugh, J.D. Schuur, S.W. Liu, M.E. McGrath, B. Liu, A. Sayah, M.C. Burke, J.H. Pope, and B.E. Landon, *Emergency Department Utilization After the Implementation of Massachusetts Health Reform.* Annals of Emergency Medicine, 2011. **58**(3): p. 225-234.e1.
83. Rubenstein, L.V., E.M. Yano, A. Fink, A.B. Lanto, B. Simon, M. Graham, and A.S. Robbins, *Evaluation of the VA's Pilot Program in Institutional Reorganization toward Primary and Ambulatory Care: Part I, Changes in process and outcomes of care.* Academic Medicine, 1996. **71**(7): p. 772-83.

84. Misky, G., H. Wald, and E. Coleman, *Post-hospitalization transitions: Examining the effects of timing of primary care provider follow-up.* J Hosp Med, 2010. **5**(7): p. 392-397.

85. Coleman, E.A., J.D. Smith, J.C. Frank, S.J. Min, C. Parry, and A.M. Kramer, *Preparing patients and caregivers to participate in care delivered across settings: the Care Transitions Intervention.* Journal of the American Geriatrics Society, 2004. **52**(11): p. 1817-1825.

86. Tabler, J., D.L. Scammon, J. Kim, T. Farrell, A. Tomoaia-Cotisel, J. Day, and M.K. Magill, *Patient Care Experiences and Perceptions of the Patient-Provider Relationship: A Mixed Method Study.* The Patient Experience Journal, 2014. **1**(1): p. 75-87.

87. Saultz, J.W. and W. Albedaiwi, *Interpersonal continuity of care and patient satisfaction: a critical review.* The Annals of Family Medicine, 2004. **2**(5): p. 445-451.

88. Adler, R., A. Vasiliadis, and N. Bickell, *The relationship between continuity and patient satisfaction: a systematic review.* Family Practice, 2010. **27**(2): p. 171-178.

89. Stange, K.C. and R.L. Ferrer, *The paradox of primary care.* The Annals of Family Medicine, 2009. **7**(4): p. 293-299.

90. Nutting, P.A., M.A. Goodwin, S.A. Flocke, S.J. Zyzanski, and K.C. Stange, *Continuity of primary care: to whom does it matter and when?* The Annals of Family Medicine, 2003. **1**(3): p. 149-155.

91. Marshall, M.N., P.S. Romano, and H.T. Davies, *How do we maximize the impact of the public reporting of quality of care?* Int J Qual Health Care, 2004. **16 Suppl 1**: p. i57-63.

92. Committee on Quality of Health Care in America, *Crossing the Quality Chasm: A New Health System for the 21st Century. Committee on Quality of Health Care in America*. 2001, Institute of Medicine National Academy Press: Washington, D.C.

93. Bodenheimer, T., *Primary care--will it survive?* N Engl J Med, 2006. **355**(9): p. 861-4.

94. Starfield, B., *Is US health really the best in the world?* Jama, 2000. **284**(4): p. 483-5.

95. World Health Organization, *Primary health care: now more than ever*. 2008, World Health Organization.

96. Miller, W.L., B.F. Crabtree, M.I. Harrison, and M.L. Fennell, *Integrating Mixed Methods in Health Services and Delivery System Research.* Health Services Research, 2013. **48**(6pt2): p. 2125-2133.

97. Hoover, K.R. and T. Donovan, *The Elements of Social Scientific Thinking*. 2001, Cengage Learning.

98. Trigg, R., *Understanding Social Science: Philosophical Introduction to the Social Sciences*. 2000, Wiley.

99. Sale, J.E.M., L.H. Lohfeld, and K. Brazil, *Revisiting the Quantitative-Qualitative Debate: Implications for Mixed-Methods Research.* Quality & quantity, 2002. **36**(1): p. 43-53.
100. Lane, D.C., *Validity is a Matter of Confidence—But Not Just in System Dynamics.* Systems Research and Behavioral Science, 2015. **32**(4): p. 450-458.
101. Lane, D.C., *Rerum cognoscere causas: Part I — How do the ideas of system dynamics relate to traditional social theories and the voluntarism/determinism debate?* System Dynamics Review, 2001. **17**(2): p. 97-118.
102. Lane, D.C., *Rerum cognoscere causas: Part II—Opportunities generated by the agency/structure debate and suggestions for clarifying the social theoretic position of system dynamics.* System Dynamics Review, 2001. **17**(4): p. 293-309.
103. Lane, D.C., *Should system dynamics be described as a 'hard' or 'deterministic' systems approach?* Systems Research and Behavioral Science, 2000. **17**(1): p. 3-22.
104. Lane, D.C., *Social theory and system dynamics practice.* European Journal of Operational Research, 1999. **113**(3): p. 501-527.
105. Lane, D.C., *The folding star: a comparative reframing and extension of validity concepts in system dynamics*, in *International Conference of the System Dynamics Society*. 1995.
106. Lane, D.C., *With a little help from our friends: how system dynamics and soft OR can learn from each other.* System Dynamics Review, 1994. **10**(2 ‐ 3): p. 101-134.
107. Forrester, J.W., *Industrial dynamics*. 1961, MITPress: Cambridge Mass.
108. Morecroft, J.D., *System dynamics: Portraying bounded rationality*, in *System Dynamics Research Conference*. 1981: Rensselaerville, NY.
109. Morecroft, J.D., *System dynamics: Portraying bounded rationality.* Omega, 1983. **11**(2): p. 131-142.
110. Morecroft, J.D., *Rationality in the analysis of behavioral simulation models.* Management Science, 1985. **31**(7): p. 900-916.
111. Peterson, D.W. and R.L. Eberlein, *Reality check: A bridge between systems thinking and system dynamics.* System Dynamics Review, 1994. **10**(2 ‐ 3): p. 159-174.
112. Richmond, B., *Systems thinking/system dynamics: let's just get on with it.* System Dynamics Review, 1994. **10**(2 ‐ 3): p. 135-157.
113. Sterman, J., *Business Dynamics: Systems Thinking and Modelling for a Complex World*. 2000, McGraw-Hill Higher Education: Boston.
114. Morecroft, J.D., *Strategic modelling and business dynamics: A feedback systems approach*. 2015, John Wiley & Sons.
115. Ford, F.A., *Modeling the Environment, Second Edition*. 2009, Island Press.

116. Yearworth, M. and L.A. White, *The uses of qualitative data in multimethodology: Developing causal loop diagrams during the coding process.* European Journal of Operations Research, 2013. **231**(1): p. 151-161.

117. Kim, H. and D.F. Andersen, *Building confidence in causal maps generated from purposive text data: Mapping transcripts of the Federal Reserve.* System Dynamics Review, 2012. **28**(4): p. 311-328.

118. Doyle, J.K. and D.N. Ford, *Mental models concepts revisited: some clarifications and a reply to Lane.* System Dynamics Review, 1999. **15**(4): p. 411.

119. Doyle, J.K. and D.N. Ford, *Mental models concepts for system dynamics research.* System dynamics review, 1998. **14**(1): p. 3-29.

120. Forrester, J.W., *System dynamics, systems thinking, and soft OR.* System dynamics review, 1994. **10**(2 - 3): p. 245-256.

121. Martinez - Moyano, I.J. and G.P. Richardson, *Best practices in system dynamics modeling.* System Dynamics Review, 2013. **29**(2): p. 102-123.

122. Martínez-Moyano, I.J. and G.P. Richardson. *An expert view of the system dynamics modeling process: Concurrences and divergences searching for best practices in system dynamics modeling*. in *20th International Conference of the System Dynamics Society, Palermo, Italy*. 2002.

123. Barlas, Y., *Formal aspects of model validity and validation in system dynamics.* System Dynamics Review, 1996. **12**(3): p. 183-210.

124. Cronholm, P.F., J.A. Shea, R.M. Werner, M. Miller-Day, J. Tufano, B.F. Crabtree, and R. Gabbay, *The patient centered medical home: mental models and practice culture driving the transformation process.* Journal of general internal medicine, 2013. **28**(9): p. 1195-1201.

125. Forrester, J.W., *Counterintuitive behavior of social systems.* Technology Review, 1971. **73**: p. 53-68.

126. Sterman, J., *Learning from Evidence in a Complex World.* American Journal of Public Health, 2006. **96**: p. 505-514.

127. Brunstein, A., C. Gonzalez, and S. Kanter, *Effects of domain experience in the stock–flow failure.* System Dynamics Review, 2010. **26**(4): p. 347-354.

128. Cronin, M.A. and C. Gonzalez, *Understanding the building blocks of dynamic systems.* System Dynamics Review, 2007. **23**(1): p. 1-17.

129. Cronin, M.A., C. Gonzalez, and J.D. Sterman, *Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens.* Organizational Behavior and Human Decision Processes, 2009. **108**(1): p. 116-130.

130. Moxnes, E., *Misperceptions of basic dynamics: the case of renewable resource management.* System Dynamics Review, 2004. **20**(2): p. 139-162.

131. Sterman, J.D., *Does formal system dynamics training improve people's understanding of accumulation?* System Dynamics Review, 2010. **26**(4): p. 316-334.

132. Sterman, J.D., *Risk communication on climate: mental models and mass balance.* Science, 2008. **322**(5901): p. 532-533.

133. Sterman, J.D., *Misperceptions of feedback in dynamic decision making.* Organizational behavior and human decision processes, 1989. **43**(3): p. 301-335.

134. Sterman, J.D., *Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment.* Management science, 1989. **35**(3): p. 321-339.

135. Sterman, J.D. and L.B. Sweeney, *Cloudy skies: assessing public understanding of global warming.* System Dynamics Review, 2002. **18**(2): p. 207-240.

136. Moxnes, E., *Not only the tragedy of the commons: misperceptions of feedback and policies for sustainable development.* System Dynamics Review, 2000. **16**(4): p. 325-348.

137. Sweeney, L.B. and J.D. Sterman, *Thinking about systems: Student and teacher conceptions of natural and social systems.* System Dynamics Review, 2007. **23**(2 - 3): p. 285-311.

138. Kopainsky, B. and L.F. Luna-Reyes, *Closing the loop: promoting synergies with other theory building approaches to improve system dynamics practice.* Systems Research and Behavioral Science, 2008. **25**(4): p. 471-486.

139. Luna-Reyes, L.F. and D.L. Andersen, *Collecting and analyzing qualitative data for system dynamics: methods and models.* System Dynamics Review, 2003. **19**(4): p. 271-296.

140. Lane, D.C., *Participative modelling and big issues: Defining features of system dynamics?* Systems Research and Behavioral Science, 2010. **27**(4): p. 461-465.

141. Sterman, J.D., *Sustaining sustainability: creating a systems science in a fragmented academy and polarized world*, in *Sustainability science*. 2012, Springer. p. 21-58.

142. Homer, J.B. and G.B. Hirsch, *System dynamics modeling for public health: background and opportunities.* Am J Public Health, 2006. **96**(3): p. 452-8.

143. Ackoff, R.L., *Re-creating the Corporation: A Design of Organizations for the 21st Century*. 1999, Oxford University Press: Oxford.

144. Oral, M. and O. Kettani, *The facets of the modeling and validation process in operations research.* European Journal of Operational Research, 1993. **66**(2): p. 216-234.

145. Lane, D.C., *It works in practice but does it work in theory?* Systems Research and Behavioral Science, 2006. **23**(4): p. 565-570.

146. Richardson, G.P. and A.I. Pugh III, *Introduction to System Dynamics Modeling with Dynamo*. 1981, MIT Press. p. 400.

147. Lynham, S.A., *The General Method of Theory-Building Research in Applied Disciplines.* Advances in Developing Human Resources, 2002. **4**(3): p. 221-241.

148. Sweeney, L.B. and D. Meadows, *The Systems Thinking Playbook: Exercises to Stretch and Build Learning and Systems Thinking Capabilities*. 2010, Chelsea Green Publishing Company: Chelsea River Junction, VT.

149. Elias, A.A., R.Y. Cavana, and L.S. Jackson, *Stakeholder Analysis to Enrich the Systems Thinking and Modelling Methodology.* Proceedings of the 2001 ISDC, 2001.

150. Mitchell, R.K., B.R. Agle, and D.J. Wood, *Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts.* Academy of management review, 1997. **22**(4): p. 853-886.

151. Hovmand, P.S., E.A.J.A. Rouwette, D.F. Andersen, G.P. Richardson, and A. Kraus. *Scriptapedia 4.0.6*. in *31st International Conference of the System Dynamics Society*. 2013. Boston, MA, Available from: www.systemdynamics.org/conferences/2013/proceed/papers/P1405.pdf.

152. Vennix, J.A.M., *Group model building: Facilitating team learning using system dynamics*. 1996, John Wiley and Sons: Chichester, England. .

153. Lane, D.C., *Friendly amendment: a commentary on Doyle and Ford's proposed re-definition of 'mental model'.* System Dynamics Review, 1999. **15**(2): p. 185-194.

154. Tomoaia-Cotisel, A., H. Kim, S. Allen, and K. Blanchet, *Causal Loop Diagrams*, in *Applied Systems Thinking for Health Systems Research: A Methodological Handbook*, D. de Savigny, K. Blanchet, and T. Adam, Editors. 2017, McGraw Hill Education.

155. Morecroft, J.D., *A Critical Review of Diagramming Tools for Conceptualizing Feedback System Models.* Dynamica, 1982. **8**(part 1): p. 20-29.

156. Lane, D.C., *Diagramming conventions in system dynamics.* Journal of the Operational Research Society, 2000. **51**(2): p. 241-245.

157. Richmond, B., *Systems thinking: critical thinking skills for the 1990s and beyond.* System Dynamics Review, 1993. **9**(2): p. 113-133.

158. Barlas, Y. and S. Carpenter, *A Fundamental Philosophy of Science Question and Validation of System Dynamic Models*, in *International System Dynamics Conference* 1987: Shanghai, China.

159. Richardson, G.P. and A.I. Pugh III, *Introduction to System Dynamics Modeling with Dynamo*. 1981, Productivity Press Inc. p. 311.

160. Barlas, Y. and S. Carpenter, *Philosophical roots of model validation: two paradigms.* System Dynamics Review, 1990. **6**(2): p. 148-166.

161. Groesser, S.N. and M. Schwaninger, *Contributions to model validation: hierarchy, process, and cessation.* System Dynamics Review, 2012. **28**(2): p. 157-181.

162. Ramirez, V., J. Mazzola, and M. Ablan, *Una revisión del uso y aplicación de pruebas de validación en dinámica de sistemas*, in *Congreso Latinoamericano de Dinámica de Sistemas*, G. Perez and E. Villasenor, Editors. 2012: Ciudad de México DF - México.

163. Tsioptsias, N., A.A. Tako, and S. Robinson, *Model validation and testing in simulation: a literature review.* 2016.

164. Morecroft, J.D., *Strategic modelling and business dynamics: A feedback systems approach*. 2015: John Wiley & Sons. 443.

165. Senge, P.M. and J.W. Forrester, *Tests for building confidence in system dynamics models.* System dynamics, TIMS studies in management sciences, 1980. **14**: p. 209-228.

166. Randers, J., *Elements of the system dynamics method*. 1980, MIT press Cambridge, MA.

167. Barlas, Y., *Model credibility (validity) and evaluation: Concepts and methods*, in *Assessment of Climate Change Mitigation Pathways and Evaluation of the Robustness of Mitigation Cost Estimates*. 2013, European Union.

168. Barlas, Y. *Model Verification*. in *System Dynamics Summer School*. 2014. Delft, Netherlands, Available from: http://simulation.tbm.tudelft.nl/SDSS/Intro.html.

169. Senge, P., *The fifth discipline: The art and science of the learning organization.* New York: Currency Doubleday, 1990.

170. Palys, T., *Purposive sampling.* The Sage encyclopedia of qualitative research methods, 2008. **2**: p. 697-698.

171. Perla, R.J. and L.P. Provost, *Judgment sampling: a health care improvement perspective.* Quality Management in Healthcare, 2012. **21**(3): p. 169-175.

172. Lewis, J. and J. Ritchie, *Generalizing from Qualitative Research.* , in *Qualitative Research Practice A Guide for Social Science Students and Researchers,* , J. Lewis and J. Ritchie, Editors. 2003, SAGE Publications: London. p. 263-286.

173. Turner, B.L., H. Kim, and D.F. Andersen, *Improving coding procedures for purposive text data: researchable questions for qualitative system dynamics modeling.* System Dynamics Review, 2013. **29**(4): p. 253-263.

174. Gray, S.A., E. Zanre, and S. Gray, *Fuzzy cognitive maps as representations of mental models and group beliefs*, in *Fuzzy Cognitive Maps for Applied Sciences and Engineering*. 2014, Springer. p. 29-48.

175. Andersen, D.L., L.F. Luna - Reyes, V.G. Diker, L. Black, E. Rich, and D.F. Andersen, *The disconfirmatory interview as a strategy for the assessment of system dynamics models.* System Dynamics Review, 2012. **28**(3): p. 255-275.

176. QSR International. *NVivo software*. Available from: http://www.qsrinternational.com/products_nvivo.aspx.

177. Ventana Systems. *Vensim software*. 2012 August 1, 2012; Available from: http://www.vensim.com/software.html.

178. Warren, K., *Why has feedback systems thinking struggled to influence strategy and policy formulation? Suggestive evidence, explanations and solutions.* Systems Research and Behavioral Science, 2004. **21**(4): p. 331-347.

179. Olaya, C. *Models that include cows: the significance of operational thinking*. in *30th International Conference of the System Dynamics Society*. 2012. St. Gallen, Switzerland, Available from: www.systemdynamics.org/conferences/2012/proceed/papers/P1375.pdf.

180. Groesser, S.N. and M. Schaffernicht, *Mental models of dynamic systems: taking stock and looking ahead.* System Dynamics Review, 2012. **28**(1): p. 46-68.

181. Hopper, M. and K. Stave. *Assessing the effectiveness of systems thinking interventions in the classroom* in *26th International Conference of the System Dynamics Society*. 2008. Athens, Greece, Available from: http://www.systemdynamics.org/conferences/2008/proceed/index.htm.

182. Stave, K. and M. Hopper. *What constitutes systems thinking? A proposed taxonomy*. in *Proceedings of the 26th International Conference of the System Dynamics Society. Athens*. 2007.

183. Tomoaia-Cotisel, A. *Redesigning a Primary Care System – Bringing Together Employees' Understanding of the Moving Parts*. in *Utah Health Services Research: Annual Conference*. 2013. Salt Lake City, UT: presented by Tomoaia-Cotisel, A.

184. Tomoaia-Cotisel, A., *Tensions in redesigning primary care: Practice clustering, mental models and hospitalization outcomes*, in *Third Global Symposium on Health Systems Research*. 2014: Cape Town, South Africa.

185. Tomoaia-Cotisel, A., S.D. Allen, K. Blanchet, Z. Chalabi, J. Day, T.W. Farrell, A. Haines, J. Kim, M.K. Magill, M. McKee, J. Neuburger, B. Rechel, and D.L. Scammon. *Exploring Task-shifting in Primary Care: The Case of the University of Utah Community Clinics*. in *INFORMS*. 2014. San Francisco, CA, Available from: https://informs.emeetingsonline.com/emeetings/formbuilder/clustersessiondtl.asp?csnno=21459&mmnno=260&ppnno=79547.

186. Tomoaia-Cotisel, A., K. Blanchet, Z. Chalabi, S. Allen, V. Olsavsky, A.C. Butu, M.K. Magill, and B. Rechel. *Panel: Unpacking Health Systems Through System Thinking: Exploring Primary Care: System Dynamics in USA and Romania*. in *Geneva Health Forum*. 2014. Geneva, CH.

187. Tomoaia-Cotisel, A., K. Blanchet, Z. Chalabi, B. Rechel, S.D. Allen, D.L. Scammon, J. Day, J. Kim, N.J. Waitzman, T. Farrell, and M.K. Magill, *Building a conceptual model of a primary care system – an iterative systems thinking approach*, in *American Public Health Association: Annual Conference*. 2013: Boston, MA.

188. Tomoaia-Cotisel, A., K. Blanchet, Z. Chalabi, B. Rechel, S.D. Allen, D.L. Scammon, M.K. Magill, J. Day, R. Day, T. Farrell, J. Kim, and N.J. Waitzman, *Systems Thinking - A Facilitator of PCMH Implementation?*, in *Utah Health Services Research Conference*. 2013, Presented by Tomoaia-Cotisel, A.: Salt Lake City, UT.

189. Tomoaia-Cotisel, A., K. Blanchet, Z. Chalabi, B. Rechel, S.D. Allen, D.L. Scammon, M.K. Magill, J. Day, R. Day, T. Farrell, J. Kim, and N.J. Waitzman, *Redesigning a Primary Care System – Bringing Together Employees' Understanding of the Moving Parts*, in *Academy Health: Annual Research Meeting*. 2013, Presented by Tomoaia-Cotisel, A.: Baltimore, MD.

190. Tomoaia-Cotisel, A., Z. Chalabi, K. Blanchet, S. Allen, D.L. Scammon, J. Day, B. Rechel, and M.K. Magill, *Tensions in Redesigning Primary Care: Practice Clustering, Mental Models and Hospitalization Outcomes*, in *Academy Health: Annual Research Meeting*. 2014: San Diego, CA.

191. Tomoaia-Cotisel, A., Z. Chalabi, K. Blanchet, J. Neuburger, S. Allen, J. Day, D.L. Scammon, J. Kim, M.K. Magill, and B. Rechel, *Variation in a Medical Home: Clusters of Practice, Disparate Outcomes and Dynamic Tensions* in *The 32nd International Conference of the System Dynamics Society*. 2014: Delft, Netherlands.

192. Tomoaia-Cotisel, A., K. Blanchet, Z. Chalabi, B. Rechel, S.D. Allen, D.L. Scammon, M.K. Magill, J. Day, R. Day, T. Farrell, J. Kim, and N.J. Waitzman, *Systems Thinking - A Facilitator of PCMH Implementation?*, in *Academy Health: Public Health Systems Research (PHSR) Interest Group*. 2013, Presented by Tomoaia-Cotisel, A.: Baltimore, MD.

193. EM University. *Level 5 New Patient Office Visit (99205)*. 2017; Available from: https://emuniversity.com/Level5NewOfficePatient.html.

194. Wasserman, M., M.A. Berninger, and J. Gerteis, *Estimating the Costs of Primary Care Transformation: A Practical Guide and Synthesis Report.* 2015.

195. Cronholm, P.F., M. Miller-Day, and R. Gabbay, *Evaluating a multipayer medical home intervention.* JAMA, 2014. **312**(4): p. 435-435.

196. Meadows, D.H., D.L. Meadows, J. Randers, and W.W. Behrens, *The limits to growth.* New York, 1972. **102**: p. 27.

197. Franco, D. *Fifty years of table functions*. in *Proceedings of the 25th International Conference of the System Dynamics Society and 50th Anniversary Celebration*. 2007.

198. Andersen, D.F. and G.P. Richardson, *Managing Rapid Growth in the Governor's Office of Regulatory Assistance.* unpublished case study.

199. Yerkes, R. and J. Dodsen, *The Relation of Strength of Stimulus to Rapidity of Habit Formation.* Journal of Comparative Neurology & Psychology, 1908. **18**: p. 459-482.

200. Chuang, H.H.-C. and R. Oliva, *Inventory record inaccuracy: Causes and labor effects.* Journal of Operations Management, 2015. **39–40**: p. 63-78.

201. Radzicki, M. and R. Taylor, *Better Before Worse or Worse Before Better*, in *US Department of energy's introduction to system dynamics, a systems approach to understanding complex policy issues, version 1*. 1997, System Dynamics Society: Albany, NY, Available from: https://www.systemdynamics.org/DL-IntroSysDyn/bwb.htm.

202. Ghaffarzadegan, N., J. Lyneis, and G.P. Richardson, *How small system dynamics models can help the public policy process.* System Dynamics Review, 2011. **27**(1): p. 22-44.

203. Graham, A.K. and S.A. Els. *System dynamics and systems thinking: it takes all kinds*. in *17th International Conference of the System Dynamics Society, Wellington, New Zealand*. 1999.

204. Lyneis, J.M. *System dynamics in business forecasting: a case study of the Commercial Jet Aircraft Industry*. in *CD-ROM Proceedings of the 1998 System Dynamics Conference*. 1998.

205. Repenning, N.P. and J.D. Sterman, *Capability traps and self-confirming attribution errors in the dynamics of process improvement.* Administrative Science Quarterly, 2002. **47**(2): p. 265-295.

206. Morrison, B., *Dynamics of Learning by Doing Under Constraints: Analysis of the Tipping Point*, in *International System Dynamics Society Conference*. 2004.

207. Lyneis, J.M. and D.N. Ford, *System Dynamics Applied to Project Management: A Survey, Assessment, and Directions for Future Research.* System Dynamics Review, 2007. **23**: p. 157-189.

208. Levine, R.L. and J.K. Doyle. *Modeling generic structures and patterns in social psychology*. in *proceeding of the 20th System dynamics conference, Italy*. 2002. Citeseer.

209. Cyert, R. and J. March, *A Behavioral Theory of the Firm*. 1963, New Jersey: Prentice Hall.
210. Simon, H.A., *Administrative Behavior, 3rd Edition*. 1976, New York: Wiley.
211. Simon, H.A., *A behavioral model of rational choice.* The quarterly journal of economics, 1955: p. 99-118.
212. Simon, H.A., *Rationality and Decision Making*, in *Models of Man*. 1957, Wiley: New York.
213. Rahmandad, H., *Effect of delays on complexity of organizational learning.* Management Science, 2008. **54**(7): p. 1297-1312.
214. Rahmandad, H., N. Repenning, and J. Sterman, *Effects of feedback delay on learning.* System Dynamics Review, 2009. **25**(4): p. 309-338.
215. Gentner, D., *Structure-Mapping: A Theoretical Framework for Analogy.* Cognitive Science, 1983(7): p. 155-170.
216. Mukherjee, S., *My Father's Body, at Rest and in Motion*, in *The New Yorker*. 2018.
217. Whetten, D.A., *What constitutes a theoretical contribution?* Academy of management review, 1989. **14**(4): p. 490-495.
218. Stange, K.C., *Holding On and Letting Go: A Perspective from the Keystone IV Conference.* The Journal of the American Board of Family Medicine, 2016. **29**(Supplement 1): p. S32-S39.
219. Oliva, R., *Tradeoffs in responses to work pressure in the service industry.* California Management Review, 2001. **43**(4): p. 26-43.
220. Altschuler, J., D. Margolius, T. Bodenheimer, and K. Grumbach, *Estimating a Reasonable Patient Panel Size for Primary Care Physicians With Team-Based Task Delegation.* The Annals of Family Medicine, 2012. **10**(5): p. 396-400.
221. Crabtree, B.F. and W.L. Miller, *Doing qualitative research*. 2nd ed. 1999, Thousand Oaks, Calif.: Sage Publications.
222. Burwell, S., *Health Care Changes*, in *Health Care Policy*. 2017, CSPAN: Washington, DC.
223. Murray, M., *Answers to your questions about same-day scheduling.* Family practice management, 2005. **12**(3): p. 59.
224. Hroscikoski, M.C., L.I. Solberg, J.M. Sperl-Hillen, P.G. Harper, M.P. McGrail, and B.F. Crabtree, *Challenges of Change: A Qualitative Study of Chronic Care Model Implementation.* The Annals of Family Medicine, 2006. **4**(4): p. 317-326.
225. Murray, M. and D.M. Berwick, *Advanced access: reducing waiting and delays in primary care.* Jama, 2003. **289**(8): p. 1035-1040.
226. Magill, M.K., J. Day, and J. Kim, *Transformed Primary Care: Care by Design™*. 2013, Agency for Healthcare Research and Quality: Rockville MD.

227. Rittenhouse, D.R., L. Schmidt, K. Wu, and J. Wiley, *Contrasting trajectories of change in primary care clinics: lessons from New Orleans safety net.* Ann Fam Med, 2013(11(Suppl 1)): p. S60-7.

228. Senge, P.M. and R. Oliva, *Developing a theory of service quality/service capacity interaction*, in *International Conference of the System Dynamics Society* 1993. p. 477.

229. Oliva, R., *A dynamic theory of service delivery: implications for managing service quality*, in *Sloan School of Management*. 1996, Massachusetts Institute of Technology: Cambridge, MA.

230. Oliva, R. and J.D. Sterman, *Cutting Corners and Working Overtime: Quality Erosion in the Service Industry.* Management Science, 2001. **47**(7): p. 894-914.

231. Quigley, D.D., Z.S. Predmore, A.Y. Chen, and R.D. Hays, *Implementation and Sequencing of Practice Transformation in Urban Practices with Underserved Patients.* Quality Management in Healthcare, 2017. **26**(1): p. 7-14.

232. Tuepker, A., D. Kansagara, E. Skaperdas, C. Nicolaidis, S. Joos, M. Alperin, and D. Hickam, *"We've Not Gotten Even Close to What We Want to Do": a Qualitative Study of Early Patient-Centered Medical Home Implementation.* Journal of General Internal Medicine, 2014. **29**(2): p. 614-622.

233. Van Cleave, J., A.A. Boudreau, J. McAllister, W.C. Cooley, A. Maxwell, and K. Kuhlthau, *Care Coordination Over Time in Medical Homes for Children With Special Health Care Needs.* Pediatrics, 2015.

234. Nutting, P.A., W.L. Miller, B.F. Crabtree, C.R. Jaen, E.E. Stewart, and K.C. Stange, *Initial Lessons From the First National Demonstration Project on Practice Transformation to a Patient-Centered Medical Home.* Annals of Family Medicine, 2009. **7**(3): p. 254-260.

235. Nutting, P.A., B.F. Crabtree, W.L. Miller, E.E. Stewart, K.C. Stange, and C.R. Jaén, *Journey to the patient-centered medical home: a qualitative analysis of the experiences of practices in the National Demonstration Project.* The Annals of Family Medicine, 2010. **8**(Suppl 1): p. S45-S56.

236. Miller, W.L., B.F. Crabtree, P.A. Nutting, K.C. Stange, and C.R. Jaén, *Primary Care Practice Development: A Relationship-Centered Approach.* The Annals of Family Medicine, 2010. **8**(Suppl 1): p. S68-S79.

237. Senge, P.M., *The fifth discipline: the art and practice of the learning organization*. 1994, Doubleday/Currency: New York City.

238. Pratt, M.G., K.W. Rockmann, and J.B. Kaufmann, *Constructing professional identity: The role of work and identity learning cycles in the customization of identity among medical residents.* Academy of management journal, 2006. **49**(2): p. 235-262.

239. Doyle, J., M. Radzicki, and W. Trees, *Measuring change in mental models of complex systems.* Complex decision making: Theory and practice, 2008: p. 269-294.

240. Ruddy, M.P., L. Thomas-Hemak, and L. Meade, *Practice transformation: professional development is personal.* Academic Medicine, 2016. **91**(5): p. 624-627.

241. Berwick, D.M., *Crossing the boundary: changing mental models in the service of improvement.* Int J Qual Health Care, 1998. **10**(5): p. 435-41.

242. Berwick, D.M., *Developing and testing changes in delivery of care.* Annals of Internal Medicine, 1998. **128**(8): p. 651-656.

243. Repenning, N.P. and J. Sterman, *Getting quality the old-fashioned way: self confirming attributions in the dynamics of process improvement.* 1997.

244. Powers, W.T., *Feedback: Beyond Behaviorism: Stimulus-response laws are wholly predictable within a control-system model of behavioral organization.* Science, 1973. **179**(4071): p. 351-356.

245. Sterman, J.D., *Learning in and about complex systems.* System Dynamics Review, 1994. **10**(2 - 3): p. 291-330.

246. Valko, G. and R. Wender, *Evaluating a multipayer medical home intervention.* JAMA, 2014. **312**(4): p. 434.

247. Jaén, C. and R. Palmer, *Shorter adaptive reserve measures.* Ann Fam Med, 2012. **8**(Suppl 1): p. 1-2.

248. Orr, R. *Periodized Programs - Part 3*. 2005 [cited 2017 December 13]; Available from: https://www.ptonthenet.com/content/articleprint.aspx?p=1&ArticleID=MjU1NSBXczNKeWwyWnJtYXYzK0h1b09PeWF3PT0=.

249. Forrester, J.W., *Information Sources for Modeling the National Economy.* Journal of the American Statistical Association, 1980. **75**(371): p. 555-566.

250. Van Maanen, J., *The fact of fiction in organizational ethnography.* Administrative science quarterly, 1979. **24**(4): p. 539-550.

251. Richardson, G.P., *Problems with causal - loop diagrams.* System dynamics review, 1986. **2**(2): p. 158-170.

252. Lane, D.C., *The emergence and use of diagramming in system dynamics: a critical account.* Systems Research and Behavioral Science, 2008. **25**(1): p. 3-23.

253. Luna-Reyes, L.F., J. Gil-Garcia, and C. CB. *Collaborative digital government in Mexico: some lessons from federal web-based inter-organisational information integration initiatives*. in *Americas Conference on Information Systems*. 2006.

254. Von Raesfeld, A., *Colaboracion y Transferencia de Conocimiento en Proyectos: Estudio de un Caso*, in *School of Business*. 2005, Universidad de las Americas-Puebla: Cholula, Mexico.

255. Maitlis, S. and M. Christianson, *Sensemaking in Organizations: Taking Stock and Moving Forward.* The Academy of Management Annals, 2014.

256. Lane, D.C., *'Till the Muddle in my Mind Have Cleared Awa': Can We Help Shape Policy Using Systems Modelling?* Systems Research and Behavioral Science, 2016. **33**(5): p. 633-650.

257. Ford, D.N. and J.D. Sterman, *Expert knowledge elicitation to improve formal and mental models.* System Dynamics Review, 1998. **14**(4): p. 309-340.

258. Richmond, B., *The strategic forum: aligning objectives, strategy and process.* System dynamics review, 1997. **13**(2): p. 131-148.

259. Kuhn, T.S., *The Structure of Scientific Revolutions*. 1970: University of Chicago Press.

260. Eckberg, D.L. and L. Hill Jr, *The paradigm concept and sociology: A critical review.* American Sociological Review, 1979: p. 925-937.

261. Masterman, M., *The nature of a paradigm*, in *Criticism and the Growth of Knowledge*, I. Lakatos and A. Musgrave, Editors. 1970, Cambridge University Press: Cambridge, England. p. 59-90.

262. Burrell, G. and G. Morgan, *Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life*. 1979: Pearson Education.

263. Deetz, S., *Crossroads—Describing differences in approaches to organization science: Rethinking Burrell and Morgan and their legacy.* Organization science, 1996. **7**(2): p. 191-207.

264. Schumacher, E.F., *A Guide for the Perplexed*. 1977: Harper Colophon Books.

265. Wilber, K., *Sex, Ecology, Spirituality: The Spirit of Evolution, Second Edition*. 2001: Shambhala.

266. Stange, K.C., *The paradox of the parts and the whole in understanding and improving general practice.* Int. J. Qual. Health Care, 2002. **14**: p. 267-268.

267. Stange, K.C., *A Science of Connectedness.* The Annals of Family Medicine, 2009. **7**(5): p. 387-395.

268. Stange, K.C., R.S. Etz, H. Gullett, S.A. Sweeney, W.L. Miller, C.R. Jaén, B.F. Crabtree, P.A. Nutting, and R.E. Glasgow, *Metrics for assessing improvements in primary health care.* Annual review of public health, 2014. **35**: p. 423-442.

269. Edwards, M.G., *The integral holon: A holonomic approach to organisational change and transformation.* Journal of Organizational Change Management, 2005. **18**(3): p. 269-288.

270. Lane, D.C., *Social theory and system dynamics practice*, in *International System Dynamics Society Conference*. 1994. p. 53-66.

271. Lane, D.C. and E. Husemann, *Steering away from Scylla, falling into Charybdis: the importance of recognising, simulating and challenging reinforcing loops in social systems*. 2002: Duncker & Humblot.

272. Lane, D.C. and E. Husemann, *Steering without Circe: attending to reinforcing loops in social systems.* System Dynamics Review, 2008. **24**(1): p. 37-61.

273. Schultz, M. and M.J. Hatch, *Living with Multiple Paradigms: The Case Of Paradigm Interplay in Organizational Culture Studies.* Academy of Management Review, 1996. **21**(2): p. 529-557.

274. Gioia, D.A. and E. Pitre, *Multiparadigm perspectives on theory building.* Academy of management review, 1990. **15**(4): p. 584-602.

275. Reichel, A., *(Re-)Structuration of System Dynamics*, in *System Dynamics Conference*. 2004: Oxford, UK.

276. Pruyt, E. *What is system dynamics? A paradigmatic inquiry*. in *Proceedings of the 2006 Conference of the System Dynamics Society*. 2006.

277. Doyle, J.K., *The cognitive psychology of systems thinking.* System Dynamics Review, 1997. **13**(3): p. 253-265.

278. Stave, K., N. Zimmermann, and H. Kim, *Exploring the nature of insight in System Dynamics*, in *Conference of the International System Dynamics Society*. 2016: Delft, Netherlands.

279. Black, L.J., *When visuals are boundary objects in system dynamics work.* System Dynamics Review, 2013. **29**(2): p. 70-86.

280. Vennix, J.A. and W. Scheper. *Modeling as organizational learning: an empirical perspective*. in *Proceedings of the 1990 International Conference of the System Dynamics Society*. 1990.

281. Lane, D.C., *Industrial dynamics by Jay W. Forrester [Invited review and reappraisal].* Journal of the Operational Research Society, 1997. **48**(10): p. 1037-1042.

282. Barton, J. *Pragmatism, systems thinking and system dynamics*. in *System Dynamics Conference*. 1999.

283. Warren, K., *Strategic management dynamics*. 2008: John Wiley & Sons.

284. Cavaleri, S.A., *System Dynamics: A Form of the Integrative System Approach*, in *International System Dynamics Conference*, J. Vennix, *et al.*., Editors. 1992: Utrecht, The Netherlands.

285. Senge, P. *Some thoughts on the boundaries of classical system dynamics*. in *CD-Rom Proceedings of the 16th International Conference of the System Dynamics Society—Québec'98*. 1998.

286. Sterman, J.D. and J. Wittenberg, *Path dependence, competition, and succession in the dynamics of scientific revolution.* Organization Science, 1999. **10**(3): p. 322-341.

287. Teddlie, C. and A. Tashakkori, *Common "core" characteristics of mixed methods research: A review of critical issues and call for greater convergence.* American Behavioral Scientist, 2012. **56**(6): p. 774-788.

288. Lane, D.C., *What Is a 'Policy Insight'?* Systems Research and Behavioral Science, 2012. **29**(6): p. 590-595.

289. Stave, K., N. Zimmermann, and H. Kim. *What are System Dynamics Insights?* in *International System Dynamics Society Conference*. 2016. System Dynamics Society.

290. Rahmandad, H. and J. Sterman, *Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models.* Management Science, 2008. **54**(5): p. 998-1014.

291. Fetters, M.D., L.A. Curry, and J.W. Creswell, *Achieving Integration in Mixed Methods Designs—Principles and Practices.* Health Services Research, 2013. **48**(6pt2): p. 2134-2156.

292. Millington, J.D. and J. Wainwright, *Mixed qualitative-simulation methods: Understanding geography through thick and thin.* Progress in Human Geography, 2017. **41**(1): p. 68-88.

293. Tashakkori, A. and C. Teddlie, *Mixed methodology : combining qualitative and quantitative approaches*, in *Applied social research methods series v. 46*. 1998, Sage: Thousand Oaks, Calif. p. xi, 185 p.

294. Mingers, J., *Real-izing information systems: critical realism as an underpinning philosophy for information systems.* Information and organization, 2004. **14**(2): p. 87-103.

295. Van Maanen, J., *Reclaiming qualitative methods for organizational research: A preface.* Administrative science quarterly, 1979. **24**(4): p. 520-526.

296. World Health Organization, *Health in 2015: from MDGs, Millennium Development Goals to SDGs*, in *Sustainable Development Goals*. 2015. p. 204.

297. Rittel, H.W. and M.M. Webber, *Dilemmas in a general theory of planning.* Policy sciences, 1973. **4**(2): p. 155-169.

298. Black, L.J. and D.F. Andersen, *Using visual representations as boundary objects to resolve conflict in collaborative model - building approaches.* Systems Research and Behavioral Science, 2012. **29**(2): p. 194-208.

299. Saeed, K., *Slicing a complex problem for system dynamics modeling.* System Dynamics Review, 1992. **8**(3): p. 251-261.

300. Schmidt, N. and J. Brown, *Evidence-Based Practice for Nurses*. 2011, Jones & Bartlett Learning.

301. Andersen, D.F., G.P. Richardson, and J.A.M. Vennix, *Group model building: Adding more science to the craft.* System Dynamics Review, 1997. **13**: p. 187–201.

302. Goles, T. and R. Hirschheim, *The paradigm is dead, the paradigm is dead… long live the paradigm: the legacy of Burrell and Morgan.* Omega, 2000. **28**(3): p. 249-268.

303. Barlas, Y., *Comments on "On the very idea of a system dynamics model of Kuhnian science".* System Dynamics Review, 1992. **8**(1): p. 43-47.

304. Tomoaia - Cotisel*, A., D.L. Scammon*, R.L. Day, J. Day, J. Kim, N.J. Waitzman, T.W. Farrell, and M.K. Magill, *Connecting the Dots and Merging Meaning: Using Mixed Methods to Study Primary Care Delivery Transformation.* Health Services Research, 2013. **48**(6pt2): p. 2181-2207

305. Radzicki, M. and R. Taylor, *A System Problem and Its Symptoms are Separated By Time and Space*, in *US Department of energy's introduction to system dynamics, a systems approach to understanding complex policy issues, version 1*. 1997, System Dynamics Society: Albany, NY, Available from: https://www.systemdynamics.org/DL-IntroSysDyn/tas.htm.

306. Shah, I., *The blind ones and the matter of the elephant.* Tales of the dervishes, 1993: p. 25-26.

307. Peikes, D., A. Zutshi, J. Genevro, K. Smith, M. Parchman, and D. Meyers, *Early evidence on the patient-centered medical home.* Am J Manag Care, 2012. **18**: p. 105-16.

308. World Health Organization, *WHO Global strategy on people-centred and integrated health service: interim report*. 2015, United Nations: Geneva, Switzerland.

309. Loewenson, R. and S. Simpson, *Strengthening Integrated Care Through Population-Focused Primary Care Services: International Experiences Outside the United States.* Annual Review of Public Health, 2017. **38**(1): p. 413-429.

310. Magill, M.K., J. Day, R. Day, J. Kim, D.L. Scammon, A. Sheets-Mervis, A. Tomoaia-Cotisel, and N.J. Waitzman, *Transformed Primary Care-Care by Design*. 2009, AHRQ: Salt Lake City, UT.

311. Jaén, C.R., B.F. Crabtree, R.F. Palmer, R.L. Ferrer, P.A. Nutting, W.L. Miller, E.E. Stewart, R. Wood, M. Davila, and K.C. Stange, *Methods for Evaluating Practice Change Toward a Patient-Centered Medical Home.* The Annals of Family Medicine, 2010. **8**(Suppl 1): p. S9-S20.

312. Glaser, B. and A. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*. 1967, Aldine de Gruyter: Hawthorne, NY.

313. Crabtree, B.F., W.L. Miller, and K.C. Stange, *Understanding Practice from the Ground Up.* J Fam Pract, 2001. **50**(10): p. 881-887.

314. Poole, M.S. and A.H. Van de Ven, *Using paradox to build management and organization theories.* Academy of management review, 1989. **14**(4): p. 562-578.

315. Weick, K.E., *What theory is not, theorizing is.* Administrative Science Quarterly, 1995. **40**(3): p. 385-390.

316. Felin, T. and W.S. Hesterly, *The knowledge-based view, nested heterogeneity, and new value creation: Philosophical considerations on the locus of knowledge.* Academy of Management Review, 2007. **32**(1): p. 195-218.

317. Billings J. *Emergency department profiling algorithm*. 2000 [cited 2012 June 6]; Available from: http://wagner.nyu.edu/faculty/billings/nyued-background.

318. Howe, K.R., *Closing Methodological Divides: Toward Democratic Educational Research*. 2003, Springer.

319. Deverka, P.A., D.C. Lavallee, P.J. Desai, L.C. Esmail, S.D. Ramsey, D.L. Veenstra, and S.R. Tunis, *Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement.* 2012.

320. Tomoaia-Cotisel, A., J. Neuburger, Z. Chalabi, B. Rechel, S.D. Allen, D.L. Scammon, J. Day, J. Kim, N.J. Waitzman, T. Farrell, and M.K. Magill, *Effects of observed variations in primary care practice on Medicare patients' hospital utilization*, in *American Public Health Association: Annual Conference*. 2013: Boston, MA.

321. Tomoaia-Cotisel, A. *Preliminary Research Findings Discussion - Presentation* in *UUCC Senior Leadership*. 2011. Salt Lake City, UT.

322. Meadows, D., *Leverage points: Places to intervene in a system.* The Sustainability Institute, 1999. **3**: p. 19pp.

323. Tomoaia-Cotisel, A., D.L. Scammon, J. Kim, J. Day, R. Day, T. Farrell, N.J. Waitzman, and M.K. Magill, *Refining & Integrating Qualitative Methods – Providing an In-depth Understanding of Care Transformation & Evaluation*, in *US Agency for Healthcare Research & Quality: Annual Conference*. 2011, Presented by Tomoaia-Cotisel, A.: Bethesda, MD.

324. Tomoaia-Cotisel, A. *Redesigning a Primary Care System – Bringing Together Our Understanding of the Moving Parts*. in *NAC/LAC -- Care by Design Research Program Annual Meeting*. 2013. Salt Lake City, UT: Presented by Tomoaia-Cotisel, A.

325. Checkland, P. and J. Scholes, *Soft Systems Methodology in Action*. 1990, Wiley.

326. Glaser, B.G., *Basics of grounded theory analysis: Emergence vs forcing*. 1992, Sociology Press.

327. Corbin, J. and A. Strauss, *Basics of qualitative research: Techniques and procedures for developing grounded theory*. 2008.

328. Strauss, A. and J. Corbin, *Basics of qualitative research: Grounded theory procedures and techniques*. 1998, London, UK: Sage. 212.

329. Dizikes, P., *The secrets of the system*, in *MIT news*. 2012, MIT: Cambridge, MA.

330. Eden, C., S. Jones, and D. Sims, *Thinking in organisations*. 1979, Macmillan.

331. Balci, O. *Validation, verification, and testing techniques throughout the life cycle of a simulation study*. in *Simulation Conference Proceedings, 1994. Winter*. 1994. IEEE.

332. Forrester, J.W., *Policies, decisions and information sources for modeling.* European Journal of Operational Research, 1992. **59**(1): p. 42-63.

333. Özesmi, U. and S.L. Özesmi, *Ecological models based on people's knowledge: a multi-step fuzzy cognitive mapping approach.* Ecological Modelling, 2004. **176**(1–2): p. 43-64.

334. Lohmann, S., J. Ziegler, and L. Tetzlaff. *Comparison of tag cloud layouts: Task-related performance and visual exploration*. in *IFIP Conference on Human-Computer Interaction*. 2009. Springer.

335. Hearst, M.A. and D. Rosner. *Tag clouds: Data analysis tool or social signaller?* in *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*. 2008. IEEE.

336. Huff, A.S., *Mapping strategic thought*. 1990, John Wiley & Sons.

337. Concannon, T.W., P. Meissner, J.A. Grunbaum, N. McElwee, J.-M. Guise, J. Santa, P.H. Conway, D. Daudelin, E.H. Morrato, and L.K. Leslie, *A New Taxonomy for Stakeholder Engagement in Patient-Centered Outcomes Research.* Journal of General Internal Medicine, 2012. **27**(8): p. 985-991.

338. Alexander, G., J. Kurlander, and M. Wynia, *Physicians in retainer ("concierge") practice. A national survey of physician, patient, and practice characteristics. .* J Gen Intern Med. , 2005. **20**(12): p. 1079–1083.

339. Laff, M., *Family Physician Salaries Up, but Still Trail Those of Subspecialists.* The Annals of Family Medicine, 2015. **13**(4): p. 390-391.

340. Wu, W.N., G. Bliss, E.B. Bliss, and L.A. Green, *A direct primary care medical home: the Qliance experience.* Health Affairs, 2010. **29**(5): p. 959-962.

341. Hyden, M. and M. Wong. *How to measure physician compensation per RVU*. MGMA Connection Plus 2013; August 14, 2013]. Available from: http://www.mgma.com/practice-resources/mgma-connection-plus/online-only/2013/august/how-to-measure-physician-compensation-per-rvu.

342. payscale.com. *Medical Assistant Salary*. 2018; Available from: https://www.payscale.com/research/US/Job=Medical_Assistant/Hourly_Rate.

343. Grantham, S., *Redesigning Primary Health Care Teams for Population Health and Quality Improvement.* 2017.

344. US Centers for Medicare and Medicaid Services. *Single HCPCS Code 99205 Office/outpatient visit new*. Physician Fee Schedule Search 2016; Available from: https://www.cms.gov/apps/physician-fee-schedule/search/search-results.aspx?Y=0&T=0&HT=0&CT=0&H1=99205&M=1.

APPENDIX A: PUBLICATIONS & PRESENTATIONS RELEVANT TO THIS DISSERTATION

Peer-reviewed Publications

Tomoaia-Cotisel A, Farrell TW, Solberg LI, Berry CA, Calman NS, Cronholm PF, Donahue KE, Driscoll DL, Hauser D, McAllister JW, Mehta S, Reid RJ, Tai-Seale M, Wise CG, Fetters MD, Holtrop JS, Rodriguez HP, Brunker CP, McGinley EL, Day R, Scammon DL, Harrison M, Genevro JL, Gabbay R, and Magill MK. *(2016).* ***Implementation of Care Management: An Analysis of Recent AHRQ Research*****.** *Medical Care Research and Review*. 1077558716673459.

Farrell TW, **Tomoaia-Cotisel A**, Scammon D, Brunisholz K, Kim J, Day J, Gren L, Wallace S, Gunning K, Tabler J, Magill MK. *(2015).* ***Impact of an Integrated Transitions Management Program in Primary Care on Hospital Readmissions***. *Journal of Healthcare Quality. 37(1):81-92.*

Tabler J, Scammon DL, Kim J, Farrell T, **Tomoaia-Cotisel A**, Day J, and Magill M K. (2014). *Patient Care* ***Experiences and Perceptions of the Patient-Provider Relationship: A Mixed Method Study.*** *The Patient Experience Journal. 1(1):75-87.*

Scammon DL, Tabler J Brunisholz K, Gren LH Kim J, **Tomoaia-Cotisel A** Day J, Farrell TW and Magill MK. (*2014*). ***Organizational Culture Associated With Provider Satisfaction***. *Journal of the American Board of Family Medicine. 27(2):219-228.*

Tomoaia-Cotisel A*, Scammon DL*, Day J, Day R, Kim J, Waitzman N, Farrell TW, and Magill MK. *(2013).* ***Connecting the Dots and Merging Meaning: Using Mixed Methods to Study Primary Care Delivery Transformation.*** *Health Services Research. 48(6 Pt 2):2181 – 2207. (*shared first authorship)*

Tomoaia-Cotisel A, Scammon DL, Waitzman N, Cronholm P, Halladay J, Driscoll D, Solberg L, Hsu C, Tai-Seale M, Hiratsuka V, Shih S, Fetters M, Wise C, Alexander J, Hauser D, McMullen C, Scholle S, Tirodkar M, Schmidt L, Donahue K, Parchman M, and Stange K. *(2013).* ***Context Matters: The Experience of 14 Research Teams in Systematically Reporting Contextual Factors Important for Practice Change.*** *Annals of Family Medicine.* **11***:S50-S59.*

Day J, Scammon DL, Kim J, Sheets-Mervis A, Day R, **Tomoaia-Cotisel A**, Waitzman N, and Magill MK. *(2013).* ***Quality, Satisfaction, and Financial Efficiency Associated with Elements of Primary Care Practice Transformation: Preliminary Findings. Annals of Family Medicine.*** **11***:S115-S12.*

Non-peer-reviewed Publications

Tomoaia-Cotisel, A, Kim, H, Allen S, Blanchet, K. *(2017). Causal Loop Diagrams.* In Don de Savigny, Karl Blanchet, and Taghreed Adam (Eds.). Applied Systems Thinking for Health Systems Research: A Methodological Handbook. **Open University Press.**

Farrell TW, **Tomoaia-Cotisel A,** Scammon DL, Day J, Day RL, Magill MK. *(2015). Care Management: Implications for Medical Practice, Health Policy and Health Services Research*. **Agency for Healthcare Research and Quality.**

Hirsch G, Homer H, **Tomoaia-Cotisel A** (Eds.) *System Dynamics Applications to Health and Health Care. (2015).* ***System Dynamics Review.*** Virtual Issue.

Webinars

Care Management: Key Elements and Implications for Medical Practice, Health Policy and Health Services Research. (April 2015). *Agency for Healthcare Research and Quality.*

> ***Presenters:*** *A Tomoaia-Cotisel and Timothy Farrell;* ***Moderator****: Jan Genevro (AHRQ)*
> ***Discussants:*** *Ed Wagner, Michael Harrison, Debra Scammon, Michael Magill*

Advanced Methods Webinars: Integrating Mixed Methods in Health Services and Delivery System Research: Webinar #5: Mixed Methods. (May 2014). *Agency for Healthcare Research and Quality.*

> ***Presenters***: Benjamin Crabtree, Michael Magill, Debra Scammon, A Tomoaia-Cotisel;
> ***Moderator***: Michael Harrison

Acknowledgements

McNellis, RJ, Genevro, JL, and DS Meyers. *(2013).* ***Lessons Learned from the Study of Primary Care Transformation****. Annals of Family Medicine.* 11*:S1-S5.* (Acknowledgements include: Andrada Tomoaia-Cotisel, for developing the framework for lessons learned)

<u>Oral Presentations</u> (ATC is first author and presenter unless otherwise indicated)

(2014) Exploring task-shifting in primary care: the case of the University of Utah Community Clinics. **INFORMS Annual Meeting**. San Francisco, CA.

(2014) **Third Global Symposium on Health Systems Research.** Cape Town, South Africa.

- *Demystifying complexity in health systems: a hands-on workshop on developing causal loop diagrams.* (ATC contributor)

(2014) **World Health Organization**. Geneva, Switzerland.

- *Applying Systems Thinking in Health Systems Strengthening: practice with causal loop diagrams* (Moderator: T. Adam; Panel: Kwamie A., Paina L., Tomoaia-Cotisel A., Varghese J.).
- *Advancing the Application of Systems Thinking in Health* (Moderator: T. Adam; Panel: Malik A., Paina L., Tomoaia-Cotisel A.).

(2014) Systems Thinking & System Dynamics methods in health system strengthening. **Geneva Health Forum.** Geneva, Switzerland.

(2013) Transitions and System Dynamics. **11th Annual Rocky Mountain Geriatrics Conference - Navigating Complexity: A Roadmap for Successful Transitions across the Care Continuum**. Snowbird, UT. (A. Tomoaia-Cotisel and T. Farrell).

(2013) Exploring the impact of perceived versus actual competition on a clinic's sensitivity to its stakeholders. **Association for Marketing & Health Care Research: Annual Conference.** Big Sky, MT. (A. Tomoaia-Cotisel and S. Allen)

<u>Conference Papers & Posters</u> (ATC is first author and presenter unless otherwise indicated)

(2014) Distilling the essential components of care management: a survey of lessons learned from the AHRQ Transforming Primary Care & ARRA grantees. **Conference on the Science of Dissemination and Implementation.** Washington, DC.

(2014) **Third Global Symposium on Health Systems Research.** Cape Town, South Africa.

- *Tensions in redesigning primary care: practice clustering, mental models and hospitalization outcomes*
- *Fostering organizational learning -researchers partnering with healthcare delivery systems in primary care transformation* (S. Allen & A. Tomoaia-Cotisel)

(2014) Variation in a Medical Home: clusters of practice, disparate outcomes and dynamic tensions. **International System Dynamics Society Annual Conference.** Delft, Netherlands.

(2014) **Academy Health: Annual Research Meeting.** San Diego, CA.

- *Tensions in redesigning primary care: practice clustering, mental models and hospitalization outcomes.*
- *Assessing patient experience with a comprehensive care management program: a pilot test of the CAHPS-PCMH Survey* (Scammon, D.L. *et al.*.)
- *The development of teamness in the context of team care: an elusive target* (Scammon, D.L. *et al.*.)

(2014) **American Society of Health Economists.** Los Angeles, CA. (Presented by Kim JW)

- *Disparities in patient experiences with care in a Patient-Centered Medical Home* (Kim, J.W. *et al.*.)
- *The effects of level of transformation in primary care on healthcare utilization and cost among a Medicare cohort* (Kim, J.W. *et al.*.)

(2014) Systems thinking - a facilitator of PCMH implementation? **NIH Conference on Complex Systems, Health Disparities & Population Health: Building Bridges.** Bethesda, MD.

(2013) **American Public Health Association: Annual Conference**. Boston, MA.

- *Effects of observed variations in primary care practice on Medicare patients' hospital utilization*
- *Building a conceptual model of a primary care system – an iterative systems thinking approach*

(2013) **Academy Health: Annual Research Meeting.** Baltimore, MD.

- *Systems Thinking - A facilitator of PCMH implementation?* (General Session & Public Health Systems Research Interest Group)
- *Redesigning a primary care system – bringing together employees' understanding of the moving parts.*
- *The use of patient reported outcomes in managing patients with chronic conditions: experience from the field.* (Scammon, D. *et al.*.)
- *Characterization and feasibility of a care manager-based transitions management program within a Patient Centered Medical Home?* (Farrell, T., *et al.*.)

(2012) **Academy Health: Annual Research Meeting.** Baltimore, MD.

- *The role of research in developing organizational learning within an academic medical center*
- *Management's role in implementing and sustaining change.*

(2011) **US Agency for Healthcare Research & Quality: Annual Conference.**

- *Refining & Integrating qualitative methods – providing an in-depth understanding of care transformation & evaluation*

APPENDIX B: ETHICAL APPROVALS

This work received ethical approval from two institutions: The London School of Hygiene and Tropical Medicine Research Ethics Committee and The University of Utah Institutional Review Board. This appendix contains a copy of the approval letters from these institutions.

There was one application to the London School of Hygiene and Tropical Medicine Research Ethics Committee and there were three applications to the University of Utah Institutional Review Board.

For the reader's convenience, the following table presents the application number, application name, data corresponding to that project, and determination. Please note that data used in the dissertation are bolded and data referenced only in the scoping study are italicized.

Table B.1. University of Utah Institutional Review Board Applications

Application Number	Application Name	Data Used in Dissertation (bold) and/or Scoping Study (*italics*)	Determination
LSHTM6516	Exploring the Impact of Primary Care Structure on Patients' Risk of Hospitalization	All data as marked below plus additional **stakeholder discussions**	Favorable ethical opinion
IRB_00042147	Transformed Primary Care – Care by Design I (**TPC I**)	**interviews**, *observations, surveys*	Exempt
IRB_00042886	Transformed Primary Care – Care by Design II (**TPC II**)	*clinic, provider and patient operations data for quantitative analysis*	Exempt
IRB_00042890	Transformed Primary Care – Care by Design III UPDB (**TPC III**)	*patient utilization data*	Minimal risk

B.1 LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE – ETHICAL APPROVAL

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Observational / Interventions Research Ethics Committee

Andrada Tomoaia-Cotisel
Research Degree Student
HSRP / PHP
LSHTM

15 January 2014

Dear Ms. Tomoaia-Cotisel,

Study Title:	**Exploring the Impact of Primary Care Structures on Patients' Risk of Hospitalization -- Starting with Medicare Beneficiaries in the State of Utah, USA**
LSHTM ethics ref:	**6516**

Thank you for your letter of 11 January 2014, responding to the Observational Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
LSHTM ethics application	n/a	
Protocol	v1	20 August, 2013

After ethical review

Any subsequent changes to the application must be submitted to the Committee via an Amendment form on the online application website. All studies are also required to notify the ethics committee of any serious adverse events which occur during the project via an Adverse Event form on the online application website. At the end of the study, please notify the committee via an End of Study form on the online application website.

Yours sincerely,

[redacted]

Professor John DH Porter
Chair
ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

B.2 TPC I – SURVEYS, INTERVIEWS & OBSERVATIONS DATA

Electronic
RESEARCH INTEGRITY AND COMPLIANCE
THE UNIVERSITY OF UTAH

IRB_00042147
PI: Michael Magill
Title: Transformed Primary Care – Care by Design

Thank you for submitting your request for approval of this study. The IRB has administratively reviewed your application and a designated IRB member has determined that your study is exempt, under 45 CFR 46.101(b), Category 2, from the Federal regulations governing human research.

It is the policy of the University of Utah that all human subject research which is exempt under this section will be conducted in accordance with (1) the Belmont report (http://ohrp.osophs.dhhs.gov/humansubjects/guidance/belmont.htm), (2) this institution's administrative procedures to ensure valid claims of exemption, and (3) orderly accounting for such activities. All research involving human subjects must be approved or exempted by the IRB before the research is conducted.

Since this determination is not an approval, it does not expire or need renewal. This determination of exemption only applies to the research study as submitted to the IRB and you are expected to follow the protocol as outlined. Before implementing any changes in the study, you must submit an amendment application to the IRB and secure either approval or a determination of exemption.

B.3 TPC II – OPERATIONS DATA

Electronic
RESEARCH INTEGRITY AND COMPLIANCE
THE UNIVERSITY OF UTAH

IRB_00042886
PI: Michael Magill
Title: Transformed Primary Care – Care by Design II

Thank you for submitting your request for approval of this study. The IRB has administratively reviewed your application and a designated IRB member has determined that your study is exempt, under 45 CFR 46.101(b), Category 4, from the Federal regulations governing human research.

It is the policy of the University of Utah that all human subject research which is exempt under this section will be conducted in accordance with (1) the Belmont report (http://ohrp.osophs.dhhs.gov/humansubjects/guidance/belmont.htm), (2) this institution's administrative procedures to ensure valid claims of exemption, and (3) orderly accounting for such activities. All research involving human subjects must be approved or exempted by the IRB before the research is conducted.

Since this determination is not an approval, it does not expire or need renewal. This determination of exemption only applies to the research study as submitted to the IRB and you are expected to follow the protocol as outlined. Before implementing any changes in the study, you must submit an amendment application to the IRB and secure either approval or a determination of exemption.

B.4 TPC III – OUTCOMES ANALYSIS

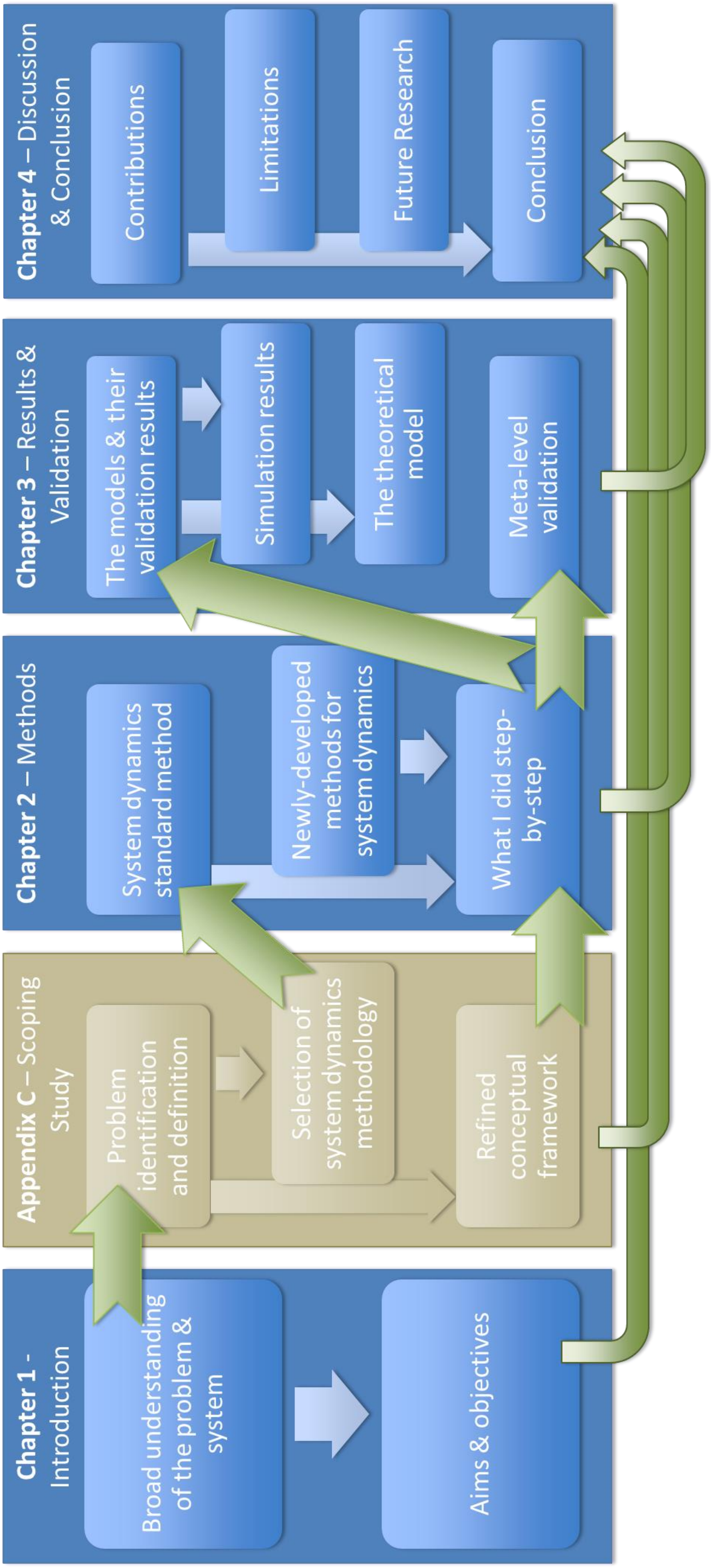
INSTITUTIONAL REVIEW BOARD
THE UNIVERSITY OF UTAH
75 South 2000 East Salt Lake City, UT 84112 | 801.581.3655 | IRB@utah.edu

IRB: IRB_00042890
PI: Michael Magill
Title: Transformed Primary Care – Care by Design III UPDB

This New Study Application qualifies for an expedited review by a designated University of Utah IRB member as described in 45 CFR 46.110 and 21 CFR 56.110. The research involves one or more activities in Category 5 (published in 63 FR 60364-60367). The designated IRB member has reviewed and approved your study as a Minimal risk study on 1/28/2011. Federal regulations and University of Utah IRB policy require this research protocol to be re-reviewed and re-approved within 1 year from the approval date.

Your study will expire on 1/27/2012.
Any changes to this study must be submitted to the IRB prior to initiation via an amendment form.

APPENDIX C: SCOPING STUDY



C.1 INTRODUCTION

> "Society's problems seem to have grown more complex and difficult to resolve … careful observation is critical [to] understanding these problems"[97] (p. vii-viii).

This appendix presents the scoping study that was undertaken to refine the problem definition and, in so doing, to inform the conceptual framework and choice of methods for the dissertation. The appendix is organized as follows: initial problem definition, methods of the scoping study, results of the scoping study, the resulting problem definition, followed by the conceptual framework.

Datasets[91][304] used in the scoping study were collected as part of the Mixed Methods Project (described in Chapter 1). The Mixed Methods Project performed both primary and secondary data collection. Primary data collection included: interviews with HSDO management and clinic personnel, in-clinic observations and measurements of PCMH implementation[92]. Secondary data included: outcome measures (e.g., quality and productivity measures), surveys of HSDO management (e.g., culture), clinic personnel (e.g., burnout, culture, teamness, satisfaction) and patients (e.g., satisfaction)[93][304].

A stepwise mixed methods approach allowed for learning from one step to inform the next step of the scoping study. The approach comprised four steps.

- In Step 1, a qualitative analysis was conducted. It pointed to a complex system of tensions operating in primary care service delivery; tensions which not only make PCMH difficult to achieve but also cause variation in the way in which primary care is practiced.
- In Step 2, the tensions were explored using mixed methods. The results further confirmed and deepened the understanding of the findings of Step 1.
- In Step 3, further qualitative and quantitative analyses were conducted, which resulted in an improved understanding of practice variation and its potential link to patient health; as well as the final choice for SDM as the appropriate methodology for the dissertation.

---

[91] These data include the interviews used in this dissertation as well as other datasets used only in the scoping study. Please see Tomoaia-Cotisel *et al.*, Table 1 for a description of the methods used in data collection and analysis, see Table 2 for a description of the datasets generated, and Table 3 for a description for how each variable (for the quantitative data) was measured.

[92] I conducted all of the interviews and led a small team in conducting the in-clinic observations.

[93] I oversaw the organization of that data: structuring filing system for the datasets, preparing templates for cataloging each dataset as well as documenting revisions to the dataset when they were made.

- In the cross-cutting step (occurring in tandem with the previous steps), these findings were verified via stakeholder dialogue. This activity provided confidence in moving forward with each step and resulted in verifying SDM as an acceptable and appropriate research approach for this study. All four steps contributed to a refined problem definition and conceptual framework.

Table C.1 below provides an orientation to the stepwise process of the scoping study. For each step in the scoping study, I identify the input(s), the methods and the result(s) (output(s)). The initial input was data from the Mixed Methods Project. The final output was the problem definition and choice of SDM as the methodology for this dissertation.

Table C.1 Scoping Study for the Purpose of Problem Definition

	Scoping Study Steps	The Mixed Methods Project data (Inputs to methods used)	Methods of the Scoping Study (Section C.3)	Results: Problem Definition & Choice of SDM as the Methodology for This Dissertation (Section C.4 and C.5)
Scoping Study	Step 1 – *Qualitative Analyses*	Semi-Structured Interviews, Reflective Journal	Reflective Journaling, Thematic Coding	***Emergent Themes*** – respondents experienced tensions in implementing PCMH ***Problem Identification*** –practice variation results from the way in which clinicians respond to tensions but the mechanism was not well-understood
	Step 2 – *Mixed Methods Case Studies*	Semi-Structured Interviews, Process & Outcome Measures	Case Studies	***Problem Definition*** – a more detailed understanding of the problem (with examples via case comparison)
	Step 3 – *Further Quantitative & Qualitative Analyses*	Process Measures, Emergency Department Data & Thematic Codes	Statistical Analyses and Review of Thematic Coding	Selecting a Complex Systems Methodology for Further Investigation – identified SDM as the methodology of choice
	Cross-cutting Step – *Stakeholder Dialogue*	Emergent Themes, Problem Definition & Methodology Selected	A Series of Stakeholder Dialogues	***Validation*** – Stakeholders verified: • The existence of tensions • The difficulty of implementing PCMH • Variation in improvement of service delivery and outcomes • The importance of complexity and approval of SDM to study PCT

C.2INITIAL PROBLEM DEFINITION

The problem statement was defined over the course of the scoping study. Initially, the problem was conceptualized broadly as a lack of understanding of the "challenges faced by primary care practices as they transform"[3]. The Mixed Methods Project approached this problem by looking at PCT's component parts independently, stating "that [while] the system as a whole must change, it is nonetheless important to focus on specific elements in relative isolation"[310].

C.3METHODS OF THE SCOPING STUDY

C.3.1STEP 1 QUALITATIVE ANALYSES

The aim of the interviews was to document the experiences of stakeholder groups (management, clinicians and clinical staff members), enabling an analysis of how critical factors contribute to the success of the transformation process[310]. The interviews focused on how and why changes occurred; thus aiding to rule out alternative hypotheses (e.g., hypotheses identified via outside observers' instincts and correlations). In so doing, these descriptions identified links suggestive of causality between process change and outcomes (e.g., a process change which enhances coordination of post-surgical care between clinicians and surgeons can maintain health by detecting and solving the problems which too often lead to hospital readmission – an outcome of interest).

Interviews were semi-structured; an interview protocol was developed and followed, but questions were modified and rearranged as needed based on the flow of the conversation. The interview protocol was informed by a protocol developed for the National Demonstration Project. Funded by the American Academy of Family Practitioners, this was the first large-scale study evaluating the implementation of PCMH principles in 36 HSDOs across the US[63, 311]. The Mixed Methods Project's interview protocol had two sections. In the first section of the interview, questions explored stakeholders' experience with transformation; specifically inquiring into factual (e.g., specific event) and interpretive (e.g., what the event meant) reflections regarding the process of change [312]. In the second section of the interview, questions were formulated to explore the competing and interacting forces that may have been at play during the transformation as components were assumed to be interdependent [313]. Therefore, topics covered in the interviews included: changes made in care delivery (i.e., changes noticed in clinic procedures, features that they perceive are working well and poorly, and their perceptions of reasons underlying their experience); management processes implemented to facilitate these changes; adaptations introduced at the local level; and perceptions of the degree to which success and/or failure has occurred.

I interviewed one primary care provider (herein, clinicians) and one staff member (herein, clinical staff members) from each care team in the HSDO's 10 community clinics (n=46 interviews) as well as the "center manager" and "medical director" for each clinic and members of the HSDO senior leadership team (n=40 interviews).

All interviews were conducted in person. Most interviews lasted about an hour (with a range of 30 to 90 minutes). I spent a few days in each clinic, interviewing those previously selected. The community clinics Quality Director as well as clinic center managers performed the selection using purposive sampling. Selection criteria were: 1) to have one clinician and one clinical staff member from each team[94] 2) to capture a range of approaches to implementation and 3) to capture a range in the length of involvement in the transformation process (e.g., initial implementation, introduction into a clinic mid-way through implementation).

Some members of the Mixed Methods Project worked together as a subunit, called the qualitative research team. This team kept reflective journals. The purpose of reflective journaling[300] was to surface qualitative team members' assumptions and preconceptions, as well as to document insights, findings, and rationale for methodological choices. I conducted all of the interviews for the Mixed Methods Project. During interviews and during in-clinic observations, I took notes. After each interview and at the end of each day of observations, when I conducted interviews, I took time to reflect on the discussions and my observations and write down my impressions especially regarding the frequently mentioned themes. Other qualitative team members also participated in observations and kept reflective journals by taking journal notes throughout the process of the research.

In meetings of the qualitative research team, reflective journal entries were reviewed and discussed. This process informed our team's efforts to achieve collective agreement on how to interpret the interviews in thematic coding and how best to carry out the Mixed Methods Project's evolving mixed methods study design. **Specifically for this dissertation, this process of reflective journaling permitted the preliminary identification of emergent themes.**

Within the Mixed Methods Project's study design, the purpose of thematic coding was to supplement quantitative analyses by linking concepts back to quantitative results, specifically to provide description of the setting, the transformation process, why and how the

[94] Teams ranged in size from 1 to 5 primary care clinicians, with clinician to clinical staff member ratios usually being between 1 and 2.

transformation occurred, and to identify facilitators and barriers of practice transformation[310]. Interviews were recorded, audio recordings were transcribed and transcripts were thematically coded in stages using NVivo qualitative data analysis software[176]. The coding plan was to identify categories and themes regarding the content of PCMH, its origins at the HSDO and variations on the PCMH model.

Initial coding focused on specific elements of PCMH (e.g., access, care teams) and key concepts from the interview tool (e.g., satisfaction, turnover) in relative isolation (e.g., the purpose was to describe each individual element, without specific attention to how it may relate to other elements). Reflective journaling emphasized the importance of tensions (i.e., mysterious interdependence among components leading to tension between two goals of the system, such as continuity of care versus access). Therefore, a second round of coding was performed (herein “thematic coding”), focusing on the tensions. An improved understanding of the tensions was gained, including the identification of emergent themes (e.g., team learning).

C.3.2STEP 2 MIXED METHODS CASE STUDIES

Convergent case studies[291] were undertaken to improve understanding of the tensions. This technique refers to the merging of findings from two parallel case studies to inform one set of conclusions. Specifically, the goal was to move from problem identification to a more concrete problem definition, including outcomes of interest.

This design had three general, theory-generating goals. The first goal relates to the “opposition” strategy proposed by Poole *et al.*, following their suggestion of “juxtaposing contradictory propositions and assumptions” to develop the outline of a theory[314]. The second goal is related to the first, and was to contrast the two types as a stepping-stone to developing theoretical arguments, such as which variables and concepts should be studied in future research, as suggested in other work[315, 316]. The third goal was an important theoretical goal established in the second stakeholder dialogue (see “Cross-cutting step” below), which was to control for context.

Since interviews often identified clinicians with vastly different approaches to dealing with the tensions, a case study comparison was used to explore two different approaches and resulting experiences. Two clinicians were selected because they typified two common and opposite approaches to dealing with the tensions. Both clinicians were practicing in the same clinic, thereby controlling for context. As clinical staff members were named in two of the tensions (see Table C.3 below), their interviews were also reviewed and compared. Specifically, this case study comparison consulted each clinician's interview as well as the interview that was conducted for a clinical staff member on their team.

These two clinicians were also compared on various quantitative measures captured as part of the Mixed Methods Project[95]. Descriptive statistics were generated on implementation, productivity, teamness, burnout, and quality for each clinician's team (please see Tomoaia-Cotisel *et al.*[304] for detailed description of measures, data collection and methods). Descriptive statistics and statements were juxtaposed across the cases [304].

C.3.3STEP 3 FURTHER QUANTITATIVE & QUALITATIVE ANALYSES

Step 2 resulted in a clearer description of the problem and in identifying the need for using complex systems methods to better understand the problem. Step 3 (described below) considers a selection of the statistical analyses performed in the Mixed Methods Project (section C.3.3.1) together with new analyses (section C.3.3.2). The objective is to identify an appropriate complex systems methodology for this dissertation research.

C.3.3.1 RELEVANT ANALYSES FROM THE MIXED METHODS PROJECT

Analyses from the Mixed Methods Project that were considered in meeting Step 3's objective were those that contributed to answering the following two questions:

First, what is the operative level of aggregation for the tensions? An exploratory analysis used correlation analysis to study clinicians' practices (e.g., degree of adherence to primary care tenets) and outcomes of interest at the clinic level (quality, patient satisfaction, clinician satisfaction and financial performance). Analysis of Variance (referred to as ANOVA) of clinic observations data explored relationship between clinicians' practices and 1) the team they belonged to and 2) the clinic where they were located. A k-means cluster analysis, followed by ANOVA analysis, compared clinicians' practices across the HSDO, rather than by team or clinic (six distinct clusters were identified).

Second, does variation in clinicians' practices impact health? If this variation does not impact health, while the tensions, as experienced in this setting, would remain important from a management standpoint, they would no longer be important from a primary care standpoint. To explore this question, patients over 65 years of age were matched to primary care clinicians using a claims database. Patients were divided into comparison groups across all six clinician

---

[95] These data were gathered as a team. My role was to design and run the analysis described here. I also share lead authorship on the corresponding peer-reviewed publication.

clusters using propensity score matching. Comparison of health by clinician practice cluster was done using t-tests and odds ratios.

Patients' health was considered using a proxy measure assumed to be potentially sensitive to differences in primary care practice for patients over age 65 during the time frame considered: emergency department visits. Patients' emergency department visits were categorized using the New York University algorithm (the four categories are: not preventable/avoidable, preventable/avoidable, primary care treatable, and non-emergent) [317] Student's t-tests were used to assess within-category differences among the six practice clusters.

The ratio of emergency department visits that were primary-care-related and those that were not (i.e., the "PC Ratio") was calculated for each cluster. Odds ratios were calculated to compare the PC Ratio for the initial and comparison clusters. This approach estimates the differences in patients' health which can be traced to primary care service delivery, across the patients of clinicians in the six practice clusters.

C.3.3.2 METHODS FOR SELECTING A COMPLEX SYSTEMS METHODOLOGY

To better understand how to study what gives rise to the tensions, I needed to select an appropriate complex systems methodology. This selection relates to the paradigm issues raised in Chapter 1. The agent/structure (aka individual/context) problem in social science can be resolved by using methods that consider the "intertwined relationship of individual and social setting" including how this relationship plays out over time through "the fact of unintended consequences"[98] (p. 207). The qualitative and quantitative analyses that were utilized to select a complex systems methodology for the dissertation are presented below.

These analyses were crafted to distinguish how much to focus on agents or social networks. Depending on the answer, this study utilized the framework in Table C.2 (adapted from descriptions in Rahmandad & Sterman[290]) to select a complex systems methodology. This framework considers two issues in identifying the essential aspects of the problem under study: heterogeneity of agents and the social network linking agents together. Rahmandad and Sterman[290] use an infectious disease model with agents placed into four compartments: susceptible, exposed, infected and recovered. Depending on the epidemiology of the disease (e.g., a disease spreading easily through a cough or touching something an infected person touched previously *versus* a disease that requires contact with infected blood), people's ability to move across compartments will differ on a spectrum between a fully connected network and one where each person is only connected to one's neighbors. In the former, people can move easily between compartments (easily allowing for mixing). In the latter, people cannot move as easily between compartments (thus compartments matter).

If neither heterogeneity of agents nor the social network linking agents together are found to be essential to the problem, then assumptions for differential equation modeling appear to be present. (e.g., system dynamics modeling). Therefore, I present a third question to confirm the importance of system structure to the problem. If the answer to this question fails to identify system structure as essential, then a return to the problem statement is warranted.

Are the characteristics of *individual agents* important to variation in clinicians' practices? If so, which agents matter beyond clinicians (i.e., clinical staff members, HSDO leadership or local clinic leadership)? Additional k-means **cluster analyses** as well as review of previous **thematic coding** of clinical staff member and clinician interviews looked for common characteristics of the individuals implementing that might explain the variation.

Are *social networks* important to variation in clinicians' practices? If so, are they formal or informal? A review of previous **thematic coding** of clinical staff member and clinician interviews (that described in Step 1 above) looked for descriptions regarding the existence and influence of social networks.

Is the *system's structure* important to this variation? If so, what is the system boundary (i.e., the most relevant system scope and time frame)? A review of previous **thematic coding** of clinical staff member and clinician interviews, as well as a **review of** the additional k-means **cluster analyses** (those done for *agents* above) was conducted to look for the existence and influence of system structures on implementation.

Table C.2 Framework for Selecting a Complex Systems Methodology

<table>
<tr><td colspan="2" rowspan="2">Selecting a complex systems methodology</td><td colspan="2">Have the characteristics of individual agents been demonstrated to be essential to understanding the problem?</td></tr>
<tr><td>No (homogeneity)</td><td>Yes (heterogeneity)</td></tr>
<tr><td rowspan="2">Has the social network linking agents been demonstrated to be essential to understanding the problem?</td><td>Yes (compartments matter)</td><td>Social Network Analysis</td><td>Agent-based Modeling</td></tr>
<tr><td>No (perfect mixing between compartments)</td><td>Differential Equation Modeling</td><td></td></tr>
</table>

The methods described in Steps 1-3 were performed sequentially, in a mixed methods fashion. Formal validation procedures and assumptions vary according to the method used. In the next section, I describe the approach I used to validate the methods and results in Steps 1-3.

C.3.4CROSS CUTTING STEP: STAKEHOLDER DIALOGUE

Stakeholder dialogue is a process whereby the conceptual knowledge of stakeholders is accessed to inform the research[2]. This was in accordance with Miller[124] that expected Mixed Methods study designs to empower research teams to engage "policy-makers, system leaders, and practitioners in dialogue about the nature of the research and findings"[318].

Devereka *et al.*.[319] present a framework for effective stakeholder engagement. I use their definition for stakeholders as well as for stakeholder engagement:

- **Stakeholders** – "individuals, organizations or communities that have a direct interest in the process and outcomes of a project, research or policy endeavor"
- **Stakeholder engagement** – " an *iterative* process of *actively* soliciting the knowledge, experience, judgment and values of individuals *selected* to represent a broad range of direct interests in a particular issue, for the dual purposes of: [1] creating a shared understanding [and 2] making relevant, transparent and effective decisions." (emphasis added)

A wide variety of methods exist. The core of any stakeholder engagement process is to create the opportunity for iterative interactions that generate rich dialogue and reciprocal learning[319].

The approach used in the scoping study includes aspects of the Mixed Methods Project as well as more informal face-to-face interactions. Stakeholders engaged were: local, national and international experts in primary care, problem owners at the HSDO, primary care clinicians and experts in health services research, health systems research and SDM in health. In the corresponding results section below, I describe several stakeholder dialogue methods used during the scoping study: (1) a presentation of preliminary findings to the HSDO management, (2) a discussion with grantees (i.e., recipients of the Transforming Primary Care grants) as well as national health services researchers, (3) poster presentations at two US national health conferences, and (4) a presentation to the Mixed Methods Project's advisory committee.

C.4RESULTS OF THE SCOPING STUDY

C.4.1STEP 1 OUTPUT: EMERGENT THEMES, PROBLEM IDENTIFICATION

In Step 1, initial coding of the semi-structured interviews provided a background to the terms and concepts in this context. Reflective journaling and thematic coding of those same interviews led to emergent themes and to the identification of the problem of interest in the HSDO. Thematic coding permitted, first, the identification of an important problem, second, descriptions of the causal structure around that problem and third, an understanding of the dynamic complexity of that system structure.

In reflective journaling, my notes indicated that clinicians implemented PCMH differently from one another. Clinicians describe reasons for this variation as having to do with feeling emotional distress from being strained or stretched by the competing forces at play in the transformation process. I termed these forces 'tensions'. The Oxford English dictionary defines tension as "(1) the state of being stretched tight… a strained state or condition resulting from forces acting in opposition to each other [or] (2) mental or emotional strain… a relationship between ideas or qualities with conflicting demands or implications".[174]

I created a list of these tensions as several binary sets of trade-offs between one goal or priority and another. Table C.3 below presents this list with a description of what each tension means.

Table C.3 Commonly Identified Tensions

Tension	Further Explanation
Continuity of care versus *Access*	Ensuring a high level of continuity of care involves patients seeing their own clinician. This conflicts with the goal of allowing patients to seek care at the time it is needed.
Generalist 'clinical staff members versus *Specialist clinical staff members*	Having all clinical staff members doing the various clinical staff member tasks so that anyone can complete a needed task, this conflicts with the desire to have clinical staff members' work be error-free (as having specialized clinical staff members was said to be better at meeting this goal).
Health maintenance versus *Chief complaint*	Treating the patient for his various needs conflicts with the desire to only address his top priority ailment (e.g., due to the amount of time allotted per visit).
In-clinic work versus *Out-of-clinic work*	There is a lot of work to do in order to complete all that a patient needs. Getting that work done conflicts with the desire to get it done during clinic hours (as it takes more time than available).
Managing to the vision versus *Managing to reality*	Ensuring that management activities support the vision of PCMH conflicts with the cognitive limitations of implementing such a complex initiative in the context of day-to-day realities.
Mandatory for clinical staff members versus *Negotiable for clinicians*	Clinical staff members' perception that implementing PCMH fully is part of their job description conflicts with clinicians' perception that they can choose their level of implementation.
Quality versus *income generation*	Ensuring optimal quality requires some un-reimbursed work which conflicts with stakeholders' need to ensure an appropriate level of revenue is generated.
Standardization versus *Deviation*	Increasing the degree of standardization in clinical staff members' jobs improves clinical staff members' ability to work with multiple clinicians, but it conflicts with clinicians' individual preferences for what they want their clinical staff members to do.
Fostering a well-functioning system versus *Getting through the day*	Ensuring the smooth operation of an often-hectic work environment conflicts with the need of the health system to ensure that every task which could help a patient be completed.
What is asked versus *What is rewarded*	Complying with what the organization and culture aspire to achieve conflicts with the incentive structure in place.

As reflective journaling emphasized the importance of the tensions, thematic coding focused on emergent themes related to these tensions. Emergent themes clarified that variation in practice was influenced by the ways in which clinicians responded to tensions. I found that when respondents could identify causal mechanisms behind a given tension (e.g., *Mandatory for clinical staff members* versus *Negotiable for clinicians)*, it was a unidirectional set of causal relationships from one concept to the other (e.g., implementing PCMH increases clinical staff members' ability to do their job, increasing clinical staff member job satisfaction). The tensions were generally resolved, but towards an undesired, suboptimal solution. However, the means for achieving both goals (the optimal solution) was uncertain. This uncertainty was the source of emotional distress, which participants expressed when describing the 'tensions'. Therefore, participants described the tensions as being very important to understand in PCT as they prevent achieving multiple goals simultaneously and PCT requires doing just that. This step uncovered the problem of interest for my dissertation.

C.4.2STEP 2 OUTPUT: PROBLEM DEFINITION

Step 2 started with the problem identification from Step 1 – the tensions (and validation from the Cross-cutting Step, indicating that indeed the identified problem was an important theoretical and operational problem). In this step, we moved from identifying the problem to giving it an explicit definition.

We do so by choosing a narrower lens: focusing on the experience of two teams, as described in methods above. As we began to connect the dots[304] between emergent themes and measurable outcomes for these two teams, we concluded that we could observe the rough outline of a complex system.

The case studies utilized quantitative data to verify the differences in clinician practices between Team A and Team B described in interviews and to explore differences in outcomes. These teams had very different experiences with PCMH implementation as they were allowed to co-exist and co-evolve with very different approaches to care. Team A and Team B responded to tensions identified in Table C.2 above differently. Team B opted for managing to reality, income generation, and getting through the day; while Team A opted for managing to the vision, quality, and fostering a well-functioning (health) system.

The main results can be summarized as follows:

- Team B provided more comprehensive care and statements indicated they were more committed to the goals of PCMH than Team A was
- Team B also had better quality outcomes for most diseases tracked than Team A
- However, Team A had higher productivity than Team B.
- Team A also had lower cynicism and lower exhaustion and statements indicate that they were more likely to have "regulars" (i.e., patients coming in frequently and for the same services) than Team B.

In addition to pointing out these differences, Team A and Team B interviews also described how these differences emerged, both in relation to changing situations and interpersonal relations. Clinical staff members often provided a complementary perspective to that of clinicians, on the choice of practices on their teams.

The insights emerging from this analysis, based on experiences of both clinical staff members and of clinicians, were useful in defining the problem more explicitly. This new problem statement is reproduced in the “resulting problem definition” section below (as well as in Chapter 1).

C.4.3STEP 3 OUTPUT: SELECTING A COMPLEX SYSTEMS METHODOLOGY

In order to identify an appropriate complex systems methodology for this dissertation research, Step 3 presents a selection of the statistical results performed in the Mixed Methods Project (section C.4.3.1) together with new results (section C.4.3.2).

Section C.4.3.1 presents results answering:

- What is the operative level of aggregation for the tensions?
- Does variation in clinicians' practices impact health?

Section C.4.3.2 presents results answering:

- Are the characteristics of individual agents important to variation in clinicians' practices?
- Are social networks important to variation in clinicians' practices?
- Is the system's structure important to this variation?

C.4.3.1 RELEVANT RESULTS FROM THE MIXED METHODS PROJECT

The Mixed Methods Project's first statistical analysis[96] explored relationships between PCMH implementation and outcomes of interest using correlation analysis at the clinic level. This analysis found some potentially interesting and some counter-intuitive relationships. Nevertheless, the mechanisms were unclear. Reflecting on opportunities for future research, we stated:

> *"The elements of care teams and continuity of care are inherent to the structure and success of the PCMH; however, the **comprehensive system as a whole is likely more important than any individual elements**. Further, PCMH implementation **is a dynamic process, with changing relationships** between individual elements. The **context** in which one operates may have considerable **impact on individual elements** as well as the **overall system design**. Full evaluation of PCMH implementation will **require complex mixed methods studies** to identify the most productive approach to primary care redesign." (emphasis added) [56]*

Further statistical analysis explored three specific questions and each conclusion informs the next analysis. First, what is the operative level of aggregation for the tensions? Our findings indicated that clinician implementation clustered based on the way clinicians practice, not their team or clinic. Second, does practice variation impact health? Our findings indicate that

[96] This is the only analysis presented in this dissertation where I held a secondary role. As a member of the Mixed Methods Project team, I participated in discussing the methods, reviewing the results and writing the paper. Commensurate with this role, I was a co-author on the peer-reviewed publication presenting this work.

indeed, this variation is important to patient outcomes (and thus a good fit for a dissertation in public health). The statistical analyses performed are described in Table C.4 below.

Table C.4 A Summary of Relevant Mixed Methods Project Statistical Analyses

Research Question	Method Summary	Results Summary	Conclusion	Reference
What is going on at the clinic level?	Correlation of PCMH implementation and outcomes of interest at the clinic level (quality, patient satisfaction, clinician satisfaction and financial performance)	Some potentially interesting and some counter-intuitive relationships, stopping short of identifying causal mechanisms	Implementation and outcomes appear to be inter-related in unexpected ways, but the causal mechanisms are unclear. Identified need to look at PCMH implementation using "more complex mixed methods" designs	Day *et al.*[56]
What is the operative level of aggregation for studying the tensions?	ANOVA analyses checking for the ability of team or clinic membership to explain individual-level variation in the way clinicians practice	When different clinicians practice in a similar manner, it was not explained by the team they belonged to or by the clinic where they were located.	The operative level of aggregation for studying the tensions is at the way individual clinicians practice, not the team or clinic to which they belong	Tomoaia-Cotisel *et al.*[190, 191, 320]
	k-means cluster analysis, followed by ANOVA analysis	Grouped clinicians across the HSDO into 6 distinct practice styles, or clusters	Yes, this variation impacts health	
Does practice variation impact health?	Patients were matched to primary care clinicians using a claims database with patients over 65 years of age. Patients were divided into comparison groups across all clinician clusters using propensity score matching. Comparison was done using t-tests and odds ratios.	The rate of primary care preventable emergency room admissions varied significantly by practice cluster, with one practice style consistently performing better than the rest in comparisons.		

C.4.3.2 RESULTS INFORMING SELECTION OF A COMPLEX SYSTEMS METHODOLOGY

To better understand how to study what gives rise to the tensions, the next step was to select an appropriate complex systems methodology by assessing the importance of social networks, agents, and system structure in understanding clinicians' variation in PCMH implementation. Three questions were considered.

First, are the characteristics of individual agents important to this variation in clinicians' practices? A review of the previous thematic coding of clinical staff member and clinician interviews found that there are some common characteristics that appear to influence implementation; for example: for clinicians, their sensitivity to changes in their salary and for clinical staff members, their ability to learn.

A new K-means cluster analysis was performed. Two key findings were made. First, 6 distinct clusters were found for each of the 4 years the HSDO had been tracking PCMH implementation. However, each year's clusters had different sets of PCMH implementation tendencies from the last. Performing k-means cluster analysis on the set of 24 clusters (6 clusters per year over 4 years) found there were 4 clusters of clusters, or "groupings", which persisted across time. Second, the members of each cluster and each grouping both changed over time. All clinicians changed groupings over time, and no clinician's pattern of change was similar to another's. The example below will illustrate what is happening in the data:

> Consider a grouping with four clinicians practicing similarly in year 1. In year 2, three of them are found in new clusters and new groupings, one stays and others join forming a new cluster in the grouping. By year 3, the last of the original four clinicians has moved. The grouping is now made up entirely of new people. The four clinicians still work at the HSDO, but now all practice differently from one another and from their former selves.

Regardless of their individual-level similarities or differences, clinicians practicing similarly would be clustered in one group one year, and over the next year, they would each make different types of choices about how to implement PCMH, ending up in different clusters and groupings.

These findings indicated that, although certain intangible characteristics of agents were influential in PCT at the HSDO, the readily-quantified similarities or differences across individuals do not appear to strongly determine their PCMH implementation.

Second, are social networks important to this variation in clinicians' practices? A review of previous thematic coding of clinical staff member and clinician interviews considered two types of social network interactions: formal and informal cross-pollination. In *formal* cross-pollination, certain staff would be assigned to act as facilitators of PCMH implementation where their interactions with others in a team or clinic would influence PCMH implementation among clinicians. In *informal* cross-pollination, ad hoc interactions among staff influence PCMH implementation among clinicians.

Regarding formal cross-pollination, formal facilitator roles are not being used. Formal social interactions could occur in the periodic clinic level meetings, or in the HSDO level meetings. Participation in these meetings was not mentioned as strongly influencing the implementation process. In terms of informal cross-pollination, a few clinical staff members do "float" across groups (to address staffing shortages in any one group on any given day). This could permit clinical staff members to bring promising practices from their experience with other clinicians. This interaction was not mentioned as strongly influencing the implementation process. Therefore, our findings indicated that social networks were not an important factor in PCT at the HSDO.

Third, is the system's structure important to this variation? The new k-means cluster analysis found certain implementation patterns which persisted over time. These are the groupings described above. Clinicians practicing similarly would be clustered in one group one year, and the next year, the behavioral tendencies of that group would persist, even with new people coming in and the existing people leaving the group over time. Considering that the groupings could be seen as "compartments", these observations indicate that clinicians' initial compartment does not determine which compartment they will be in the next year or the year after.

A review of previous thematic coding of clinical staff member and clinician interviews found that various structural elements were mentioned as influencing care team decisions to implement, for example: the staffing level and the incentive structure. Together with the observation of mixing across compartments, these findings indicate that system structure strongly influenced PCT at the HSDO.

This leads to the follow-up question: what is the system boundary? For system scope, structural elements mentioned as influencing decisions relating to implementation and to the variation in clinician practices, including the tensions need to be considered to understand PCT. The time frame under study should span multiple years and consider change at a more granular level.

Table C.5 below identifies these questions and presents a summary of the methods, their results, and the conclusions drawn.

Table C.5 Quantitative & Qualitative Analyses to Inform Selection of Complex Systems Methodology

Research Question	Method Summary	Results Summary	Conclusion
Are the characteristics of individual agents important to variation in clinicians' practices?	k-means cluster analyses	No, clinicians practicing similarly would be clustered in one group one year, and the next year, those clinicians would be found in different clusters.	To some extent
	Review of previous thematic coding of clinical staff member and clinician interviews	Yes, there are some common characteristics that appear to influence implementation, for example: for the clinician, their sensitivity to changes in their salary and for the clinical staff member, their ability to learn.	
Are social networks important to variation in clinicians' practices?	Review of previous thematic coding of clinical staff member and clinician interviews	Informal cross-pollination: A few clinical staff members do "float" across groups (to address staffing shortages in any one group on any given day) but they were not mentioned as strongly influencing the implementation process. Formal cross-pollination: Clinical staff members and clinicians have periodic clinic level meetings. Clinicians also have less frequent HSDO level meetings. These were also not mentioned as strongly influencing the implementation process.	No
Is the system's structure important to this variation?	Review of previous thematic coding of clinical staff member and clinician interviews	Various structural elements were mentioned as influencing care team decisions to implement, for example: the staffing level and the incentive structure.	Yes
	k-means cluster analyses (same as for agents above)	Certain implementation patterns persisted over time. Clinicians practicing similarly would be clustered in one group one year, and the next year, the behavioral tendencies of that group would persist with new people coming in, as the original cluster would dissolve with members moving to different clusters and often to different groups.	

As described in Section C.3.4 above, validation consisted of obtaining expert opinion via several means including: (1) a presentation of preliminary findings to the HSDO management, (2) a discussion with other Agency for Healthcare Research and Quality grantees as well as national health services researchers, (3) poster presentations at two US national health conferences, and (4) a presentation to the Mixed Methods Project's advisory committee.

(1) A presentation of preliminary findings to the HSDO management[321]

I presented the results of step 1 to the HSDO management as preliminary findings and engaged with them in discussion. This presentation included the list of tensions in Table C.2, above. HSDO management verified the existence of the 'tensions' (practice variation was a recognized problem already). Also, they suggested that my research on this subject would be most useful if it were able to help inform HSDO policies where a small change could impact all of the tensions (i.e., to look for *high-leverage* interventions)[322].

(2) A discussion with other US Agency for Healthcare Research and Policy grantees as well as national health services researchers

I also presented the results of step 1 at a US health services research conference hosted by the US Agency for Healthcare Research and Quality. The poster[323] focused on the evolving role of qualitative methods in the Mixed Methods Project and included some illustrative findings on the tensions and practice variation. Conference attendees with whom I discussed the poster corroborated the observation that implementation of PCMH was harder than had been expected. PCMH implementation did not result in the expected uniform improvements of health services delivery or outcomes.

At this conference, the US Agency for Healthcare Research and Quality also hosted a special closed-door meeting among Transforming Primary Care grantees. During this meeting, study teams reported a great deal of variation in the design of PCMH across their project sites and their perception that it would be impossible to generalize results - summed up by participants as 'once you've seen one PCMH, you've seen one.' They expressed concern that the intended users would not be able to extract "lessons learned" from the complexity introduced by issues of variation in designs and context across studies.

At this meeting, two participants (myself included) identified complex systems methods as potential tools to address this limitation. There was no resistance to using these approaches.

Grantee discussions continued over the following years. Efforts to wrestle with this complexity resulted in sharing of data and experiences, and ultimately in the publication of now two articles illustrating the linkage between variation in practice and context across clinics studied by these grantees[16, 57].

The Mixed Methods Project also grappled with this complexity independent of the other grantees, resulting in the process (and results) presented in Step 2[304].

This grantees meeting and follow-on efforts affirmed the acceptability and appropriateness of complex systems methods for studying PCT.

(3) Poster presentations at US health conferences

Step 3 results were presented at two national conferences[190, 320]. I engaged in dialogue with conference attendees from public health and health services research disciplines. Conclusions from these discussion included that using SDM would be an acceptable and appropriate approach.

(4) A presentation to the Mixed Methods Project's advisory committee[324]

Mixed Methods Project team members presented preliminary findings and plans for future research to the project's advisory committee: an audience consisting of both local and national stakeholders (policy and research experts, as well as management and front-line clinicians working in the US). At this meeting, I presented steps 1-3, focusing on the problem definition and proposal of using SDM as the complex systems method of choice (with preliminary results). Mixed Methods Project advisory committee members verified that using SDM to address the identified problem would be an acceptable and appropriate course of action.

C.5RESULTING PROBLEM DEFINITION AND IMPLICATIONS FOR THE DISSERTATION STUDY DESIGN

Based on the findings of the scoping study, this dissertation conceptualizes the problem as follows:

> *Primary care transformation has been and continues to be an elusive target. In the short term, implementation is hard and failure abounds. In the long term, some practices reach successful implementation. We lack sufficient understanding of the structure of primary care, and of the policies that can impact this structure.*

This dissertation seeks to develop a grounded, dynamic theory of PCT in order to build understanding of the key structures generating these hoped-for and feared observed behaviors.

Results from the scoping study inform the dissertation's approach to meeting this objective. Table C.6 below lists the scoping study steps and summarizes the results obtained as well as their contributions to the dissertation's study design.

This problem statement is the cumulative result of all phases of this dissertation. As presented in this appendix, tensions were found to exist within the structure of primary care such that the four tenets of primary care and the context of PCMH implementation at the HSDO influenced each other, where the hoped for levels of implementation could not be reached in all tenets at the same time. These tensions involve complex interactions within the underlying causal structure of primary care which contribute to the observed failure and success modes.

While the value of the primary care tenets is well understood, current theory lacks an understanding of the complex interactions between them as well as their interaction within the system of care already in place. This understanding is necessary in order to realize the aspirations of health care systems worldwide.

Table C.6 Scoping Study Findings and Implications for the Dissertation Study Design

Scoping Study	Scoping Study Results	Implications for the Dissertation Study Design
Step 1 – *Qualitative Analyses*	• Identified tensions • Individual quotations identified fragments of causal mechanisms for the tensions • Variation in practice was influenced by the ways in which clinicians responded to tensions (including suboptimal solutions) • Clinicians and clinical staff members experienced emotional distress in dealing with the tensions	• Identified problem/purpose for research • Identified participants' hopes and fears about how their practices would change over time (the "reference modes" in SDM parlance) • Variation results from the dynamics among causal mechanisms over time (the "dynamic hypothesis" in SDM parlance)
Step 2 – *Mixed Methods Case Studies*	• Quantitative data identified ways in which care teams differed • Clinician and clinical staff member descriptions linked their team's outcomes to how they addressed the tensions	• Provided further information for articulating the problem statement and developing the *dynamic hypothesis* • Identified possible existence of causal relationships between different ways of dealing with the tensions and resulting practice • Identified the clinician and clinical staff member perspectives as important to understanding the problem
Step 3 – *Further Quantitative & Qualitative Analyses*	Mixed Methods Project statistical analyses • Raised interesting questions about possible interactions between primary care tenets • Identified clinicians' way of practicing as appropriate level of aggregation (not clinic or team) • Practice variation matters to health	Mixed Methods Project statistical analyses • The identified problem matters to health • Further study of variation needs to focus on answering questions about *why/how* these variations arise
	Further Qualitative & Quantitative Analyses • Social networks were not important to this variation, because mixing across key compartments (clusters and groupings) was observed • Characteristics of individual agents were somewhat important to this variation (e.g., clinicians' salary preferences and clinical staff members' ability to learn) • System structure was important to this variation (and was described both by clinicians and clinical staff members)	Further Qualitative & Quantitative Analyses • Prioritize analysis on system structure, rather than social networks or agent heterogeneity (variation) • Take into account variation in important characteristics of agents • Consider a time frame that spans multiple years • Focus on clinician and clinical staff member interviews as the data source for building the model
Cross-cutting Step – *Stakeholder Dialogue*	• Research on the tensions should identify small policy changes which can have a large impact • Complex systems methods, and SDM specifically, are acceptable and appropriate for studying PCT	• Use SDM to build understanding on the structure of the system which gives rise to the tensions. • Use SDM to find policies that are high leverage

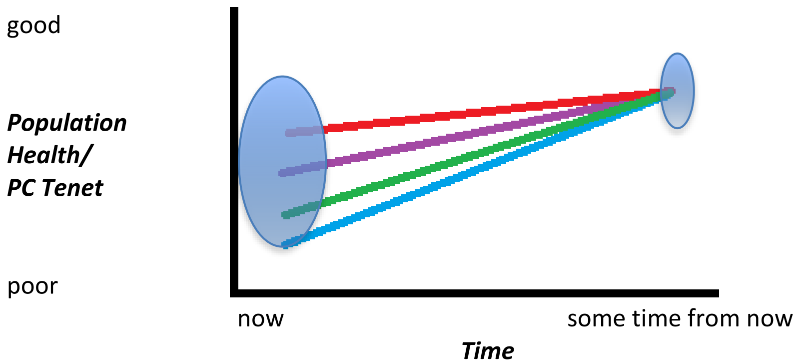
C.6 CONCEPTUAL FRAMEWORK

So far, this appendix has described how I arrived at the problem definition and selection of SDM as the methodology for this dissertation. The results providing this clarity also facilitated a clearer conceptual framework that has guided this dissertation research.

Primary care theory recognizes that the dominant perspective in US health system transformation is optimization; which has had important unintended consequences: "driven by efforts to optimize parts rather than the whole, the US health care system spins out of control, with rising costs and declining value, despite rapid technological advances"[2]. This dominant perspective in research, policy and service delivery recognizes all four primary care tenets, and that it is hard for primary care practice to reach the ideal in each tenet, but fails to take into account that they may be inter-related.

Figure C.1 presents a graphical representation of this dominant perspective[97]. The x-axis is time. The four colors represent the four tenets of primary care where the goal of PCMH implementation is, over time, to achieve a system where all tenets are at 100%. The bubbles correspond to the state of population health; where, the taller the bubble, the wider the variation in health, and the higher the center of the bubble, the better the average health. The left of the graph presents one configuration of the tenets (a sub-optimal structure of primary care). On the right, we have all four tenets successfully attained (i.e., successful PCT). *In between*, there is a linear progression where (from left to right) the primary care system (i.e., team, HSDO, etc.) transforms optimizing each tenet in the desired manner and optimizing health by simultaneously reducing variation and raising its average level, thereby attaining the desired PCT and population health outcomes in the optimal manner.

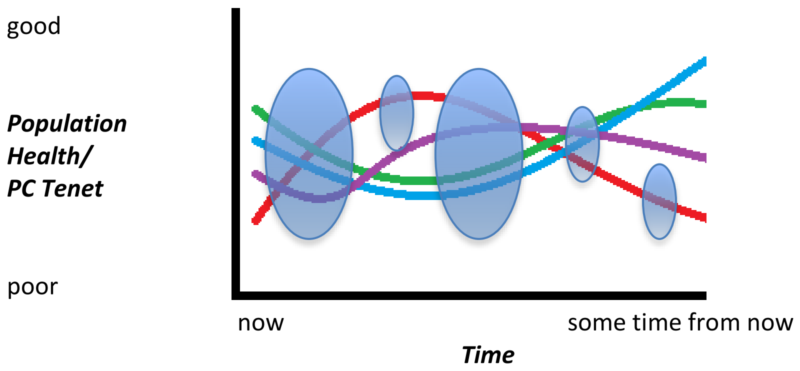
Figure C.1 Dominant Perspective of US Health System Transformation – Optimization



[97] In later chapters, this kind of qualitative graphical representation of the behavior of a variable over time is referred to as a reference mode.

Figure C.2 represents a graphical representation, based on what has been learned in the scoping study, of the impact had by tensions between the tenets on implementation and outcomes over time. Again, the four colors represent the four tenets of primary care where the goal of PCMH implementation is, over time, to achieve a system where all tenets are at 100%. However, due to the tensions between tenets, a focus on optimizing one tenet (as visually represented by increasing the red line) influences the implementation of the other tenets (as visually represented by different trajectories of the blue, green and purple lines) such that optimization of all the tenets is not reached (as visualized by never having all four lines crossing at the top). Again, the bubbles correspond to the state of population health; where the taller the bubble the wider the variation in health, and where the higher the center of the bubble the better the average health. In this graph, there are different tenet configurations (lines) and different impacts on outcomes (bubbles); however, unlike in Figure C.1, all configurations and impacts are sub-optimal (in that none reaches the target on all counts) but some are preferable to others. Also, the progression is not linear[98]. This graphical representation can be interpreted at various levels of analysis; from the team, to the clinic, to the HSDO, and to the health system as a whole. Based on the scoping study findings, the interpretation presented focuses at the team level of implementation.

Figure C.2 Alternative Perspective of US Health System Transformation – Dynamic Tension



This alternative perspective considers how the use of the dominant perspective plays out in a situation of dynamic tension between the primary care tenets. Interaction between the tenets resists overly narrow attempts to improve the system. This resistance emerges *endogenously* from the system's structure resulting in the variations observed in the success of PCT and population health. Therefore, my theoretical aim is to develop a grounded, dynamic theory of PCT; in order to build understanding of the key structures generating these varied behaviors.

[98] This figure can be interpreted at various levels of analysis; from the team, to the clinic, to the HSDO, and to the health system as a whole. Based on the scoping study findings, the interpretation presented focuses at the team level of implementation.

C.7 CONCLUSION

This appendix describes the first phase of this dissertation's journey, arriving at the problem definition, selection of SDM as the approach and a refined conceptual framework. To summarize, Step 1 pointed to a complex system of tensions operating in primary care service delivery; tensions which not only make PCMH difficult to achieve but also cause variation in the way in which primary care is practiced. Steps 2 explored the tensions, further confirming the finding of complexity. Step 3 resulted in an improved understanding of practice variation and its importance for health; as well as the choice of SDM as the methodology of choice for the dissertation. Step 4 (occurring in tandem with the previous steps) verified these findings and provided confidence in moving forward with SDM.

The goal of the modeling process is to enable "actors or stakeholders [to] build up their 'intelligence', a valuable asset to better cope with opportunities and threats as they may present themselves later." [144] It is important to note that methods selection contains an element of interpretation. Oral and Kettani describe methods selection as depending on *'managerial situation'* which is "the way that *real world events* are *interpreted* by decision makers and policy formulators" (emphasis added)[144]. As an organizationally embedded research team, we uncovered the managerial situation via the scoping study. This situation, also referred to as 'Appreciation of the Situation'[105, 325], is more than a problem to be solved, as it may also require assessment, prediction or analysis. Quantitative data were useful for confirming some aspects, but we relied most heavily on the interviews.

While this description of the scoping study started and ended with the tensions, it should be noted that it was not always clear how important they might be or where I would need to start in studying them. In this phase, I developed an understanding of the tensions' importance: first, to PCMH implementation, then, to practice variation and finally, to health outcomes. I also developed an improved understanding of the nature of the tensions and of the specific methods that might be needed to study them.

In observing the causal chain of *system* to *tensions* to *practice variation* to *health outcomes*, these analyses made it clear that the tensions mattered. The tensions are a challenge that is both an important concern for and poorly understood by key stakeholders in primary care service delivery. In exploring the sources of data available in the Mixed Methods research project, I found that, although useful for testing the hypotheses identified in qualitative analysis and building my understanding of the problems that the tensions present, numerical data on their own were unable to provide an improved understanding of the system structure giving rise to the tensions. The best source of information about the system causing the tensions was to be found in the interviews I had conducted, specifically, the interviews with clinicians and clinical staff members.

APPENDIX D: VALIDATION PURPOSE AND TYPES IN SYSTEM DYNAMICS

As part of my methodological aim, I developed some theory which applies to validation in system dynamics. Specifically, this contribution relates to the validity types described in Section D.2 below. Before this, I present a brief description of the purpose of validation in similar disciplines in the social sciences and in SDM.

D.1 THE PURPOSE OF VALIDATION IN SYSTEM DYNAMICS

Social scientists use validation tests to assess when research is ready for dissemination. These methods are applied more or less formally across different disciplines, and the aspirations can range from describing the true nature of reality to describing subjective opinions.

Groesser *et al.*[161] provides two reference points for thinking about validation. These come from quantitative and qualitative research methodologies that are widely used to study issues with multiple interacting variables: multivariate regression modeling and Grounded Theory. Both methods involve an iterative process of refinement from initial data analysis to a "final" model (of variables) in the former case and theory (of categories) in the latter case. In multivariate regression modeling, confidence relies upon the adjusted R^2 statistic.

System dynamicists have developed model validation tests which permit the use of statistical measures (e.g., the various statistics used for assessing model fit to data) for system dynamics models[113] (p. 875).

In Grounded Theory, confidence in findings relies upon three rules for determining *theoretical saturation* – that is, (a) the category is well developed in terms of its properties and dimensions demonstrating variation, (b) the relationships among categories are well established and validated, (c) no new or relevant data seem to emerge regarding a category[326-328].

System dynamicists have looked to Grounded Theory and the Case Study Method for theory-building concepts which aid them in developing system dynamics models as theories[113] (p.157-158,890)[138, 139]. This interaction with qualitative fields has influenced model building (see, Kim & Andersen and Yearworth[116, 117]) more so than model validation. I used the grounded theory saturation rules in developing my contributions to validity and validation in SDM.

SDM model validation has two objectives: "Relevance to and Consistency with Appreciation of the Situation" and "Suitability for Purpose"[105] (p. 123).

Typical studies begin with a stakeholder group[99] that has determined that SDM is needed to answer a challenging problem, sometimes referred to as the *Appreciation of the Situation*[100][105, 144, 325]. For the first validation objective, "Relevance to and Consistency with Appreciation of the Situation"[105] (p. 123), validation tests assess the extent to which the model is faithful to this real-world situation. The basis for understanding this situation comes in the earliest phase of SDM research, which can best be described as qualitative data collection. In other words: to be useful, a model should reflect this situation, as it is understood in stakeholders' mental models; the model needs to simulate the right behavior and for the right reasons[105].

For the second objective, *Suitability for Purpose*, validation tests focus on the boundary of the model relative to the boundary identified in the problem statement. In other words: having all the structures that are relevant to the problem, and not having additional ones; and when taking the model to extremes, having them produce behaviors that are within the limits of the situation[105].

Both of these objectives focus on structure, behavior and culture. Their focus on structure and behavior is evident in the descriptions above. The culture focus relates to: supporting debate among stakeholders and representing what is within the cognitive limits of stakeholders. More specifically, Lane's tests of cultural aspects ask the following questions:

> "Do the models support debate on different perspectives in the [Appreciation of the Situation?] …
>
> "Do the models support debate concerning, and represent the behavior of the relevant actors…
>
> "Are the feedback links in the models consistent with the abilities of current actors in the system to access, interpret and employ information?"[105] (p. 123).

[99] SDM is most widely used in management consulting and by academics in the field of management science.

[100] Oral and Kettani describe *managerial situation* as "the way that real world events are interpreted by decision makers and policy formulators" (p. 218). This situation, also referred to by Checkland as *Appreciation of the Situation*, is more than a problem to be solved, as it may also require assessment, prediction or analysis.

D.2 VALIDITY TYPES CONSIDERED IN SYSTEM DYNAMICS MODELING

This section defines the four types of validity[101][105, 144, 163, 166] considered in the SDM literature: conceptual, formulational, experimental and data. I describe each validity type broadly and then provide specific definitions for its *subtypes*[102].

D.2.1 CONCEPTUAL VALIDITY

Adapted from Lane[105], Conceptual Validity (CptV) in SDM entails the extent to which stakeholders agree on two aspects:

I. **Model aspects**
   a. Any materials[103][105] that would inform the conceptualizing of the simulation model
      i. draw on the mental models of the group to: express the problem statement and begin to make sense of the problem statement

II. **Cultural aspects**
   a. SDM is felt to be an *appropriate approach* (clients see their ideas *in a model*)
   b. The materials informing the conceptualization of the simulation model express
      i. the *acceptable and useful social and cultural aspects* of the situation, meaning: feelings of individuals, group's ideas, and group's values
      ii. the *goals* the group wishes to reach
      iii. the *policies* the group finds acceptable
      iv. the *role restrictions* that apply to group members and might become relevant if policy changes are suggested
   c. A simulator based on the simulation model helps stakeholders interacting with it to *learn*[104][329] (i.e., to change their mental models)

---

[101] Theoretical work on validity types in SDM is presented in Lane (1995). A 2016 review of validity in *modeling and simulation disciplines* performed by Tsopiapsis *et al.* identified additional types. However, these were also considered by Lane and Oral & Kettani who considered (based on the work of Randers) that these other types are combinations of the four *base types*.

[102] Lane uses subtypes to refer to different aspects of validity types. Being specific in this way helps to design methods.

[103] Lane refers to these things as the *Communicated Conceptual Model* which is assumed to include all of the information in the mental, written and numeric databases and everything generated from them in the problem statement, and in the pre-simulation phase of modeling, including CLDs, the reference mode, dynamic hypothesis, problem statement, model boundaries, and mental model elicitation objects.

[104] For example, does an online simulator based on the *Beer Game* help participants understand the pitfalls of the *fundamental attribution error* (its learning objective) or do they only learn 'this is hard'?

Lane's definition is useful but insufficient in understanding how new validation methods should test for conceptual validity. In particular, the *model aspects* portion does not offer enough guidance for system dynamics models. Therefore, I looked to validity concepts used in qualitative research, specifically Grounded Theory[105][116, 117, 138, 161] for help with the *model aspects* and to Lane's proposed tests for culture to better understand the *cultural aspects*. Specifically, I found the *three rules* for *theoretical saturation* used in Grounded Theory[106][326-328] to be helpful in considering the specific conceptual validity claims of system dynamics *models*. These three rules informed three of this dissertation's proposed subtypes[107][105] for conceptual validity in SDM.

A 2016 literature review of model validation in Operations Research, Modeling and Simulation, and Computer Science, Tsioptsias *et al.*[163] was useful in considering whether I had considered an adequate set of factors. With regard to conceptual validity, the authors reported agreement across these disciplines for using three[108] contributing factors that are roughly analogous to the factors addressed by the *three rules*.

The fourth proposed subtype carries forward the *cultural aspects* part of Lane's definition. To help the reader to better understand these aspects, I present here tests which Lane suggests address both conceptual validity *and* the cultural focus of SDM validation.

The first test, **Perspectives Boundary Adequacy,** concerns the avoidance of what Eden & Sims refer to as *coercion* – any approach that fails to address the diversity of perspectives in the target group[105, 330]. This can occur in a situation where researchers force: 1) the choice of modeling approach, 2) the choice of problem statement, 3) a set of goals to be achieved and/or 4) the policies for achieving them. This test was developed to promote the acquisition of the necessary skills for implementing SDM's *non-coercive approach***.**

The second test, **Norms/Values Boundary Adequacy,** concerns the development of models that recognize the social and cultural realities of the target group. SDM assumes that the goals and policies are formulated within the norms and values of the target group, and are therefore

---

[105] This interest in Grounded Theory came because several resources I used for my model development and theoretical reflection were informed by this field.
[106] As previously mentioned, in Grounded Theory, confidence in findings relies upon theoretical saturation – that is, (a) the category is well developed in terms of its properties and dimensions demonstrating variation, (b) the relationships among categories are well established and validated, (c) no new or relevant data seem to emerge regarding a category.
[107] This is following Lane's lead as he used subtypes to distinguish the different contributing factors for formulational and experimental validity.
[108] A fourth contributing factor was also mentioned: "each problem should be validated with a specific purpose in mind and compared with the real world" (p. 4). This factor relates to the objectives of validation, not conceptual validity alone.

culturally acceptable. This test aims to ensure that the behavior of agents in the model is not *against the rules*—that the goals and the policies are acceptable within the culture of the target group.

The third test, **Roles Boundary Adequacy**, concerns the assumption that agents in the model have bounded rationality[109]. In SDM, this is implemented by ensuring that the decisions represented in the model (decision functions) are based upon information (variables) that agents can actually access, interpret and use.

Thus, this exploration led me to propose four subtypes for conceptual validity: three for model aspects and one for the cultural aspects.

D.2.1.1 NEWLY-PROPOSED SUBTYPES FOR CONCEPTUAL VALIDITY IN THIS DISSERTATION

Conceptual Validity Subtype #1 (CptV1) concerns the extent to which the variables and boundaries in the model are well-developed and validated.

For variables to be well-developed, they should be "precise"[107] (p. 57). In all types of SDM models, variable definitions should be meaningful in practice[107] (p. 57-59) and/or theory. For variables to be precise in CLDs means that they conform to CLD grammar (e.g., variable names are nouns or noun phrases; see Section 2.2.2.1.2) [107] (p. 59). Precision in simulation models means that variables should be (1) defined considering their units[107] (p. 59), their time-varying status as constant, variable, stock or flow, and (2) mutually exclusive, internally consistent and available to decision makers[107] (p. 103)[113] (p. 516-518).

For model boundaries to be well-developed, they should be suitable to the study purpose. That is, the time frame and the system scope (i.e., all of the variables) should be suitable to the study purpose.

The validity of well-developed variables and boundaries is explored by comparing the model to the real world in some way (e.g., by eliciting additional stakeholder mental models) and addressing any discrepancies (the same applies for the relationships[109] considered in CptV2). In so doing, the model's *accuracy* is explored, not with the goal of labeling the model accurate or inaccurate, but with the goal of raising (and finding satisfactory ways of answering)

[109] In comparing relationships to the real world, it is possible for the evidence to refer to a specific relationship. Other times, it may refer to a sequence of variables and links (i.e., a causal chain). What the data contain depends on the data source.

questions regarding choices that had been made in defining the variables and setting the system boundaries (i.e., via either revising the model or explaining why the difference exists).

Conceptual Validity Subtype #2 (CptV2) concerns the extent to which the relationships (i.e., links) in the model are well-established and validated.

For relationships to be well-established, they should be *precise*. For relationships in CLDs to be precise means that they are assigned their necessary attributes as defined in CLD grammar (see Section 2.2.2.1.2). These attributes include the direction and polarity of causality as well as the existence or not of a time delay. Precision in simulation models means that each relationship is implemented in an equation according to the attributes considered in the mental models.

Conceptual Validity Subtype #3 (CptV3) concerns the extent to which new and relevant data regarding the elements in the model have ceased to emerge.

The emergence of new data regarding the elements (i.e., variables, links, delays, feedback loops) would necessitate reformulation. Therefore, in order to have confidence in a model's formulation, it is necessary to consider the extent to which new reformulations are needed. When the collection and/or analysis of data has ceased to require reformulation of the elements in the model, *saturation* of new data has been reached.

The emergence of relevant data regarding the elements serve to further clarify the precision of each element. For example, relevant data emerging for a variable might result in a change to the variable's name, where the relationships would stay the same. Another example may be that the variable name stays the same, but that what it is conceptualized to mean has evolved. When the emergence of relevant data regarding the elements serve only to provide further evidence for the conceptualizations already in place, *saturation* of relevant data has been reached.

Conceptual Validity Subtype #4 (CptV4) concerns the extent to which the relevant actors feel that the cultural aspects of the research are addressed (i.e., cultural acceptability, bounded rationality, and non-coercive approach).

Conceptual and cultural aspects of models are considered; namely, the cultural acceptability of the goals and policies in, and the bounded rationality of decisions (i.e., decision functions only use information that real decision-makers would have available, in the time they would have it[113] (p. 516-520); and of this information, that they consider some of it to be the relevant *signal* and ignore the rest[114] (p. 219)). The non-coercive approach of the SDM research endeavor is also addressed.

D.2.2 FORMULATIONAL VALIDITY

According to Lane[105], Formulational Validity (FV) entails the extent to which the simulation model is consistent with the conceptual model. The subtypes presented below are based on Lane[105] who draws from Balci[331]. I extend two of these to consider qualitative models.

D.2.2.1 SUBTYPES FOR FORMULATIONAL VALIDITY – WITH PROPOSED ADDITIONAL ASPECTS OF RELEVANCE TO THE MODEL IN THIS DISSERTATION

Formulational Validity Subtype #1 (FV1) concerns the extent to which the limitations of language impacted the development of the simulation model. In other words: have concepts in the conceptual model been omitted or distorted in this translational process?

This dissertation extends this type of validity to include the limitations of language impacting the development of the conceptual model from the descriptive written and/or mental data.

Formulational Validity Subtype #2 (FV2) concerns the extent to which the simulation model is shown to be consistent with the structure and behavior described in the conceptual model and problem statement. In other words, is there a satisfactory representativeness between the simulation and the conceptual understanding?

This dissertation does not extend this type of validity, it is used as in Lane[105].

Formulational Validity Subtype #3 (FV3) concerns the extent to which the simulation model conforms to SDM construction guidelines (e.g., not having artificial min/max functions to fix a bug in the model).

This dissertation extends this type of validity to include the extent to which the conceptual model (in this dissertation's case a CLD) conforms to SDM construction guidelines (e.g., each variable name should be a noun or noun phrase and they must have a sense of direction in terms of quantity or degree[113] (p.152)).

D.2.3 EXPERIMENTAL VALIDITY

According to Lane, Experimental Validity (EV) entails having a simulation model that "generate[s] useful insights … [where] those insights are rigorously supported by runs with the model and have been demonstrated to be robust by sensitivity analysis"[105] (p. 119). This dissertation uses these subtypes unaltered.

D.2.3.1 SUBTYPES FOR EXPERIMENTAL VALIDITY

Experimental Validity Subtype #1 (EV1) concerns the design and carrying out of simulation runs to test the structural design of the model. When results of simulation runs challenge the current structure, alternative structures are considered (i.e., the trial and error of equation writing referred to by Peterson[111]).

Experimental Validity Subtype #2 (EV2) concerns the usefulness, accuracy, and robustness of the insight gained from simulation model runs. For small policy models, the requirement is having qualitatively similar behavior (between simulation model and problem statement). For detail-calibrated models, a greater level of precision is required.

D.2.4 DATA VALIDITY

According to Lane[105], Data Validity (DV) entails the extent to which the mental, written and numerical data[332] required for research in SDM (in all the model development and validation processes) is reliable/appropriate, accessible, and sufficient. When describing which validity types are attached to each test (refer to Table 2.9), Lane marks *Parameter Confirmation* and *Reference Mode Reproduction* as relevant to DV and states that all other tests are assumed to have relevance to DV.

This dissertation *explicitly* segregates DV subtypes and identifies *Structure Confirmation – Empirical* as a third test that is relevant to DV, to DV1 and DV2.

D.2.4.1 NEWLY-PROPOSED SUBTYPES FOR DATA VALIDITY IN THIS DISSERTATION

Data Validity Subtype #1 (DV1) concerns the extent to which the mental data required for model development is reliable/appropriate, accessible and sufficient. Below are questions specific to the first two aspects; for *sufficient*, refer to the definition of CptV3 (saturation) above.

- Reliable/appropriate
    - Were the right kind of questions asked?
    - Is there grand-standing? (In other words, does the person stand to gain anything by sharing this information? If so, they may be grand-standing. For example, consider the difference between a press conference and confidential conversation.)
    - Is it purposive text[117]?
- Accessible
    - Do the *participants* have first-hand experience, over enough time, with a range of potential dynamic behaviors and in an aspect of the system inside the model boundaries to speak about system structure and behavior?

Data Validity Subtype #2 (DV2) concerns the extent to which the written data required for model development is reliable/appropriate, accessible, sufficient. Below are questions specific to the first two aspects; for *sufficient*, refer to the definition of CptV3 (saturation) above.

- Reliable/appropriate
    - Did descriptions consider the right kind of things?
    - Is the content aspirational rather than descriptive?
    - Is it purposive text[117]?
- Accessible
    - Do the *authors of the written data* have first-hand experience, over enough time, with a range of potential dynamic behaviors and in an aspect of the system inside the model boundaries to speak about system structure and behavior?
    - If not, did their *sources* have this kind of experience?

Data Validity Subtype #3 (DV3) concerns the extent to which the numeric data required for research in SDM is reliable/appropriate, accessible, and sufficient. Below are questions specific to these three aspects.

- Reliable/appropriate
    - Have the right kind of things been measured, in the right time intervals?
    - Have the measurements been validated?
    - Are there uncertainty bounds?
- Accessible
    - Do *researchers* have access to the raw and/or appropriately-processed data?
- Sufficient
    - Of the variables in the system boundaries, how many have been measured?

APPENDIX E: NEWLY-DEVELOPED VALIDATION METHODS

This section presents the seven newly-developed validation methods in detail. These tests are not pass/fail per se. Rather, they identify strengths and weaknesses of the model. Specifically, they identify areas for potential improvement and build confidence in the model's validity. These tests are performed by asking the questions relevant to each test and demonstrating how the model addresses them. It is then up to the audience to judge.

E.1 SHARED MENTAL MODEL SATURATION

Model name	Participant CLDs	Team CLDs, Clinic CLDs	SMM1	SMM2	SMM3 / Conceptual Model	SIM1	SIM2	SIM3	Theoretical Model

Shared Mental Model Saturation (SMM-S) has been attained when it has been adequately shown that (1) the system structure in the shared mental model adequately describes the system underlying the problem statement and (2) the addition of an additional interview or batch of interviews (from the saturation reserve) appears unlikely to modify the shared mental model.

The determination of SMM-S is made by answering two questions, referred to as the SMM-S Test:

- Do the variables identified meet the model's purpose?
- Has conceptual saturation been reached?

SDM attempts to capture the causal structure of complex phenomena, which are understood best by the stakeholders who directly experience these phenomena. This is because these stakeholders' experiences have impacted their mental models in long-term memory. The term *Shared Mental Model* is reserved for the aggregated understanding obtained from combining CLDs representing individual mental models, assuming that the resulting cognitive structure could be obtained from any of the individuals when specifically/appropriately probed.

The sequence of steps in SMM-S is designed to enable the researcher, firstly, to see if their first draft shared mental model (SMM1, the output of CLD Combination) has the potential to tell a story about their research which suits their original purpose well; secondly, to examine to what extent does the model convey the complete story participants told; and thirdly, to evaluate this story for its ability to bring together the disparate understandings of participants into one whole without distortion or implausible formulations.

E.1.1 COMPARISON OF SHARED MENTAL MODEL AND PROBLEM STATEMENT

The modeler first checks if things mentioned in the problem statement have been omitted from the shared mental model. This is a qualitative activity. CLD Combination already funneled shared mental model variables to be specific to problem statement. The problem statement is read and the shared mental model (SMM1) – which is the result of CLD Combination – is reviewed.

During this review, the modeler reads SMM1, develops written definitions and descriptions of its elements (i.e., its variables, links and feedback loops, including the possible consequences of time delays and of link and loop polarities) and uses these descriptions to tell stories using SMM1 (i.e., of scenarios involving possible shifts in loop dominance, a sudden change in one variable percolating through the system, etc.), and checks if these stories are comprehensive of the dynamics expressed in the problem statement.

A discrepancy between the stories told using SMM1 and the problem statement could be found when SMM1 fails to account for important aspects of the problem statement (e.g., in a hypothetical study with a problem statement about malaria control, the SMM1 fails to include any variables relating to the use of insecticide treated bed-nets). If a discrepancy is uncovered, additional analysis is required: either re-analysis of model development set or drawing on the saturation reserve. Alternatively, perhaps a re-consideration of the problem statement is needed.

E.1.2 CREATION OF SHARED MENTAL MODEL SATURATION CURVES & DIAGRAMS

Once no more discrepancies exist, the modeler proceeds to the next portion of the SMM-S Test by creating and using the visualizations described below.

This method compares all of the CLDs at the penultimate level of CLD Combination (in this case, the clinic level CLDs) with the shared mental model produced by their combination (SMM1). This method consists of two procedures: (1) clinic CLD pair-wise comparisons (producing SMM-S Curves) and (2) clinic CLDs to SMM1 comparisons (producing SMM-S Diagrams).

E.1.2.1 CLINIC CAUSAL LOOP DIAGRAM PAIR-WISE COMPARISONS

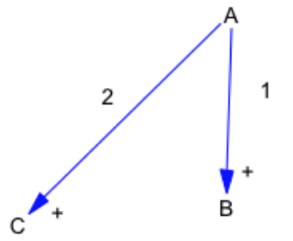
Clinic CLD pair-wise comparisons are used to generate SMM-S Curves. SMM-S Curves are adapted from the *accumulation curve* used in the Fuzzy Cognitive Maps literature[174, 333]. SMM-S Curves are used to assess whether eliciting one more participant's mental model will cease to provide new elements; in other words, that additional data collection is likely to yield no new elements, and/or that additional data analysis from a saturation reserve is likely to yield no new validation findings. This is analogous to the concept of theoretical saturation in qualitative theory development methods such as Grounded Theory and Case Study Research[138]. The difference here is that instead of looking to see whether existing thematic categories are capable of absorbing what additional participants say (i.e., that an additional participant does not add new categories), one looks to see whether the existing shared mental model (SMM1) is able to absorb the elements that additional participants identify (i.e., variables, links, delays and loops).

Since in system dynamics, CLDs have variables, links, delays and loops, a separate set of SMM-S Curves is generated to observe whether saturation has been reached for each of these elements. Each set of SMM-S Curves contains two lines: one showing the inflow of new elements and the second line showing the accumulation of elements with the addition of each clinic CLD (i.e., cumulative count). The accumulation represents the level of conceptual saturation – where saturation is indicated by a flat line (i.e., zero slope).

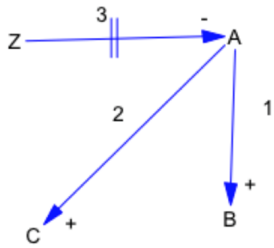
The data for these curves were generated by comparing clinic CLDs. First, clinic CLDs are visually inspected, and the one that appears to be most detailed is selected as the anchor (in this research, Clinic E was selected). Second, counts are generated for that clinic, specifically: number of variables, links, delays and loops. Third, the clinic CLDs of the remaining clinics are compared to that of Clinic E, in a pair-wise fashion using Vensim software[177]. Each time a comparison is made, both total and new element counts (new elements identified by the additional clinic CLD) are marked and counted for: variables, links, delays and loops.

For example, in Figure E.1 below, Clinic E and Clinic F counts are generated. Then, Clinic E is compared to Clinic, F and counts of new elements are generated. It is helpful to use color to label comparison results. In the comparison result CLD on the right of this figure, the variables in common are colored orange. The links in common are colored green. The new variable is bolded and the new link is left blue. There are no new loops. The counts are then generated from this comparison figure: Clinic F adds one variable, one link, and one delay (all related to link 3).

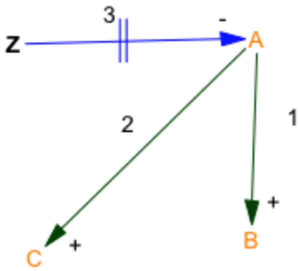
Figure E.1 Example of Generating the Data for SMM-S Variables Curves



Clinic E CLD --- the anchor



Clinic F CLD --- the comparison clinic



Clinic F CLD Comparison Result

What is Counted	**Total Number of elements found in:**		**Count of New Elements Found in Clinic F**
	Clinic E	**Clinic F**	
Variables	3	4	1
Links	2	3	1
Delays	0	1	1
Loops	0	0	0

Figure Key:

Blue = link not mentioned in the anchor clinic
Green = link mentioned in the anchor clinic
Orange = variable mentioned in the anchor clinic
Black & Bold = variable not mentioned in anchor clinic

Then the combination of Clinic E and Clinic F is compared to another clinic and again, counts are generated; and so on until all of the clinics have been compared in this fashion (these steps are not shown in the figure above as the visual process would be similar).

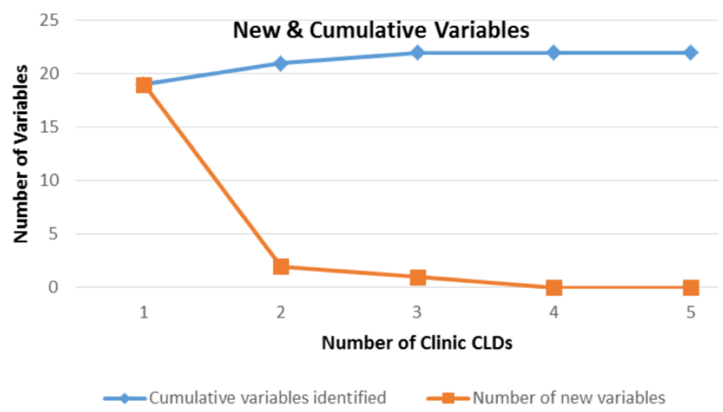
From these counts, the variables, links, delays, and feedback loops SMM-S Curves are generated. The x-axis is the number of clinic mental models. The y-axis is the number of elements for each element: (1) clinic-specific total element counts are sorted in ascending order and graphed (blue line) and (2) clinic-specific new element counts are sorted in ascending order and graphed (orange line). Figure E.2 below presents an example table of variable counts and the resulting set of SMM-S Curves for variables[110].

Figure E.2 Example of Generating the Variables SMM-S Curves

Table Counting Variables for an Example SMM-S Analysis

CLDs	Number of new variables	Cumulative variables identified
1	19	19
2	2	21
3	1	22
4	0	22
5	0	22

Variables SMM-S Curves



In this example, the cumulative variables SSM-S Curve flattens out (i.e., zero slope). This indicates that variable saturation has been reached. If the curve had not flattened out, additional data from the saturation reserve would need to be analyzed to prepare a new shared mental model.

E.1.2.2 CLINIC CAUSAL LOOP DIAGRAMS TO SHARED MENTAL MODEL COMPARISONS

Comparison of clinic CLDs to SMM1 is used to generate SMM-S Diagrams**.** SMM-S Diagrams visualize saturation by aggregating the clinic-level information into a single CLD designed to highlight the relative level of conceptual saturation for each link. Two versions of the SMM-S Diagram are generated: 1) visualizing the percentage of clinics identifying each relationship and 2) the percentage of clinics explicitly identifying each relationship (where the denominator is the number of clinics that mentioned each relationship) (see Figure E.3 below). Each figure uses darker shades of color and thicker lines to indicate a higher percentage of clinic mentions of (i.e., level of saturation in) these relationships.

The data for these diagrams are generated by comparing SMM1 to each clinic CLD using Vensim software[177]. First, SMM1 is duplicated, such that there is one copy for use with

[110] The example presented in Figure E.1 is for a different dataset than that presented in Figure E.2.

each clinic CLD. This time, when a link identified in SMM1 is also identified in the clinic CLD, that link is numbered and the corresponding number is placed on that arrow in the clinic CLD as well as on that arrow (or set of arrows) in that clinic's copy of SMM1.

Four rules are provided below:

1. When the clinic CLD identified a link (e.g., A->C) that was represented in more than one step and less than four steps (e.g., A->B->C) in SMM1, that link was dashed (since its causal mechanism was not explicitly identified).
2. When the clinic CLD identified a link (e.g., A->Z) that was represented as more than four steps on SMM1 (e.g., A->…->Z), that link was not marked as present on the clinic's corresponding CLD. This is because *greater than four links* was deemed as not identifying the causal mechanism, but possibly indicating a hypothesized correlation.
3. When a clinic CLD contains a relationship where one of the variables had been merged with other variables during the process of creating SMM1, the corresponding relationship was marked as present. See for example, in Figure 2.9, relationships F->C, G->C, H->C, J->C, and K->C would all be credited as the clinic CLD identifying B->C on SMM1.
4. When a clinic CLD contains a relationship where both variables had been merged with other variables during the process of creating SMM1, the corresponding relationship is inside one variable in the SMM1. Thus, it cannot be marked.

At the end of this process, all clinic CLD relationships will have been accounted for in SMM1.

Variables, links, delays, and loops were labeled and counted in each clinic's copy of SMM1. For each element, MS Excel was used to make a count of the number of markings in each clinic CLD.

Figure E.3 below presents a simplified example of how an SMM-S Diagram is generated. It is described in detail as follows.

The first column houses the SMM1 produced previously via CLD Combination. The second column houses two of the clinic CLDs[111]. In each clinic CLD, links are numbered from 1 to n, where n is the total number of links in that CLD. Each clinic CLD (second column) is compared to SMM1 (first column) and a copy of the SMM1 is made for each clinic (third column, in green), with links solid or dashed according to how they are mentioned in the two clinic CLDs.

For example, for Clinic E, link A->B in SMM1 is mentioned twice: once via a direct link (A->B) and a second time via an indirect link (A->C). Since both links are mentioned by this clinic, both are colored green (third column); however, A->B would have a solid arrow as it is mentioned directly and B->C would have a dashed arrow as this link is mentioned only indirectly(via the A->C).

Similarly, for Clinic B, since links A->B and B->C are mentioned (via the A->C link the clinic CLD), and they are both colored green and are both made dashed arrows as they are both mentioned only indirectly. Again, the corresponding link number is also placed on the *Clinic B Links on Shared Mental Model* diagram. If one or more SMM1 links remain unmentioned in a clinic CLD (second column), they would remain blue in the clinic's corresponding shared mental model (third column).

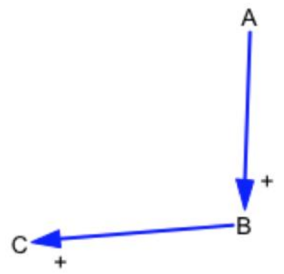
The fourth column presents the SMM-S Diagram for number of clinic mentions. A table is generated to help prepare this diagram. The table lists each link and percent of times that each link was mentioned. Since all clinics mentioned both links (either directly or indirectly) the percentage for each link is 100%. If there had been some variation, for example if there were three clinic CLDs and one link was mentioned by all and another link was mentioned by only two of the clinic CLDs, then the first link would have 100% and the second link would have 66%. This difference in gradation would be represented by a thicker line for the first link and a thinner line for the second link.

The fifth column presents the SMM-S Diagram for *explicit* clinic mentions. Again, a table is generated to help prepare this diagram. The table lists each link and percent of times that each link was mentioned *explicitly*. Since all clinics mentioned A->B

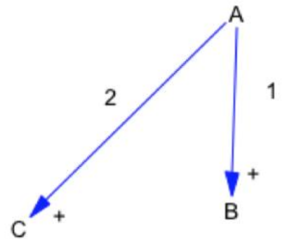
[111] The example is simplified in that to have a completed SMM-S Diagram, the SMM-S Diagram would show the result of comparing all of the clinic CLDs to the shared mental model. This example only shows to of those clinic CLDs, as their information would be combined to form the SMM-S Diagram.

explicitly (i.e., directly), the percentage for A->B is 100%. Since none of these two clinics mentioned B->C explicitly (although it was mentioned), the percentage for B->C is 0%. On the SMM-S Diagram (fourth column), A->B is marked by a solid arrow (representing explicit mentions) and B->C is marked by a dashed arrow (representing only indirect mentions). If a link is expressed explicitly by at least one Clinic CLD, then it is represented by a solid arrow. Again, when you have more than two clinic CLDs being compared to SMM1, there is the opportunity for gradation of explicit mentions – in this case, the higher the number of explicit mentions across clinic CLDs, the thicker the link. The thicker lines thus indicate a higher level of agreement among clinic CLDs that this link does exist. This pattern is not followed for implicit mentions, since they are implicit.

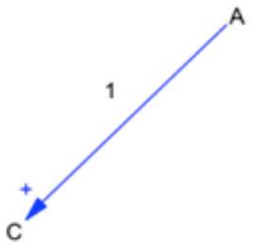
Figure E.3 Example of Generating SMM-S Diagrams



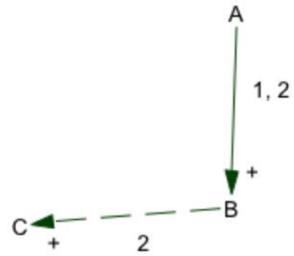
Shared Mental Model (SMM1)



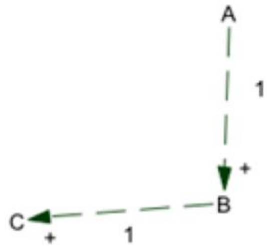
Clinic E CLD



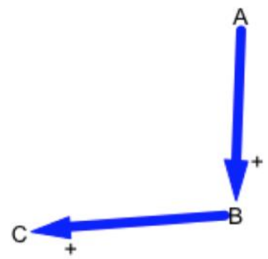
Clinic B CLD



Clinic E Links on Shared Mental Model

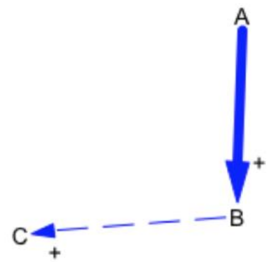


Clinic B Links on Shared Mental Model



Final SMM-S Diagram Clinic Mentions

Links	Mentions
A->B	100%
B->C	100%



Final SMM-S Diagram Explicit Clinic Mentions

Links	Mentions
A->B	100%
B->C	0%

Figure Key:

Solid line = explicit mention
Dashed line = implicit mention

Thicker lines = more mentions

Thicker lines = more mentions
Solid line = explicit mention
Dashed line = implicit mention

E.1.2.3 THE SHARED MENTAL MODEL SATURATION TEST

The SMM-S Test involves answering the following questions using the procedures (in bold) below. The sections above present the sequence for carrying out the procedures. Box E.1 presents the procedures as they relate to specific validity types and questions specific to the SMM-S Test.

Box E.1 Shared Mental Model Saturation Test (SMM-S Test)

Do the variables identified meet the model's purpose?

- Are the variables and boundaries of the Shared Mental Model well-developed and validated? (CptV1 - variables & boundaries)
    - **Check if portions of the problem statement (e.g., important variables) are being omitted from SMM1**
    - **Check the Variable SMM-S Curve for saturation**

Has conceptual saturation been reached?

- Are the relationships in the Shared Mental Model well-established and validated? (CptV2 - links)
    - **The process of developing the SMM-S Diagrams checks if portions of SMM1 require reformulation**
    - **Check SMM-S Diagram for number of mentions -** consider how well-established each relationship is, for the less well-established relationships (thin lines) consider
        - Does the relationship in SMM1 distort aspects of the clinic CLDs? (FV1 - language limitations)
        - Are bounded rationality and cultural acceptability evident in the relationships? (CptV4 - culture)
    - **Check SMM-S Diagram for number of explicit mentions -** consider how well-established each relationship is, for the less well-established relationships (dashed lines) consider
        - Does the relationship in SMM1 distort aspects of the clinic CLDs? (FV1 - language limitations)
        - Are bounded rationality and cultural acceptability evident in the relationships? (CptV4 - culture)
- Have new and relevant data regarding the elements in SMM1 ceased to emerge? (CptV3 - saturation)
    - **Check Link, Delay and Feedback Loop SMM-S Curves for saturation**
    - **The process of developing the SMM-S Diagrams checks if portions of SMM1 require renaming and/or reformulation**

The researcher then reflects on the adequacy of the evidence provided for SMM-S. If the results are inadequate, then additional interviews from the saturation reserve need to be consulted in revising the shared mental model and then the SMM-S test needs to be run again.

E.2 CONCEPTUAL MODEL SATURATION

Model name	Participant CLDs	Team CLDs, Clinic CLDs	SMM1	SMM2	SMM3 / Conceptual Model	SIM1	SIM2	SIM3	Theoretical Model

Conceptual Model Saturation (CM-S) has been reached when it has been demonstrated that the addition of one interview is not likely to modify the model. I start with the existing shared mental model (that was just verified via SMM-S; i.e., SMM2). That model is revised (SMM2.x, where x denotes version number[112]) as needed during CM-S, until the addition of one more interview is not likely to modify it. At this point, once the CM-S Test is passed, the CLD is referred to as the *Conceptual Model* (i.e., SMM3).

The determination of CM-S is made by answering the following questions, referred to as the CM-S Test:

- Do identified elements exist in SMM2 and are the identified elements formulated correctly? (And if not correct, then revise/remove incorrect ones; creating SMM2.x)
- Have I reached saturation in existing elements?

When the researcher is able to use a variable or relationship in SMM2 to interpret raw interview transcript data from the *model validation set*, then I can have more confidence that SMM2 has exposed the variables and relationships of interest to the problem statement. This analysis of mental data occurs in the coding of interview transcripts and generates large amounts of data supporting the precision and accuracy of the variables and relationships in SMM2. The presentation of *Rigorously-Interpreted Quotations- for Causality* (Causal RIQs) provides evidence of this for the target audience.

When more participants describe an element in SMM2 in the same way, then I have more confidence in that element as being part of the *true system*. Fewer participants describing an element could mean an element is not part of the true system, or that it is only visible to a limited number of participants. This does not invalidate the element. Elements can be invalidated if participants describe them differently than was conceived in SMM2, using challenge statements. Challenge statements result in immediate changes to the Conceptual Model, starting this process over again. Support statements are recorded in terms of the variables and relationships. When a relationship is mentioned in multiple causal statements by multiple participants representing multiple stakeholder groups, it is considered well-established and validated.

112 As this section deals with how revisions are made, I identify those revised versions of the model with one name SMM2.x. For simplicity, these are referred to as SMM2 in the body of the dissertation.

Model development has thus far has focused on those implementing and thus most closely experiencing PCT (i.e., the clinician and MA interviews). It was assumed that this would be enough for capturing the dynamics in the problem statement. In this step, this assumption is tested by also including one layer further removed from implementation of PCT (i.e., interviews from clinic management). Surfacing and checking this assumption allows the researcher to ask: "if managers' viewpoints had been considered in model development, would that have influenced the shared mental model's structure?"

The CM-S methods (described below) generate the following three visualizations:

- CM-S Diagrams
- CM-S Curves

While the CM-S Diagrams identify elements validated in the validation set interviews, the CM-S Curves visualize whether the conditions for assuming saturation have been met for those validated elements (via accumulation flattening). In other words, I have heard about everything and I have heard about it sufficiently. In so doing, CM-S Diagrams and CM-S Curves provide a visual answer to the two CM-S Test questions above.

The quotations provide rigorously interpreted textual documentation for the elements found in the conceptual model. The selection of quotations is informed by the data analysis that supports the CM-S Diagrams and CM-S Curves.

The sequence of steps in CM-S is designed to enable the researcher first, to check if the story that their shared mental model (SMM2, the output of SMM-S) tells checks out when compared with stories told by participants whose interviews were not used in developing it; second, to revise the model as needed; third to examine to what extent does the newly-revised model convey the complete story these participants told; fourth, to evaluate this new model for its ability to bring together the disparate understandings of participants into one whole without distortion or implausible formulations; and fifth, to show how the model represents the stories that participants tell.

E.2.1 INTERVIEW CODING FOR CONCEPTUAL MODEL SATURATION

All raw interview transcripts from the model validation set (i.e., clinician, MA, and clinic management) are coded for use in the CM-S, SIM-S, and SD-S validation methods. The transcripts are coded in MS Word using the comment function (rather than using a qualitative analysis software).

Statements related to the problem statement are coded into one of two types:

- **Causal statements** - those identifying cause and effect between at least two variables
- **Conceptual statements** - those providing detailed definitions of meanings attached to concepts or assumptions (e.g., system boundary, time step, time horizon)

When these statements include **system dynamics-related items** they are marked using an additional code. For causal statements, they include descriptions of time delays, feedback loops or nonlinearities; for conceptual statements, they show awareness of mental models, that they matter to the problem or how emotional engagement will be required to make changes to the system or mental models.

The rest of the statements were considered to include information that is **extraneous** because it was not relevant to the problem statement.

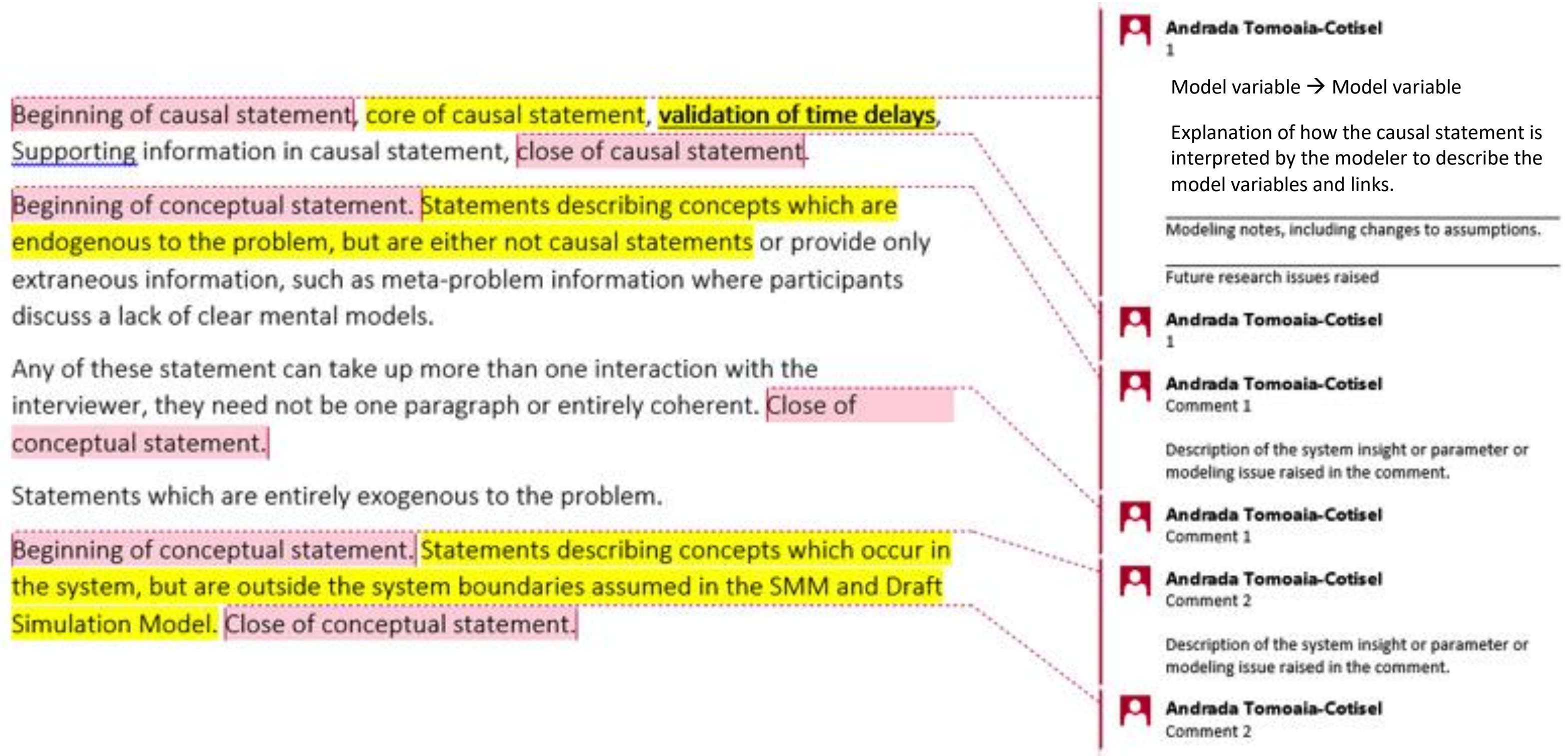
The coding is applied as in Box E.2 below.

Box E.2 Coding Transcripts for the Conceptual Model Saturation Test

- For causal statements
  - Each causal statement is numbered.
  - The opening of the statement is marked by a comment containing:
    - The statement number (e.g., "1")
    - Any causal chains expressed in that statement (using elements as expressed in SMM2 along with any new elements mentioned in the statement)
    - The researcher's interpretation of the statement (explanation of how the causal statement is interpreted by the modeler to describe the model concepts)
      - Mark as possible challenge or supporting statement but do not address challenge statements until section E.2.2. "Creation of CM-S Diagrams"
    - Modeling notes, including changes to assumptions
  - The closing of the statement is marked by a comment containing:
    - The statement number
- For conceptual statements
  - Each conceptual statement is numbered
  - The opening of the statement is marked by a comment containing:
    - The statement number (e.g., "Comment 1")
    - The researcher's interpretation of the concept
  - The closing of the statement is marked by a comment containing:
    - The statement number
- For system dynamics-related items
  - The text is marked using bold text and underlining

A generic example is provided in Figure E.4 below.

Figure E.4 Example Coding for Saturation Test



E.2.2 CREATION OF CONCEPTUAL MODEL SATURATION DIAGRAMS

I start with the existing shared mental model (that was just verified via SMM-S; i.e., SMM2). The data for the CM-S Diagrams are generated by comparing SMM2 to each coded interview from the model validation set. For each causal statement coded in the interview, the associated links in SMM2 are labeled with the statement number. These labels catalog the confirming and disconfirming causal statements. This is done on copies of SMM2 - one per interview. Guidelines are provided below; for each causal statement in the model validation interviews:

1. When the causal statement identified a direct link in one step (e.g., A->C) that was represented as two to four steps (e.g., A->B->C) in SMM2, the corresponding links on SMM2 were labeled using the corresponding interview statement number.
2. When the causal statement identified a direct link in one step (e.g., A->Z) that was represented as more than four steps on SMM2 (e.g., A->…->Z), that link was not marked as present on the shared mental model. This is because *greater than four links* was deemed as not identifying the causal mechanism, but possibly indicating a hypothesized correlation.
3. When the causal statement identified a possible reformulation, comments were added into SMM2 via a dashed purple line linking each comment to the variable to which it belonged.
4. When a reformulation is confirmed (i.e., by another causal statement by the same participant and/or by another participant) the new formulation was drawn into the diagram (in maroon) and the comments were removed. This formulation was used as the gold standard for all cataloging going forward (i.e., the modified shared mental model, e.g., SMM2.x, is used for subsequent cataloguing).
5. When a new formulation of the shared mental model is confirmed, the relevant causal statements in previously-catalogued validation set interviews are then reviewed and the previously-coded analysis is revised considering the new formulation.

Then, these catalogued shared mental models are combined into CM-S Diagrams by transferring the labels link by link from the catalogued shared mental models, one for each interview, to a *conceptual spreadsheet* – described in more detail below.

CM-S Diagrams show the relative levels of identification of links among participants in the model validation set. They visualize saturation by aggregating the individual-level information into a single CLD designed to highlight the relative level of conceptual saturation (among model validation interviews) for each link.

Two versions of the CM-S Diagram are generated: 1) visualizing the percentage of participants who identify each link and 2) visualizing the number and type of professions identifying a particular link. Each figure uses thicker lines to indicate a higher level of saturation for each by link.

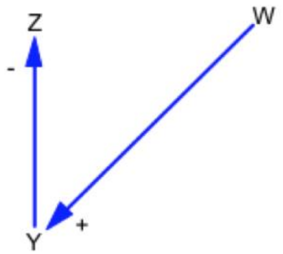
Figure E.5 below presents an example of how CM-S Diagrams are generated.

> The first two columns identify the inputs for the analysis generating CM-S Diagrams. The first column contains SMM2 (i.e., the shared mental model produced previously via SMM-S). The second column presents causal statements from two of the model validation set interviews (as previously described in Section E.2.1above). The top row in the second column summarizes the information from one interview (MAO3). The bottom row in this column summarizes the information from a second interview (CL03). Both participant introduced a new variable X which is not in SMM2. Each row also contains the linkages between the variables as inferred from each participant.
>
> The third column contains the shared mental model (modified as needed during this step, i.e., SMM2.x), with links labeled according to how they are mentioned in the two causal statements. For example, CS03's statement #1 mentions a more detailed causal chain (W->X->Y) for what the shared mental model previously represented as (W->Y). CL03's statement #2 also mentions this more detailed causal statement (W->X->Y). Both are colored marron since they confirm the need for reformulation of that link. The original shared mental model is reformulated to include the more detailed causal chain. CS03 also validates an additional piece to this causal chain (Y->Z) while CL03 omits this piece.
>
> The fourth column presents the CM-S Diagram that visualizes the percentage of participants who identify each link. A table is generated to help prepare this diagram. The table lists each link and percent of participants that mentioned that link. Since both participants mentioned links W->X and X->Y, the percentage for those two links is 100%. Since only one of the two participants mentioned links Y->Z, the percentage for that link is 50%. This difference in gradation is represented by a thicker line for the first two links and a thinner line for the third link.
>
> The thicker arrows thus indicate a higher level of agreement among participants regarding the existence of those specific arrows on the shared mental model. The thinner arrows show potential blind spots.

The fifth column presents the CM-S Diagram that visualizes the number and type of professions identifying a particular link. Again, a table is generated to help prepare this diagram. The table lists each link and the number and type of professions identifying that link. Since both the clinician and the MA identified W->X and X->Y, both professions are marked as having recognized that link. Since only the MA identified Y->Z, only MAs are marked as having recognized this link. On the CM-S Diagram, W->X and X->Y are colored brown indicating that both professions identified them and Y->Z is colored green indicating that only MAs identified it. When you have more than two professions, you have more possible combinations of professions identifying links (a different color is used for each combination). In addition, when you have more types of professions identifying, the arrow is made thicker (to indicate the number of professions identifying a link).

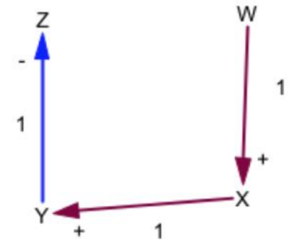
The thicker arrows thus indicate a higher level of agreement among professions regarding the existence of those specific arrows on the shared mental model. The thinner arrows (and different colors) show potential blind spots for particular professions.

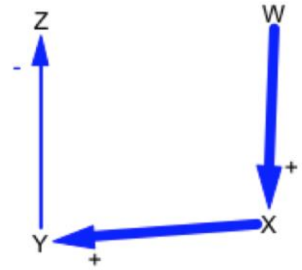
Figure E.5 Example of Generating CM-S Diagrams

<table>
<tr>
<td rowspan="2">

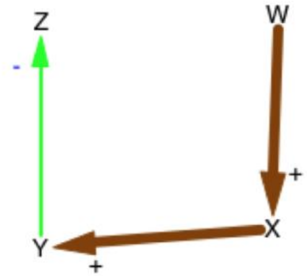
SMM-S Verified Shared Mental Model (SMM2)</td>
<td>Text = W Text = X
Text = Y Text = Z

W →+ X →+ Y →- Z

Interview MA03, quote #1 (Validation Set)</td>
<td>

Quote CS03-1 Links on SMM2.x (the modified shared mental model)</td>
<td rowspan="2">

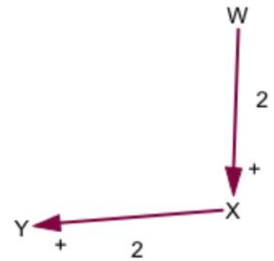
CM-S Diagram - Percent of Participants Identifying

<table>
<tr><th>Links</th><th>Mentions</th></tr>
<tr><td>W->X</td><td>100%</td></tr>
<tr><td>X ->Y</td><td>100%</td></tr>
<tr><td>Y ->Z</td><td>50%</td></tr>
</table></td>
<td rowspan="2">

CM-S Diagram - Professions Identifying

<table>
<tr><th>Links</th><th>Professions</th></tr>
<tr><td>W->X</td><td>(2) Clinician & MA</td></tr>
<tr><td>X ->Y</td><td>(2) Clinician & MA</td></tr>
<tr><td>Y ->Z</td><td>(1) MA</td></tr>
</table></td>
</tr>
<tr>
<td>Text = W Text = Y

W →+ X →+ Y

Interview CL03, quote #2 (Validation Set)</td>
<td>

Quote CL03-2 Links on SMM2.x (the modified shared mental model)</td>
</tr>
<tr>
<td colspan="2">Figure Key:</td>
<td>Blue = SMM2 links
Purple = SMM2.x links</td>
<td>Thicker lines = more mentions</td>
<td>Each color represents a different combination of professions</td>
</tr>
</table>

E.2.3 CREATION OF CONCEPTUAL MODEL SATURATION CURVES

The CM-S Curves show the inflow of new validated elements (variables, links, delays, and feedback loops) and their accumulation, across participants; analogous to the SMM-S Curves (section E.1.2).

The CM-S Curves are constructed from data in the conceptual spreadsheet. The spreadsheet is the collection of all participant conceptual tables. Four conceptual tables are constructed for each participant (one for each element: variables, links, delays, and feedback loops). Table E.1 presents a conceptual table for links for one participant. It is generated as in Box E.3 below.

Box E.3 Steps for Constructing the Conceptual Model Saturation Curves

- Each causal relationship in the modified shared mental model (SMM2.x) is numbered.
- A table is then created.
  - Causal relationship numbers are listed in the first column (titled "SMM2.x Link ID" column).
  - Causal statement numbers are listed as headings for the subsequent columns (those under the "Causal statement ID" column).
- When the causal statement (column) validates the causal link (row), the corresponding cell is marked with an "x". For example, causal statement 2 validates SMM2.x links #1, 3, 5, and 22.
- The causal statements are then tallied to calculate the frequency of mentions for each causal relationship (in the "Frequency" column).

There are 30 links (each with an ID, labeled in ascending order from 1 to 30). Eight of the links (1, 2, 3, 5, 6, 7, 14, and 22) are validated in the seven causal statements belonging to this participant. Whereas statement #3 validates only link #3, statement #5 validates links #5, 6, and 7. There are 17 total instances of individual links being validated by these seven statements.

Table E.1 Developing Statistics for the Links CM-S Curves - example table for one Participant

SMM2.x Link ID	Causal statement ID							Frequency	Mention Y/N
	1	**2**	**3**	**4**	**5**	**6**	**7**		
1		x						1	Y
2				x				1	Y
3		x	x					2	Y
4								Not mentioned	
5	x	x		x	x	x		5	Y
6	x				x	x		3	Y
7	x				x			2	Y
…								Not mentioned	
14						x		1	Y
…								Not mentioned	
22		x					x	2	Y
…								Not mentioned	
30								Not mentioned	
SUM								**17**	**8**

The CM-S Curves operate across the participants, one CM-S curve for each element-specific conceptual tables for each participant. CM-S Curves show the saturation of CM-S validated elements. They are very similar to the SMM-S Curves; however, this time, the unit for the x-axis is *individual participants* instead of clinics.

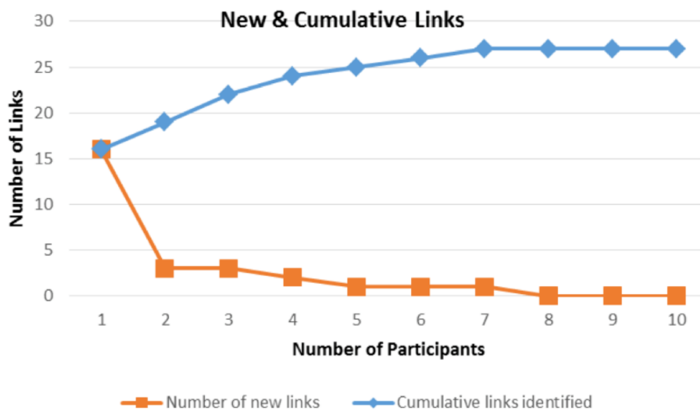
Figure E.6 below presents hypothetical CM-S Curves for links; where the participant in Table E.1 is the second participant on the curve. This participant identifies three new links beyond what the first participant had identified. The x-axis shows the number of participants included at that point (not the participant ID). The unit for the y-axis continues to be the number of elements (variables, links, delays, feedbacks). The two lines on the CM-S curve continue to be: accumulation and new elements.

Figure E.6 Example of Generating the Links CM-S Curves from Participant CLDs

Table Counting Links for an Example CM-S Analysis

CLDs	Number of new links	Cumulative links identified
1	16	16
2	3	19
3	3	22
4	2	24
5	1	25
6	1	26
7	1	27
8	0	27
9	0	27
10	0	27

Links CM-S Curves



The data for the CM-S Curves requires comparison of the results in Table E.1 above for all the participants. The "Mentions Y/N" columns from all participants are collated and compared to find the first participant mentioning each link. The ID for the participant who first identified each link is recorded in a new column. A count is then made for each participant of the number of new variables they identified. Participants are sorted on this list in descending order. The first point for both CM-S Curves is the highest number of new variables mentioned by a single participant. Subsequent points show the contribution of new variables from participants in descending order.

The same process is followed for the other elements.

When the curve flattens (i.e., no new elements are identified by additional participants), it is assumed that saturation has been reached for that element.

E.2.4 CREATION OF “RIGOROUSLY-INTERPRETED QUOTATIONS - FOR CAUSALITY”

A sub-set of the validation set interviews is selected for *rigorous analysis of causal statements* to produce what this dissertation refers to as Causal RIQs. One interview is selected from each of the clinical professions and both interviews from managers in the model validation set.

For clinicians and MAs, this decision is determined by considering statistics generated for each interview and presented in Table E.2. These statistics consider various aspects of individuals’ capability for systems thinking[182]. This table presents the method and source of information for calculating each of these statistics. This method for interview selection is supported by recommendations to focus on descriptions from participants with *rich* mental models in conducting research on complex systems[256].

Table E.2 Sources for the Statistics in the Systems Thinking Table

Systems Thinking Measure	Information Source
Count of statements validating a causal relationship	Sum of the “Frequency” column of Table E.1
Total causal relationships in SMM2.x marked for that individual	Sum of the “Mentions Y/N” column of Table E.1
SMM2.x coverage level	“Total causal relationships in SMM2.x marked for that individual” divided by the total count of links in “SMM2 Link ID” column of Table E.1
Percent of words pertaining to SMM2.x	See row “Percent pertaining to SMM2.x” in Figure E.7 Step 5
Percent of words validating SMM2.x	See row “Percent validating causality and dynamics” in Figure E.7 Step 5
Percent of words pertaining broadly to SMM2.x	See row Percent validating variables only” in Figure E.7 Step 5
Percent of words not pertaining to SMM2.x	See row “Percent not pertaining to SMM2.x” in Figure E.7 Step 5
Feedback loops	See SMM2.x marked for each participant (and count the feedback loops that are identified by that participant)
Delays	See SMM2 marked for each participant (and count the delays that are identified by that participant)

This analysis consisted of creating a series of tables, called *rigorously-interpreted quotations for causality* (Causal RIQs) Tables. An example is provided in Table E.3 below. In these tables, a comparison is made between a single causal statement in the transcript (left column) and the model variables and causal links which it describes in SMM2 (blue areas) generating an *interpretation* (right column lower half). This comparison involves producing a *quotation* (by truncating the statement) which tells the story represented by a portion of SMM2. When a specific phrase in the quotation refers to a model variable it is underlined and copied to the *Phrases needing interpretation* section (light blue area). In this section, phrases are placed in bullet lists adjoining the model variable(s) they refer to, in descending order along the variables' causal chain in SMM2 (dark blue area). This causal chain is then presented (e.g., Model Variable →+ Model Variable) (yellow area). This is repeated for all of the causal chains in the quotation. Finally, the quotation is interpreted (green area). Note, these colors are for guidance and not used in the actual CM-S results tables.

Table E.3 below shows an example of what this table would look like for one quotation. Appendix D presents these Causal RIQs by profession and in tables, where each table focuses on a different piece of the shared mental model.

Table E.3 Example Causal RIQ

<table>
<tr><td>Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="5">MA03-1) "It would be really good if we could implement the care [team] … but, when we have multiple doctors that all have patients at the same time … that are double booked. You are seeing anywhere from [x] patients in an hour and you have 2 or 3 MAs working … it just doesn't, you don't have the opportunity to stay with the same patient the entire time." (71/161)</td><td>• multiple doctors that all have patients at the same time … that are double booked
• You are seeing anywhere from [x] patients in an hour</td><td>Volume of patient visits</td></tr>
<tr><td>• 2 or 3 MAs</td><td>Capacity</td></tr>
<tr><td>• Implement the care team
• stay with the same patient the entire time</td><td>Task Shifting</td></tr>
<tr><td colspan="2">Volume of patient visits →- Capacity →+ Task Shifting</td></tr>
<tr><td colspan="2">A very high workload level reduces the opportunities for Task Shifting. The mechanism for this problem is Capacity. A high level of workload overwhelms the care team capacity, so Task Shifting cannot work.</td></tr>
</table>

In quote 1, above, MA03 uses several phrases which identify variables in SMM2.x. These phrases are underlined on the left. On the right, the phrases are linked to the variables they identify. MA03 identifies *Volume of patient visits* when describing "multiple doctors that all have patients at the same time … that are double booked" and "you are seeing anywhere from [x] patients in an hour". This quote identifies a steady level of *Capacity* at "2 or 3 MAs". Next, the quote identifies *Task Shifting* as "implement the care team" and "stay with the same patient the entire time". MA03 also identifies causal relationships between these variables (see causal chain and interpretation in bottom right of table).

Box E.2 below provides a comprehensive guide to the symbols and structure of the table.

Box E.4 Detailed Guide to Symbols and Structure of Rigorously-Interpreted Quotations for Causality

- On the left of the table
    a. A quote number and a participant number.
    b. The truncated quotation.
    c. Phrases referring to variables are underlined in the quotation. When the same phrase refers to more than one variable, it is double underlined.
    d. Phrases referring to system-dynamics-related items are underlined in the quotation.
    e. The word count of the quotation (truncated for confidentiality) over the total words in the original statement.
- On the right of the table
    a. The phrases and their corresponding variables are first presented, in the same order as their causal sequence in SMM2.x.
        i. The phrases that were underlined in the left hand column are presented in the right hand column, this time sorted by the variable or system dynamics-related item to which they are assigned.
        ii. Variables which are indirectly mentioned are shown with a zero (e.g., the quote says A->C, but SMM2.x has A->B->C, so "0" is used in place of a phrase).
        iii. When more than one phrase describes the same variable, each phrase is listed as a separate bullet.
        iv. The variable to which these phrases are assigned is then identified in a column to the right of the one containing the corresponding phrases.
        v. Variables are ordered according to their progression in the causal chain (not order of appearance in the text).
    b. When a system dynamics-related item is mentioned a similar process is followed. The relevant phrases are copied, each with a separate bullet, and the issue (e.g., Time Delay) is then identified in a column to the right of the one containing the corresponding phrases.
    c. Causal chains from the SMM2.x which are used to interpret the statement are in bold.
    d. A brief comment shows how the statement is interpreted using the causal chains.

E.2.5 THE CONCEPTUAL MODEL SATURATION TEST

The CM-S Test involves answering the following questions using the procedures (in bold) below. The sections above present the sequence for carrying out the procedures. This section presents the procedures as they relate to specific validity types and questions specific to the CM-S Test.

Box E.5 Conceptual Model Saturation Test (CM-S Test)

Do elements identified in Causal RIQs for validation set interviews exist in SMM2.x and are the identified elements formulated correctly? (And if not correct, then revise/remove incorrect ones)

- Are the variables and boundaries of the Shared Mental Model well-developed and validated? (CptV1 - variables & boundaries)
    - **Check Causal RIQs -** subjecting the *variables* in SMM2 to raw interview transcript data (from the model validation set) demonstrates how capable SMM2.x is of exposing the variables in participants' mental models
    - **Check against the problem statement -** that the elements revised via CM-S in SMM2 meet the model's purpose
    - **Check variable CM-S Curve for saturation**
- Are the relationships in the Shared Mental Model well-established and validated? (CptV2 - links)
    - **The process of developing the CM-S Diagrams checks if portions of SMM2 require reformulation**
    - **Check CM-S Diagram** by visualizing the percentage of participants who identify each relationship. Consider how well-established each relationship is, for the less well-established relationships (thin lines), specifically:
        - Does the relationship in SMM2.x distort aspects of the model validation set participants' statements? (FV1 - language limitations)
        - Are bounded rationality and cultural acceptability evident in the relationships? (CptV4 - culture)
    - **Check CM-S Diagram** by visualizing the number and type of professions identifying a particular link. Consider how well-established each relationship is, for the less well-established relationships (thin lines), specifically:
        - Does the relationship in SMM2.x distort aspects of the model validation set participants' statements? (FV1 - language limitations)
        - Are bounded rationality and cultural acceptability evident in the relationships? (CptV4 - culture)
    - **Check Causal RIQs --** subjecting the *relationships* in SMM2 to raw interview transcript data (from the model validation set) demonstrates how capable SMM2.x is of exposing the relationships in participants' mental models

Is there saturation in existing elements?

- Have new and relevant data regarding the elements in the Shared Mental Model ceased to emerge? (CptV3 – saturation)
    - **Check the Variables, Link, Delay and Feedback Loop CM-S Curves for saturation**
    - **The process of developing the CM-S Diagrams checks if portions of SMM2 require renaming and reformulation**

The researcher then reflects on the adequacy of the evidence provided for CM-S. If the results are inadequate, then additional interviews from the saturation reserve need to be consulted in revising SMM2 and then the CM-S test needs to be run again.

E.3 SIMULATION MODEL SATURATION

Model name	Participant CLDs	Team CLDs, Clinic CLDs	SMM1	SMM2	SMM3 / Conceptual Model	SIM1	SIM2	SIM3	Theoretical Model

Simulation Model Saturation (SIM-S) has been reached when it has been demonstrated that the simulation model does not contain radical departures from the participants' interviews. Specifically, this is demonstrated by verifying the simulation model (SIM1) does not contain radical departures from the causal statements in Causal RIQs (from the model validation set interviews). These are used here to validate the structure *and behavior* of the simulation model.

The determination of SIM-S is made by answering the following question, referred to as the **SIM-S Test:**

- Does the simulation model contain radical departures from the causal statements in the Causal RIQs?

This question is considered from the perspectives of structure, behavior and culture in the sections below.

The sequence of steps in SIM-S is designed to enable the researcher first, to check if the simulation model is within scope of the problem statement; and second, to check that the story that the simulation model (SIM1) tells checks out when compared with stories about the problem (its structure, behavior, the operating policies) as told by participants whose interviews were reserved for validation testing. Throughout this process, SIM1 is revised as needed (i.e., SIM1.x) and clarifications/revisions that this process identifies for the conceptual model are also noted – these changes are taken into account when developing the Theoretical Model.

E.3.1 STRUCTURE

The simulation model (SIM1) is subjected to the concepts described in (1) the problem statement, (2) the structural assumptions made in the dynamic hypothesis (i.e., the elements found in the conceptual model) and (3) the structural aspects of Causal RIQs. For each of these three *inputs*, the following process is followed:

Box E.6 Steps for Evaluating the Structural Concepts from Modeling Inputs for the Simulation Model Saturation Test

1. The structure of SIM1 is compared to the structural concepts referred to in the input.
    a. If these structures match, then the modeler repeats the process.
    b. If they do not match, then additional reflection is needed. Is it because:
        i. an aspect of structure is outside the scope of the input? Then document and repeat the process.
        ii. the quotation refers to an important problem aspect which is not in the model's current structure? Then, reformulation is warranted. Once completed, repeat the process.

In a simulation model, the structure includes more than just variables and relationships; specifically, it includes assumptions regarding the time frame, time step and how to initialize the model (in equilibrium or in a system undergoing growth).

E.3.2 BEHAVIOR

SIM1 is subjected to the reference behavior modes described in the problem statement and the Causal RIQs developed in CM-S. This method provides a formal way of reflecting on SIM1 (e.g., its purpose, structure, parameters and behavior) by way of comparison with the Causal RIQs. For each quotation, the modeler attempts to implement the described scenario using the two steps in Box E.7 below.

Box E.7 Steps for Evaluating the Behavior of SIM1 with Respect to Rigorously-Interpreted Quotations for Causality under the Simulation Model Saturation Test

1. If it can be implemented, the behavior of SIM1 is compared to the behavior described in the quotation.
    a. If the correspondence between these behaviors is reasonable, then the modeler moves on to the next quotation.
    b. If they do not match, then additional reflection is needed. Is it because:
        i. there are other aspects of the real scenario which the quotation does not cover, but which could be implemented in the model to generate a behavior match? If so, then implement the revised scenario and check if the behaviors match, document and return to step 1 here.
        ii. an aspect of scenario mentioned in the quotation is outside the scope of the problem statement? Then document and return to step 1.
        iii. the quotation refers to an important problem aspect which is not in the model's current structure? Then, reformulation is warranted. Once completed, return to step 1.
2. If the scenario described in the quotation cannot be implemented in the model, then additional reflection is needed (see 1.b. above).

E.3.3 CULTURE

SIM1 is subjected to information on the *culture* of the target group found in Causal RIQs. The question is the following: does SIM1 address cultural acceptability (i.e., that goals and policies are in terms of goals that people would actually desire and policies that they could actually implement) and cognitive limitations (i.e., that decision functions are in terms of things people actually see, in accordance with SDM's views on *bounded rationality*[108].)

For cultural acceptability, the modeler checks that the goals and policies described in the quotations are also found in the simulation model. For cognitive limitations, the modeler checks that the cognitive limitations inferable from the full quotations are also found in the simulation model (e.g., if the quotation refers to a clinician perceiving staff members' capabilities by observing how many errors they make, and the simulation model shows clinicians observing the current, true value of a staff members' capabilities, we know there is a mismatch, and in fact the simulation model does not represent the cognitive limitations indicated in this quotation).

E.3.4 THE SIMULATION MODEL SATURATION TEST

The SIM-S Test involves answering the following questions using the procedures (in bold) below. The sections above present the sequence for carrying out the procedures. Box E.8 presents the procedures as they relate to specific validity types and questions specific to the SIM-S Test.

Box E.8 Simulation Model Saturation Test (SIM-S Test)

Does the simulation model contain radical departures from the causal statements in the Causal RIQs?

- Are the variables and boundaries in the Simulation Model well-developed and validated? (CptV1 - variables & boundaries)
    - **Check that the simulation model's <u>variables</u> make sense in terms of the descriptions provided in the quotations and interpretations.** Does the treatment of variables in the simulation model as stocks, flows, auxiliaries or constants make sense? Does the simulation model involve less or more detail on variables, the time frame or the system scope than is warranted by the descriptions?
- Are the relationships in the Simulation Model well-established and validated? (CptV2 - links)
    - **Check that the simulation model's <u>relationships</u> make sense in terms of the descriptions provided in the quotations and interpretations.** Does the treatment of relationships in the simulation model as flows or information links make sense? Does the simulation model involve less or more detail on relationships than is warranted by the descriptions?
- Have important elements been omitted or distorted because of the procedures used in the creation of the Simulation Model (i.e., has "the discipline of the programming language" [105] (p. 120)) caused the model to take on an inaccurate form)? (FV1 - language limitations)
    - **Check that the causality in the full quotations is covered in the model.** If quotes talk a lot about social networks, agents or discrete processes rather than system structure, then consider the model construction (e.g., whether it accounts sufficiently for network structure or agents).
- How consistent is the Simulation Model with the structure and behavior that it was meant to represent? (FV2 - conceptual equals simulation)
    - **For structure, check that (1) the problem statement and (2) the structural assumptions** made in the dynamic hypothesis (i.e., the elements found in the conceptual model) match those found in SIM1
    - **For behavior, check that the reference behavior modes** described in the problem statement can be reproduced in SIM1
- Do experiments in the Simulation Model challenge or support its structure? (EV1 - structural design)
    - **Run simulations for the structural aspects** described in the Causal RIQs and compare run results with behavioral expectations mentioned in those statements
- Are bounded rationality and cultural acceptability evident in the Simulation Model? (CptV4 - culture)
    - **Check that the goals and policies described in the full quotations** are also found in the simulation model (i.e., that goals and policies are in terms of goals that people would actually desire and policies that they could actually implement)
    - **Check that the cognitive limitations inferable from the full quotations** are also found in the simulation model (i.e., that decision functions are in terms of things people actually see, in accordance with SDM's views on *bounded rationality*[108].)

SIM1 is revised as needed during SIM-S. The researcher then reflects on the adequacy of the evidence provided for SIM-S. If the results are inadequate, then additional interviews from the saturation reserve and/or additional stakeholders (see Table 2.12 FV1 and FV2) need to be consulted in revising SIM1 and then the SIM-S test needs to be run again.

At this point, once the SIM-S Test is passed, the simulation is referred to as SIM2.

E.4 SYSTEM DYNAMICS SATURATION

System Dynamics Saturation (SD-S) has been reached when the extent to which the validation set interviews verify that the target group sees SDM research as a useful way of addressing the issue has been demonstrated (CptV4 - culture). This is demonstrated using coded transcripts from the model validation set interviews, which are summarized by participant and then segmented using the **system dynamics-related items** code. The results are presented using *Rigorously-Interpreted Quotations – for Cognition* (Cognitive RIQs) for conceptual statements) and findings which demonstrate perceptions of causality, including additional Causal RIQs using causal statements which consider tensions, feedbacks and delays.

This step verifies that the model validation set participants' mental models[113][119] contain causality and that they perceive system structure as causing their problem. It verifies that they understand their mental models to have the attributes they are assumed to have in SDM, that mental models matter that they can (and indeed must) change, and that doing so involves emotional engagement.

Table E.4 below presents the evidence which is considered when assessing SD-S. Together, all of this information demonstrates the extent to which participants are likely to feel that, when presented with it, SDM research is appropriate and acceptable even though they were not directly exposed to it prior to their interview.

Table E.4 Conceptual Validity and the System Dynamics Saturation Test

<table>
<tr><th colspan="2">What is validated?</th><th>Methods</th><th>SD-S Test</th></tr>
<tr><td rowspan="4">Conceptual Elements of Mental Models</td><td>Variables</td><td>SMM-S, CM-S</td><td rowspan="5">Participants have mental models with causality and time delays
The shared mental model has a structure with time delays and feedbacks
The structure of the entire shared mental model is not seen by individual participants</td></tr>
<tr><td>Causal links</td><td rowspan="4">SMM-S, CM-S, SIM-S</td></tr>
<tr><td>Significant delays</td></tr>
<tr><td>Feedback loops</td></tr>
<tr><td colspan="2">Models</td></tr>
<tr><td colspan="2">SDM as a useful way to address the issue</td><td>SD-S</td><td>Participants see themselves as having mental models which are characteristic of Mental Models of Dynamic Systems
Changing mental models is part of the perceived solution to the problem</td></tr>
</table>

"Conceptual Elements of Mental Models" and "Models" are validated in the SD-S Test using the results of tests already performed to this point (SMM-S, CM-S, and SIM-S). Methods carried out expressly for the SD-S Test are performed to check that SDM is seen as a useful

[113] Social scientists approach these using ideas similar to what is considered in SDM. Nevertheless, there are definitional differences (e.g., some are retrospective while others could include past present and future).

way to address the issue (the final row of this table). These methods involve the creation of: first, Information Accumulation Graphs; second, Causal RIQs for SD-S analysis; and third, Cognitive RIQs.

E.4.1 CREATION OF INFORMATION ACCUMULATION GRAPHS

Information Accumulation Graphs are generated from the transcripts of interviews in the model validation set using the methods described below (see next page). Information Accumulation Graphs (developed for use in this dissertation) visualize the mental database of participants to show how their *mental data* is used in a CLD. In interview coding, mental data is segmented into three types: causal, conceptual and extraneous (see Section E.2.1 above). Viewing the relative density of words providing each kind of data allows one to assess the extent to which participants perceive the system as being more causal or conceptual, or important. A graph dominated by causal data indicates a participant who perceives the system mostly as a *system* with a causal structure made up of elements which generate dynamically complex behavior (causal links with polarity, delays, feedback loops). A graph dominated by conceptual data, on the other hand, indicates a participant who perceives the system mostly as *a set of disconnected variables* and a graph dominated by extraneous data indicates a participant who perceives the system's structure as being a minor issue compared to other issues.

Other techniques for summarizing the content of interviews have been developed for use in qualitative research and systems modeling. Qualitative researchers may use *tag clouds* to visualize the relative frequency of words in text data or of categories of themes in coded text data, by displaying the relative frequency of mentions of the words or categories in the data (could be text from an individual or a group or other) using relative font size of words or categories[334, 335]. Systems modelers may use diagrams (broadly known as *mental maps*[117, 336] or *cognitive maps*[118]) in modeling software to visualize the relationships between themes using words, arrows and possibly other symbols. These diagrams are also often used to represent understanding gleaned from groups, not always through interviews. When used to represent mental models elicited from text data, only one of these types, the tag cloud for words, could show what *else* was in the data beyond what was elicited. However, a tag cloud for words is not necessarily well-suited to visualizing the *goodness of fit* between the contents of an individual's mental model found in text data and the contents of a model which intends to represent a group's shared understanding of the system being perceived by that mental model.

Information Accumulation Graphs were designed to be a more granular, micro version of the *Accumulation Curve* concept from Fuzzy Cognitive Mapping[333] and the logic of the *Behavior Over Time Graph* used in SDM[151, 278]. Information Accumulation Graphs are designed to communicate to what extent does the mental model expressed by an individual during interview (i.e., the information flow over time *during* the interview) use causal or conceptual thinking about the system under study and to what extent is the system important to a given participant (i.e., the information accumulation). This tells us to what extent the participants' mental models can be said to be *in* the CLD (an important aspect of "CptV4 Culture").

Box E.9 describes how these graphs are created for each transcript.

Box E.9 Steps for Creation of Information Accumulation Graphs

Step 1. Bring the coded transcript (see Section E.2.1 for a description of how this data is collected) into MS Excel. This step entails the following:

- Copy the transcript into MS Excel. Check that each row is a separate paragraph. Each paragraph should be a different statement.
- Filter the transcript text so that only participant responses are considered (i.e., no interview comments or questions).
- Transfer the coding label for each statement from MS Word.
- Create a word count for each row.

Step 2. Create a pivot table to tabulate word counts by label over the span of the interview.

Step 3. Use the MS Excel pivot table features to calculate how the words accumulate in the different categories during the interview by adding up the new information in each interaction and accumulating it, by type.

Step 4. Display results obtained in the table documenting inflows and accumulations in the Information Accumulation Graph.

Step 5. Calculate several summary statistics that can be calculated for comparison across individuals.

The Information Accumulation Graph shows the cumulative word count on the y-axis and the progression of the interview on the x-axis, with a count of the new paragraphs. New paragraphs can be back and forth interactions between the researcher and a participant or new ideas. This graph can show indirectly the amount of time spent in the interview on describing causal statements, discussing related concepts, and extraneous issues. Using the generic example transcript that was coded in Section E.2.1, Figure E.7 below illustrates how the Information Accumulation Graphs are created.

Figure E.7 Example Participant Analysis Related to Generation of Accumulation Graph

Step 1: Bring the coded interview transcript into excel

Order	Text	Word coun	Code
1	Beginning of causal stater	21	1
2	Beginning of conceptual s	37	Comment
3	Any of these statement ca	27	Comment
4	Statements which are ent	8	

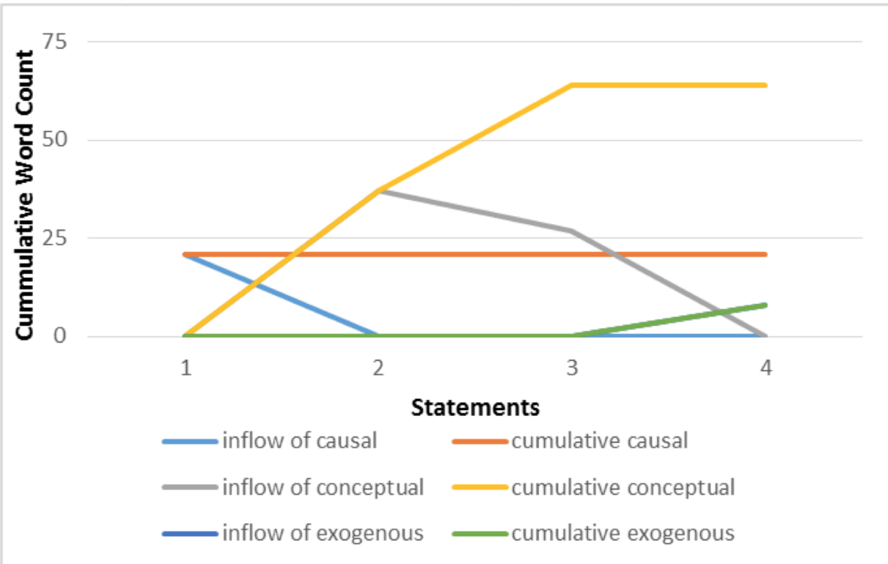
Step 2: Tabulate word counts by information type

Row Lab	1	Comment	(blank)	Grand Total
1	21			21
2		37		37
3		27		27
4			8	8
Grand Tota	21	64	8	93

Step 3: Documenting inflows and accumulations

inflow of causal	cumulative causal	inflow of conceptual	cumulative conceptual	inflow of exogenous	cumulative exogenous
21	21	0	0	0	0
0	21	37	37	0	0
0	21	27	64	0	0
0	21	0	64	8	8

Step 4: Generating Information Accumulation Graph



Step 5: Generating summary statistics of the share of information by validation type

	% words
Percent pertaining to SMM2.x	91%
Percent validating causality and dynamics	23%
Percent validating variables only	69%
Percent not pertaining to SMM2.x	9%

E.4.2 CREATION OF "RIGOROUSLY-INTERPRETED QUOTATIONS – FOR CAUSALITY" FOR SYSTEM DYNAMICS SATURATION

Whereas a subset of interviews is selected for CM-S, all interviews can be to produce SD-S Causal RIQs. All statements are selected which contain both **causal statements** and **system dynamics-related items** in the model validation set.

Rigorous *analysis of causal statements* is performed the same way as in CM-S (see E.2.4), except for in the case of feedback loops. It should be noted that SDM research does not require an assumption that participants understand that they are perceiving feedback loops as a pre-condition to beginning research in a participatory setting. I assumed that perceiving the type of feedback with which SDM is concerned, known as *information-feedback*, would be unlikely. People's mental models may contain feedback through their use of causal chains which, when connected, produce feedback loops. I do not assume that finding a loop after connecting chains mean that a person *perceives* information-feedback, only that the causal structure of their mental model *contains feedback loops*. That said, SDM does not assume that information-feedback is imperceptible. Similarly, the shifts in loop dominance which influence the behavior over time of dynamically-complex systems are another, still higher, level of abstraction. SDM researchers are still looking for ways to represent these changes in diagrams[180].

Considering these assumptions, Causal RIQs can show perception of feedback loops or information-feedback, with the following modifications.

In the results chapter, these Causal RIQs are presented for any quote showing perception of information feedback, or interacting feedback loops. A quote which shows a participant perceiving the chain of links which make up a single feedback loop does not qualify for inclusion in SD-S. Also, a similar table is not created for perceptions of significant time delays. Time delays and single loops are perceived much more easily and more often than information feedback or interacting feedback loops. Instead, all phrases for time delays in all Causal RIQs in this research (CM-S and SD-S) are brought together in a single table and used to explore how time delays are perceived.

Table E.5 below shows an example of what this table would look like (and Figure E.8 is also presented to visualize the links present and missing in PT07's quotations). Rather than use a real quote, as was done in Table E.3 for CM-S, the content of this table describes how the pieces of the table in these Causal RIQs are different from those created for CM-S. This includes a description of how a variable may sometimes be inferred when it is part of the longer causal chains being described in the quote, but was not directly mentioned in the quote. This is useful if the quote refers to a feedback loop (e.g., describing a sequence of shifting loop dominance driving changing conditions over time), but fails to explicitly mention one of the variables.

Figure E.8 Participant 7 example CLD

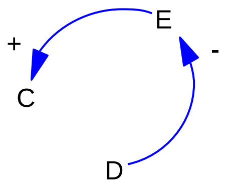
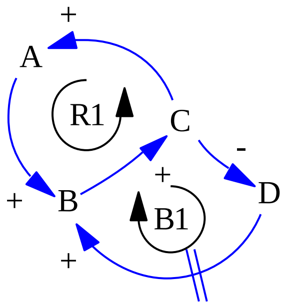
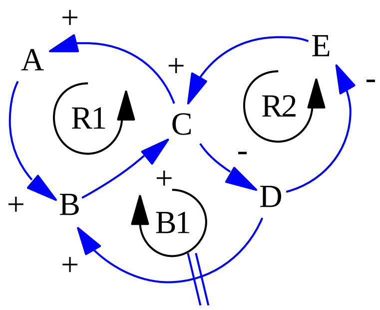
<table>
<tr><td>PT07-1
A previous quote's links (may not be shown in any Causal RIQ)
</td><td>PT07-3
The current quote's links
</td><td>Group CLD (i.e., SMM2.x)
</td></tr>
<tr><td colspan="3">Implementing validation tests may results in periodic model revisions, these are labeled accordingly (e.g., I marked models made during the SMM-S Test as SMM2.1, SMM2.2 … SMM2.x). These are referred to as SMM2 in the body of this dissertation.</td></tr>
</table>

Table E.5 Example Causal RIQ for Feedback

<table>
<tr><td>Participant-Quote number) "Quote … <u>phrases referring to variables</u>" (word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="16">PT07-3) "This quote refers to Loops R1, R2 & B1 in Figure E.8 above. It uses phrases which refer to variables A, B, C, D and E. It also includes a phrase which indicates the participant is aware of the workings of information-feedback in this system.
"The interpretation includes two comments. The first describes how the loop works, and the second, how the causal links play out in the quote, and how the loop works in general. When applicable it also shows how the quotation describes the loops' interaction over time with other loops.
"Occasionally, inferring one variable can add another feedback loop. If this decision is supported by the quote, or the participant referencing the inferred variable and or links in a previous quote, then another loop will be added, and a "0" will be used in place of a phrase." (8/8)</td><td>• phrase</td><td>Feedback Loop</td></tr>
<tr><td>• phrase</td><td>Time Delay</td></tr>
<tr><td>• phrase</td><td>A</td></tr>
<tr><td>• phrase</td><td>B</td></tr>
<tr><td>• phrase</td><td>C</td></tr>
<tr><td colspan="2">A →+ B →+ C →+ A</td></tr>
<tr><td colspan="2">Loop R1 This comment describes how the quote is interpreted to show the existence of a reinforcing feedback loop (R1).

This interpretation describes how the loop works, at a level of abstraction higher than the experience in the quote.</td></tr>
<tr><td>• phrase</td><td>D</td></tr>
<tr><td colspan="2">B →+ C →- D --||→- B</td></tr>
<tr><td colspan="2">Loop B1 This comment describes the same thing, but for a balancing loop (B1) which interacts with the feedback loop (R1) cited above.

This interpretation also describes how the loop works, at a level of abstraction higher than the experience in the quote.</td></tr>
<tr><td>0</td><td>E</td></tr>
<tr><td colspan="2">C →- D →- E →+ C</td></tr>
<tr><td colspan="2">Loop R2 This comment describes how the existence of this loop (R2) is supported, despite the lack of direct evidence for perceiving "E". It also tells how the balancing loop (B1) and the reinforcing loop (R1) appear in the quote to interact. This might mean they create the conditions under which this new reinforcing loop (R2) can exist and perhaps even dominate behavior of the system.

This interpretation describes how the loop works and any evidence that "E" can be inferred, as it was not in the original quote.</td></tr>
</table>

Box E.10 below provides a comprehensive guide to the symbols and structure of the table specific to when a *system dynamics-related item* is mentioned showing perception of multiple feedback loops, or actual perception of information-feedback.

Box E.10 Detailed Guide to Symbols and Structure of Rigorously-Interpreted Quotations for Causality for Perception of Information-Feedback or Multiple Feedback Loops

- On the left of the table
    a. (all aspects are the same as in CM-S)
- On the right of the table
    a. When perception of information-feedback is mentioned, "Feedback Loop" is identified in the right-hand side cell, with the corresponding phrases on the left-hand side cell shown in bullet points.
    b. The phrases and their corresponding variables are presented for each feedback loop perceived using the same notation as in CM-S.
        i. Variables are ordered according to their progression in the feedback loop
        ii. One feedback loop is presented at a time.
    c. Two brief comments for each feedback loop in the statement show how the statement is interpreted.
        i. The first references the name of the feedback loop and interprets the quote in the context of the participant and in terms of the loop's polarity, as well as its relationship to other feedback loops in the quote.
        ii. The second interprets the statement in generalizable terms.

E.4.3 CREATION OF "RIGOROUSLY-INTERPRETED QUOTATIONS – FOR COGNITION"

When many participants perceive the same system, it validates the assumption that mental models are *internal*, but also *enduring* and *accessible*. When participants perceive only part of the system, it validates the assumption that individual mental models are *limited*. When multiple participants validate the same elements of system structure, it validates the assumption that the shared mental model is *analogous to perceived structure* of the real *external system*.

When participants see themselves as having mental models which are characteristic of Mental Models of Dynamic Systems, they verify that a research approach which studies system dynamics is likely to be culturally acceptable. This likelihood is bolstered when participants recognize that changing mental models is part of the perceived solution to the problem. Finally, when participants express awareness that the needed mental model changes will require engaging the target group on a sub-conscious emotional level, the conditions for cultural acceptability of SDM research are saturated. In so doing, they verify that mental models matter and that emotional engagement matters to the problem at hand.

Statements (from the entire model validation set) previously coded[114] as *conceptual statements* containing *system dynamics related items* are analyzed using the table described below. This analysis results in Cognitive RIQs.

An example is shown in Table E.6 below. Gray cells are part of the template for the table. White cells are to be filled in for each quotation. The first column presents the quotation, underlining phrases referring to system-dynamics related items.

The second group of columns identifies whether each system dynamics-related item is present or absent in that quotation. An "x" marking the presence of an item in the:

- First category (i.e., "shows awareness…") indicates that the participant describes their cognitive process in a way that agrees with aspects of Mental Models of Dynamic Systems.
- Second category (i.e., "mental models matter") indicates that the participant considers mental models to be important with respect to the problem statement.
- Third category (i.e., "emotional engagement") indicates that the participant describes himself/herself or others as being emotionally engaged in this problem.

The final row presents the interpretation of the quotation specific to system-dynamics related items. This table is designed to select a subset of quotations with system-dynamics-related items for presentation in the results section. A modified version of this table is used for presentation of results.

Table E.6 SD-S Rigorous Analysis of Conceptual Statements with System Dynamics-Related Items

<table>
<tr><td rowspan="2">Participant-Quote number)
"Quote … phrases referring to system dynamics-related items"
(word count/total words in causal statement)</td><td colspan="3">Presence of System Dynamics-Related Items
An "x" indicates the presence of the system dynamics-related item in this quotation.</td></tr>
<tr><td>Mark</td><td>Items</td><td>Item Categories</td></tr>
<tr><td rowspan="11">NM01-47) "I don't see the whole picture here, I only see my part, so I trust my managers and my administrators that they are going to see the whole picture. So when I give suggestions, if it is something that may not work then I know that it may be [because of] something that I am not seeing." (57/57)</td><td>x</td><td>Relatively enduring</td><td rowspan="7">Shows awareness of having a mental model, describing it in accordance with the aspects of Mental Models of Dynamic Systems</td></tr>
<tr><td>x</td><td>Accessible</td></tr>
<tr><td>x</td><td>Limited</td></tr>
<tr><td>x</td><td>Internal</td></tr>
<tr><td>x</td><td>Conceptual</td></tr>
<tr><td>x</td><td>Representation of an external system</td></tr>
<tr><td>x</td><td>Whose structure is analogous to the perceived structure of that system</td></tr>
<tr><td>x</td><td>Existence of MMs in the problem</td><td rowspan="2">Mental models matter</td></tr>
<tr><td></td><td>Changes to MMs matter for the problem</td></tr>
<tr><td></td><td>Changes to the system involve emotion</td><td rowspan="2">Emotional engagement</td></tr>
<tr><td></td><td>Changes to MMs involve emotion</td></tr>
<tr><td colspan="4">Interpretation: NM01 recognizes that MMs are important to negotiating change across stakeholders in PCT.</td></tr>
</table>

Box E.11 below provides a comprehensive guide to the symbols and structure of this table.

[114] See Appendix E for the detailed methods for coding the model validation set interviews.

Box E.11 Detailed Guide to Symbols and Structure of the Rigorously-Interpreted Quotations for Cognition

- In the first column:
  a. Lists the items for analysis
- In the second column, first cell:
  a. A quote number and a participant number.
  b. The truncated quotation.
  c. Phrases referring to system-dynamics-related items are underlined.
  d. The word count of the quotation (truncated for confidentiality) over the total words in the original statement.
- In the second column, subsequent cells:
  a. An x is placed in each cell corresponding to a SD-S related item that is mentioned in the quotation.
- In the second column, last cell:
  a. A brief comment shows how the statement is interpreted to show a system dynamics-related item.

E.4.4 THE SYSTEM DYNAMICS SATURATION TEST

The SD-S Test involves answering the following questions using the procedures (in bold) below. The sections above present the sequence for carrying out the procedures. This section presents the procedures as they relate to specific validity types and questions specific to the SD-S Test.

Box E.12 System Dynamics Saturation Test (SD-S Test)

To what extent do the validation set interviews verify that the target group sees SDM research as a useful way of addressing the issue?

- Does the target group see SDM research as a useful way of addressing the issue? (CptV4 - culture)
  - **Check Causal RIQs –** whether participants have mental models with causality
  - **Check Information Accumulation Graphs –** for the extent to which participants' interviews contained causal statements
  - **Check CM-S Diagram** by noting the percentage of participants who identify each relationship. Consider:
    - whether the shared mental model has a structure with time delays and feedbacks
    - whether the structure of the entire shared mental model is not seen by individual participants
  - **Check Cognitive RIQs** – consider:
    - whether participants see themselves as having mental models which are characteristic of Mental Models of Dynamic Systems
    - whether changing mental models is part of the perceived solution to the problem
    - the extent to which the needed mental model changes will require engaging the target group on a sub-conscious emotional level

The researcher then reflects on the adequacy of the evidence provided for SD-S. If the results are inadequate, then additional interviews from the saturation reserve need to be consulted to run the SD-S test again. If supporting evidence is still not found, then findings to date deserve review.

E.5 DATA SUITABILITY

The determination of **Data Suitability** is made by reflecting on the three types of data that are possible (mental, written and numerical), the data available, and their limitations. Specifically, the purpose of this test is to answer the following question:

- Are the data upon which the model is based used within the scope of their limitations?

After reflection on the data's attributes with respect to validity (i.e., reliability/appropriateness, accessibility, and sufficiency), a judgment is made regarding the data validity of these data. This test relates to DV1 (mental data), DV2 (written data), and DV3 (numerical data), (please see their definitions in Section D.2.4 for example questions).

Table E.7 below presents an organized way for capturing and reporting on the suitability of the data used in model development. The first column lists the data types. The second column lists the data sources. The third column, has a comment that describes the data validity ascribed to these data (reliable/appropriate, accessible, and sufficient). The fourth column, documents the determination that the modeler/researcher made with respect to the potential of these data for use in the model development. The fifth and final column, documents the way(s) in which each data type was used in model development. During model development and upon model completion, columns four and five are compared. The purpose of this comparison is to answer the Data Suitability Test. A populated version of this table is presented in the results section.

Table E.7 Example Table for Data Suitability

Data Types	Specific Data Available (examples provided)	Description of Data Validity	Determination	Use in Model
Mental	Discussions (recorded, transcribed, field notes…)	<<to fill in>>	<<to fill in>>	<<to fill in>>
Written	Meeting notes, other models	<<to fill in>>	<<to fill in>>	<<to fill in>>
Numerical	Administrative data, publicly available data, numerical data collected specifically for the research	<<to fill in>>	<<to fill in>>	<<to fill in>>

This test relates both to model documentation and model validation. It should be started prior to model development and updated during model development.

E.6 METHODS SUITABILITY

The determination of **Methods Suitability** is made by reflecting on the modeling process and the modeling language(s) used in the research process. Specifically, the purpose of this test is to answer the following question:

- What are the trade-offs of the specific modeling approach used in this research?

After reflection on the approach's attributes with respect to validity, a judgment is made regarding the methods suitability of the approach. This test relates to conceptual validity (CptV4 - culture) and formulational validity (FV1 - language limitations).

- Check the methods
  - Does the modeling process address cultural acceptability and cognitive limitations and use a non-coercive approach? (CptV4 - culture)
  - Have important elements or relationships been omitted or distorted because of the procedures used in the modeling process? (FV1 - language limitations)

Table E.8 below presents an organized way for capturing and reporting on the suitability of the methods used in model development. This table is based on the trade-offs approach presented for use in shared mental models research by Gray *et al.*[174], and expands it for the SDM research context. At the top of the table, the methods are described briefly in four categories: sampling design, elicitation, aggregation and validation. Sampling design includes discussing how the researcher identifies the stakeholders whose mental models will be accessed in this research. Elicitation includes methods used in capturing individuals' mental models. Aggregation includes methods used in making assumptions about shared mental models. The lower portion of the table contains descriptions of the key trade-offs of the methods (in terms of pros and cons) used in each of the four categories. A populated version of this table is presented in the results section.

Table E.8 Methodological tradeoffs of the Methodology in Phase 1-3

Sampling Design: <<Brief description>>

Elicitation methods: <<Brief description>>

Aggregation methods: <<Brief description>>

Validation methods: <<Brief description>>

Methods	Pros	Cons
Sampling Design	<<List pros>>	<<List cons>>
Elicitation Methods	<<List pros>>	<<List cons>>
Aggregation Methods	<<List pros>>	<<List cons>>
Validation Methods	<<List pros>>	<<List cons>>

E.7 STAKEHOLDER DIALOGUE SUITABILITY

The determination of **Stakeholder Dialogue Suitability** is made by reflecting on the research project and the extent to which stakeholders have been engaged.

- Have stakeholders been appropriately engaged throughout the research project?

This question exists because stakeholder dialogue is considered important in both model development and model validation. Model validation in SDM includes inward-focused methods for structuring the researcher's thinking on trial and error in model development as well as outward-focused methods for engaging with stakeholders (see section 2.2.2.3). In both cases, the purpose is building stakeholders' confidence in the model. Because models are based on stakeholder mental models for the purpose of improving stakeholder mental models, stakeholder engagement is *also* conceived of as being tightly tied to model development.

Forrester described the process of model validation with stakeholders as an: "examination, by the managers familiar with the actual system, of the model assumptions and its over-all performance" which was carried out in order to check whether it would "reveal implausible structure or policy assumptions or behavior that is judged to degrade the model for its intended uses"[107] (p. 263,268). To be suitable, then, stakeholder engagement for research in SDM should touch on all aspects of model validity including the *objectives* of relevance to, and consistency with, the appreciation of the situation, and suitability for purpose, and the *foci* of structure, behavior and culture.

Concannon *et al.*[337] present a framework for effective stakeholder engagement. They define the term *stakeholder* and they also identify important considerations for which, how and when stakeholders should be engaged. A wide variety of methods exist. The core of any stakeholder engagement process is to create the opportunity for iterative interactions that generate rich dialogue and reciprocal learning[319].

While the exact approach is tailored to the specific research project, a process of stakeholder dialogue draws on the knowledge of stakeholders to validate the *usefulness* of the model and its congruence with stakeholders' experience with this problem in a variety of contexts. Verifying with stakeholders that the research project's progress has taken into account their perspectives eventually adds up to contributing to the assessment of validity at most validity types (i.e., CptV1, CptV2, CptV4, FV1, FV2, FV3, EV1 and EV2).

APPENDIX F: DETAILED RESULTS OF CONCEPTUAL MODEL SATURATION & SIMULATION MODEL DOCUMENTATION

This appendix first presents the Conceptual Model - Saturation results that informed the revision of the shared mental model, producing Shared Mental Model 3 (referred to as the Conceptual Model). The Conceptual Model served as the blueprint for the simulation model. Then, this appendix presents the constants and equations for the simulation model.

F.1 RIGOROUSLY-INTERPRETED QUOTATIONS FOR CONCEPTUAL MODEL SATURATION

Rigorous analysis of causal statements consisted of using Causal RIQs ("rigorously-interpreted quotations – for causality") to subject the *variables and relationships* in SMM2.x[115] to interview transcripts set aside for validation.

This section presents: 1) the information upon which a sub-set of interviews was selected for rigorous analysis of causal statements and 2) a sampling of the Causal RIQs generated. These RIQs are first summarized in paragraph form, followed by a reference to the corresponding table. These tables are organized as described previously. The Causal RIQs presented here demonstrate that SMM2.x is capable of exposing variables and relationships in validation-set participants' mental models**.** In so doing, this evidence builds confidence in the causal structure of SMM3.

The interviews are semi-structured, meaning they touch on different subjects, like a survey, but they are also allowed to flow like a conversation. In so doing, participants do not make specific definitions of variables and relationships every time that they discuss a specific subject. Instead they share experiences where, at times, they take different aspects of causal mechanisms for granted. In these instances, a "0" is used to indicate that these portions of the causal mechanism were not referred to explicitly in that statement.

F.1.1 SELECTION OF INTERVIEWS FOR RIGOROUS ANALYSIS OF CAUSAL STATEMENTS

One clinician (CL03) and one MA (MA03) from the validation-set interviews were selected for analysis using measures presented in Table F.1 below. These measures capture various aspects of systems thinking by profession and participant. The bolded text identifies the clinician and MA selected for rigorous analysis of causal statements. Causal statements from the center manager (CM01) and nurse manager (NM01) interviews in the validation-set were also included in this rigorous analysis. Therefore, their systems thinking measures are also presented in this table.

[115] Implementing validation tests may results in periodic model revisions, these are labeled accordingly (e.g., I marked models made during the SMM-S Test as SMM2.1, SMM2.2 … SMM2.x). These are referred to as SMM2 in the body of this dissertation.

Table F.1 Systems Thinking by Participant

	Providers				MAs				Managers	
	CL01	CL02	**CL03**	CL04	MA01	MA02	**MA03**	MA04	**NM01**	**CM01**
Count of statements validating a causal relationship	36	17	**20**	21	11	37	**39**	9	**18**	**30**
Total causal relationships in SMM2.x marked	16	8	**13**	11	8	8	**17**	5	**10**	**18**
SMM2.x coverage level	53%	27%	**43%**	37%	27%	27%	**57%**	17%	**33%**	**60%**
Percent of words pertaining to SMM2.x	44%	73%	**86%**	62%	31%	59%	**66%**	71%	**24%**	**88%**
Percent of words validating SMM2.x	19%	58%	**69%**	45%	20%	19%	**52%**	47%	**16%**	**29%**
Percent of words pertaining broadly to SMM2.x	25%	15%	**17%**	17%	11%	41%	**14%**	24%	**8%**	**59%**
Percent of words not pertaining to SMM2.x	56%	27%	**14%**	38%	69%	41%	**34%**	29%	**76%**	**12%**
Feedback loops	2	0	**2**	2	1	1	**4**	1	**3**	**2**
Delays	4	1	**2**	4	2	2	**5**	1	**4**	**1**

CL03's interview is selected for presenting clinician illustrative quotations since CL03 spends most of the interview time discussing causal statements (86% of words pertaining to SMM2.x). While CL01 identifies three more causal relationships in SMM2.x, CL03 is selected for this analysis because CL01 spends less time explaining those causal relationships (19% of words pertaining to SMM2.x). Therefore, it is hoped that CL03's interview would do a better job of clarifying formulations.

Except for a bit of personal background at the beginning, MA03 spent more of the interview discussing causal information than any other type of information. MA03 also identified the largest number of feedbacks and delays of anyone in the clinic. This indicates that MA03 has a well-developed capability for systems thinking. MA03 provided information pertaining broadly to and information not pertaining to SMM2.x at about the same level asMA04 and less than MA01 and MA01 (who each spent about 80% of the time discussing this type of information). MA03's interview is selected for presenting the MA illustrative quotations.

Mangaement interviews from the model-validation-set are automatically included in this analysis. While NM01 spends 76% of the interview discussing things not pertaining to SMM2.x, NM01 does identify three feedback loops and four delays in the short amount of time spent on relationships in SMM2.x. The 76% portion is mostly spent providing normative statements on variables in the SMM2.x, praise or criticim without causal or conceptual descriptions. CM01 is the only participant to spend most (59%) of the interview time describing variables (providing conceptual information broadly related to SMM2.x). At the same time, this interview spends the least amout of time outside SMM2.x (12% of words).

F.1.2 MEDICAL ASSISTANT CONCEPTUAL MODEL – SATURATION RESULTS

SMM2.x was capable of exposing the variables and relationships in MA03's mental model regarding PCT. MA03's mental model of PCT centers on the relationship "Capacity →+ Task-shifting"[116]. Specifically, MA03 indicated that a decrease in "Capacity" causes a decrease in "Task-shifting". Quotes 1-3 below describe this relationship as the mechanism for a tension between "Volume of patient visits" and "Task-shifting". Quote 3 (MA03-3) gives the most explicit description of this tension, which shows MA03's frustration. See Table F.2 below.

Quote 4 describes an unintended consequence of maintaining "Capacity" persistently below what is required for task-shifting. It reduces task-shifting even after it is "*going and working*". This was done on purpose through a policy of moving MAs around the clinic to address

[116] This is the second-most-often mentioned relationship in the validation-set interviews.

capacity issues (rather than ensuring sufficient capacity on each team), which disrupted task-shifting. Quote 5 below describes how this policy was justified. Its purpose was to "*accommodate*" teams, by giving temporary relief to their capacity problems. In so doing, it disrupts their daily routine, making "*a kink in the day*", task-shifting is central to what they do and taking away capacity disrupts it. See Table F.3 below.

In Quotes 6-7, MA03 describes two situations relating to "MA satisfaction". There is a constant drag on all growth in "MA satisfaction" coming from the external fact that it is seen as a boring, low-skilled, (and unlicensed and low-paying) job. This impacts turnover. However, pairing clinicians and MAs appropriately can fight against this drag. Quote 7 describes two variables that increase MA satisfaction by describing a situation where task-shifting makes MAs happier in their job as they learn more about providing primary care. See Table F.4 below.

In quotes 8-9 MA03 describes the mechanism whereby pairing clinicians and MAs appropriately permits an increase in "MA satisfaction". MA03's clinician has placed trust in MAs to make choices about how they perform tasks that have been shifted to them, in this case: listening to the patient. With practice, MAs' ability to perform this task improves over time, permitting them to make fewer errors and be flexible. These three things (trust, new tasks and improved capabilities) increase job satisfaction for MA03. In quote 9, MA03 juxtaposes their team's process for addressing task-shifting issues with the dysfunctional process used on some other teams. When there is no trust between clinicians and MAs, "*nothing gets done"* meaning task-shifting issues arise but are not addressed. On MA03's team, however there is an ongoing dialogue between MA and clinician where a care problem is defined, a solution is drafted up and then tested. When management approval is needed, the time delay is lengthened. See Table F.5 below.

Table F.2 Causal RIQs – MA03 “Capacity →+ Task-shifting”

<table>
<tr><th>Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)</th><th colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</th></tr>
<tr><td rowspan="5">MA03-01) “It would be really good if we could implement the care [team] … but, when we have multiple doctors that all have patients at the same time … that are double booked. You are seeing anywhere from [x] patients in an hour and you have 2 or 3 MAs working … it just doesn’t, you don’t have the opportunity to stay with the same patient the entire time.” (71/161)</td><td>• multiple doctors that all have patients at the same time … that are double booked
• You are seeing anywhere from [x] patients in an hour</td><td>Volume of patient visits</td></tr>
<tr><td>• 2 or 3 MAs</td><td>Capacity</td></tr>
<tr><td>• implement the care team
• stay with the same patient the entire time</td><td>TS</td></tr>
<tr><td colspan="2">Volume of patient visits →- Capacity →+ TS</td></tr>
<tr><td colspan="2">A very high workload level reduces the opportunities for TS. The mechanism for this problem is Capacity. A high level of workload overwhelms the care team capacity, so TS cannot work.</td></tr>
<tr><td rowspan="5">MA03-02) “The one reason that we can’t do it is that a lot of the times we don’t have the staff. (…so having an additional MA would help?)
“It would. … If there were any ways to eliminate the double-books and things like that. I think it would make it a lot easier too… The providers get overwhelmed with patients … it’s like we are just trying to motor through them as fast as possible which isn’t good.”(77/120)</td><td>• the double books
• overwhelmed with patients
• just trying to motor through them as fast as possible</td><td>Volume of patient visits</td></tr>
<tr><td>• have the staff</td><td>Capacity</td></tr>
<tr><td>• we can’t do it</td><td>TS</td></tr>
<tr><td colspan="2">Volume of patient visits →- Capacity →+ TS</td></tr>
<tr><td colspan="2">A high level of workload overwhelms the care team capacity, so TS cannot work.</td></tr>
<tr><td rowspan="5">MA03-03) “It seems like, like we are being told … ‘get the patients through faster’ but we need to sit down with them and have a conversation with them. But we need to be fast. And, we don’t want to keep the patients waiting. But we need to double book these patients. And you know… so there are a lot of conflicts.” (61/349)</td><td>• get the patients through faster’
• need to be fast
• don’t want to keep the patients waiting
• need to double book</td><td>Volume of patient visits</td></tr>
<tr><td>0</td><td>Capacity</td></tr>
<tr><td>• we need to sit down with them and have a conversation with them</td><td>TS</td></tr>
<tr><td colspan="2">Volume of patient visits →- Capacity →+ TS</td></tr>
<tr><td colspan="2">Without enough MAs, PCMH requires too many things of them at any given moment. Maintaining high levels of access overwhelms the care team capacity which is necessary for TS to work.</td></tr>
</table>

Table F.3 Causal RIQs – MA03 Shifting the Burden of "Capacity"

Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable** →+ **Model Variable** (→ = causal link, --||→+ = with delay, +/- = positive or negative polarity) Interpretation	
MA03-04) "Sometimes we will go to (another department) … just depending on the staffing for that day… It affects the team. Obviously, that puts us down a person, and then things kind of get stressful … Honestly, I would rather that it didn't happen, because we have our own little system that is going and working. And when we get pulled to other departments it's a kink in the day."(69/247)	• Sometimes we will go to (another department) • down a person • when we get pulled	**Capacity**
	• the team • we have our own little system • it's a kink in the day	**TS**
	Capacity →+ TS	
	• kind of get stressful	**MA satisfaction**
	Capacity --||→+ MA satisfaction	
	Without enough staff, MAs move across departments to fill scheduling gaps. This reduces TS.	
MA03-05) "I know that they have tried to provide more staffing under our many, many requests. … But they have come down and said that "it's just not in the budget and (so) it's not feasible to get any more help." So, they have, as much as they can, they have tried to accommodate the staffing. Getting floats from other places, or people from other departments to help out when needed. So, in that respect, that has been better." (78/102)	• it's just not in the budget	**Clinic profit margin**
	• provide more staffing • it's not feasible to get any more help	**Hiring new MAs**
	• staffing	**Capacity**
	Clinic profit margin →- Hiring new MAs →+ Capacity	
	A very high workload reduces the opportunities for TS. The mechanism for this problem is the number of MAs and the "care team". A high level of workload overwhelms the care team capacity, so TS cannot work	

Table F.4 Causal RIQs – MA03 Things that Impact "MA Satisfaction"

Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable →+ Model Variable** (→ = causal link, --||→+ = with delay, +/- = positive or negative polarity) Interpretation	
MA03-06) "Honestly, since we have done some re-arranging with the medical assistants it has been a lot better. There used to be a lot of turnover, because MAs were just not happy. Um, either there was MA drama on a team, or there was provider/MA relationships that just were not working out. … Since we have made the changes and were aware of what was going on that way it has been a lot better. Everyone seems a lot more happy and [to] like their job a little more." (88/113)	• Honestly since • there used to be • since we have made the changes and were aware of what was going on that way	**Time Delay**
	• a lot more happy and [to] like their job a little more	**MA as dead-end job**
	• not happy • a lot more happy and [to] like their job a little more	**MA satisfaction**
	• turnover	**MA retention**
	MA as dead-end job --||→- MA satisfaction --||→+ MA retention	
	• provider/MA relationships	**Clinician - MA relationship (Trust)**
	0	**TS**
	0	**MA capabilities**
	Clinician - MA relationship (Trust) →+ TS --||→+ MA capabilities →+ MA satisfaction	
	Increasing MA satisfaction "a lot more" only makes them like their job "a little more." The drag on MA satisfaction comes from the external fact that being an MA is a dead-end job. Increasing MA satisfaction does have a noticeable impact on turnover. TS requires clinicians to trust their MAs. In this clinic, MDs were allowed to "re-arrange" the MAs to find ones they could trust. After a time delay, which is evidenced here by the certainty that it has passed, MAs have been happier and more satisfied with their job, although only slightly.	

<table>
<tr>
<td>Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)</td>
<td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = positive or negative polarity)
Interpretation</td>
</tr>
<tr>
<td rowspan="7">MA03-07) “I love the patient care aspect of what I do … Even if they go on and on about every little thing… it’s interesting... So, having the chance to kind of sit down more casually and having a conversation with the patient, I love that. ‘Cause, you learn a lot, … about all the problems that they have … how they are dealing with it and what the doctor has been doing for them … it makes the job more interesting than just sitting at the computer and doing the same questions over and over again.” (96/155)</td>
<td>• having a conversation</td>
<td>TS</td>
</tr>
<tr>
<td>• you learn a lot… what the doctor has been doing for them</td>
<td>MA capabilities</td>
</tr>
<tr>
<td>• I love the patient care aspect
• I love that
• the job more interesting</td>
<td>MA satisfaction</td>
</tr>
<tr>
<td colspan="2">TS --||→+ MA capabilities →+ MA satisfaction</td>
</tr>
<tr>
<td>• more interesting than just sitting … and doing the same …[thing] over and over again</td>
<td>MA as dead-end job</td>
</tr>
<tr>
<td colspan="2">MA as dead-end job --||→ - MA satisfaction</td>
</tr>
<tr>
<td colspan="2">TS expands the set of tasks performed by MAs, including some of the MD’s “listening” work. Gaining the capability to do this in a conversational style rather than just as data entry takes time. This pays off for the MA in increased job satisfaction and providing more stimulating, higher-skilled work experience, which is needed to overcome the fact of MA usually being a boring low-skilled job.</td>
</tr>
</table>

Table F.5 Causal RIQs – MA03 The Influence of "MD/MA Relationship (Trust)"

Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable** →+ **Model Variable** (→ = causal link, --||→+ = with delay, +/- = link polarity) Interpretation	
MA03-08) "After you have been here for a while and you have seen the same type of thing over and over again, you almost have the questions memorized. And so, it gives you a little bit better, I mean you can focus more on what the patient is actually saying, (rather than just thinking) 'what the next question is going to be' … (or) fumbling … to get some kind of answer out of them. "But there are times when you can't really go by (scripts). Because they come in for some really obscure thing, or they have got so many issues that they really just came in more to talk …. In instances like that, you have to be … flexible and that's what I like about our department… Our doctor [says], 'if the situation doesn't warrant it … don't worry about it.' … I will figure out myself … having that flexibility is really good." (155/246)	• After you have been here for a while and you have seen the same type of thing over and over again	**Time Delay**
	• Our doctor [says] 'if the situation doesn't warrant it … don't worry about it.'	**Clinician - MA relationship (Trust)**
	• I will figure out myself • you have to be … flexible • flexibility	**TS**
	• have the questions memorized • you can focus • you have to be flexible	**MA capabilities**
	• what I like about our department • really good	**MA satisfaction**
	Clinician - MA relationship (Trust) →+ TS --||→+ MA capabilities →+ MA satisfaction	
	In TS, the MD places trust in MAs to make some limited decisions about patient care. MAs gain the ability to make decisions gradually as they master simpler tasks by making mistakes, seeing similar situations and memorizing scripts. The MD delegating decision authority to the MA, and their improved performance over time makes MAs feel more satisfied about their job.	

<table>
<tr>
<th>Participant-Quote number) “Quote … phrases referring to variables”
(word count/total words in causal statement)</th>
<th colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</th>
</tr>
<tr>
<td rowspan="11">MA03-09) “In our department and in our team, the providers are very open and … always want to hear suggestions from the MAs, and kind of work with things that way. I know that in other departments the providers will go to the nursing supervisor, and say ‘This isn’t working. So and so is doing this.’ And it’s almost… that they are venting and they just need to vent. And then nothing gets taken care of. I really think that it depends on the provider and the MAs…
“If there is a problem, and it’s brought about in a positive, open way, I think that changes are made quickly… [Venting] just makes the problem worse. The way that our department, I love my department. I mean we all talk openly, if there is a problem, then we all talk about it and we try and figure out a way to solve it. And, I wish that all of the other departments could … (but) there is this… providers [up here] and then the MAs down here. And there, and they don’t really like to associate on that kind of level… Where I am at, if there is any issue then (the) MAs will talk about it first and figure out what is going on… think of some solutions and then bring it to the doctor… then we ... see if we can figure it out, and then we will bring in the nursing supervisors and say, ‘This is what we talked about, this is what our solution is, what do you think about it?’.
And then, it goes up from there, and if it’s approved or ok’d or anything, then it’s all good. And that’s how we do things, and I would like to see that throughout the rest of clinic <<laughter>>.”
(381/511)</td>
<td>• If there is any issue then … and then … and then … and then … and if</td>
<td>Time Delay</td>
</tr>
<tr>
<td>• they are venting
• it depends on the provider and the MAs
• there is this… providers [up here] and then the MAs down here. And there, and they don’t really like to associate on that kind of level</td>
<td>Clinician - MA relationship (Trust)</td>
</tr>
<tr>
<td>• suggestions from the MAs
• work with things
• nothing gets taken care of
• we all talk about it and we try and figure out a way to solve it
• then we ... see if we can figure it out
• it goes up from there</td>
<td>TS</td>
</tr>
<tr>
<td>• If there is a problem, and it’s brought about in a positive, open way</td>
<td>MA capabilities</td>
</tr>
<tr>
<td>• I love my department</td>
<td>MA satisfaction</td>
</tr>
<tr>
<td colspan="2">Clinician - MA relationship (Trust) →+ TS --||→+ MA capabilities →+ MA satisfaction</td>
</tr>
<tr>
<td>• then we ... see if we can figure it out</td>
<td>Visit time spent training MAs</td>
</tr>
<tr>
<td colspan="2">TS →+ Visit time spent training MAs</td>
</tr>
<tr>
<td colspan="2">TS is continuous process improvement. The capabilities development delay is not just due to the time needed for learning from experience. There is also an approval process for TS because new TS requires formal changes to the patient care process. Usually, it is the MA’s capabilities which earn the trust of MDs. In this case, MDs’ venting releases the pressure to address problems, which slows TS. Alternatively, MDs can hasten TS by demonstrating good management skill.
To achieve a high level of TS, a high level of trust in the MD/MA relationship must be established. The converse is also true, when there is not trust, TS falls away.</td>
</tr>
</table>

F.1.3 CLINICIAN CONCEPTUAL MODEL – SATURATION RESULTS

SMM2.x was also capable of exposing the variables and relationships in CL03's mental model of PCT. CL03 begins with the same relationship discussed by MA03: "Capacity →+ Task-shifting". In the HSDO, interview participants saw this link as where PCT begins. Quote 14 shows that CL03's understanding of PCT begins with this relationship and ends up with improving efficiency (both during the visit and during the day). Quote 15 describes how "Capacity" influences "Task-shifting": attempting task-shifting with a 1:1 ratio of MAs to clinicians would require clinicians to perform tasks which are the exclusive domain of MAs, like rooming patients. Therefore, when the care team ratio is 1:1, task-shifting fails. In quote 16, CL03 describes an early policy which illustrated the potential of task-shifting by shifting one key team task to a specialist MA. See Table F.6 below.

In quote 17, CL03 describes the challenge of keeping a schedule which can handle a combination of planned and unplanned visits. Providing high levels of access to visits for patients of both types is described as being very important to guaranteeing coordinated, comprehensive care. Not doing this is a source of worry for CL03. How to resolve it is unknown. In quote 18, CL03 describes the challenges of keeping a schedule which can handle a combination of in-visit and out-of-visit work. Out-of-visit work is necessary to ensure that in-visit work is as comprehensive as possible. For this issue, care team "Capacity" is perceived to be the main limiting factor. See Table F.7 below.

In quote 19, CL03 discusses the use of a pool of MAs instead of assigned teams. The "MA capabilities" needed for task-shifting can be specific to each clinician. A pool of generalist MAs then, by definition, has a lower level of capabilities. Shifting tasks in this context impacts clinicians because it increases errors. This can mean rework, more training or worse. This policy is similar to the one described in MA03-5, both are examples of "*accommodations*" used to address the issue of insufficient "Capacity".

In quote 20, the care team policy is introduced. For CL03, Capacity is greater than 1:1, but less than 2:1. And yet, it is enough to qualify as a "*huge improvement*" and "*a lot better*", because it allowed MAs time to develop the right set of capabilities to earn the clinician's trust. This does not mean that task-shifting has reached a maximum, only that an initial phase where trust is a limiting factor, has been passed.

In quote 21, CL03 describes what added Capacity could achieve, and juxtaposes this with the reality of current capacity. With more capacity, useful preparatory work could be done permitting the team to deliver high quality patient care and improved patient experience. In reality, demand exceeds capacity and MAs are often unable to meet the clinician's task-shifting expectations. CL03 then concludes this thought experiment by reiterating the importance of understanding these dynamics (e.g., time delay) for implementation to work.

In quote 22, CL03 returns to the subject of rework. Only this time, the context is in a care team (not in the pool). This context gives CL03 the expectation that the root cause of MA errors could be addressed. This cause is seen as a mismatch between the current level of "MA capabilities" and the expectations of "Task-shifting". Nevertheless, there is hope that somehow the team can work it out. CL03 is confident that, with better communication skills, the performance of care teams could improve. See Table F.8 below.

In quote 23, CL03 describes how making change through task-shifting seems to take more time than CL03 has available. In quote 24, CL03 describes how task-shifting involves openness on the team. See Table F.9 below.

Quote 23 makes clear that a clinician's opinion that the PCMH concepts are good (i.e., buy-in) is a necessary but insufficient condition to enable implementation. Clinicians also need scheduled, paid teamwork time and specific training for making the teams work. There is still difficult mental effort required. This work is described here as "*how to make it work*" "*making it work*" "*make it happen*" and "*trying to figure out how to add these things*". This is part of task-shifting and having a good level of "Clinician - MA Relationship (Trust)" – the idea that the MA and the clinician work together to figure out how best to divide responsibilities, to solve problems and build capabilities as a team.

In quote 24, CL03 indicates that the problem of an insufficient level of "*team approach*"[117] is at the root of problem-solving issues which make task-shifting so difficult. CL03 sees clinicians working in PCT as needing training in and support for using a "*communication* process" for PCT. This abstract concept in CL03's mental model of PCT is captured in SMM2.x in the elements of "Task-shifting", the capabilities development delay, "MA capabilities" and "Clinician - MA Relationship (Trust)".

[117] It is interesting to note at this stage that although CL03 sees this as a problem, MA03 perceives CL03 as having solved many of these problems on their team already, in comparison to other teams in the clinic.

Table F.6 Causal RIQs – CL03 "Capacity →+ Task-shifting"

<table>
<tr><th>Participant-Quote number) "Quote … <u>phrases referring to variables</u>"
(word count/total words in causal statement)</th><th colspan="2">Phrase needing interpretation = variable
Model concepts →+ Model concepts
Interpretation</th></tr>
<tr><td rowspan="7">CL03-14) "My sense is … my understanding was that (PCMH) was … an attempt to have a <u>team-based approach</u> to providing patient care and <u>having the medical assistants be more involved in streamlining the visit and participating in more of the visit and making things run more smoothly and efficiently</u>. And also… having patient-centered care. So that is my sense of … what the goal was." (65/84)</td><td>• team-based approach</td><td>Capacity</td></tr>
<tr><td>• having the medical assistants be more involved in streamlining the visit and participating in more of the visit and making things run more smoothly and efficiently</td><td>TS</td></tr>
<tr><td>0</td><td>Visit time spent training MAs</td></tr>
<tr><td>• streamlining the visit</td><td>Having sufficient time allotted for the visit</td></tr>
<tr><td>• things run more smoothly and efficiently</td><td>Visits on schedule</td></tr>
<tr><td colspan="2">Capacity →+ TS →+ Visit time spent training MAs →- Having sufficient time allotted for the visit →+ Visits on schedule</td></tr>
<tr><td colspan="2">As opposed to an MA to MD ratio of 1:1, a "team" consists of more MAs than MDs. PCMH is perceived as beginning with team-based care which uses TS to improve efficiency of the work and to create patient-centered care which means more effective primary care.</td></tr>
<tr><td rowspan="4">CL03-15) "I really <u>only had 1 MA working with me</u> so … the initial idea that that MA would room the patient, <u>then stay in the room and scribe for us, and then would, you know, draw the patient's labs and walk them to radiology</u>. But that wasn't really an option for me, just because of <u>staffing issues</u>. It seemed like if we did that, then I would never, the next patient would never be in a room." (77/102)</td><td>• only had 1 MA working with me
• staffing issues</td><td>Capacity</td></tr>
<tr><td>• then stay in the room and scribe for us and then would, you know, draw the patient's labs and walk them to radiology</td><td>TS</td></tr>
<tr><td colspan="2">Capacity →+ TS</td></tr>
<tr><td colspan="2">With a 1:1 ratio, the MD would need to perform tasks which belong exclusively to MAs, like rooming patients, in order to permit TS to occur. If Capacity = 1:1, then attempts to TS fail.</td></tr>
</table>

Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)	Phrase needing interpretation = **variable** Model concepts →+ Model concepts Interpretation	
CL03-16) “We had a … designated phone call returner and she would just … power through the phones calls... she was a very high-functioning (MA) … all I had to do … was review it … and (then) send it back. … I feel like there is a lot of work that I am doing that I probably don’t [need to, perhaps] somebody else could be doing.” (58/383)	• designated phone call returner	**Capacity**
	• all I had to do … was review it … and (then) send it back • work that I am doing that… somebody else could be doing	**TS**
	• high-functioning (MA) • all I had to do … was review it … and (then) send it back	**MA capabilities**
	0	**Clinician - MA relationship (Trust)**
	Capacity →+ TS --||→+ MA capabilities →+ Clinician - MA relationship (Trust) →+TS	
	A specialist MA is an addition to the team, which increases capacity and permits more TS. This policy reduces the length of the capabilities development delay, since it is just one task to master. Task mastery earns MD trust and thus increases the MD’s desire for more TS.	

Table F.7 Causal RIQs – CL03 Things that Impact “Visits with coordinated, comprehensive care”

<table>
<tr><th>Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)</th><th colspan="2">Phrase needing interpretation = variable
Model concepts →+ Model concepts
Interpretation</th></tr>
<tr><td rowspan="6">CL03-17) “I really, really, struggle [with having my patients get scheduled when they need to] … I feel like I never have available appointments for my patients and I’m constantly getting phone calls like ‘so and so just got out of the hospital and they need to be seen but you don’t have any appointments available, so can we double book them?’ … some days I have like X patients show up in the morning and then other days like all Y people come and it’s totally crazy … I could use more help figuring out a way to make my schedule more accessible. And maybe it’s just not possible … to accommodate everybody’s desire to be seen … [while] also making sure that people have [the appointments they need. For example,] for our diabetic patients, I really do want them to schedule an appointment in 3 months so that they don’t forget and fall through the cracks. ... So, I would say the Access thing, I just feel that my patients can’t get in to be seen when they want to be and I don’t really know how to address that issue” (68/84)</td><td>• I never have available appointments
• make my schedule more accessible
• my patients can’t get in to be seen when they want to be</td><td>Visits on schedule</td></tr>
<tr><td>• I really do want them to schedule an appointment in 3 months
• they don’t forget and fall through the cracks
• making sure that people have [the appointments they need]</td><td>Continuity of care (visit to visit)</td></tr>
<tr><td>• so and so just got out of the hospital
• they don’t forget and fall through the cracks
• making sure that people have [the appointments they need]</td><td>Visits with coordinated, comprehensive care</td></tr>
<tr><td>• double book
• some days I have like X patients show up in the morning and then other days like all Y people come and it’s totally crazy
• everybody’s desire to be seen
• making sure that people have [the appointments they need]</td><td>Volume of patient visits per hour</td></tr>
<tr><td colspan="2">Visits on schedule →+ Continuity of care (visit to visit) →+ Visits with coordinated, comprehensive care →- Volume of patient visits per hour</td></tr>
<tr><td colspan="2">To be accessible, the MD’s schedule needs to accommodate both planned and unplanned visits. The current strategy works for some days, but leaves the practice vulnerable to shocks; and leaves the clinician struggling to provide continuity for patients needing planned visits and comprehensiveness/coordination for all patients.</td></tr>
</table>

<table>
<tr><td>Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)</td><td colspan="2">Phrase needing interpretation = variable
Model concepts →+ Model concepts
Interpretation</td></tr>
<tr><td rowspan="5">CL03-18) “I don’t know if [the problem is] that the MAs are not using their time efficiently or if it’s just that we are asking them to do too many different things. And that it’s very hard to return phone calls when you’re simultaneously watching the schedule. I feel like … I’m at my desk and I’m trying to do a couple phone calls and I see that there’s 2 patients arrived and I can’t really complete the phone calls or [else] I don’t really have the time to spend talking to the patient. So, so, that’s an area where I do feel like if we had a designated person to just do phone calls for a few hours every day that the team, it would work better.” (58/383)</td><td>• MAs are not using their time efficiently or if it’s just that we are asking them to do too many different things
• if we had a designated person to just do phone calls</td><td>Capacity</td></tr>
<tr><td>• hard to return phone calls</td><td>Having sufficient time allotted for the visit</td></tr>
<tr><td>• I’m trying to do a couple phone calls
• talking to the patient</td><td>Visits with coordinated, comprehensive care</td></tr>
<tr><td colspan="2">Capacity --||→+ Having sufficient time allotted for the visit →+ Visits with coordinated, comprehensive care</td></tr>
<tr><td colspan="2">Both the MD and the MAs are overwhelmed by the volume of in and out-of-visit work. Because of a too low capacity on the care team, neither has the time they would like in order to prepare for and attend to each visit sufficient to provide the highest quality care.</td></tr>
</table>

Table F.8 Causal RIQs – CL03 "Task-shifting --||→+ MA Capabilities"

<table>
<tr><td>Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)</td><td colspan="2">Phrase needing interpretation = variable
Model concepts →+ Model concepts
Interpretation</td></tr>
<tr><td rowspan="10">CL03-19) "[So,] there was sort of a shift to try and have more of a team… we started trying to have huddles and team meetings. … It was challenging because … having different people rotate through and having a kind of a constantly changing pool of MAs was really, for me, it was hard. I felt like, (tone shows embarrassment) 'I have certain ways that I like doing things' … having to have a different person showing up who didn't know how to do that … every time they would … try to help, it ended up being [that] I would have to repeat all the [work] … because they just didn't know how to do it." (110/239)</td><td>• challenging
• for me it was hard</td><td>Time Delay</td></tr>
<tr><td>• a shift to try and have more of a team
• having different people rotate through and having a kind of a constantly changing pool of MAs</td><td>Capacity</td></tr>
<tr><td>• a shift to try and have more of a team
• have huddles and team meetings
• I have certain ways I like doing things
• try to help</td><td>TS</td></tr>
<tr><td>• who didn't know how to do that
• they just didn't know how</td><td>MA capabilities</td></tr>
<tr><td>• having to have a different person showing up
• they just didn't know how</td><td>Clinician - MA relationship (Trust)</td></tr>
<tr><td colspan="2">Capacity →+ TS --||→+ MA capabilities →+ Clinician - MA relationship (Trust) →+ TS</td></tr>
<tr><td>• have huddles and team meetings
• I would have to repeat all the work</td><td>Visit time spent training MAs</td></tr>
<tr><td colspan="2">TS →+ Visit time spent training MAs</td></tr>
<tr><td colspan="2">Initially, teamwork is new capacity and meetings. MDs figure out how TS might work in their practice, and give the MAs a chance to try and help out. With too short a time on the team, MAs are not given enough time to learn how to do the tasks the MD prefers and commit errors. These errors create a new task (rework) for the MD. Rework takes up the MD's time; causing frustration to the MD; and reducing the MD's willingness to TS. Rework is part of on-the-job training.</td></tr>
</table>

<table>
<tr>
<th>Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)</th>
<th colspan="2">Phrase needing interpretation = variable
Model concepts →+ Model concepts
Interpretation</th>
</tr>
<tr>
<td rowspan="7">CL03-20) “There [are] at least 4 MAs … rotat[ing] through … that actually has been a huge improvement, because we all sort of work together as a team and they kind of know what [to do]… there are 2 or 3 providers that they work with, but I think it’s, we all have fairly consistent things that we want done certain ways and I think they’re able to accommodate that. So, I think that’s gotten a lot better.
And some of the MAs who really had no interest in [our department] I think the idea of forcing them to do it has been given up which is much better. Because it was just really challenging for everyone.” (115/125)</td>
<td>• 4 MAs rotating through
• 2 or 3 providers</td>
<td>Capacity</td>
</tr>
<tr>
<td>• we all sort of work together as a team
• fairly consistent things that we want done</td>
<td>TS</td>
</tr>
<tr>
<td>• they kind of know what [to do]
• they’re able to accommodate</td>
<td>MA capabilities</td>
</tr>
<tr>
<td>• gotten a lot better
• the idea of forcing them to do it has been given up which is much better</td>
<td>Clinician - MA relationship (Trust)</td>
</tr>
<tr>
<td colspan="2">Capacity →+ TS --||→+ MA capabilities →+ Clinician - MA relationship (Trust) →+ TS</td>
</tr>
<tr>
<td colspan="2">TS is working better because MAs were allowed to choose a department, dividing the pool into separated teams. This choice improved the clinician - MA relationships in the department and on this team. By assigning them to one team, MAs can learn the TS tasks these clinicians want done, reinforcing the relationship of trust.</td>
</tr>
</table>

<table>
<tr><th>**Participant-Quote number)** “Quote … phrases referring to variables” (word count/total words in causal statement)</th><th colspan="2">Phrase needing interpretation = **variable**
Model concepts →+ Model concepts
Interpretation</th></tr>
<tr><td rowspan="10">**CL03-21)** “The potential is there to … make a huge impact on how we take care of our patients… in terms of the quality measures. … If we had a designated person … going through and calling every single patient with diabetes … to come in and get their labs … and reminding … going through my schedule … and [saying] ‘these 5 patients need [x]’ and giving the MA a little note saying, ‘could you please order these [tests]’… That’s what the MAs are supposed to be doing but they are doing so many other things when they are going in there, when that … reminder pops up, they’re just like “oh, ok [the doctor] will do it.”…
“It would make a huge difference in quality and I think our patients would feel more satisfied … it just makes a difference when people are working together and are focused on your needs and someone is there to follow up with you and call you back and walk you through the clinic and show you where you should go. Instead of just sort of sending you off.
“So I think that all of those things are super important and I think it’s a great concept and it’s actually how to make it work that is challenging.” (215/294)</td><td>• it’s actually how to make it work that is challenging</td><td>**Time Delay**</td></tr>
<tr><td>• a designated person
• they are doing so many other things</td><td>**Capacity**</td></tr>
<tr><td>• going through and calling every single patient with diabetes … to come in and get their labs
• reminding… going through my schedule… [saying] ‘these 5 patients need [x]’
• giving the MA a little note saying, ‘could you please order these [tests]’
• they are doing so many other things when they are going in there
• what the MAs are supposed to be doing
• when people are working together
• make it work</td><td>**TS**</td></tr>
<tr><td>0</td><td>**Visit time spent training MAs**</td></tr>
<tr><td>• what the MAs are supposed to be doing but they are doing so many other things</td><td>**Having sufficient time allotted for the visit**</td></tr>
<tr><td>0</td><td>**Visits on schedule**</td></tr>
<tr><td>• to follow up with you and call you back</td><td>**Continuity of care (visit to visit)**</td></tr>
<tr><td>• going through and calling
• reminding … going through my schedule … and [saying] ‘these 5 patients need [x]’ … ‘could you please order these [tests]
• quality
• focused on your needs</td><td>**Visits with coordinated, comprehensive care**</td></tr>
<tr><td colspan="2">**Capacity →+ TS →+ time spent training MAs→- Having sufficient time allotted for the visit →+ Visits on schedule →+ Continuity of care (visit to visit) →+ Visits with coordinated, comprehensive care**</td></tr>
<tr><td colspan="2">If the team had more capacity, then MAs would have the time to do the necessary preparatory work for the visits. This would improve the clinician’s ability to provide the best care possible and the team’s ability to provide the best experience possible.</td></tr>
</table>

<table>
<tr><th>Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)</th><th colspan="2">Phrase needing interpretation = variable
Model concepts →+ Model concepts
Interpretation</th></tr>
<tr><td rowspan="3"></td><td>• they’re just like “oh, ok (the Dr.) will do it”</td><td>Clinician - MA relationship (Trust)</td></tr>
<tr><td colspan="2">Clinician - MA relationship (Trust) →+ TS</td></tr>
<tr><td colspan="2">But, at current capacity, MAs do not have time to perform TS as expected. This situation limits the growth in TS. Overworked MAs are less productive and push some TS work back to the MD, thus signaling that no new TS is possible.</td></tr>
<tr><td rowspan="6">CL03-22) “I just struggle. I just find it very hard. There is a lot of things that I feel like we could be doing better … [For example, in] meetings … I say what I want to have happen and the MAs nod but it doesn’t necessarily consistently get implemented.
“I feel like we need more training to make it a team approach. … I don’t really know if everybody is very honest about giving their feedback, about what’s working and what’s not working. And I kind of get the sense that I’m this ‘authority figure’…
“There are times …they just don’t do x-files, and it’s … painful to go back over and over again and say, ‘I really need you to do x-files, it saves me so much time with charting and please do x-files’. … I think … the patient just says ‘oh, I’m just here for med re-fills’ but I wish that the MAs would be a little more pro-active … [if] they have diabetes, I’m probably going to be talking … about that stuff…
“I get a little bit [frustrated]. It’s just hard. You know, that’s one part of it where I wish, I feel like, we could use more, I don’t know, <<sigh>> just figuring about a better way to have it really be a team-generated set of goals, instead of me just saying ‘I want you guys to do things this way’ and then having them sort of try to do it and then … after a week or two [it] slips again. (257/294)</td><td>• I just struggle. I just find it very hard. There is a lot of things that I feel like we could be doing better
• I get a little bit frustrated
• It’s just hard</td><td>Time Delay</td></tr>
<tr><td>• in meetings… I say what I want to have happen and the MAs nod
• what’s working and what’s not working
• it saves me so much time with charting
• ‘I really need you to do x-files…’</td><td>TS</td></tr>
<tr><td>• doesn’t necessarily consistently get implemented
• having them sort of try to do it and then … after a week or two it slips again</td><td>MA capabilities</td></tr>
<tr><td>• to make it a team approach
• I don’t really know if everybody is very honest
• feedback
• authority figure
• I wish that the MAs would be a little more pro-active
• to have it really be a team-generated set of goals</td><td>Clinician - MA relationship (Trust)</td></tr>
<tr><td colspan="2">TS --||→+ MA capabilities →+ Clinician - MA relationship (Trust) →+ TS</td></tr>
<tr><td colspan="2">The decision of how to increase TS includes work in team meetings. When TS choices are made that go beyond what MAs are capable of doing, it generates two types of tasks: 1) rework task for the MD and 2) a new item on the TS meeting agenda for MDs and MAs. This results from the capabilities development delay and can be frustrating for the MD. When the team relationship is such that MAs do not fully participate in meetings, it limits the team’s ability address TS problems.</td></tr>
</table>

Participant-Quote number) "Quote … <u>phrases referring to variables</u>" (word count/total words in causal statement)	Phrase needing interpretation = **variable** Model concepts →+ Model concepts Interpretation	
	• it's … painful to go back over and over again	**Visit time spent training MAs**
	• it saves me so much time with charting	**Having sufficient time allotted for the visit**
	TS →+ Visit time spent training MAs→- Having sufficient time allotted for the visit	
	While frustrating, rework is a significant portion of on-the-job training, and an important indicator that capabilities remain insufficient for a given task. When TS is implemented, it makes the MD more efficient.	

Table F.9 Causal RIQs – CL03 on Management Training for PCT

<table>
<tr><th>Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)</th><th colspan="2">Phrase needing interpretation = variable
Model concepts →+ Model concepts
Interpretation</th></tr>
<tr><td rowspan="10">CL03-23) "I just have a general feeling that we haven't had enough time, or focus, or energy put into 'how to actually implement this on the ground.' … I really believe in a lot of the concepts. Maybe I've become a little bit cynical about actually making it work … just getting through the day and seeing our patients is so much work already and then trying to figure out how to add these things…
"If we were doing them, it would make our lives a lot more efficient but it just, I think in order to make it happen you need to have a concentrated effort, and training, and have some administrative support and I just, I'm just not sure that we've really had that." (78/102)</td><td>• haven't had enough time
• I've become a little bit cynical
• concentrated effort
• trying to figure out how to add these things</td><td>Time Delay</td></tr>
<tr><td>• actually making it work
• trying to figure out how to add these things
• doing [these things] … would make our lives a lot more efficient
• in order to make it happen</td><td>TS</td></tr>
<tr><td>0</td><td>MA capabilities</td></tr>
<tr><td colspan="2">TS --||→+ MA capabilities</td></tr>
<tr><td>• trying to figure out how to add these things</td><td>Visit time spent training MAs</td></tr>
<tr><td>0</td><td>Having sufficient time allotted for the visit</td></tr>
<tr><td>• just getting through the day and seeing our patients is so much work already
• make our lives a lot more efficient</td><td>Visits on schedule</td></tr>
<tr><td colspan="2">TS →+ Visit time spent training MAs →- Having sufficient time allotted for the visit →+ Visits on schedule →+TS</td></tr>
<tr><td colspan="2">Making TS work requires time and tests MDs' patience. TS is not a machine to turn off and on. It has to be made to work. This requires the care team's mental effort in 1) thinking through which processes to improve using TS, 2) implementing the change and then 3) sustaining it into the future. Taking the time to make TS work trains the MAs so that they are capable of performing well in the new tasks. Having the workload on schedule is a pre-requisite for spending the time necessary to make TS work.</td></tr>
</table>

<table>
<tr><th>Participant-Quote number) “Quote … <u>phrases referring to variables</u>” (word count/total words in causal statement)</th><th colspan="2">Phrase needing interpretation = variable
Model concepts →+ Model concepts
Interpretation</th></tr>
<tr><td rowspan="10">CL03-24) “<u>More and more</u> we need to have better ways to have <u>really open communication as a team</u> and to have each person in the team <u>feel really accountable</u> and get rewarded when they do well and <u>get really appropriate feedback</u> when things don’t go well.
“Is there a way that you think of [doing] that, do you have ideas for how that can be done? I don’t know. I mean, yeah, I really don’t know. I mean, I think some sort of <<long pause>> yeah, I’m not really sure.
“Some sort of <u>communication process</u>, I mean the problem is that we’re supposed to have <u>huddles and … team meetings</u> and <u>they are time consuming. I never have a lunch break as it is [so it’s hard] to try and find the time</u> … I guess I’m not really sure. I think just having some sort of <u>exercises where people really openly say what they think is working and isn’t working</u>, and <u>how they want</u>, like <u>setting personal goals</u> for what they want to do, and <u>how they want to make things better</u>, and then <u>having someone sort of help and support them with those goals, and doing that all together as a team</u>.
“And I say that in theory, but when it actually comes down to it, <u>it just feels like a lot more work to do [than] that</u>.
But I do think that that would make things better.”
(78/102)</td><td>• more and more
• it just feels like a lot more work to do than that</td><td>Time Delay</td></tr>
<tr><td>• huddles and … team meetings
• exercises where people really openly say what they think is working and isn’t working
• how they want to make things better</td><td>TS</td></tr>
<tr><td>0</td><td>MA capabilities</td></tr>
<tr><td>• really open communication as a team
• feel really accountable
• get really appropriate feedback
• communication process
• really openly say what they think
• setting personal goals
• having someone sort of help and support them with those goals, and doing that all together as a team</td><td>Clinician - MA relationship (Trust)</td></tr>
<tr><td colspan="2">TS --||→+ MA capabilities →+ Clinician - MA relationship (Trust) →+ TS</td></tr>
<tr><td>• huddles and … team meetings
• having someone sort of help and support them with those goals, and doing that all together as a team</td><td>Visit time spent training MAs</td></tr>
<tr><td>• they are time consuming. I never have a lunch break as it is [so it’s hard] to try and find the time</td><td>Having sufficient time allotted for the visit</td></tr>
<tr><td>• they are time consuming. I never have a lunch break as it is [so it’s hard] to try and find the time</td><td>Visits on schedule</td></tr>
<tr><td colspan="2">TS →+ Visit time spent training MAs →- Having sufficient time allotted for the visit →+ Visits on schedule →+ TS</td></tr>
<tr><td colspan="2">After an initial TS trial period is passed, TS gets progressively harder to do.
Growing TS always requires the MD to have trust. This becomes more difficult over time, as a high level of TS requires MAs to also have a high level of trust (accountability, open communication and correction) in their clinician.
TS requires time, which can be considered as training for MAs. When this work takes time away from visits it disrupts the daily schedule, thereby reducing the amount of TS.</td></tr>
</table>

F.1.4 MANAGER CONCEPTUAL MODEL – SATURATION RESULTS

SMM2.x was capable of exposing the elements in the mental models of NM01 and CM01. These two managers primarily contributed to overall model development by adding depth to the concepts surrounding "Capacity". CM01 and NM01 both see "*appropriate staffing*" for PCMH as the correct focus of their attention. This section presents this understanding, beginning with quotations from the nurse manager.

F.1.4.1 MANAGERS: NURSE MANAGER

NM01 observes care teams up close, working constantly with clinicians and MAs to address problems. In quotes 32 and 33, NM01 describes PCT as being fraught with problems surrounding MA capabilities. Quote 32 describes a situation where on-the-job training was used to develop MA capabilities in order to enable all of the MAs in the clinic to implement task-shifting for the specific task of drawing patients' blood for lab tests. In quote 33, we learn that this success is the exception rather than the rule. In most situations, there is a wide variation in MA capabilities, which makes it difficult for teams to develop the trust which is required to succeed in task-shifting. See Table F.10 below.

In quote 34, NM01 refers to a situation which was common under the 1:1 care team model. One clinician's MA could be overwhelmed by demand at the same time that nearby clinicians' MAs were idle, available to help, but unable to do so. Team separation meant this fact could not even be communicated. Fluctuations in the level of demand mean that MAs would take turns idly watching each other struggle. A care team model which allows MAs to work for more than one MD utilizes this excess capacity to even out these fluctuations. MAs enjoy this working environment much more because of this cooperation and the more stable workload.

In quote 35, NM01 describes the links between "MA capabilities" and "capacity". A team's capacity is more than just the number of team members, it is also a function of those MAs' capabilities. NM01 assigns a zero contribution to capacity from each "*low functioning*" MA. The same end result could be had by simply averaging capabilities across a team.

Quote 36 describes "Capacity" by adding the consideration of PCMH's purpose. This is viewed by NM01 as ensuring efficiency in terms of "Visits on schedule". Poor MA attendance ("MA Retention") reduces this efficiency as it decreases "Capacity", limiting the ability to stay on schedule. This also decreases "MA satisfaction" as people are more stressed trying to keep up. See Table F.11 below.

Table F.10 Causal RIQs – NM01 on Developing Capabilities

Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable →+ Model Variable** (→ = causal link, --||→+ = with delay, +/- = link polarity) Interpretation	
NM01-32) "Well one of the changes that … [we implemented] a long time ago [was to get] the MAs to draw their own labs… and it took a while for everybody to catch up and to learn how to draw blood and how to be good at it. But it worked really well. Now the patients… don't need to go … to the labs to wait." (65/176)	• a long time ago • it took a while • to learn • but it worked really well	**Time Delay**
	• get the MAs to draw their own labs	**TS**
	• how to draw blood • and how to be good at it	**MA capabilities**
	TS --||→+ MA capabilities	
	• for everybody to catch up • to learn	**Visit time spent training MAs**
	TS→+ Visit time spent training MAs	
	MAs were able to all develop an equal level of capabilities for one TS task, after a time delay. This was accomplished through on-the job training for MAs.	

<table>
<tr><td>Participant-Quote number) “Quote … <u>phrases referring to variables</u>”
(word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="9">NM01-33 “<u>In the beginning</u> … <u>things slip through the cracks</u> because of <u>ownership problems</u>. [In a 1:1 team], <u>you are feeling that ownership</u>. Where there is <u>a team [>1 MAs]</u>, there is not necessarily one person that [the provider can] go to.
“And that is probably one thing that is difficult … on the teams we have [MAs who are] <u>high performers, medium and low performers</u>. <u>It’s kind of hard to get them all functioning at the same speed</u>. The <u>high performers</u> are the ones that <u>everybody goes to for everything</u> and then the <u>low performers</u>; they kind of <u>don’t do as much</u> … <u>When the doctor … wants</u> to make sure <u>[something] gets done</u>, then <u>they always go to</u> the <u>high performers</u>.” (119/158)</td><td>• it’s kind of hard to get them all functioning at the same speed
• In the beginning</td><td>Time Delay</td></tr>
<tr><td>• things slip through the cracks
• high performers, medium and low performers
• get them all functioning at the same speed</td><td>MA capabilities</td></tr>
<tr><td>• ownership problems
• you are feeling that ownership
• when the doctor wants
• they always go to</td><td>Clinician - MA relationship (Trust)</td></tr>
<tr><td>• everybody goes to for everything
• don’t do as much
• something gets done
• they always go to</td><td>TS</td></tr>
<tr><td colspan="2">MA capabilities →+ Clinician - MA relationship (Trust) →+ TS --||→+ MA capabilities</td></tr>
<tr><td>• a team [>1 MAs]</td><td>Capacity</td></tr>
<tr><td>• to get them all functioning at the same speed</td><td>Visit time spent training MAs</td></tr>
<tr><td colspan="2">Capacity →+ TS→+ Visit time spent training MAs</td></tr>
<tr><td colspan="2">In order for the MD to trust MAs such that they will shift tasks to them, they need to see the MAs as being equally and sufficiently capable to perform those tasks. A low level of MA capabilities negatively impacts trust, and therefore TS, because of errors. On-the-job training, while effective, is seen as difficult.</td></tr>
</table>

Table F.11 Causal RIQs – NM01 on MA Maintaining Capacity

Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable →+ Model Variable** (→ = causal link, --||→+ = with delay, +/- = link polarity) Interpretation	
NM01-34) "I haven't really seen [MA turnover because of PCMH]. I think that people really like that, working mostly with the docs <<sic>> and the MAs, they really like it. They like having more help. The MAs like having the team concept, being able to communicate with each other, getting help when they need [it]. Before it was like, every man for themselves *<<sic>>*. They had their doctor and they had to do what their doctor wanted and they didn't have a lot of time to move over and help somebody else. But now, they are more free to help each other. So now it is more like a team than it used to be." (112/122)	• having more help • the team concept • getting help • every man for themselves • they had their doctor • move over and help somebody else • more free to help each other • more like a team	**Capacity**
	• people really like • they really like it • MAs like	**MA satisfaction**
	• I haven't really seen MA turnover because of PCMH	**MA retention**
	Capacity --||→+ MA satisfaction --||→+ MA retention	
	The "team concept" consists of having MAs working for more than one MD, so at least a 2:2 ratio. Thus MAs have help in that 8the workload is spread across multiple MAs and there is the opportunity for communicating and meeting each other's needs.	
NM01-35) "Let's say you have 6 MAs, and 3 doctors, and if 2 of the MAs are more low functioning then you are really only basically working with 4 MAs. So there is a breakdown of the team when that happens. "Or if someone calls in sick or if they go home early … that messes up the Care Team model… you are almost back to the old way again, [of] when you don't have the staffing." (76/119)	• low functioning	**MA capabilities**
	• that messes up • a breakdown of the team	**MA satisfaction**
	• if someone calls in sick or if they go home early	**MA retention**
	• that messes up • a breakdown of the team • 6 MAs, and 3 doctors • only basically working with 4 • the staffing	**Capacity**
	• breakdown of the team • messes up the Care team model … almost back to the old way	**TS**
	MA capabilities →+ MA satisfaction --||→+ MA retention →+ Capacity →+ TS	
	A low level of average MA capabilities on a team, and/or absenteeism, can reduce effective capacity. This causes MAs to get overworked, so TS and MA satisfaction both decline.	

<table>
<tr><th>Participant-Quote number) “Quote … <u>phrases referring to variables</u>” (word count/total words in causal statement)</th><th colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</th></tr>
<tr><td rowspan="8">NM01-36) “[On a] good day … <u>nobody has called in sick</u>, so we know that our <u>teams are all full</u> … [and] <u>people’s attitudes [are positive]</u>. <u>If they are liking their jobs</u>, if <u>things are flowing well</u>, if doctors’ <u>patients are getting roomed [in a] timely [manner]</u>. I think that when our real <u>contention</u> comes is <u>when people get behind</u>. And so, <u>if we can keep up with the patients, and keep one step ahead of the doctors</u>, <u>that day flows a lot better</u>.” (83/91)</td><td>• nobody has called in sick</td><td>MA retention</td></tr>
<tr><td>• teams are all full
• when people get behind</td><td>Capacity</td></tr>
<tr><td>• people’s attitudes [are positive]
• if they are liking their jobs
• contention</td><td>MA satisfaction</td></tr>
<tr><td colspan="2">MA retention →+ Capacity --||→+ MA satisfaction</td></tr>
<tr><td>• patients are getting roomed in a timely manner
• if we can keep up with the patients, and keep one step ahead of the doctors</td><td>Having sufficient allotted for the visit</td></tr>
<tr><td>• when people get behind
• things are flowing well
• that day flows a lot better</td><td>Visits on schedule</td></tr>
<tr><td colspan="2">Capacity --||→+ Having sufficient allotted for the visit →+ Visits on schedule</td></tr>
<tr><td colspan="2">Poor MA attendance at work can function like turnover- by limiting the level of effective capacity. When all MAs attend work as scheduled, then team capacity is sufficient to prevent being overworked – this contributes to MA job satisfaction.
Having sufficient capacity also permits teams to keep up with the day’s visits.</td></tr>
</table>

F.1.4.2 MANAGERS: CENTER MANAGER

SMM2.x does a good job of exposing CM01's mental model. In addition to providing causal information, the center manager also provides a large amount of conceptual information (as shown in Table F.1 above). SMM2.x exposes the variables which CM01 describes and the causal structure which CM01 sometimes avoids linking. Grey text is used to mark portions of SMM2.x that CM01 did not explicitly identify but that appear to be taken for granted.

In quote 37, CM01 emphasizes that PCT is more than increasing the capacity of the care team; with that increased capacity there is the expectation for the delivery of higher quality care to patients. In quote 38, CM01 juxtaposes teamwork in PCMH and traditional primary care. CM01 also explains clinicians' authority for deciding how much to engage MAs in task-shifting, which contributes to variations in task-shifting across departments. In quote 39, CM01 points to an important problem involving "Capacity" and patients' wait times. Without a sustained, sufficient level of effective capacity, PCMH implementation seems to cause longer wait times.

In quote 40, CM01 describes challenges MAs face. The change in workload from "Task-shifting" is felt as an increase in workload (decreasing "Capacity"), rather than a shifting of the workload to new tasks. This causes MAs to experience stress, and turnover. While quote 40 proposes a new link "Task-shifting →- Capacity", the causal links in SMM2.x clarify that this link operates through a longer causal chain, which reflects the negative polarity this quote references. See Table F.12 below.

In quote 41, CM01 describes how MA absenteeism ("MA retention") negatively impacts "Capacity", causing task-shifting to break down. CM01 also identifies a second reason for the breakdown in task-shifting: clinicians only pursue it if the care team is/remains "efficient". CM01 believes that clinicians' tendency to base the level of "Task-shifting" on "Capacity" or efficiency is short-sighted and fails to recognize "*different solutions*".

In quote 42, CM01 describes how having sufficient capacity on a team allows clinicians to give each visit a sufficient amount of time. Again, CM01 emphasizes that "MA retention" issues strain this capacity. CM01 also recognizes that PCMH has many aspects that need to be learned, and some are harder than others. The time it takes to learn how to successfully implement these aspects is the "Capacity --||→+ having sufficient time allotted for the visit" delay in SMM2.x. See Table F.13 below.

In quote 43, CM01 indicates that incentive pay impacts PCT. In this case, incentive pay was based on the number of visits seen. A clinician left the HSDO because of the combination of the difficulty of implementation and the incentive pay policy that, together, impeded that clinician's ability to reach salary goals. This quote also introduces several concepts that are not found in SMM2.x: clinician capabilities, clinician satisfaction and clinician turnover. These variables are considered indirectly in SMM2.x. Clinician capabilities include their ability to balance "Having sufficient time allotted for the visit" and "Volume of patient visits" can vary, as well as their skill as a teacher in task-shifting, which requires simultaneously training MAs, fostering teamwork and perceiving "MA Capabilities". In the quote, the lack of clinician capabilities, caused failure to implement PCT efficiently. This led to clinician dissatisfaction and turnover since the clinician was unable to minimize the "Difference between desired and actual pay".

Including these variables would require changing the problem statement which is focused on success and failure of those clinicians who stay in the system. The problem statement was not changed (see discussion for related thoughts on future research). These concepts are taken into account in SMM2.x, to the extent that they influence clinicians who stay. SMM2.x shows what is required for clinicians to successfully navigate these tensions, keeping their pay at an acceptable level, their satisfaction high and avoiding turnover: either they figure out how to successfully implement by working through the delays or they reduce the level of task-shifting. The simulation model also treats these specific issues indirectly, and does so in a more comprehensive manner (see Section 3.5.5.2.2).

In quote 44, CM01 explains that when the clinic is busier, it also generates more revenue. See Table F.14 below.

Quotes 45 and 46 discuss pre-visit planning and planned care – for example, having test done prior and results available during the visit. These activities lead to more comprehensive care for the patient. See Table F.15 below.

Table F.12 Causal RIQs – CM01 on "Capacity → X"

<table>
<tr><th>**Participant-Quote number)** "Quote … phrases referring to variables" (word count/total words in causal statement)</th><th colspan="2">• Phrase(s) needing interpretation = **Model Variable**
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</th></tr>
<tr><td rowspan="9">**CM01-37)** "One of the key components of [PCMH] is that we never leave the patient alone. The patient is followed through. The hand-offs should seem seem-less. … We design the practice and the structure of how that patient receives care … to increase quality, to increase efficiency… There's a lot of different components … a lot of people just feel that "it's just 2 MAs per provider"… it's really about, in my impression … delivering services to the patient and having specific ways of doing that to increase efficiency and increase the quality of care for the patient and it is also to give the patient a better experience. And by doing all of those things it would also impact our financial[s], because we would retain the loyalty of the patient coming back." (132/296)</td><td>• 2 MAs per provider</td><td>**Capacity**</td></tr>
<tr><td>• the patient is followed through
• hand-offs
• the structure of how that patient receives care
• having specific ways of doing that</td><td>**TS**</td></tr>
<tr><td>• efficiency</td><td>**Having sufficient time allotted for the visit**</td></tr>
<tr><td>• efficiency</td><td>**Visits on schedule**</td></tr>
<tr><td colspan="2">**Capacity →+ TS →+ Visit time spent training MAs →- Having sufficient time allotted for the visit →+ Visits on schedule**</td></tr>
<tr><td>• quality
• quality of care</td><td>**Visits with coordinated, comprehensive care**</td></tr>
<tr><td colspan="2">**Having sufficient time allotted for the visit →+ Visits with coordinated, comprehensive care**</td></tr>
<tr><td colspan="2">In its most basic form, PCMH adds to clinicians' capacity to deliver services to patients. This added capacity is allocated through the purposeful application of TS toward two goals: improving efficiency and improving quality.</td></tr>
</table>

<table>
<tr><td>Participant-Quote number) “Quote … phrases referring to variables”
(word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="7">CM01-38) “There have been changes … since the start of [PCMH], definitely… The huddles make a big difference [by helping] the communication. [It] has … changed …to ‘what staff do we have?’ and ‘do we need to work on doing a better job doing this, this and this?’ Versus a doctor, and a MA sitting down and saying ‘ok, we will need more time for these’ and ‘let’s get these people in for labs next time.’
…
“It is [not] consistently happening the same at each … department. They are all at different levels … and the MAs have different roles in each department as well … the expectations are different…
“[So,] the amount of workload is not the same … Providers have a big piece to that… We [also] don’t have the full team staffing, talking about 2 MAs to each provider for [some departments] because they hadn’t typically wanted to do that type of a model … they do little pieces of that model … [while other] MAs tend to be a lot busier, doing just different things.” (179/366)</td><td>• the huddles make a big difference [by helping] the communication
• expectations
• providers have a big piece
• they hadn’t typically wanted to</td><td>Clinician - MA relationship (Trust)</td></tr>
<tr><td>• the huddles
• we need to work on doing a better job doing this, this and this
• ok, we will need more time for these
• let’s get these people in for labs next time
• [not] consistently happening the same at each … department. They are all at different levels
• roles
• expectations
• the amount of workload
• do little pieces of that model
• busier, doing just different things</td><td>TS</td></tr>
<tr><td>• the huddles
• doing a better job doing this, this and this</td><td>Visit time spent training MAs</td></tr>
<tr><td colspan="2">Clinician - MA relationship (Trust) →+ TS →+ Visit time spent training MAs</td></tr>
<tr><td>• staff
• a doctor and a MA
• staffing
• 2 MAs to each provider</td><td>Capacity</td></tr>
<tr><td colspan="2">Capacity →+ TS</td></tr>
<tr><td colspan="2">TS involves an increasingly close working relationship and an increasingly high degree of shared responsibility on a scale where, at zero, one MA performs scheduling tasks for one MD and, at a high level, MDs collaborate with MAs on delivering services, process improvement and managing the team. At zero, MAs are told what to do, at a high level, MAs are expected to provide solutions to service delivery problems. To get from zero to the high level, clinicians spend time training MAs. Clinicians choose how TS is done, by defining roles and setting expectations.</td></tr>
</table>

<table>
<tr><th>Participant-Quote number) “Quote … phrases referring to variables”
(word count/total words in causal statement)</th><th colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</th></tr>
<tr><td rowspan="7">CM01-39) “Features of [PCMH] that are … not working well … wait times, they are a problem … the staffing ratio, it seems to be a major barrier for doing [PCMH] consistently. When we have perfect staffing, and the staffing ratios are there, then it’s a lot easier to get it done, and have it flow well.” (56/287)</td><td>• The staffing ratio
• perfect staffing
• the staffing ratios are there</td><td>Capacity</td></tr>
<tr><td>• doing PCMH consistently</td><td>TS</td></tr>
<tr><td>0</td><td>Visit time spent training MAs</td></tr>
<tr><td>• it’s a lot easier to get it done, and have it flow well</td><td>Having sufficient time allotted for the visit</td></tr>
<tr><td>• wait times
• it’s a lot easier to get it done, and have it flow well</td><td>Visits on schedule</td></tr>
<tr><td colspan="2">Capacity --||→+ TS →+ Visit time spent training MAs →- Having sufficient time allotted for the visit →+ Visits on schedule</td></tr>
<tr><td colspan="2">When effective capacity is at its ideal level, then it is easier to have the practice operate efficiently while implementing TS. The difficulty in achieving and sustaining sufficient effective capacity is an important issue for PCT.</td></tr>
</table>

<table>
<tr><td>Participant-Quote number) “Quote … <u>phrases referring to variables</u>” (word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="13">CM01-40) “So, <u>stress</u>, definitely has increased. I don’t know if we are talking about the disadvantages. <u>The MA workload</u>, that we have expected the MAs to do, has increased quite a bit, and <u>it causes</u> a little bit more <u>burnout</u>. They <u>used to just room a patient and be done, now they are much more involved in the care</u>. Some MAs really, <u>really enjoy</u> … having <u>that kind of involvement</u>. And some don’t <u>and then end up leaving</u>.
“Overall in general, <u>we are asking them to do quite a bit, plus check phone messages, check this, plus this, plus this, plus this</u>. <u>Another thing here and there</u> and then <u>eventually</u> they just say, ‘<u>I don’t have enough time in the day</u>. <u>I don’t feel like going to work</u>.’ So they <u>call in sick</u>.” (58/140)</td><td>• and then end up
• eventually</td><td>Time Delay</td></tr>
<tr><td>• the MA workload
• I don’t have enough time in the day</td><td>Capacity</td></tr>
<tr><td>• stress
• burnout
• really enjoy
• I don’t feel like going to work</td><td>MA satisfaction</td></tr>
<tr><td>• end up leaving
• call in sick</td><td>MA retention</td></tr>
<tr><td colspan="2">Capacity --||→+ MA satisfaction --||→+ MA retention</td></tr>
<tr><td colspan="2">Increased workload is felt as reduced capacity. This negatively impacts MAs’ sense of job satisfaction which in turn influences their decision of how much work to supply the HSDO, and of whether or not to leave the job.</td></tr>
<tr><td>• used to just room a patient and be done, now they are much more involved in the care
• that kind of involvement
• we are asking them to do quite a bit, plus check phone messages, check this, plus this, plus this, plus this</td><td>TS</td></tr>
<tr><td>0</td><td>Visit time spent training MAs</td></tr>
<tr><td>0</td><td>Having sufficient time allotted for the visit</td></tr>
<tr><td>0</td><td>Visits with coordinated, comprehensive care</td></tr>
<tr><td>0</td><td>Volume of patient visits per hour</td></tr>
<tr><td colspan="2">TS →+ Visit time spent training MAs →- Having sufficient time allotted for the visit →+ Visits with coordinated, comprehensive care →- Volume of patient visits per hour→- Capacity</td></tr>
<tr><td colspan="2">In this quotation, the element of TS is clearly identified and described as different from standard practice. A high level of TS will mean that each MA’s workload is heavier than it would have been otherwise but the causal links are less clear.</td></tr>
</table>

Table F.13 Causal RIQs – CM01 on "X → Capacity"

<table>
<tr><td>Participant-Quote number) "Quote … phrases referring to variables"
(word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="9">CM01-41) "[Clinicians] don't have any resistance for those ideas that will make them more efficient, … before we had [2:1 we had 1:1] … So, when they [hear about PCMH], the first thing they … [say is] 'Oh, I get 2 MAs, I can do so much with that!' Well, now it's … 'I don't have full Care Team staffing, I need to have 2 MAs to run the Care Team the way it should be run.' Reality is, staff are tough to control that way. You have people that call in sick, they have things that come up, they have vacation, and sometimes you can't find replacements. Problem is the norm seems to be there are always staffing issues.
"So the excuse is that 'if we don't have the staffing necessary then we can't efficiently run the Care Team, so we revert back to the previous way of doing things,' rather than trying to find different solutions."
(172/211)</td><td>• staffing issues
• people that call in sick, they have things that come up, they have vacation, and sometimes you can't find replacements</td><td>MA retention</td></tr>
<tr><td>• before we had [2:1 we had 1:1]
• I get 2 MAs
• I can do so much with that
• full Care Team staffing
• 2 MAs
• staffing issues
• staffing necessary</td><td>Capacity</td></tr>
<tr><td>• Clinicians don't have any resistance
• I can do so much
• to run the Care Team the way it should be run
• run the Care Team
• previous way of doing things</td><td>TS</td></tr>
<tr><td colspan="2">MA retention →+ Capacity →+ TS</td></tr>
<tr><td>• more efficient
• efficiently</td><td>Having sufficient time allotted for the visit</td></tr>
<tr><td>• more efficient
• efficiently</td><td>Visits on schedule</td></tr>
<tr><td colspan="2">Having sufficient time allotted for the visit →+ Visits on schedule →+ TS</td></tr>
<tr><td colspan="2">The PCMH team approach employs TS. MDs perceive Capacity as limiting the level of TS they can implement. Yet, providing care teams with any constant level of Capacity is made difficult by the unpredictability of work attendance among MAs.

MDs control the level of TS based in part on the perceived level of efficiency in their practice.</td></tr>
</table>

<table>
<tr><td>Participant-Quote number) “Quote … <u>phrases referring to variables</u>”
(word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="8">CM01-42) “[Our hope was PCMH would]… <u>make [patients’] visit as quick and as efficient as possible</u>, and <u>allowing more time for the providers to spend face to face with the patient</u>, versus <u>to try and rush in between</u>. “<u>Some [aspects of PCMH] were harder than others and</u> [for] <u>some we are still struggling</u> … If we have <u>an MA or 2 that are out</u>, we will have <u>one provider that may suffer</u>, while <u>others may not be busy that could be helping out</u>. The idea is [changing x] would [address]… the constant complaint from the providers [which] is that they are <u>short</u> <u>staffed</u>, which they are. [MAs] <u>take a week off, they are sick, they are on vacation</u>, and you try to <u>staff as appropriately as you can</u>.” (139/391)</td><td>• Some [aspects of PCMH] were harder than others and [for] some we are still struggling</td><td>Time Delay</td></tr>
<tr><td>• an MA or 2 that are out
• take a week off, they are sick, they are on vacation</td><td>MA retention</td></tr>
<tr><td>• one provider that may suffer
• others may not be busy that could be helping out
• short staffed
• staff as appropriately as you can</td><td>Capacity</td></tr>
<tr><td>• make [patients’] visit as quick and as efficient as possible
• allowing more time for the providers to spend face to face with the patient
• to try and rush in between</td><td>Having sufficient time allotted for the visit</td></tr>
<tr><td colspan="2">MA retention →+ Capacity --||→+ Having sufficient time allotted for the visit</td></tr>
<tr><td>• staff as appropriately as you can</td><td>Hiring new MAs</td></tr>
<tr><td colspan="2">Hiring new MAs →+ Capacity</td></tr>
<tr><td colspan="2">Retaining a sufficient level of Capacity on a team can even out the workload and allows clinicians to give each visit a sufficient amount of time. Because of MA work attendance issues, and a significant time delay, this ideal state is difficult to attain. When appropriate, hiring MAs can also increase capacity.</td></tr>
</table>

Table F.14 Causal RIQs – CM01 on Finances

Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable →+ Model Variable** (→ = causal link, --||→+ = with delay, +/- = link polarity) Interpretation	
CM01-43) "There was a [clinician] that couldn't … get the concept up [and running], and he just got frustrated that he wouldn't be able to continue on his guarantee, making the money that he needed to be, not for lack of desire … it just wasn't for him. He had a really hard time with … maintaining an efficient visit so he couldn't see the numbers that he felt that he needed to succeed. He ended up leaving." (77/137)	• he couldn't see the numbers	**Volume of patient visits per hour**
	• making the money	**Amount of clinician pay**
	• he wouldn't be able to continue on his guarantee • making the money that he needed to be • he felt that he needed to succeed	**Difference between desired and actual pay**
	Volume of patient visits per hour →+ Amount of incentive pay →+ Amount of clinician pay →+ Difference between desired and actual pay	
	• get the concept up [and running]	**TS**
	0	**Visit time spent training MAs**
	• maintaining an efficient visit	**Having sufficient time allotted for the visit**
	0	**Visits with coordinated, comprehensive care**
	TS →+ Visit time spent training MAs →- Having sufficient time allotted for the visit →+ Visits with coordinated, comprehensive care →- Volume of patient visits per hour	
	In the current incentive structure, MDs earn more when they see more patients. MDs use the incentive structure to decide how many patients to see. MDs not implementing TS cannot afford to spend a long time with their patients.	
CM01-44) "Our [x] area has grown a lot, especially this last year. It has gotten much, much busier. We are significantly over budget with our visits to primary care. … Because we have budgeted for this number of visits and then we ended up having more… Our revenues are also much, much higher as a result of that." (57/388)	• has gotten much, much busier • over budget with our visits	**Volume of patient visits per hour**
	• significantly over budget • our revenues	**Clinic profit margin**
	Volume of patient visits per hour →+ Clinic profit margin	
	A forecast is made in terms of visits and expenses. If the number of visits is higher than the forecast clinic finances are better than they otherwise would be.	

Table F.15 Causal RIQs – CM01 on Comprehensiveness

<table>
<tr><td>Participant-Quote number) “Quote … phrases referring to variables”
(word count/total words in causal statement)</td><td colspan="2">• Phrase(s) needing interpretation = Model Variable
Model Variable →+ Model Variable (→ = causal link, --||→+ = with delay, +/- = link polarity)
Interpretation</td></tr>
<tr><td rowspan="6">CM01-45) “[We are] continually trying to do more pre-visit planning and have people come in early to get labs and things. That does improve patient flow, because [when] they come in, their labs are done. They are here, they are done, and the doc can talk about it.” (47/47)</td><td>• the doc can talk about it</td><td>Having sufficient time allotted for the visit</td></tr>
<tr><td>• pre-visit planning</td><td>Visits with coordinated, comprehensive care</td></tr>
<tr><td colspan="2">Having sufficient time allotted for the visit →+ Visits with coordinated, comprehensive care</td></tr>
<tr><td>• patient flow</td><td>Visits on schedule</td></tr>
<tr><td colspan="2">Having sufficient time allotted for the visit →+ Visits on schedule</td></tr>
<tr><td colspan="2">If the team can work prior to and parallel to the doctor-patient encounter, this permits the MD to give the patient more attention in the visit. It also permits the patients to flow through more easily.</td></tr>
<tr><td rowspan="6">CM01-46) “Planned care, just even one aspect of planned care, has huge potential, to get everybody doing that. Just the efficiency factor there, plus, the experience of having, of being able to go into a doctor’s office, I think about my own [experience], going in to a doctor’s office, and having your results right there to discuss with their doctor. They are in the room… So, the perception that the visit is more productive and you get your questions answered would be greater.” (82/178)</td><td>• discuss with their doctor
• the visit is more productive</td><td>Having sufficient time allotted for the visit</td></tr>
<tr><td>• efficiency factor</td><td>Visits on schedule</td></tr>
<tr><td colspan="2">Having sufficient time allotted for the visit →+ Visits on schedule</td></tr>
<tr><td>• planned care
• having your results right there
• you get your questions answered</td><td>Visits with coordinated, comprehensive care</td></tr>
<tr><td colspan="2">Having sufficient time allotted for the visit →+ Visits with coordinated, comprehensive care</td></tr>
<tr><td colspan="2">Planned care improves the efficiency of in-visit discussions, because patient results are available and patient questions are more easily answered.</td></tr>
</table>

F.2 MODEL IMPROVEMENTS RESULTING FROM CONCEPTUAL MODEL SATURATION

The CM-S Test identified elements in SMM2 that were either not mentioned or re-conceptualized by validation-set interviews. Variables and links that were not mentioned by the validation-set interviews were kept. This is because no single clinic came upon the entire shared mental model in the model development set, so it was not expected that the model-validation-set clinic would be able to do so. These variables and links are flagged as less certain and therefore merit special consideration as regards remaining model development and validation activities.

Variables and links were re-conceptualized when validation-set interviews described them in a different manner than represented in SMM2. Statements marked as challenge statements during interview coding were reviewed during CM-S Diagram generation and the new formulation was kept as a “re-conceptualized” relationship or variable.

Table F.16 below presents the elements that were not mentioned as well as those that were corrected because they were re-conceptualized by participants in the validation-set interviews. Variables in the model use “MA” to refer to clinical staff members and “wRVUs” to refer to the work Relative Value Unit (money paid per predefined task).

Table F.16 Elements Not Mentioned or Elements that were Re-conceptualized by Clinic 6 Interviews

Element Type	Treatment Type	Specific Element	Specific Treatment
Variables	Re-conceptualized	"MA - patient relationship"	Removed, aggregated into "MA capabilities"
		"Capacity"	Renamed, previously called "Care Team Ratio"
		"Off-the-job MA training"	Renamed, previously called "Training MAs"
	Not mentioned	"Clinician develops relationship with patient"	Kept, with caveat that they were not mentioned in CM-S
		"wRVUs per visit"	
Links	Re-conceptualized	The links involving "Continuity of care (from visit to visit)"	Removed old link, new links are from "Visits on schedule", and to "Visits with coordinated, comprehensive care"
		The cause of "Clinician - MA relationship (Trust)"	Removed old link, new link shows cause as "MA capabilities"
		The links involving "MA - patient relationship (Trust)"	Removed old links, since this variable is aggregated into "MA capabilities", the cause link is still from "Task-shifting to MAs" and the effect link is now in the longer chains from "MA capabilities" to "Task-shifting to MAs"
	Not mentioned	The links involving the not mentioned variables (i.e., "Clinician develops relationship with patient" and "wRVUs per visit")	Kept, with caveat that they were not mentioned in CM-S
Delays	Re-conceptualized	"Task-shifting to MAs" to "MA - patient relationship (Trust)"	This link is preserved as the cause link from "Task-shifting to MAs" to "MA capabilities" also has a delay
		"Task-shifting to MAs" to "Clinician - MA relationship (Trust)"	Removed old link, new link shows two step causal chain through "MA capabilities"
	Not mentioned	"Continuity of care (from visit to visit)" to "MA - patient relationship (Trust)"	Removed
		"Capacity" to "Clinician develops relationship with patient"	Kept
Feedback Loops	Revised feedback loops were observed based upon the changes in variables, links and delays above.		

The re-conceptualized elements were reviewed in relation to the problem statement. Again, no portions of the problem statement were found omitted from SMM2.x. This indicates that the boundaries of the shared mental model continue to be well-developed.

The Causal RIQs supporting re-conceptualizations involving "Clinician - MA relationship (Trust)" and "Off-the-job MA training" are presented in Table F.17 below. The tables include seven elements, shown in Box F.1.

Box F.1 The Elements of Rigorously-Interpreted Quotations for Causality Tables

- On the left of the table
  a. A quote number and a participant number (CL = clinician, MA = medical assistant NM = nurse manager, and CM = center manager).
  b. The quotation.
  c. Phrases referring to variables are underlined in the quotation.
  d. The word count of the quotation over the total words in the original statement (truncated for confidentiality).
- On the right of the table
  a. The underlined phrases, this time on the right of the table with the variable to which they are assigned.
  b. Concepts being validated in the statement are in bold.
  c. A brief comment shows how the statement is interpreted.

SMM2 presented trust as two separate chicken-and-the-egg problems. One where MAs build up a relationship of trust with clinicians by participating in "Task-shifting", but "Task-shifting" only happens when the clinician trusts the MA in the first place. (Note: "Task-shifting" is abbreviated TS in all Causal RIQ tables.) Parallel to this, the relationship between MAs and patients was also seen as influencing "Task-shifting", but could not develop without "Task-shifting" occurring.

In quote 29, CL04 describes a worst-case scenario where a clinician can legitimately lose trust in MAs to perform tasks competently because of a high rate of turnover. In quote 30, CL04 describes how when MAs lapse on performing "Task-shifting" it impacts on patients and causes CL04 to lose trust in the MAs. Because this seems to occur consistently, CL04 chooses to maintain a very low level of "Task-shifting". In quote 31, CL01 describes how "MA capabilities" are built up over time using a consistent approach to giving timely face-to-face feedback on task performance.

Early in the coding step of CM-S, it became clear that MAs mostly earn clinicians' trust by showing their "MA capabilities" after a trial period with "Task-shifting". In clinician descriptions, the trust between MA and patient was one of many "MA capabilities" and not a separate factor. The chicken-and-the-egg relationship between "Task-shifting" and trust was actually about how well "MA Capabilities" were demonstrated in performing some basic "Task-shifting" tasks. The time delay can distort a clinician's perception of the true level of "MA Capabilities" or a MA's potential capabilities. The chicken-and-the-egg problem still exists. The new model shows it is a consequence of clinicians' boundedly-rational expectations of MA learning rather than as a rational hesitance to place trust for patient care with MAs, or an aversion to teamwork on the part of clinicians. Far from it, clinicians recognize that they need the help, but they do not want it if it creates more problems than it solves.

SMM2 mentions two mechanisms that involve MA training. Task-shifting causes an increase in "Visit time spent training MAs". A second variable "Training MAs" increases capabilities.

In quote 29, CL04 mentions two different types of training for MAs. One is on-the-job training where MAs are trained to perform shifted tasks *as specifically desired by clinicians on the team*. This is one aspect of "Task-shifting to MAs". The second type of training refers to one that it is provided by the HSDO, one that is not *specific to the clinicians on the team.* These are off-the-job trainings. Both types increase capabilities; however, unlike the on-the-job trainings, these trainings are not part of feedback loops and they do not relate to team-specific ways of doing things.

Causal RIQs for additional items are not presented here as these quotations are presented above: for re-conceptualizations involving "Capacity" see NM01-35, MA04-10 and CL01-25 and for re-conceptualizations involving "Continuity of care (from visit to visit)" see CL03-17 and CL03-21.

Table F.17 Causal RIQs – Clinicians on Trust

Participant-Quote number) "Quote … phrases referring to variables" (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable →+ Model Variable** (→ = causal link, --||→+ = with delay, +/- = link polarity) Interpretation	
CL04-29) "But when I was at [another clinic] … it was kind of working… some MAs worked better with certain providers than others so, that was kind of nice because certain MAs you just clicked with and you could get things done a lot quicker. They knew how to … [do] things … while we were talking and they could start your note for you… If we could train all the MAs to do that and have enough staffing to do it, then that would be great. But it seems like that takes a lot of work because some of them aren't trained or willing to learn... "Staff turnover … is a big issue … we've changed … the MAs ... [None of the 7 on the team have] worked here [more than 2 years] "I have to tell MAs little things about what I would like to have done [for x, prepare y]… I just feel like I continually have to tell people that. So, it does take a lot more time and patience… "Yeah, they get plenty of training but they don't get training on specifically what I like… because I just have to tell them constantly." (174/444)	• it seems like that takes a lot of work • it does take a lot more time and patience	**Time Delay**
	• get things done a lot quicker • they knew how • some of them aren't trained	**MA capabilities**
	• some MAs worked better with certain providers than others so • certain MAs you just clicked with • willing to learn • I just feel like I continually have to tell people • I just have to tell them constantly	**Clinician - MA relationship (Trust)**
	• get things done • do things while we were talking and they could start your note for you • to do it	**TS**
	MA capabilities →+ Clinician - MA relationship (Trust) →+ TS --||→+ MA capabilities	
	The choice to place trust in an MA is sensitive to MD's assessment of the MA's ability to learn. Without passing this test, CL04 will not initiate TS, even with experience in the model and substantial buy-in. CL04 admits the true time required to train is unknown, but it seems long. Because of the high rate of turnover, CL04 experiences it to be infinitely long.	

Participant-Quote number) "Quote … <u>phrases referring to variables</u>" (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable →+ Model Variable** (→ = causal link, --||→+ = with delay, +/- = link polarity) Interpretation	
	• If we could train all the MAs • they get plenty of training	**Off-the-job MA training**
	Off-the-job MA training --||→+ MA capabilities →+ Clinician - MA relationship (Trust)	
	• I have to tell MAs little things about what I would like to have done [for x, prepare y]… I just feel like I continually have to tell people • but they don't get training on specifically what I like	**Visit time spent training MAs**
	TS→+ Visit time spent training MAs	
	There are two types of training: training done by the clinician on-the-job and training done by the HSDO. The clinician sees HSDO led training as an easier faster way to develop capabilities such that the MA can be trusted to perform at a higher level of task-shifting. Otherwise, task-shifting feels burdensome because it requires so much time spent in on-the-job training.	
	• Staff turnover … we've changed … the MAs	**MA retention**
	• have enough staffing	**Capacity**
	MA retention →+ Capacity →+ TS	
	The choice to shift tasks is sensitive to the amount of staffing, and to the length of tenure of MAs. High turnover has been a big issue for this team.	

Participant-Quote number) “Quote … phrases referring to variables” (word count/total words in causal statement)	• Phrase(s) needing interpretation = **Model Variable** **Model Variable →+ Model Variable** (→ = causal link, --||→+ = with delay, +/- = link polarity) Interpretation	
CL04-30) “If I send out a note to call this patient … sometimes I will go back and look … days later and no one’s called them. That worries me quite a bit. “Sometimes I feel like I do more things than I probably should just because I would just rather just do it myself and get it done correctly… I try and stick with basic things for the MAs to do and if [it is at all] complicated … then I just do it… “I guess I can mostly rely on people but some things I don’t trust them with and I would rather just do it myself.” (51/188)	• no one’s called them	**MA capabilities**
	• That worries me quite a bit • I would just rather just do it myself and get it done correctly • I guess I can mostly rely on people but some things I don’t trust them with	**Clinician - MA relationship (Trust)**
	• If I send out a note to call this patient • I do more things than I probably should • stick with basic things for the MAs to do and if [it is at all] complicated … then I just do it • just do it myself	**TS**
	MA capabilities →+ Clinician - MA relationship (Trust) →+ TS	
	The MD makes the TS decision by considering how much to trust MAs. When MA performance on assigned tasks is poor, then trust is low. There is a basic level of tasks that MAs can do which do not require the MD’s trust.	
CL01-31) “I certainly like to give feedback when something wasn’t done up to my standards and normally that helps… I am a stickler about always [doing x, for example,] … and usually, with new MAs they never [do x]. And then, after a couple of times, that usually changes… If I can find them … [then] I do it face to face. [It] normally [takes] a day or two… “[In this department] we all like things done a different way… So usually it takes [MAs] a little bit … to get used to that. But normally they learn.” (51/188)	• after a couple of times • It normally takes a day or two • it takes MAs a little bit	**Time Delay**
	• something wasn’t done up to my standards • with new MAs they never do x … that usually changes • they learn	**MA capabilities**
	• I certainly like to give feedback • If I can find them, then I do it face to face	**Clinician - MA relationship (Trust)**
	• I am a stickler about • we all like things done a different way	**TS**
	MA capabilities →+ Clinician - MA relationship (Trust) →+ TS --||→+ MA capabilities	
	Capabilities development takes time. This includes 1) the time for the MD to find the MA to bring up a deficiency and 2) the time for the MA to learn to get it right.	

F.3 MODEL CONSTANTS

This section focuses on the parameters entered into the model.

Table F.18 presents model constants (exogenous variables), as well as their assumed values and sources. Sources include:

- Literature (i.e., published papers)
- Interviews (either specifically stated or referred to broadly)
- The model itself (i.e., in order for it to be in equilibrium under normal conditions, the initial, perceived and current values for a variable need to be equal; also some constants' equilibrium values are calculated from the values of a subset of other model constants – these calculations are described in the model documentation).

Table F.18 Constants Used in the Model

Variable Name	Value	Units	Source or Reason
Adjustment time for avg clinic net profit	3	Month	Interviews
Adjustment time for MA to leave	6	Month	Interviews
Adjustment time how long it takes for the provider to revise his thinking re Capabilities	2	Month	Interviews
Adjustment time how long it takes for the provider to revise his thinking re MD salary	2	Month	Interviews
Adjustment time how long it takes the MD to revise their thinking on pt satisfaction	2	Month	Interviews
Adjustment time to feel the change in MA satisfaction	2	Month	Interviews
Adjustment time to hire MA	2	Month	Interviews
Adjustment time to shed MD Tech tasks	1	Month	Interviews
Average panel size per MD	2000	Patients / MDs	[338]
Avg number of visits per patient per month	0.25	Visit / patients / Month	[76, 220, 339]
Benefits per MA per month	1458.5	USD / (Month*MAs)	HR practice
Benefits per MD per month	7875	USD / (Month*MDs)	HR practice
Clinic overhead cost per month	100000	USD / Month	Calculation
Desired time to complete MA tasks	3	Month	To set desired equal to initial
Encounter length	20	Minutes / encounter	Interviews & [76, 220]
Initial fraction nonTech tasks per visit	0.25	Dimensionless	[76, 220]
Initial fraction Tech tasks per visit	0.25	Dimensionless	[76, 220]
Initial patient satisfaction	0.8	Reputation	Interviews
Initial perception of MA capabilities	50	Capabilities / MA	To set desired equal to initial
Initial ratio of MD to MA	1	MDs/MAs	Interviews

Variable Name	Value	Units	Source or Reason
MA satisfaction acceptable minimum	0.75	Satisfaction	Interviews
Desired MD monthly Salary	12332.3	USD / (MDs*Month)	To set desired equal to initial
monthly salary per MD initial perception	12332.3	USD / (Month*MDs)	Calculation
normal productivity per MA	25000	Tasks/(Month*MAs)	Calculation
normal productivity per MD	25000	Tasks/ Month*MDs)	Calculation
Normal workload per MA	75000	Tasks/MAs	Calculation
Normal workload per MD	75000	Tasks/MDs	Calculation
Normal time to complete MD Tech tasks	3	Month	Interviews
Normal time to complete nT for MD	3	Month	Interviews
Number of MDs	2	MDs	Smallest clinician team
Potential MA only tasks per visit	100	Tasks/visit	Assumption to track tasks
Potential MD tasks per visit	100	Tasks/visit	Assumption to track tasks
Proportion of potential tasks that are T	0.50	Dimensionless	[76, 220]
proportion of staff work that is in-visit work	0.800169	Dimensionless	Interviews
scheduling out	3	Month	To set desired equal to initial
time per task	0.2	Minutes / tasks	Calculation
MD time to develop willingness	4	Month	Interviews
training per unit fraction change per MA	50,000	Tasks / MAs	Calculation
training tasks needed to gain capability	150	Tasks / Capabilities	Calculation
TS start date	10	Months	Set to see behavior pre and post
visit relative complexity	0.783	Dimensionless	[340] & Calculation
wRVUs per nonTech task whether completed by MD or MA	0.01	wRVU / tasks	[193, 341] & Calculation
wRVUs per Tech task	0.03	wRVU / tasks	[193, 341] & Calculation

In addition to these model constants, there are also constants for conversions (e.g., three months per quarter). There are also constants that represent maximum values (e.g., "maximum willingness" is 1). Finally, there are constants making up the "pink noise" structure (e.g., "mean").

Policies have also been built into the model. Constants are used to designate specific reimbursement rates for the various policies Table F.19 as well as to turn policies on and off (see next section).

Table F.19 Policy Constants

Category	Policy	Value	Units	Source or Reason
Clinic Policy	Clinic policy Incentive pay per encounter	78.73	USD/encounter	Calculation to initially equal fixed salary
	Clinic policy Median MD compensation per wRVU	31.5	USD/wRVU	Calculation to initially equal fix salary
	USD per MA Only task	0.11667	USD/task	Calculation to initially equal fix salary
	Clinic policy flat rate salary per MA per month	2917	USD / (Month*MAs)	Literature[342, 343] & Calculation
	Clinic policy flat rate salary per MD per month	15750	USD / (Month*MDs)	Literature[339] & Calculation
	Clinic policy training per unit fraction change per MA	50,000	Tasks / MAs	Calculation
	Clinic policy initial capability of MA new hire	50	Capabilities / MAs	Calculation
CMS Policy	CMS policy clinic overhead compensation per wRVU	38	USD/wRVU	Literature[344] & Calculation
	CMS policy MD compensation per wRVU	170	USD/wRVU	Literature[344] & Calculation

F.3.1 POLICIES & ENVIRONMENTAL CONDITIONS

Participants described many policies and environmental conditions[118] that impact their ability to implement task-shifting and their overall experience with PCT. For example, for policies, MA03 and CL03 both indicated that, at times, management does not hire more MAs in response to their requests for increased capacity. CM01 explained that hiring is considered unacceptable if net profit is not above zero. This policy is found in the simulation model ("Clinic Policy hire regardless of clinic net profit"). For environmental conditions, CM01 indicated that how much task-shifting is done depends on specific preferences of the clinicians on the team – these are found in table functions described below.

All policies in the simulation model are decisions that management makes. Three policies relate to how the MA is paid. Another three policies relate to how the clinician is paid. Two policies relate to center manager decisions. First, when hiring MAs, one policy allows for hiring MAs that are more or less capable. Second, when starting the implementation of task-

[118] Environmental conditions are contextual features of the system that are not changed within the system boundary (i.e., alternative realities). Most commonly, these are preferences that are hard to change and therefore they act as environmental conditions. They also include things like uncertainty in revenue generation (amount of uncertainty in revenue to be received for services delivered) because they are also unchanging within the model scope.

shifting, clinic managers can decide how much task-shifting to request that clinicians begin implementing.

Table F.20 below lists the policies in the simulation model. The first column lists the policy category (who the policy is about). The second column lists the variable name. The third column provides a brief description of the policy. The fourth and fifth columns list the base-case value and value range. For a complete description of how each variable is used in the simulation model (model views showing the variable and equations using the variable), please see Appendix F.

Table F.20 Policies in the Simulation Model

Category	Variable Name	Description	Value Range	Base-case Value
MAs Salary	Clinic Policy Salaried MAs	Clinical staff members are paid a fixed salary.	1 = on 0 = off	0
	Clinic Policy MA paid for advanced capabilities	Clinical staff members are compensated when performing tasks requiring a higher level of capability (the shifted tasks).	1 = on 0 = off	1
MDs Salary	Clinic Policy Salaried MDs	Clinicians are paid a fixed salary.	1 = on 0 = off	1
	Clinic Policy MD encounter based compensation	Clinicians are paid based on the number of encounters that they complete (i.e., a fixed amount per encounter).	1 = on 0 = off	0
	Clinic Policy MD wRVU based compensation	Clinicians are paid based on the amount of work that they accomplish (i.e., by the work relative value units "wRVUs" that they complete).	1 = on 0 = off	0
MA Hiring	Clinic Policy hire regardless of clinic net profit	0 = off, you only hire when you have a positive clinic net profit 1 = on, you hire when you need to (regardless of clinic net profit situation)	1 = on 0 = off	1
Initial Task-shifting Requirement	kickstart amount	The amount of task-shifting willingness that the management requires on the part of the clinicians at the beginning of implementation (units = willingness/month).	[0-0.20]	0.10

Environmental conditions in the simulation model relate to the clinic, clinicians, MAs, and patients. They are preferences held by these individuals which determine their behavior. These can be modeled using constants or table functions.

Four of these use constants: 1) the amount of uncertainty in revenue generated each month, 2) the MD's desired salary 3) the clinicians' speed at changing their mind with respect to willingness to shift tasks and 4) the MAs' speed at learning on the team (Table F.21). Sensitivity analysis is used to explore the impact of varying these constants.

Table F.21 Environmental Conditions in the Simulation Model - Constants

Category	Variable Name	Description	Value Range	Base-case Value	Source or Reason
Clinic	Standard Deviation of revenue generation	This is a measure of the amount of uncertainty in total revenue generated per month. It is the standard deviation for pink noise.	[0-0.125]	0	Calculation
Clinicians	Desired MD monthly Salary	The monthly salary that the clinicians expect/desire to receive.	[6100-1850]	12332.3	Literature & Calculation
	MD time to develop willingness	Amount of time that it takes the clinicians to change their willingness to shift tasks once a change is warranted.	[2-6]	4	Interviews
Clinical Staff Members	training tasks needed to gain capability	For each unit increase in MA capabilities, this variable lists the number of training tasks that the MA needs to complete. This value varies, depending on the clinician and MA's skills at teaching/learning, problem solving, dialogue.	[75-225]	150	Calculation

The remaining environmental conditions use table functions. Each table function represents the response to different values of a particular variable ("decision input"). For example, for "effect of MA workload ratio on MD's desired MA staffing level", the shape of the table function is based on clinicians' sensitivity to MAs' overworked-ness. Sensitivity analysis is used to explore the impact of varying the shape of these table functions. The base-case table function is a mid-range environmental condition, where alternative table functions explore the impact of having decision makers with different preferences regarding the decision input.

For clinicians, one table function relates to MA hiring, two relate to the attempted comprehensiveness, four relate to the clinician's willingness to shift tasks, and two relate to task shedding. For MAs, two table functions relate to their satisfaction, another one relates to their tolerance of low satisfaction, and the final one relates to their desire for learning. The patient table function relates to their desire to receive clinical services in a timely manner.

Table F.22 below presents the scenarios that use table functions in the simulation model. It is formatted similarly to the two tables above; where the difference is in the fourth column, which lists the decision.

Table F.22 Preferences in the Simulation Model – Table Functions

Category	Variable Name	Description	Decision Input
Clinicians	effect of MA workload ratio on MD's desired MA staffing level	The MDs make hiring requests when the MAs are overworked. "Workload ratio per MA per MA-only task" measures the ratio between the current and normal workload for MA only tasks. When this variable is above 1, the MAs are overworked.	workload ratio per MA per MA-only task
	effect of MA Backlog on inflow of nonTech tasks	The MDs observe how well the MAs are able to keep up with the tasks that they have been shifted. As the backlog of shifted tasks becomes significantly higher than the backlog of MA-only tasks, the clinician becomes concerned, pulling back on the number of shiftable (non-technical) tasks that the clinician attempts to have the team address.	proportion of MA workload that is MA Advanced tasks
	effect of proportion of Tech tasks on inflow of Tech tasks	The MD decides how many technical tasks he expects to get done during the visit based on his current workload ratio (proportion of Tech tasks out of the total tasks for the MD). The MD decides not to allow himself to book more technical tasks until he has more capacity to do them. When he sees that he has more capacity (via the decision input), he expects to get more tech tasks done and thus allows them in.	proportion of MD workload that is Tech
	effect of MA Capability on MD willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is their perception of the MAs' ongoing level of capabilities. The more capable the MAs, the more willing the MD.	Fraction of MA capabilities perceived to be attained
	effect of MD monthly salary on MD's willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is the ratio between their desired and actual monthly salary. When at or above their desired salary, the clinicians are more willing.	Ratio of perceived to desired monthly salary
	effect of perceived patient satisfaction on MD willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is the ratio between patients' ongoing and initial satisfaction, as perceived by the clinician. When at or above the initial patient satisfaction, the clinicians are more willing	Ratio of perceived to initial patient satisfaction
	effect of willingness ratio on further changes to willingness	An aspect influencing clinicians' ongoing willingness to shift tasks is the ratio between their current level of willingness and the max level of willingness the clinicians have determined that they have.	Ratio of willingness to max willingness
	effect of time to complete backlog Tech tasks on shedding	In the initial equilibrium, the normal time to complete Tech tasks is 3 months. The decision input compares the normal time to complete these tasks to the actual time to complete these tasks. If the ratio is above 1 (actual time to complete is greater than the normal time to complete these tasks), there may be some shedding of these tasks. If the ratio is less than 1 then there is no shedding of these tasks.	Ratio of current time to complete MD Tech to normal time to complete MD Tech tasks

Category	Variable Name	Description	Decision Input
Clinicians (continued)	effect of MD time to complete nT tasks ratio on shedding of those tasks	In the initial equilibrium, the normal time to complete nTech tasks is 3 months. The decision input compares the normal time to complete these tasks to the actual time to complete these tasks. If the ratio is above 1 (actual time to complete is greater than the normal time to complete these tasks), there may be some shedding of these tasks. If the ratio is less than 1 then there is no shedding of these tasks.	Ratio of current time to complete nT tasks to normal time to complete nT tasks
Clinical Staff Members	effect of MA Capabilities on MA satisfaction	An aspect influencing MA satisfaction is their level of capabilities. As they become more capable, they become more satisfied. They enjoy learning and using their capabilities to do higher level tasks.	Ratio of avg to initial capability of MA
	effect of MA capacity on MA satisfaction	An aspect influencing MA satisfaction is their level of capacity. As they have more capacity, they become more satisfied. MAs expect to be able to do their job. As they shed more MA only tasks (i.e., their basic duties), they become less satisfied.	Ratio of shed MA only tasks to MA only tasks completed
	MA willingness to stay in MA job	MAs choose to leave depending on their relative level of satisfaction (current perceived level to initial/expected level), but the exact tolerance varies across MAs.	MA satisfaction ratio
	effect of MA capab ratio on change in MA capab	Learning generally begets more learning, but the exact shape varies across MAs.	MA capabilities to max capabilities ratio
Patients	effect of time to complete backlog of Tech tasks on Patient Satisfaction	Patients come to the doctor to get treated, to receive services that only the doctor can provide. Patients are satisfied as long as they are able to receive the services that they need in a timely manner. Therefore, the model uses the ratio of current time to complete to normal time to complete technical tasks. When that ratio is greater than 1, patients become less satisfied. When it is equal to or less than 1 patients are satisfied.	ratio of time to complete to normal time to complete MD Tech tasks

F.4SIMULATION MODEL EQUATIONS

Table F.24 below lists all model equations (i.e., for stocks, flows and auxiliary variables). The model uses Vensim software[177]. Most equations use addition, subtraction, multiplication and division – friendly algebra. Where this is not the case, Vensim notation is used (these functions are described in Table F.23 below).

Table F.23 Vensim Functions Used in the Model of Primary Care Transformation

Name	**Description**[177]	**Function Code**
Numerical INTEGration	"Returns the integral of the rate. The rate is numerically integrated. The initial value is the value of the variable on the left-hand side of the equation at the start of the simulation."	INTEG (rate, initial value)
MAXimum of Two Alternatives	"Returns the larger of A and B."	MAX(A,B)
MINimum of Two Alternatives	"Returns the smaller of A and B."	MIN(A,B)
PULSE	"Returns 1.0, starting at time start, and lasting for interval width; 0.0 is returned at all other times. If width is passed as 0 it will be treated as though it were the current value of TIME STEP. "	PULSE(start,width)
Traditional IF-THEN-ELSE Statement	"Returns first value (tval) if condition (cond) is true; second value (fval) if condition is false. cond must be a Boolean expression or an expression or variable that can be interpreted as Boolean (i.e., taking a value of 0 or 1). Only the value returned is evaluated, so the other value could be an expression that would lead to an error."	IF THEN ELSE(cond, tval, fval)
X If Divided by Zero (otherwise A/B)	"Returns **A** divided by **B**. If **B** is zero, then returns **X**. XIDZ is normally used to express some limit of **A**/**B**, as **B** approaches 0 (which would normally be undefined for **B** = 0)."	XIDZ(A,B,X)
Zero If Divided by Zero (otherwise A/B)	"Divide **A** by **B**. If **B** is zero (actually smaller than 1E-6), then return 0.0. ZIDZ is normally used to express the special case where the limit of **A**/**B**, as **B** approaches 0, is 0."	ZIDZ(A,B)

Table F.24 Simulation Model Equations

Variable	Equation	Units	Item type
Actual adherence to clinical guidelines	total wRVU tasks completed/total number of potential wRVU tasks per month	Dimensionless	Equation
actual number of nonTech tasks per month	total number of potential Tech and NonTech tasks per month*Fraction nonTech Tasks	tasks/Month	Equation
additions to cumulative revenue	net monthly profit or loss	USD/Month	Equation
adjusted benefits per MA per month	benefits per MA per month*visit relative complexity	USD/(Month*MAs)	Equation
adjusted benefits per MD per month	benefits per MD per month*visit relative complexity	USD/(MDs*Month)	Equation
adjusted clinic overhead compensation per wRVU	visit relative complexity*CMS Policy clinic overhead compensation per wRVU	USD/wRVU	Equation
adjusted clinic overhead cost per month	clinic overhead cost per month*visit relative complexity	USD/Month	Equation
adjusted flat rate salary per MA per month	flat rate salary per MA per month*visit relative complexity	USD/(Month*MAs)	Equation
adjusted flat rate salary per MD per month	Clinic Policy flat rate salary per MD per month*visit relative complexity	USD/(MDs*Month)	Equation
adjusted incentive pay per encounter	visit relative complexity*Clinic Policy incentive pay per encounter	USD/encounters	Equation
adjusted MD compensation per wRVU	CMS Policy MD compensation per wRVU*visit relative complexity	USD/wRVU	Equation
adjusted wRVUs per nonTech task whether completed by MD or MA	visit relative complexity*wRVUs per nonTech task whether completed by MD or MA	wRVU/tasks	Equation
adjusted wRVUs per Tech task	visit relative complexity*wRVUs per Tech task	wRVU/tasks	Equation
average backlog of MA training tasks per MA	Backlog of OnTheJob Training Tasks for MA/Number of MAs	tasks/MAs	Equation

Variable	Equation	Units	Item type
average backlog of training tasks per MA	Backlog of OnTheJob Training Tasks for MD/Number of MAs	tasks/MAs	Equation
average capability per MA	ZIDZ (MA Capabilities , Number of MAs)	Capabilities/MAs	Equation
average wait time to complete MA Only tasks	ZIDZ (Backlog of MA Only Tasks , MA Only tasks completed by MA)	Month	Equation
average wait time to receive MA services	ZIDZ (Total MA Backlog of Tasks , Total MA Tasks Completed per month)	Month	Equation
average waiting time to receive MA advance task services	ZIDZ (Backlog of MA Advanced Tasks , MA Advanced tasks completed by MA)	Month	Equation
Avg monthly clinic net profit	INTEG (net change in avg clinic profit,net monthly profit or loss)	USD/Month	Stock
Avg number of MA Only tasks completed by MA	INTEG (net change of avg number of MA Only tasks comp by MA,MA Only tasks completed by MA)	tasks/Month	Stock
Avg number of nonTech tasks completed by MD	INTEG (net change of avg number of nonTech tasks comp by MD,nonTech tasks completed by MD)	tasks/Month	Stock
Avg number of Tech tasks completed by MD	INTEG (net change of avg number of tech tasks comp by MD,Tech tasks completed by MD)	tasks/Month	Stock
avg wait time to complete MD Tech Tasks	ZIDZ (Backlog of Tech Tasks MD Only , Tech tasks completed by MD)	Month	Equation
Backlog of MA Advanced Tasks	INTEG (inflow of MA advanced tasks to MA-MA Advanced tasks completed by MA-MA Advanced tasks not completed ,0)	tasks	Stock
Backlog of MA Only Tasks	INTEG (inflow of MA Only tasks-MA Only tasks completed by MA-MA Only tasks not completed,150000)	tasks	Stock
Backlog of MD nonTech Tasks	INTEG (inflow of nonTech tasks to MD-nonTech tasks completed by MD-shedding nT tasks,75000)	tasks	Stock
Backlog of OnTheJob Training Tasks for MA	INTEG (inflow of training for existing staff due to upshifting of tasks to be done by MA+training needed due to MAnewhire for MA -MA Training tasks shed due to MA turnover-on the job training completedMA-outflow of training for existing staff due to downshifting of tasks to be done by MA ,0)	tasks	Stock
Backlog of OnTheJob Training Tasks for MD	INTEG (inflow of training for existing staff due to upshifting of tasks to be done by MD+training needed due to MAnewhire for MD to do -on the job training tasks completed by the MD-outflow of training for existing staff due to downshifting of tasks to be done by MD -Training tasks shed due to MA turnover,0)	tasks	Stock
Backlog of Tech Tasks MD Only	INTEG (inflow of Tech tasks-shedding Tech Tasks-Tech tasks completed by MD,75000)	tasks	Stock

Variable	Equation	Units	Item type
change in Fraction nonTech Tasks shifted to MA	(Fraction nonTech Tasks shifted to MA-Cumulative Fraction of nonTech Tasks shifted to MA) /time to change fraction of nT tasks shifted to MA	Dimensionless/Month	Equation
change in perceived MA Capabilities	Difference between actual and perceived MA Capabilities/Adjustment time how long it takes for the provider to revise his thinking re Capabilities	Capabilities/(MAs*Month)	Equation
change in perceived MD monthly salary	Difference between actual and perceived MD monthly salary/adjustment time how long it takes for the provider to revise his thinking re MD salary	USD/(MDs*Month*Month)	Equation
change in perceived patient satisfaction	difference between actual and perceived patient satisfaction/Adjustment time how long it takes the MD to revise their thinking on pt satisfaction	reputation/Month	Equation
Change in Pink Noise	(White Noise - Pink Noise)/Correlation Time	1/Period	Equation
compensation for MA advanced tasks completed by MA per month	Clinic Policy median MD compensation per wRVU*MA Advanced tasks completed by MA*adjusted wRVUs per nonTech task whether completed by MD or MA	USD/Month	Equation
Cumulative Fraction of nonTech Tasks shifted to MA	INTEG (change in Fraction nonTech Tasks shifted to MA,0)	Dimensionless	Stock
Cumulative number of encounters	INTEG (number of encounters per month,0)	encounters	Stock
Cumulative number of MD encounters	INTEG (number of MD encounters per month,0)	encounters	Stock
Cumulative revenue	INTEG (additions to Cumulative revenue,0)	USD	Stock
Cumulative salary per MA	INTEG (monthly salary per MA,0)	USD/MAs	Stock
Cumulative salary per MD	INTEG (monthly salary per MD,0)	USD/MDs	Stock
Cumulative Tech and NonTech tasks completed	INTEG (Tech and NonTech Tasks completed per month,0)	tasks	Stock
Cumulative tasks completed	INTEG (tasks completed per month,100000)	tasks	Stock
desired MA completion rate for MAadvanced tasks	Backlog of MA Advanced Tasks/desired time to complete MA tasks	tasks/Month	Equation
desired MA completion rate of MAonly tasks	(Backlog of MA Only Tasks/desired time to complete MA tasks)	tasks/Month	Equation
Difference between actual and perceived MA Capabilities	average capability per MA-Perceived MA Capabilities per MA	Capabilities/MAs	Equation
Difference between actual and perceived MD monthly salary	monthly salary per MD-Perceived MD monthly salary	USD/(MDs*Month)	Equation

Variable	Equation	Units	Item type
difference between actual and perceived patient satisfaction	Patient Satisfaction-Perceived Patient Satisfaction	reputation	Equation
effect of MA hiring policy on MA hiring	IF THEN ELSE(Avg monthly clinic net profit<0, Clinic Policy hire regardless of clinic net profit , 1)	Dimensionless	Equation
facility revenue	((Tech tasks completed by MD*adjusted wRVUs per Tech task)+((nonTech tasks completed by MD +MA Advanced tasks completed by MA)*adjusted wRVUs per nonTech task whether completed by MD or MA))*adjusted clinic overhead compensation per wRVU	USD/Month	Equation
Fraction nonTech Tasks	initial fraction nonTech tasks per visit*effect of MA Backlog on inflow of nonTech tasks	Dimensionless	Equation
Fraction nonTech Tasks shifted to MA	Fraction nonTech Tasks*willingness to Task Shift/max willingness	Dimensionless	Equation
fraction of MA capabilities perceived to be attained	Perceived MA Capabilities per MA/max MA Capabilities per MA	Dimensionless	Equation
fraction of potential MA only tasks completed out of the total tasks	MA Only tasks completed by MA/total number of potential wRVU tasks per month	Dimensionless	Equation
fraction of potential MA only tasks completed	MA Only tasks completed by MA/total number of potential MA Only tasks per month	Dimensionless	Equation
fraction of potential nT tasks completed out of the total tasks	nT tasks completed per month whether by MD or MA/total number of potential wRVU tasks per month	Dimensionless	Equation
fraction of potential nT tasks completed	nT tasks completed per month whether by MD or MA/total number of potential nT tasks per month	Dimensionless	Equation
fraction of potential T tasks completed out of the total tasks	Tech tasks completed by MD/total number of potential wRVU tasks per month	Dimensionless	Equation
fraction of potential T tasks completed	Tech tasks completed by MD/total number of potential T tasks per month	Dimensionless	Equation
Fraction Tech Tasks	initial fraction Tech tasks per visit*effect of proportion of Tech tasks on inflow of Tech tasks	Dimensionless	Equation
Gap between real time and perception of MA satisfaction	real time MA satisfaction-perception of MA Satisfaction	MA satisfaction	Equation
hiring of MA	MAX((MA workforce sought-Number of MAs)/adjustment time to hire MA,0)	MAs/Month	Equation
inflow in rate of willingness to task shift	willingness to Task Shift*((MAX(effect of MA Capability on MD willingness, 0)*MAX(effect of MD monthly salary on MD's willingness , 0)*effect of willingness ratio on further changes to	willingness/Month	Flow

Variable	Equation	Units	Item type
	willingness*MAX(effect of perceived patient satisfaction on MD willingness , 0)))/MD time to develop willingness		
inflow of MA advanced tasks to MA	actual number of nonTech tasks per month*willingness to Task Shift/max willingness	tasks/Month	Flow
inflow of MA Only tasks	total number of potential MA Only tasks per month*MD willingness to adhere to clinical guidelines	tasks/Month	Flow
inflow of nonTech tasks to MD	(1-willingness to Task Shift)/max willingness*actual number of nonTech tasks per month	tasks/Month	Flow
inflow of Tech tasks	total number of potential Tech and NonTech tasks per month*Fraction Tech Tasks	tasks/Month	Flow
inflow of training for existing staff due to upshifting of tasks to be done by MA	inflow of training for existing staff due to upshifting of tasks to be done by MD	tasks/Month	Flow
inflow of training for existing staff due to upshifting of tasks to be done by MD	Number of MAs*training per unit fraction change per MA*MAX(change in Fraction nonTech Tasks shifted to MA , 0)	tasks/Month	Flow
initial number of MAs	Number of MDs/initial ratio of MD to MA	MAs	Equation
MA Advanced tasks completed by MA	productivity per MA for MA Advanced Tasks*Number of MAs	tasks/Month	Equation
MA Advanced tasks not completed	MAX(desired MA completion rate for MAadvanced tasks-MA Advanced tasks completed by MA , 0)	tasks/(Month)	Equation
MA capabilities to max capabilities ratio	MA Capabilities/total max MA Capabilities	Dimensionless	Equation
MA Capabilities	INTEG (new hire capabilities-turnover capab loss+rate of MA Capability gain,Clinic Policy initial capability of MA new hire * Number of MAs)	Capabilities	Stock
MA compensation for MA Only tasks per month	MA Only tasks completed by MA*Clinic Policy USD per MA Only task*visit relative complexity	USD/Month	Equation
MA cost per month based on fixed salary	salary and benefits per MA per month*Number of MAs	USD/Month	Equation
MA cost per month based on wRVUs	compensation for MA advanced tasks completed by MA per month+MA compensation for MA Only tasks per month +(adjusted benefits per MA per month*Number of MAs)	USD/Month	Equation
MA cost per month	(Clinic Policy Salaried MAs*MA cost per month based on fixed salary)+(Clinic Policy MA paid for advanced capabilities *MA cost per month based on wRVUs)	USD/Month	Equation
MA Only tasks completed by MA	Number of MAs*productivity per MA for MA Only Tasks	tasks/Month	Equation

Variable	Equation	Units	Item type
MA Only tasks not completed	MAX(desired MA completion rate of MAonly tasks-MA Only tasks completed by MA,0)	tasks/Month	Equation
MA satisfaction ratio	perception of MA Satisfaction/MA satisfaction acceptable minimum	Dimensionless	Equation
MA Training tasks shed due to MA turnover	average backlog of MA training tasks per MA*turnover of MA	tasks/Month	Equation
MA workforce sought	Number of MAs*effect of MA workload ratio on MD's desired MA staffing level*effect of MA hiring policy on MA hiring	MAs	Equation
MA workforce wanting to stay	Number of MAs*MA willingness to stay in MA job	MAs	Equation
MD compensation for nonTech tasks completed by MD per month	nonTech tasks completed by MD*adjusted wRVUs per nonTech task whether completed by MD or MA *Clinic Policy median MD compensation per wRVU	USD/Month	Equation
MD compensation for Tech tasks per month	Tech tasks completed by MD*adjusted wRVUs per Tech task*Clinic Policy median MD compensation per wRVU	USD/Month	Equation
MD cost per month based on encounters	(number of MD encounters per month*adjusted incentive pay per encounter)+(adjusted benefits per MD per month*Number of MDs)	USD/Month	Equation
MD cost per month based on fixed salary	salary and benefits per MD per month*Number of MDs	USD/Month	Equation
MD cost per month based on wRVUs	MD compensation for Tech tasks per month+MD compensation for nonTech tasks completed by MD per month +(adjusted benefits per MD per month*Number of MDs)	USD/Month	Equation
MD cost per month	(Clinic Policy Salaried MDs*MD cost per month based on fixed salary)+(Clinic Policy MD encounter based compensation *MD cost per month based on encounters)+(Clinic Policy MD wRVU based compensation*MD cost per month based on wRVUs)	USD/Month	Equation
MD willingness to adhere to clinical guidelines	Fraction nonTech Tasks+Fraction Tech Tasks	Dimensionless	Equation
MDpatient tasks per month	Total MD Tasks Completed per month-on the job training tasks completed by the MD	tasks/Month	Equation
MDtime for all patient tasks per month	time per task*MDpatient tasks per month	minutes/Month	Equation
monthly salary per MA	(MA cost per month/Number of MAs)-adjusted benefits per MA per month	USD/(Month*MAs)	Equation
monthly salary per MD	(MD cost per month/Number of MDs)-adjusted benefits per MD per month	USD/(MDs*Month)	Equation
net change in avg clinic profit	(net monthly profit or loss-Avg monthly clinic net profit)/adjustment time for avg clinic net profit	USD/Month/Month	Equation
net change in MA satisfaction	Gap between real time and perception of MA satisfaction/adjustment time to feel the change in MA satisfaction	MA satisfaction/Month	Equation

Variable	Equation	Units	Item type
net change of avg number of MA Only tasks comp by MA	(MA Only tasks completed by MA-Avg number of MA Only tasks completed by MA)/avg time for net change in MA Only task comp by MA	tasks/Month/Month	Equation
net change of avg number of nonTech tasks comp by MD	(nonTech tasks completed by MD-Avg number of nonTech tasks completed by MD)/avg time for net change in nonTech task comp by MD	tasks/Month/Month	Equation
net change of avg number of tech tasks comp by MD	(Tech tasks completed by MD-Avg number of Tech tasks completed by MD)/avg time for net change in tech task comp by MD	tasks/Month/Month	Equation
net monthly profit or loss	total revenue generated per month-total expenses generated per month	USD/Month	Equation
new hire capabilities	Clinic Policy initial capability of MA new hire*hiring of MA	Capabilities/Month	Equation
nonTech tasks completed by MD	Number of MDs*productivity for nonTech Tasks	tasks/Month	Equation
Normal workload for training tasks per MA	Normal workload per MA*proportion of MA workload that is MA training tasks	tasks/MAs	Equation
Normal workload for training tasks per MD	Normal workload per MD*proportion of MD workload that is on the job training	tasks/MDs	Equation
Normal workload per MA for MA Advanced Tasks	Normal workload per MA*proportion of MA workload that is MA Advanced tasks	tasks/MAs	Equation
Normal workload per MA for MA Only Tasks	Normal workload per MA*proportion of MA workload that is MA Only tasks	tasks/MAs	Equation
Normal workload per MD for nonTech Task	Normal workload per MD*proportion of MD workload that is nTech	tasks/MDs	Equation
Normal workload per MD for Tech Tasks	Normal workload per MD*proportion of MD workload that is Tech	tasks/MDs	Equation
nT tasks completed per month whether by MD or MA	MA Advanced tasks completed by MA+nonTech tasks completed by MD	tasks/Month	Equation
number of encounters per month	time for all patient tasks per month/encounter length*proportion of staff work that is in visit work	encounters/Month	Equation
Number of MAs	INTEG (hiring of MA-turnover of MA,initial number of MAs)	MAs	Stock
number of MD encounters per month	MDtime for all patient tasks per month/encounter length*proportion of staff work that is in visit work	encounters/Month	Equation
on the job training completedMA	Number of MAs*productivity per MA for MA training tasks	tasks/Month	Equation

Variable	Equation	Units	Item type
on the job training tasks completed by the MD	Number of MDs*productivity of training tasks per MD	tasks/Month	Equation
operating margin	Avg monthly clinic net profit/total expenses generated per month	Dimensionless	Equation
outflow in rate of willingness to task shift	willingness to Task Shift*(-1*(MIN(MIN(MIN(effect of MA Capability on MD willingness, effect of MD monthly salary on MD's willingness), effect of perceived patient satisfaction on MD willingness), 0)))*effect of willingness ratio on further changes to willingness/MD time to develop willingness	willingness/Month	Flow
outflow of training for existing staff due to downshifting of tasks to be done by MD	MIN(change in Fraction nonTech Tasks shifted to MA, 0)*Backlog of OnTheJob Training Tasks for MD	tasks/Month	Flow
outflow of training for existing staff due to downshifting of tasks to be done by MA	outflow of training for existing staff due to downshifting of tasks to be done by MD	tasks/Month	Flow
overall effect of MA stuff on MA satisfaction	effect of MA Capabilities on MA satisfaction*effect of MA capacity on MA satisfaction	Dimensionless	Equation
Patient Satisfaction	initial patient satisfaction*effect of time to complete backlog of Tech tasks on Patient Satisfaction	reputation	Equation
patient tasks per month	tasks completed per month-(on the job training completedMA+on the job training tasks completed by the MD)	tasks/Month	Equation
Patients	average panel size per MD*Number of MDs	patients	Equation
Perceived MA Capabilities per MA	INTEG (change in perceived MA Capabilities,initial perception of MA capabilities)	Capabilities/MAs	Stock
Perceived MD monthly salary	INTEG (change in perceived MD monthly salary,monthly salary per MD initial perception)	USD/(MDs*Month)	Stock
Perceived Patient Satisfaction	INTEG (change in perceived patient satisfaction,initial patient satisfaction)	reputation	Stock
perceived yearly MD salary at current monthly rate	monthly salary per MD*"12 months in a year"	USD/(MDs*year)	Equation
perception of MA Satisfaction	INTEG (net change in MA satisfaction,MA satisfaction acceptable minimum)	MA satisfaction	Stock
Pink Noise	INTEG (Change in Pink Noise,Mean)	Dimensionless	Stock
policy kickstart	PULSE(TS start date, 1)*kickstart amount	willingness/Month	Equation
practice compensation for MA Advanced tasks completed by MA per month	MA Advanced tasks completed by MA*adjusted wRVUs per nonTech task whether completed by MD or MA *adjusted MD compensation per wRVU	USD/Month	Equation

Variable	Equation	Units	Item type
practice compensation for nonTech tasks completed by MD per month	nonTech tasks completed by MD*adjusted wRVUs per nonTech task whether completed by MD or MA *adjusted MD compensation per wRVU	USD/Month	Equation
practice compensation for Tech tasks per month	Tech tasks completed by MD*adjusted wRVUs per Tech task*adjusted MD compensation per wRVU	USD/Month	Equation
productivity for nonTech Tasks	normal productivity per MD*effect of workload ratio for nonTech Tasks on productivity *proportion of MD workload that is nTech	tasks/(Month*MDs)	Equation
productivity of training tasks per MD	effect of ratio of MD training tasks on productivity*normal productivity per MD*proportion of MD workload that is on the job training	tasks/(Month*MDs)	Equation
productivity per MA for MA Advanced Tasks	normal productivity per MA*effect of workload ratio for MA Advanced Tasks on productivity *proportion of MA workload that is MA Advanced tasks	tasks/(Month*MAs)	Equation
productivity per MA for MA Only Tasks	normal productivity per MA*effect of workload ratio for MA Only Tasks on productivity *proportion of MA workload that is MA Only tasks	tasks/(Month*MAs)	Equation
productivity per MA for MA training tasks	effect of ratio of MA training tasks on productivity*normal productivity per MA*proportion of MA workload that is MA training tasks	tasks/(Month*MAs)	Equation
Productivity per MD for Tech Tasks	effect of workload ratio for Tech Tasks on productivity*normal productivity per MD*proportion of MD workload that is Tech	tasks/(Month*MDs)	Equation
proportion of MA workload that is MA Advanced tasks	Backlog of MA Advanced Tasks/Total MA Backlog of Tasks	Dimensionless	Equation
proportion of MA workload that is MA Only tasks	Backlog of MA Only Tasks/Total MA Backlog of Tasks	Dimensionless	Equation
proportion of MA workload that is MA training tasks	Backlog of OnTheJob Training Tasks for MA/Total MA Backlog of Tasks	Dimensionless	Equation
proportion of MD tasks that are nTech	inflow of nonTech tasks to MD/total inflow of tasks to the MD	Dimensionless	Equation
proportion of MD tasks that are Tech	inflow of Tech tasks/total inflow of tasks to the MD	Dimensionless	Equation
proportion of MD workload that is nTech	Backlog of MD nonTech Tasks/Total MD Backlog of Tasks	Dimensionless	Equation
proportion of MD workload that is on the job training	Backlog of OnTheJob Training Tasks for MD/Total MD Backlog of Tasks	Dimensionless	Equation

Variable	Equation	Units	Item type
proportion of MD workload that is Tech	Backlog of Tech Tasks MD Only/Total MD Backlog of Tasks	Dimensionless	Equation
rate of MA Capability gain	(on the job training completedMA/training tasks needed to gain capability)*effect of MA capab ratio on change in MA capab	Capabilities/Month	Equation
ratio of Avg to Initial Capability of MA	average capability per MA/Clinic Policy initial capability of MA new hire	Dimensionless	Equation
ratio of current time to complete MD Tech to normal time to complete MD Tech tasks	time to complete backlog of MD Tech tasks/normal time to complete MD Tech tasks	Dimensionless	Equation
ratio of current time to complete nT tasks to normal time to complete nT tasks	time to complete nT tasks/normal time to complete nT for MD	Dimensionless	Equation
ratio of perceived to desired MD monthly salary	Perceived MD monthly salary/Desired MD monthly Salary	Dimensionless	Equation
ratio of perceived to initial patient satisfaction	Perceived Patient Satisfaction/initial patient satisfaction	Dimensionless	Equation
ratio of shed MA Only tasks to MA Only tasks completed	MA Only tasks not completed/MA Only tasks completed by MA	Dimensionless	Equation
ratio of time to complete Tech tasks	avg wait time to complete MD Tech Tasks/scheduling out	Dimensionless	Equation
ratio of time to complete to normal time to complete MD Tech tasks	time to complete backlog of MD Tech tasks/normal time to complete MD Tech tasks	Dimensionless	Equation
ratio of willingness to max willingness	willingness to Task Shift/max willingness	Dimensionless	Equation
real time MA satisfaction	MA satisfaction acceptable minimum*overall effect of MA stuff on MA satisfaction	MA satisfaction	Equation
Revenue generated by MA per month	practice compensation for MA Advanced tasks completed by MA per month	USD/Month	Equation
revenue generated by MD per month	practice compensation for Tech tasks per month+practice compensation for nonTech tasks completed by MD per month	USD/Month	Equation
salary and benefits per MA per month	adjusted flat rate salary per MA per month+adjusted benefits per MA per month	USD/(Month*MAs)	Equation

Variable	Equation	Units	Item type
salary and benefits per MD per month	adjusted flat rate salary per MD per month+adjusted benefits per MD per month	USD/(Month*MDs)	Equation
shedding nT tasks	effect of MD time to complete nT tasks ratio on shedding of those tasks*Backlog of MD nonTech Tasks	tasks/Month	Equation
shedding Tech Tasks	effect of time to complete backlog Tech tasks on shedding*Backlog of Tech Tasks MD Only /adjustment time to shed MD Tech tasks	tasks/Month	Equation
Tech and NonTech Tasks completed per month	Total Tech and NonTech tasks completed	tasks/Month	Equation
tasks completed per month	Total MA Tasks Completed per month+Total MD Tasks Completed per month	tasks/Month	Equation
Tech MD Tasks per month	total number of potential Tech and NonTech tasks per month*Fraction Tech Tasks	tasks/Month	Equation
Tech tasks completed by MD	Number of MDs*Productivity per MD for Tech Tasks	tasks/Month	Equation
test ratio for nonTech backlog	Backlog of MD nonTech Tasks/total MD patient tasks	Dimensionless	Equation
time for all patient tasks per month	time per task*patient tasks per month	minutes/Month	Equation
time to complete backlog of MD Tech tasks	ZIDZ (Backlog of Tech Tasks MD Only, Tech tasks completed by MD)	Month	Equation
time to complete nT tasks	Backlog of MD nonTech Tasks/nonTech tasks completed by MD	Month	Equation
total actual MA productivity	productivity per MA for MA Advanced Tasks+productivity per MA for MA Only Tasks+productivity per MA for MA training tasks	tasks/(Month*MAs)	Equation
total actual MD productivity	productivity for nonTech Tasks+productivity of training tasks per MD+Productivity per MD for Tech Tasks	tasks/(Month*MDs)	Equation
total expenses generated per month	adjusted clinic overhead cost per month+MA cost per month+MD cost per month	USD/Month	Equation
total inflow of tasks to the MD	inflow of nonTech tasks to MD+inflow of Tech tasks	tasks/Month	Equation
Total MA Backlog of Tasks	Backlog of MA Only Tasks+Backlog of MA Advanced Tasks+Backlog of OnTheJob Training Tasks for MA	tasks	Equation
Total MA Tasks Completed per month	MA Advanced tasks completed by MA+MA Only tasks completed by MA+on the job training completedMA	tasks/Month	Equation
total max MA Capabilities	max MA Capabilities per MA*Number of MAs	Capabilities	Equation
Total MD Backlog of Tasks	Backlog of MD nonTech Tasks+Backlog of OnTheJob Training Tasks for MD+Backlog of Tech Tasks MD Only	tasks	Equation
total MD patient tasks	Backlog of MD nonTech Tasks+Backlog of Tech Tasks MD Only	tasks	Equation

Variable	Equation	Units	Item type
Total MD Tasks Completed per month	nonTech tasks completed by MD+Tech tasks completed by MD+on the job training tasks completed by the MD	tasks/Month	Equation
total number of potential MA Only tasks per month	total number of visits per month*potential MA Only tasks per visit	tasks/Month	Equation
total number of potential nT tasks per month	total number of potential Tech and NonTech tasks per month*(1-proportion of potential tasks that are T)	tasks/Month	Equation
total number of potential Tech and NonTech tasks per month	total number of visits per month*potential MD tasks per visit	tasks/Month	Equation
total number of potential T tasks per month	total number of potential Tech and NonTech tasks per month*proportion of potential tasks that are T	tasks/Month	Equation
total number of potential wRVU tasks per month	total number of potential MA Only tasks per month+total number of potential Tech and NonTech tasks per month	tasks/Month	Equation
total number of visits per month	Patients*Avg number of visits per patient per month	visit/Month	Equation
total practice compensation for nonTech tasks	practice compensation for MA Advanced tasks completed by MA per month+practice compensation for nonTech tasks completed by MD per month	USD/Month	Equation
total revenue generated per month	(facility revenue+Revenue generated by MA per month+revenue generated by MD per month)*Pink Noise	USD/Month	Equation
Total Tech and NonTech tasks completed	MA Advanced tasks completed by MA+nonTech tasks completed by MD+Tech tasks completed by MD	tasks/Month	Equation
total wRVU tasks completed	wRVU tasks completed by MA+wRVU tasks completed by MD	tasks/Month	Equation
training needed due to MAnewhire for MA	training needed due to MAnewhire for MD to do	tasks/Month	Equation
training needed due to MAnewhire for MD to do	Fraction nonTech Tasks shifted to MA*training per unit fraction change per MA*hiring of MA	tasks/Month	Equation
Training tasks shed due to MA turnover	average backlog of training tasks per MA*turnover of MA	tasks/Month	Equation
turnover capab loss	average capability per MA*turnover of MA	Capabilities/Month	Equation
turnover of MA	(Number of MAs-MA workforce wanting to stay)/adjustment time for MA to leave	MAs/Month	Equation
White Noise	Mean + Standard Deviation*((24*Correlation Time/TIME STEP)^0.5*(RANDOM UNIFORM(-0.5, 0.5, Noise Seed)))	Dimensionless	Equation

Variable	Equation	Units	Item type
willingness to Task Shift	INTEG (inflow in rate of willingness to task shift+policy kickstart-outflow in rate of willingness to task shift ,0)	willingness	Stock
workload for MAs training tasks	XIDZ(Backlog of OnTheJob Training Tasks for MA, Number of MAs, 0)	tasks/MAs	Equation
workload of MDs for training tasks	XIDZ(Backlog of OnTheJob Training Tasks for MD, Number of MDs, 0)	tasks/MDs	Equation
workload per MA for MA only task	XIDZ(Backlog of MA Only Tasks, Number of MAs, 0)	tasks/MAs	Equation
workload per MA for MAadv task	XIDZ(Backlog of MA Advanced Tasks, Number of MAs, 0)	tasks/MAs	Equation
workload per MD for nonTech task	XIDZ(Backlog of MD nonTech Tasks, Number of MDs, 0)	tasks/MDs	Equation
workload per MD for tech task	XIDZ(Backlog of Tech Tasks MD Only, Number of MDs, 0)	tasks/MDs	Equation
workload ratio for MAs training tasks	XIDZ(workload for MAs training tasks , Normal workload for training tasks per MA, 0)	Dimensionless	Equation
workload ratio for MDs training tasks	XIDZ(workload of MDs for training tasks , Normal workload for training tasks per MD , 0)	Dimensionless	Equation
Workload ratio per MA for MA Advanced task	XIDZ(workload per MA for MAadv task , Normal workload per MA for MA Advanced Tasks, 0)	Dimensionless	Equation
Workload ratio per MA for MA Only task	XIDZ(workload per MA for MA only task , Normal workload per MA for MA Only Tasks, 0)	Dimensionless	Equation
Workload ratio per MD for nonTech task	XIDZ(workload per MD for nonTech task, Normal workload per MD for nonTech Task, 0)	Dimensionless	Equation
Workload ratio per MD for Tech task	XIDZ(workload per MD for tech task , Normal workload per MD for Tech Tasks , 0)	Dimensionless	Equation
wRVU tasks completed by MA	MA Advanced tasks completed by MA+MA Only tasks completed by MA	tasks/Month	Equation
wRVU tasks completed by MD	nonTech tasks completed by MD+Tech tasks completed by MD	tasks/Month	Equation

APPENDIX G: BEHAVIOR & POLICY SENSITIVTY ANALYSIS SUMMARY

G.1 ALTERNATIVE PREFERENCES

Much of the variation in PCT can be explained by clinician preferences. Therefore, it was important to study what impact each preference has on system behavior and how important would changes in these somewhat subtle policy functions be in terms of actual PCT success. Changing preferences in the model involves using different table functions than the original ones – the ones which together permit PCT's successful trajectory. In the four figures below, the original (left side blue line) and alternative table functions (right side, multiple colors) are presented. Figure G.1 presents the preferences of MAs. Figure G.2 shows patients' preferences. Figure G.3 shows clinicians' preferences. Figure G.4 shows table functions for productivity.

In Appendix F Table F.22 there is a description of each of these preferences, and a list of the number of alternatives for each preference.

Figure G.1 MA Table Functions Original and Alternative

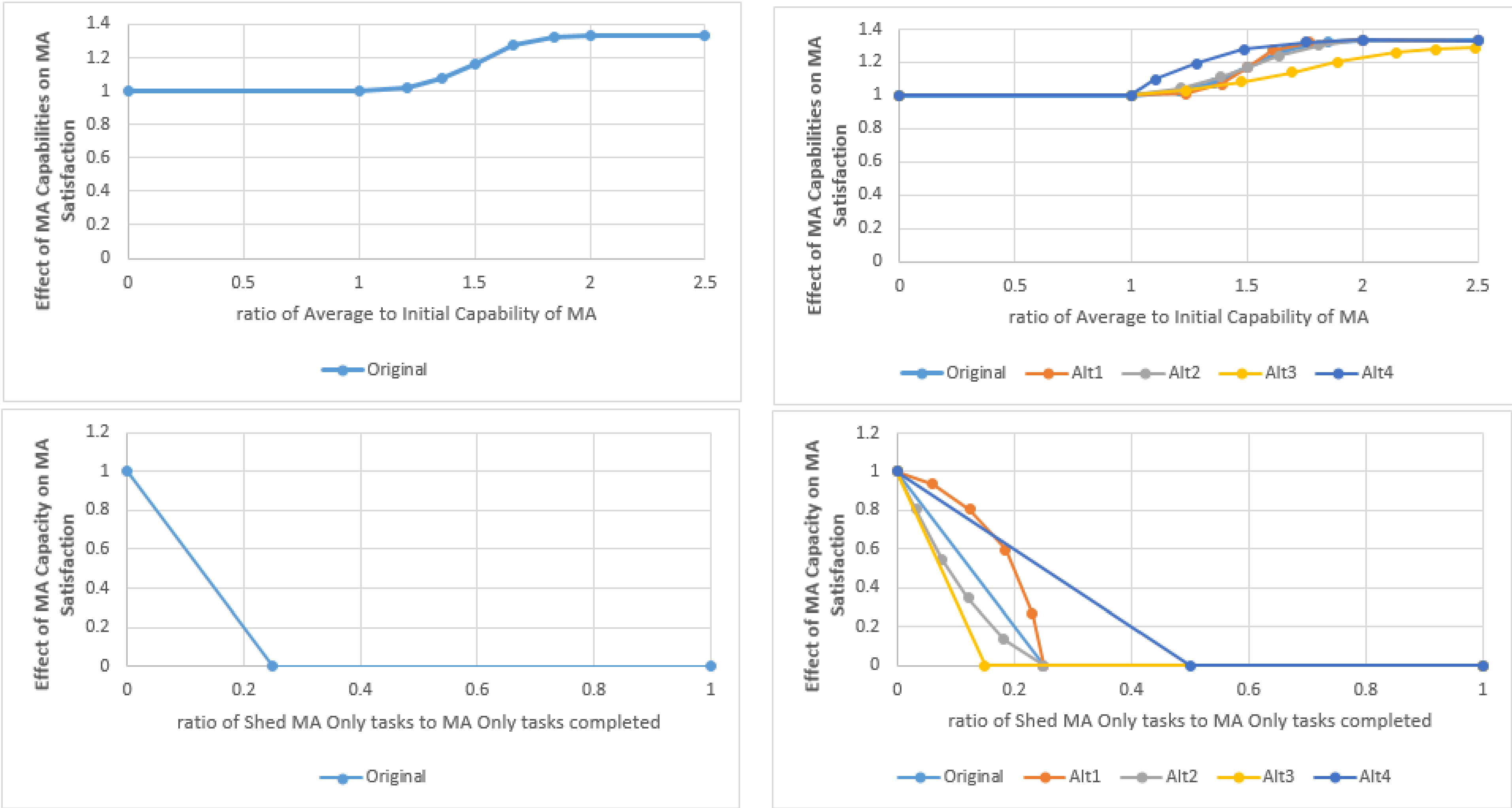


Figure G.1 MA Table Functions Original and Alternative (continued)

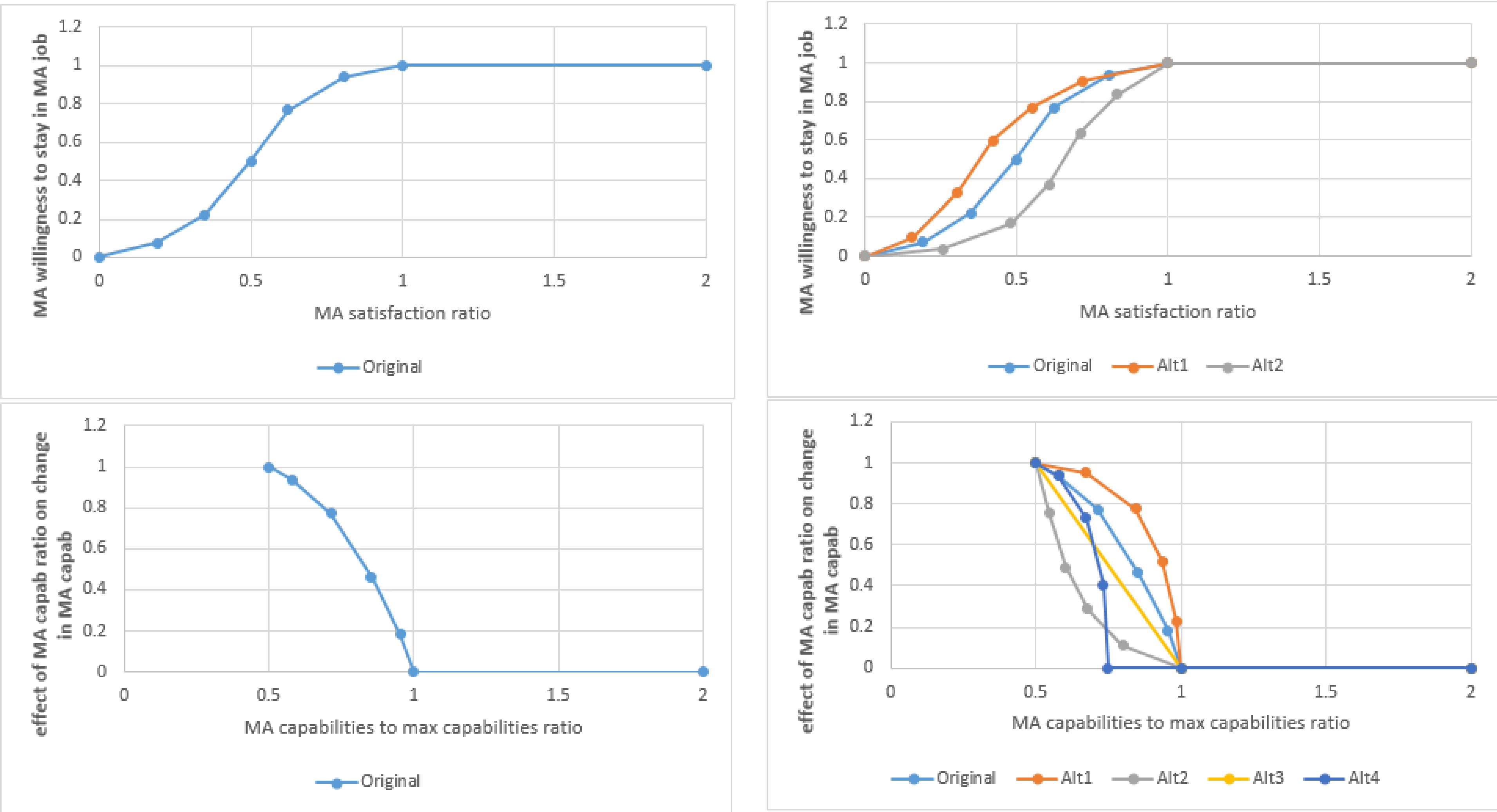
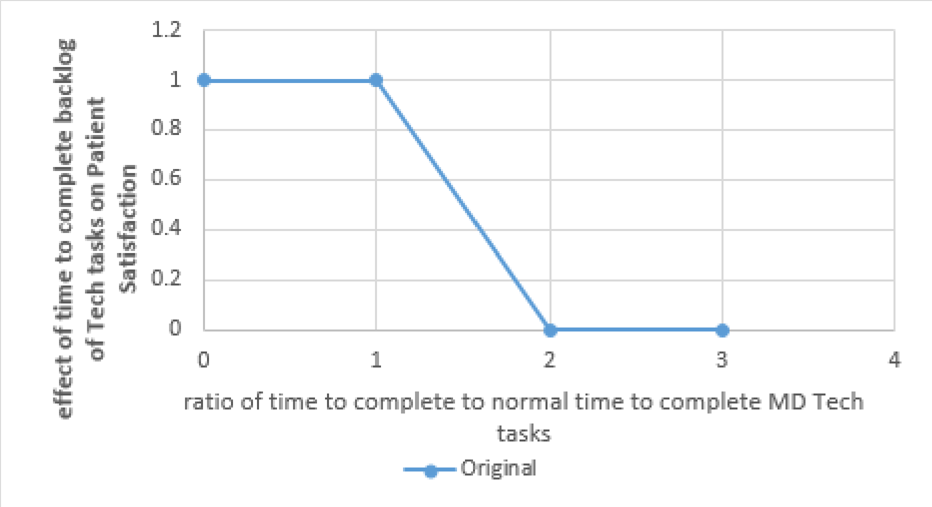


Figure G.2 Patient Table Functions Original and Alternative



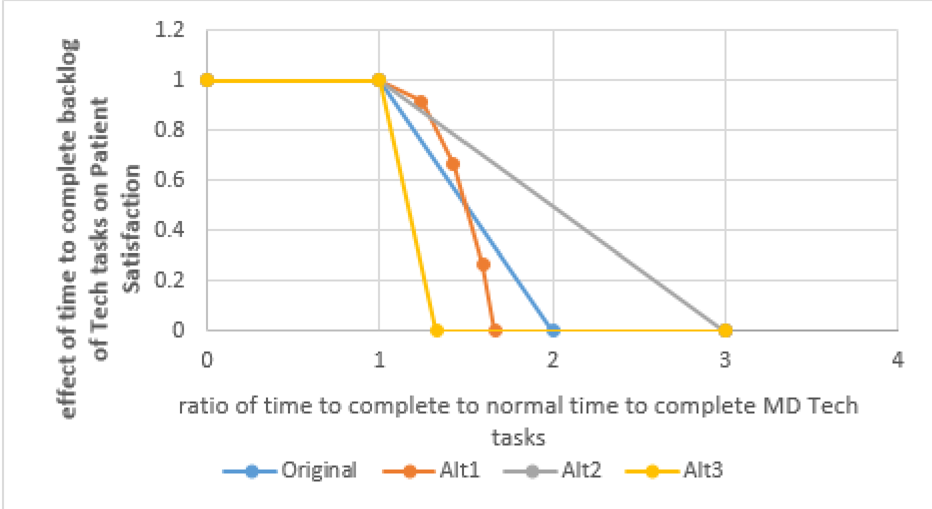


Figure G.3 MD Table Functions Original and Alternative

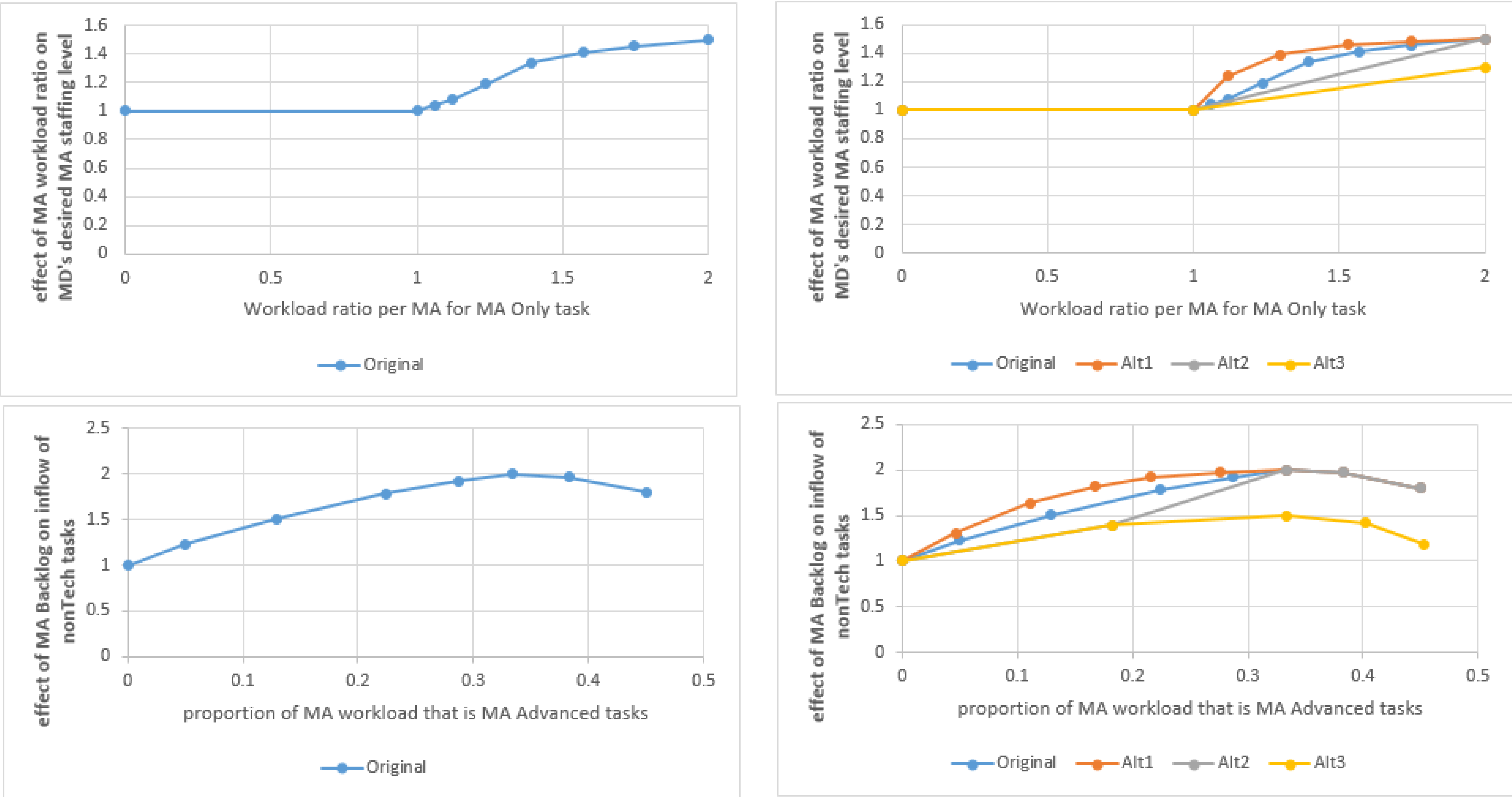


Figure G.3 MD Table Functions Original and Alternative (continued)

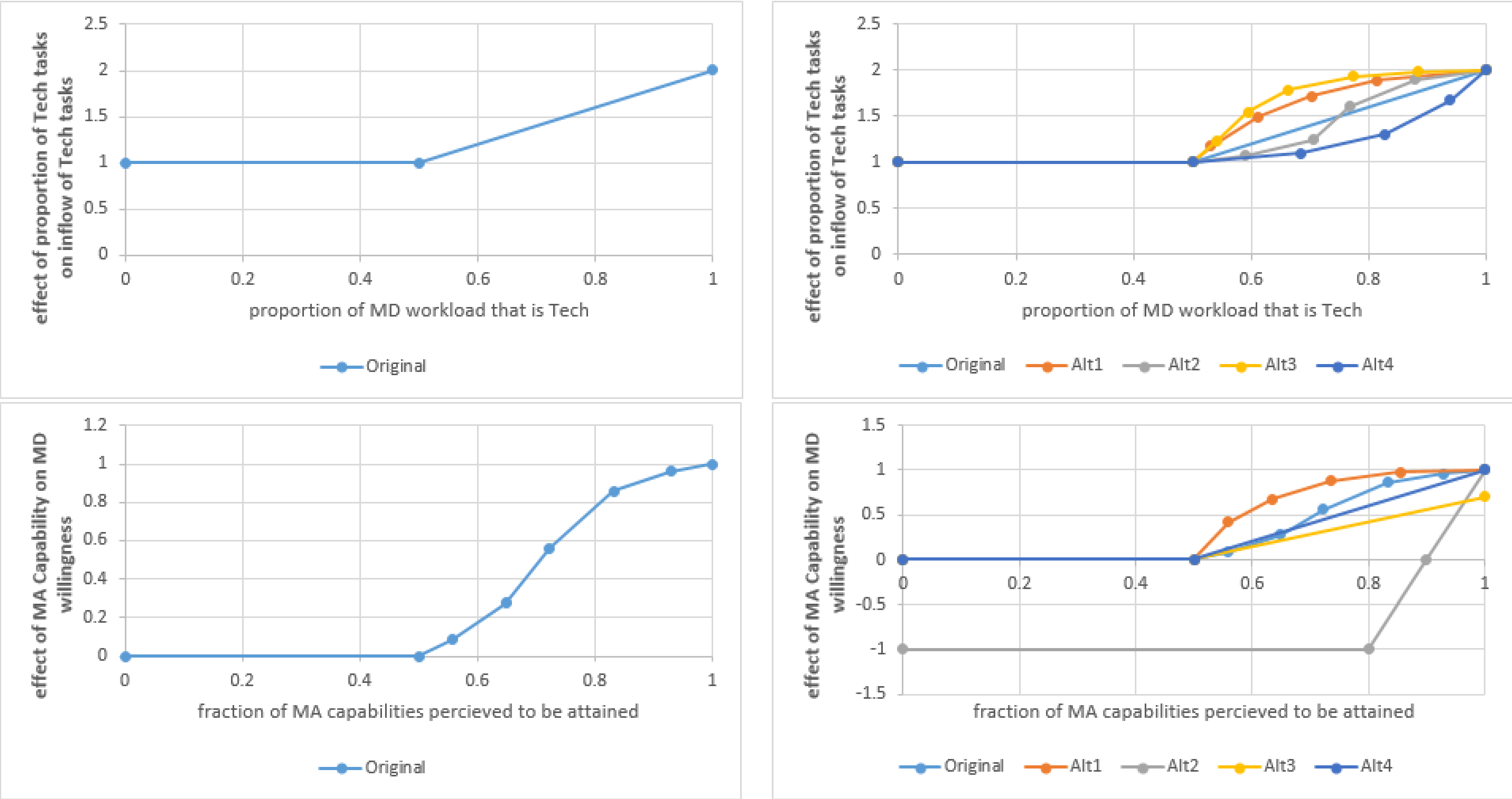


Figure G.3 MD Table Functions Original and Alternative (continued)

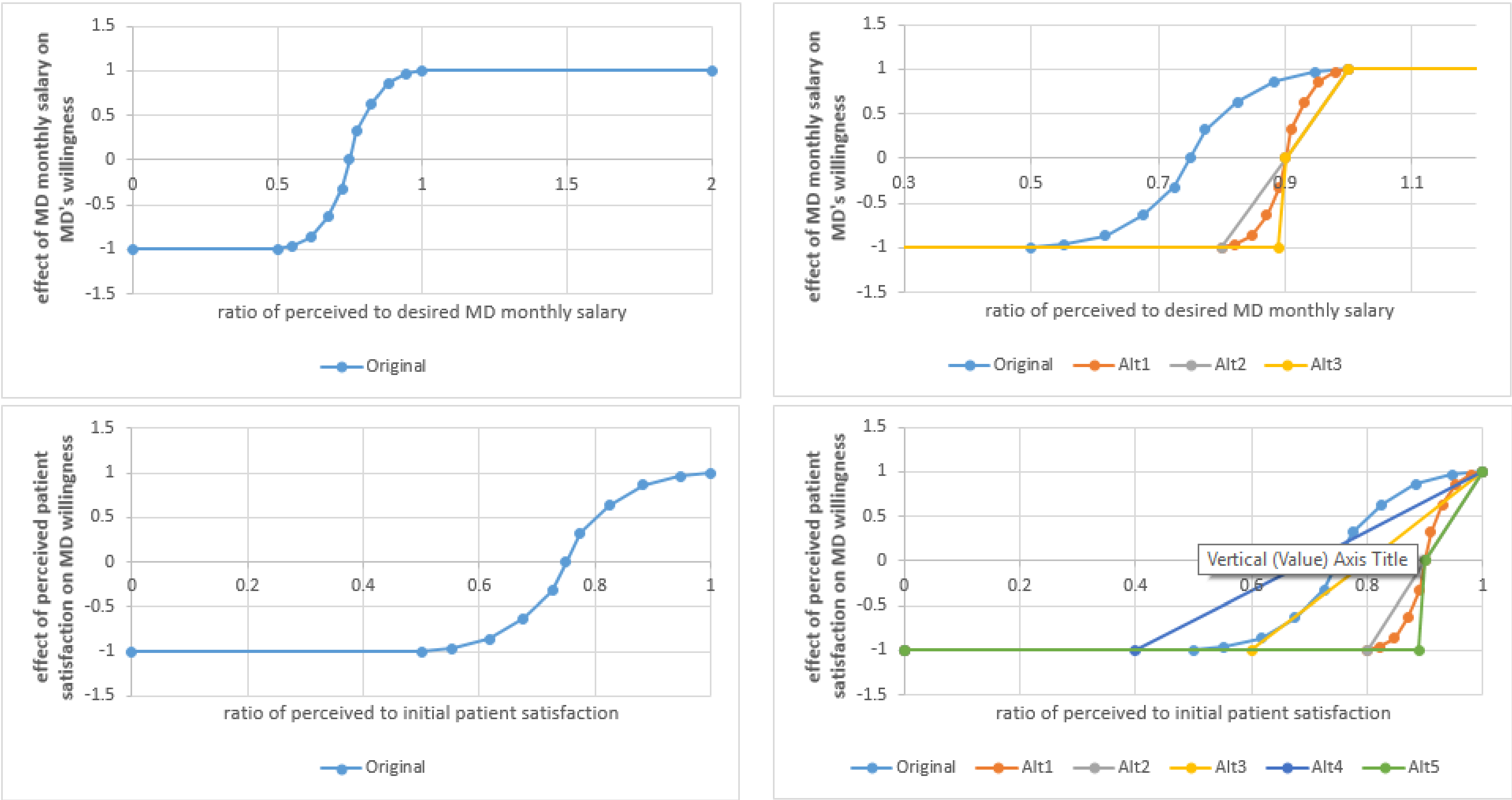


Figure G.3 MD Table Functions Original and Alternative (continued)

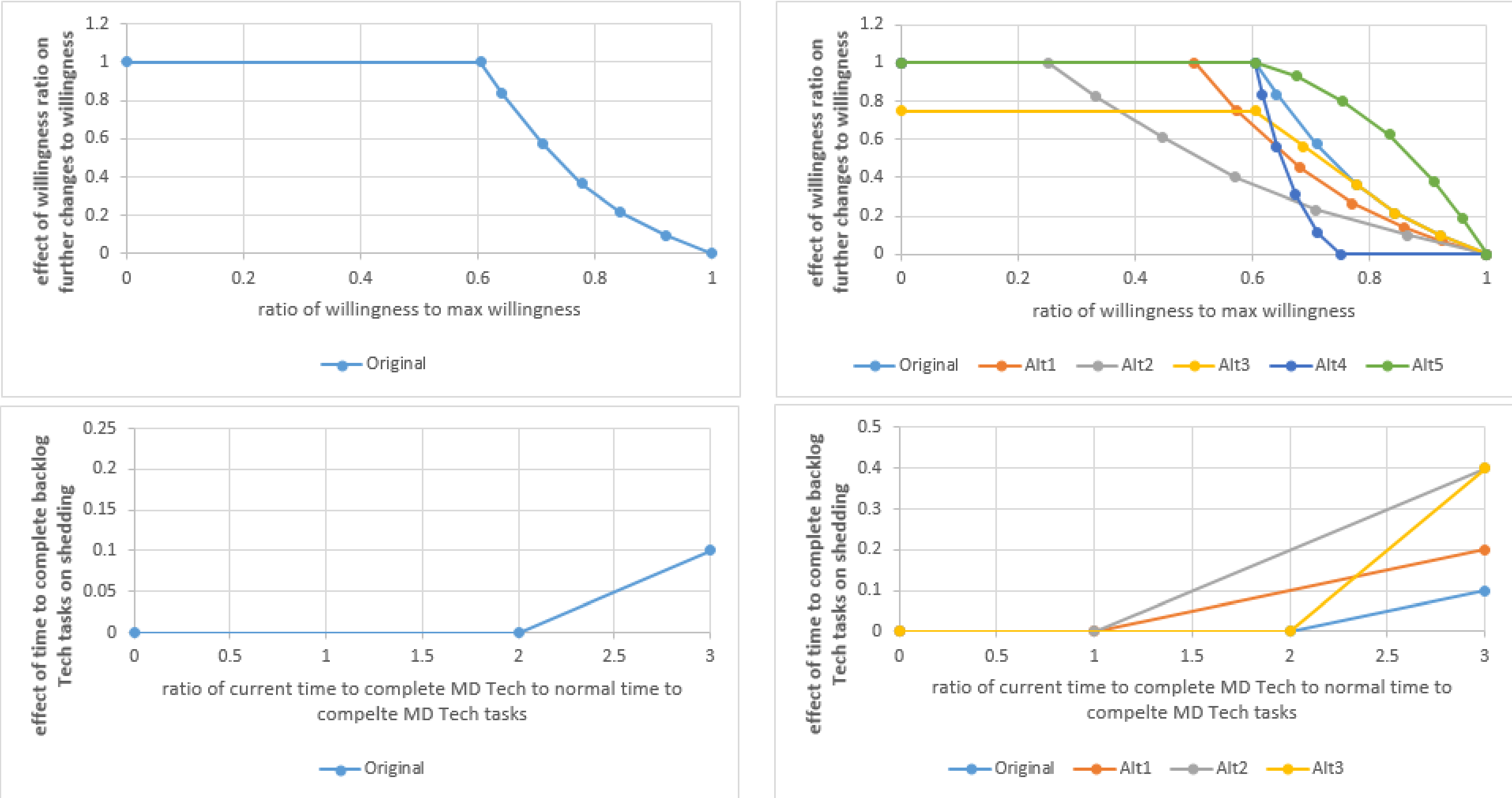
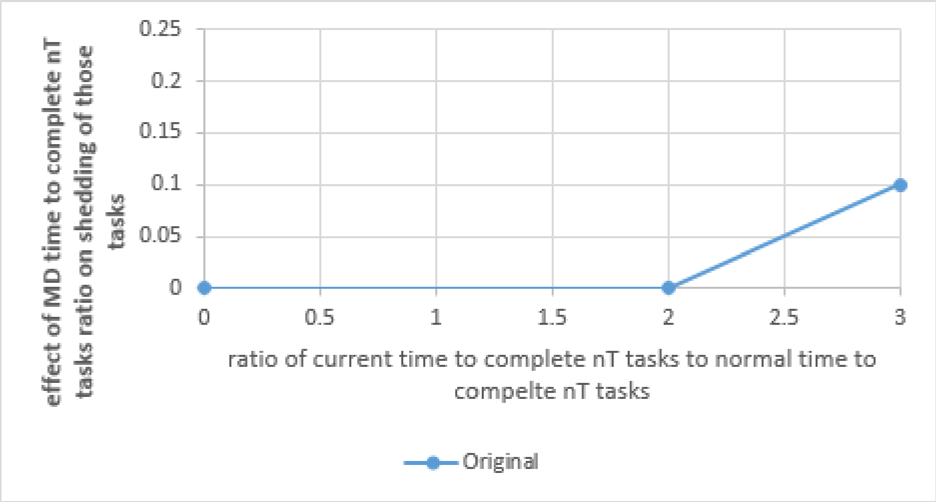


Figure G.3 MD Table Functions Original and Alternative (continued)



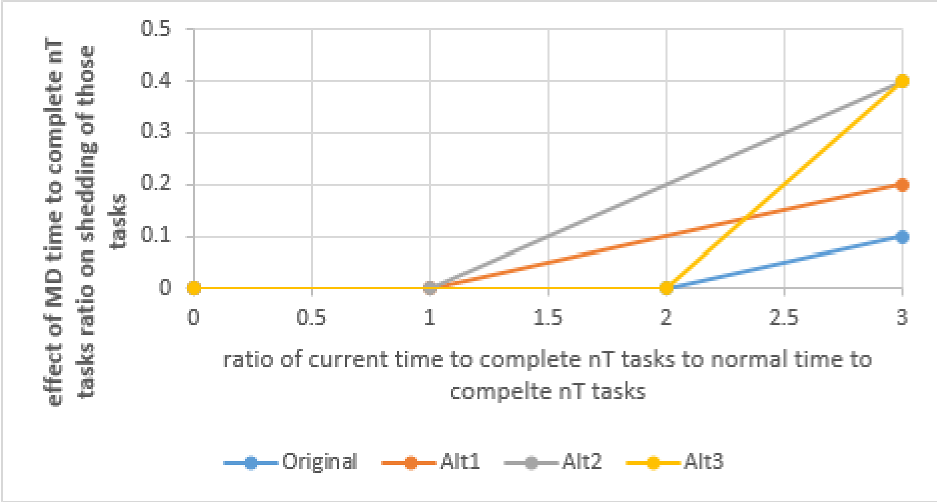
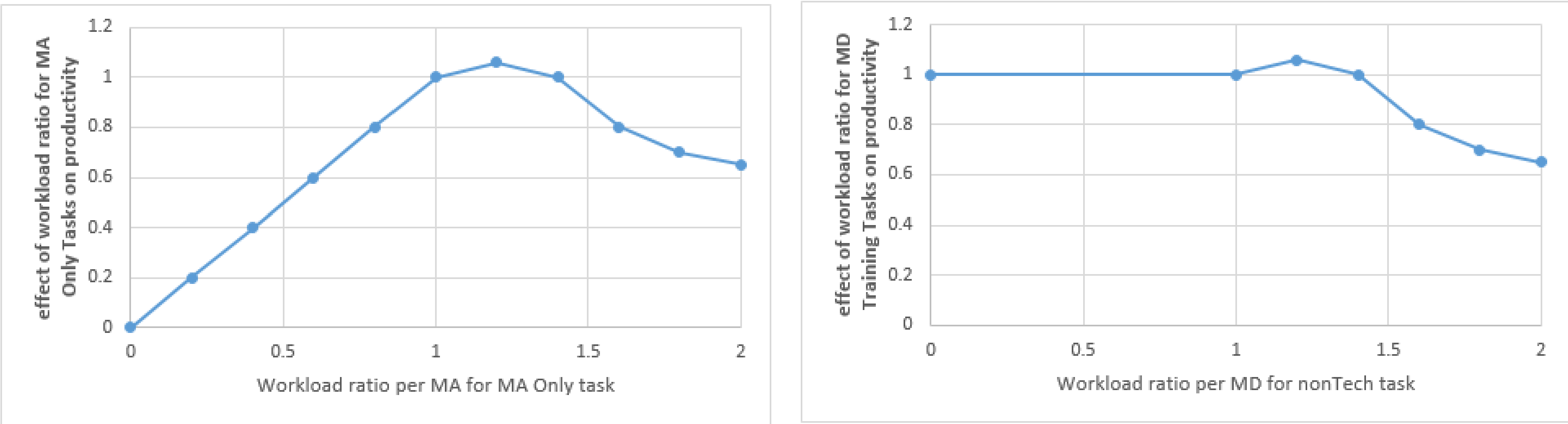


Figure G.4 Productivity Table Functions - Originals



NOTE: The table functions for MA-Advanced and MD-Technical tasks are the same shape as the one on the left of this figure. The table functions for productivity specific to training tasks are the same shape as the one on the right of this figure. Alternative table functions – and alternative formulations – were not tested for productivity because it was deemed out of scope of the problem statement.

G.2 SENSITIVITY RESULTS OF EQUILIBRIUM & TASK-SHIFTING MODELS

The results of sensitivity analysis for any one variable were assigned to one of the following four categories described below. This assignment was done by visual inspection. Delayed behavior was not considered as a change unless it exceeded *five months*. Here, I provide a description of each category:

- **No behavior change**. Sensitivity runs for this variable did not change the values for the outcome of interest.
- **Behavior and policy sensitivity was observed; as expected.** Sensitivity runs for this variable changed the values for the outcome of interest, both in the absence of task-shifting and in the presence of task-shifting. These are variables for which additional data collection is warranted to reduce the uncertainty in their parameter estimates.
- **Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters**. Doing so (in order to test whether the model is sensitive to them) was outside the scope of the current work. Future research should do so and see if they result in behavior change. If they result in no change to the outcome of interest, then the variable is only sensitive because the model was not placed in analytic equilibrium. If they do result in a behavior change, then the model is sensitive to changes in this variable and additional data collection is warranted to reduce the uncertainty in this parameter estimate.
- **Policy sensitivity was observed; as expected.** Sensitivity runs for this variable changed the values for the outcome of interest in the presence of task-shifting.

Variable-specific results are sorted by result category; then in alphabetical order by variable name in Table G.1 below. I also tested the sensitivity of the model to alternative values for multiple variables at one time and found that results were as expected (within the categories above). These results are not presented here.

The first column of Table G.1 identifies the variable. The second column indicates whether it is an initial value to a stock (or like variable). The third through sixth columns indicate the results of behavior (blue) and policy (green) sensitivity: an "x" indicates that behavior change was observed. The seventh column summarizes the findings from the preceding columns as an "x" is marked if sensitivity was found in either behavior or policy sensitivity analysis. The final column provides a brief interpretation of the result.

Table G.1. Behavior & Policy Sensitivity Results

Variable Name	Is this an initial value to a stock?	Equilibrium Only	Task-Shifting & Payment Policy			Any change?	Interpretation
			Fixed Salary	Value Payment	Volume Payment		
Adjustment time for avg clinic net profit							no behavior change
Adjustment time how long it takes for the provider to revise his thinking re Capabilities							no behavior change
Adjustment time how long it takes for the provider to revise his thinking re MD salary							no behavior change
Adjustment time how long it takes the MD to revise their thinking on pt satisfaction							no behavior change
Adjustment time to shed MD Tech tasks							no behavior change
Benefits per MA per month							no behavior change
Benefits per MD per month							no behavior change
Clinic overhead cost per month							no behavior change
Clinic Policy USD per MA Only task							no behavior change
CMS Policy clinic overhead compensation per wRVU							no behavior change
CMS Policy MD compensation per wRVU							no behavior change
Flat rate salary per MA per month							no behavior change
Initial patient satisfaction	x						no behavior change
Initial perception of MA capabilities	x						no behavior change
MA satisfaction acceptable minimum							no behavior change

Variable Name	Is this an initial value to a stock?	Equilibrium Only	Task-Shifting & Payment Policy			Any change?	Interpretation
			Fixed Salary	Value Payment	Volume Payment		
Monthly salary per MD initial perception	x						no behavior change
Normal time to complete nT for MD							no behavior change
Scheduling out							no behavior change
wRVUs per nonTech task whether completed by MD or MA							no behavior change
Desired time to complete MA tasks		x	x	x	x	x	Behavior and policy sensitivity was observed; as expected.
Kickstart amount		x	x	x	x	x	Behavior and policy sensitivity was observed; as expected.
Average panel size per MD	x	x	x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Avg number of visits per patient per month	x	x	x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Encounter length	x				x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Flat rate salary per MD per month	x		x			x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Initial capability of MA new hire	x		x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.

Variable Name	Is this an initial value to a stock?	Equilibrium Only	Task-Shifting & Payment Policy			Any change?	Interpretation
			Fixed Salary	Value Payment	Volume Payment		
Initial fraction nonTech tasks per visit	x	x	x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Initial fraction Tech tasks per visit	x	x	x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Initial ratio of MD to MA	x	x	x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Normal productivity per MA	x	x	x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Normal productivity per MD	x	x	x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Normal time to complete MD Tech tasks	x		x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Normal workload per MA	x	x	x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Normal workload per MD	x	x	x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.

Variable Name	Is this an initial value to a stock?	Equilibrium Only	Task-Shifting & Payment Policy			Any change?	Interpretation
			Fixed Salary	Value Payment	Volume Payment		
Number of MDs	x	x	x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Visit relative complexity	x		x	x	x	x	Behavior change was observed because the model was not placed in analytic equilibrium. Changes in these parameters require changes in the initial values for other parameters.
Adjustment time for MA to leave			x	x	x	x	Policy sensitivity was observed; as expected.
Adjustment time to feel the change in MA satisfaction			x	x	x	x	Policy sensitivity was observed; as expected.
Adjustment time to hire MA			x	x	x	x	Policy sensitivity was observed; as expected.
Clinic Policy incentive pay per encounter					x	x	Policy sensitivity was observed; as expected.
Clinic Policy median MD compensation per wRVU				x		x	Policy sensitivity was observed; as expected.
MD Policy Desired MD monthly Salary			x	x	x	x	Policy sensitivity was observed; as expected.
Proportion of staff work that is in visit work					x	x	Policy sensitivity was observed; as expected.
Time per task					x	x	Policy sensitivity was observed; as expected.
Time to develop willingness			x	x	x	x	Policy sensitivity was observed; as expected.
Training per unit fraction change per MA			x	x	x	x	Policy sensitivity was observed; as expected.
Training tasks needed to gain capability			x	x	x	x	Policy sensitivity was observed; as expected.
wRVUs per Tech task				x		x	Policy sensitivity was observed; as expected.