Using systems thinking to optimise health system interventions for improved maternal and child health in low-resource settings

RACHEL CASSIDY

Thesis submitted in accordance with the requirements for the degree of
Doctor of Philosophy
of the University of London

August 2023

Department of Global Health and Development

Faculty of Public Health and Policy

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

No funding received

Research group affiliation(s):
Centre for Health Economics in London (CHIL)
Health Systems and Policy Unit (HSPU)

Declaration

I, Rachel Cassidy, 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.

SIGNED: [redacted]

DATE: 4th January 2023

Abstract

Payment for performance (P4P) initiatives have been employed in low and middle-income (LMIC) countries as a means to improve the delivery and coverage of maternal and child health (MCH) services. Despite widespread implementation, there is still a lack of consensus on whether P4P is an effective initiative that leads to positive, sustained improvement in delivery of these services. There is a need to employ methods that can evaluate the pathways through which P4P alters health systems without diminishing the complex behaviour exhibited by health systems in the evaluation.

Two methods for evaluation of complex systems were used to model the impact of a P4P programme on the delivery and uptake of MCH services in Tanzania: causal loop diagrams (CLDs) and system dynamics modelling (SDM). The CLD represents relationships between variables that are important when we consider how the health system responds and transforms under P4P. The CLD was developed using qualitative data from a process evaluation of a P4P programme in Tanzania and stakeholder consultation. The CLD was then used to build a quantitative SDM, using primary (stakeholder consultation) and secondary (impact evaluation of P4P programme, official statistics and reports) data sources. In the SDM, changes in design, implementation, and context (availability and supply of drugs, access to alternative sources of funding, staffing) were tested to explore the impact on key outcomes (percentage of women who received two doses of intermittent preventive treatment during antenatal care and percentage of women who had a facility-based delivery) and the effectiveness of the programme.

The CLD pinpoints the key mechanisms underpinning provider achievement of P4P targets, reporting of health information by providers, and care seeking by the population, and identifies those mechanisms affected by P4P. For example, the availability of drugs and medical commodities was critical not only to provider achievement of P4P targets (supply of MCH services) but also to demand of services and was impacted by P4P through the availability of additional facility resources. In the SDM, severe delays in payment and change in allocation of payments (between staff and operations) impacted key outcomes, with changes in contextual factors (particularly provision of medicine) facilitating or hindering facility performance.

Recommendations for programme design must consider the impact on the holistic system, to avoid suboptimal programme impact or unintended, negative consequences. Our study shows how secondary data from an impact and process evaluation can be used to model the health system and its response to P4P, to improve our understanding of programme mechanisms and

inform the design of more effective future P4P programmes. This work will not only be relevant for P4P in Tanzania but also generate policy relevant recommendations for LMICs.

Acknowledgements

Firstly, I would like to thank my supervision team, Josephine Borghi, Neha Singh and Karl Blanchet. I will forever be thankful that I was lucky enough to have such a supportive team around me, helping me to grow as a researcher and guiding me through the PhD journey. To my advisory committee, Agnes Rwashana Semwanga, Andrada Tomoaia-Cotisel and Zaid Chalabi and fellow PhD candidate Sharif Ismail, thank you for your continued support along the winding systems thinking road. To my upgrading examiners, Natalie Savona and James Milner, for your support and confidence in the early stages of thesis development.

To the entire COSMIC team, Peter Binyaruka, Anna Foss, Chitalu Chama-Chiliba, Nkenda Sachingongu, John Maiba, Nicholaus Mziray, for your advice and encouragement the last four years. To those who participated in stakeholder interviews, thank you for giving up your precious time to support this work.

A huge thank you and my endless gratitude to my family and friends. To my parents Clare and Eamonn, and my sisters Sarah and Grace, for your unwavering support and love on this lengthy journey. To my nieces and nephews, for the very welcome distractions. To the Fysons, for welcoming me into the clan. To Pippa and Rajiv, who have been there forever, observing and loudly supporting my never-ending studenting. To TP 315, Camilla, Julia, Neha, Ros, Shefali, Silke and Sonia, your conversations, snacks and laughter offered a much-needed release from PhD pressure.

And finally, to my husband Ben, taking into your stride the chaotic events of the last four years, and ensuring I ate, slept and remained sane throughout.

This thesis is dedicated to my grandparents, whose love and strength will always inspire me.

Table of Contents

List of Tables

Table 1: Sensitivity analysis to examine the sensitivity of key outcomes (percentage of women who receive at least two doses of IPT during ANC, percentage of women who seek facility-based delivery) to changes in model parameters.

Table S1.1: Facility coverage, content of care and HMIS strengthening indicators and performance targets set during the Pwani P4P programme in Tanzania.

Table S1.2: Council Health Management Team and Regional Health Management Team performance indicators set during the pilot P4P programme in Tanzania.

Table S4.1: Population sector variable descriptions and data sources.

Table S4.2: Demand and Services sector variable descriptions and data sources.

Table S4.3: Demand and Services sector variable descriptions and data sources.

Table S4.4: Facility Commodities sector variable descriptions and data sources.

Table S4.5: Facility Commodities sector variable descriptions and data sources.

Table S4.6: Facility Operations sector variable descriptions and data sources.

Table S4.7: Facility Funding sector variable descriptions and data sources.

Table S4.8: District Manager Operations sector variable descriptions and data sources.

Table S5.1: Performance at intervention and control sites for targeted services (at least two doses of IPT during ANC (%), institutional delivery rate (%)) at baseline, short term and long term evaluation.

Table S5.2: The difference in performance between baseline and short term, and short term and long term evaluations in the intervention and control sites.

Table S5.3: The goal for model calibration for both intervention and control model runs.

Table S8.1: Sensitivity analyses for population sector parameters, change in key outcomes.

Table S8.2: Sensitivity analyses for population sector parameters, change in key outcomes (% change).

Table S8.3: Sensitivity analyses for demand and services (antenatal care) sector parameters, change in key outcomes.

Table S8.4: Sensitivity analyses for demand and services (antenatal care) sector parameters, change in key outcomes (% change).

Table S8.5: Sensitivity analyses for demand and services (facility-based delivery care) sector parameters, change in key outcomes.

Table S8.6: Sensitivity analyses for (facility-based delivery care) sector parameters, change in key outcomes (% change).

Table S8.7: Sensitivity analyses for facility commodities (antenatal care) sector parameters, change in key outcomes.

Table S8.8: Sensitivity analyses for facility commodities (antenatal care) sector parameters, change in key outcomes (% change).

Table S8.9: Sensitivity analyses for facility commodities (facility-based delivery care) parameters, change in key outcomes.

Table S8.10: Sensitivity analyses for facility commodities (facility-based delivery care) parameters, change in key outcomes (% change).

Table S8.11: Sensitivity analyses for facility operations sector parameters, change in key outcomes.

Table S8.12: Sensitivity analyses for facility operations sector parameters, change in key outcomes (% change).

Table S8.13: Sensitivity analyses for facility funding sector parameters, change in key outcomes.

Table S8.14: Sensitivity analyses for facility funding sector parameters, change in key outcomes (% change).

Table S8.15: Sensitivity analyses for district manager operations sector parameters, change in key outcomes.

Table S8.16: Sensitivity analyses for district manager operations sector parameters, change in key outcomes (% change).

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Abbreviations

2 doses of intermittent preventative treatment (IPT2)

Accident and Emergency Department (A&E)

Accountable care organisation (ACO)

Agent-based modelling (ABM)

Alberta Continuing Care System (ACCS)

Antenatal care (ANC)

Antiretroviral therapy (ART)

Caesarean delivery (CD)

Cardiac catheterization (CC)

Cardiothoracic surgery (CTS)

Catharina Hospital Eindhoven (CHE)

Causal loop diagram (CLD)

Community health fund (CHF)

Community health system (CHS)

Community health worker (CHW)

Congestive heart failure (COHF)

Council Health Management Team (CHMT)

Direct Health Facility Financing (DHFF)

Discrete-event simulation (DES)

District Executive Director (DED)

Emergency department (ED)

Focus group discussion (FGD)

Focused Antenatal Care (FANC)

General internal medicine (GIM)

Group Model Building (GMB)

Hampshire County Council (HCC)

Health Facility Governing Committee (HFGC)

Health Management Information System (HMIS)

Health Systems Research Initiative (HSRI)

Human Immunodeficiency Virus (HIV)

Ifakara Health Institute (IHI)

Integrated care system (IC)

Intensive Care Unit (ICU)

Health Service Executive (HSE)

London School of Hygiene and Tropical Medicine (LSHTM)

Long-term care (LTC)

Low- and middle-income country (LMIC)

Maternal and child health (MCH)

Maternal, newborn and child health (MNCH)

Medical Research Council (MRC)

Medical Stores Department (MSD)

Medication administration process (MAP)

Ministry of Health (MoH)

Ministry of Health and Social Welfare (MoHSW)

Mobile Stroke Unit (MSU)

National Institute for Medical Research (NIMR)

Payment for performance (P4P)

Percutaneous Coronary Intervention Centre (PCI)

Performance-based payment system (PBPS)

Pilot Management Team (PMT)

Oral polio vaccine dose at birth (OPV0)

Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA)

Prospective Health Technology Assessment (ProHTA)

Regional Health Management Team (RHMT)

Results-Based Financing (RBF)

ST-segment elevation myocardial infarction (STEMI)

Supportive living (SL)

Sustainable Development Goals (SDGs)

System dynamics model (SDM)

Total knee arthroplasty (TKA)

Vaginal delivery (VD)

Venous thromboembolism (VTE)

World Health Organisation (WHO)

BACKGROUND TO THE THESIS

1 Introduction

Abbreviations: Agent-based modelling (ABM); Causal loop diagram (CLD); Low- and middle-income country (LMIC); Maternal and child health (MCH); Payment for performance (P4P); System dynamics modelling (SDM).

In this opening chapter, the background information to the thesis is presented, with overall goal, objectives, ethics clearance and role of the candidate in the wider study that forms the basis for the thesis.

1.1 Maternal and child health and payment for performance

Globally, maternal and under-five mortality is estimated to have decreased by 45% (Alkema *et al.* 2016; UNICEF 2021a) and 61% (UNICEF 2021b) respectively since 1990. Despite this advancement, there are wide disparities in health outcomes between world regions; 86% of maternal deaths worldwide occur in South Asia and sub-Saharan Africa (UNICEF 2021a). In this latter region, 1 in 13 children die before their fifth birthday, and the risk of under-five mortality is 15 times greater than for a child born in a high-income country (UNICEF 2021b).

Payment for performance (P4P) programmes have been employed globally to promote quality and demand for health services with an aim of improving maternal and child health (MCH) outcomes (Das *et al.* 2016). P4P uses financial incentives to promote certain behaviour in order to achieve service or health targets (Mannion and Davies 2008). The general mechanisms for system change in P4P are that giving health workers, facilities and managers this incentive will increase their extrinsic motivation to achieve set targets, which in turn will increase the quality and delivery of health services. A higher quality of service will stimulate a higher volume of patients attending the facility, leading to a continued and sustained increase in patient health and access to services and improved patient outcomes (Meessen *et al.* 2011; Njuki *et al.* 2012).

1.2 Evaluation of payment for performance in low- and middle-income countries

Evaluations of P4P in low- and middle-income countries (LMICs) have focussed on exploration of programme effect on key health system outcomes, such as health worker motivation (Bhatnagar and George 2016; Engineer *et al.* 2016; Shen *et al.* 2017) and job satisfaction (Engineer *et al.* 2016; Shen *et al.* 2017), availability of facility resources (Bhatnagar and George 2016; Das *et al.* 2016; Engineer *et al.* 2016), quality of care (Bhatnagar and George 2016; Das *et al.* 2016; Engineer *et al.* 2016), health worker supervision by managers (Paul *et al.* 2014; Bhatnagar and George 2016; Mayumana *et al.* 2017) and community engagement in service delivery (Engineer *et al.* 2016; Mayumana *et al.* 2017). In 2012, a Cochrane review exploring P4P programme impact on health system functioning (Witter *et al.* 2012) found that the size and direction of programme effect was mixed, with 'very low' quality and quantity of evidence. Implementation was found to vary widely between settings, making it difficult to determine the impact of this intervention in LMICs.

A decade later, an update to the original Cochrane review has now been released (Diaconu *et al.* 2021). Whilst the evidence base for programme effect in LMIC settings has increased, there has only been a gradual increase in study quality and evidence on effect. Consistent with earlier review findings, considerable heterogeneity was identified in programme implementation across different contexts. Crucially, the effect of the programme was found to be dependent on a wide range of co-existing, interacting elements such as P4P programme design (payment per service (Witvorapong and Foshanji 2016) vs. target thresholds (Gertler and Vermeersch 2013)), incentives for facility health workers and district managers (Binyaruka *et al.* 2015) vs. community health workers (Kliner *et al.* 2015) etc.), any additional provider held funding (Lagarde *et al.* 2021), health system and programme context. The authors call for a shift in research focus towards further understanding on the contextually sensitive pathways to effect for P4P programmes using dynamic approaches (approaches that account for health system and intervention complexity in the evaluation). Other reviews that focus on the effect of P4P on quality of MCH services in LMICs make similar calls for analysis of pathways to effect that reflect the complexity of the system within which P4P is embedded, together with the complexity of the programme itself. There have also been calls to explore the influence of varied design and implementation processes on outcomes. This has been driven by a need to better understand why there has been mixed effects of P4P on healthcare outcomes (Das *et al.* 2016) and monitoring of effects beyond those services that were directly incentivised (Patel 2018).

There is a need for an 'open box' evaluation of P4P, exploring pathways to impact for the programme and how changes in the design, implementation and context affect programme success. A whole systems perspective is required to identify facilitators or barriers to programme implementation and how these factors become more or less important over time, to gain knowledge that is critical for optimising further implementation of the programme.

1.3 Systems thinking for health systems research

As complex adaptive systems (Kitson *et al.* 2017), health systems exhibit extraordinary intricacy in relationships among highly heterogenous groups (e.g. health workers, patients, managers) (Paina and Peters 2012). System processes that consist of many inter-connections, self-organising and emergent behaviour, non-linearity and linked feedback loops with system tipping points make health system behaviour difficult to predict (Lipsitz 2012). Interventions that aim to optimise health system behaviour are themselves complex; programme design may aim to incentivise stakeholders operating at different levels of the health system, where they make decisions based on their local context and environment, with data reporting often required for programme evaluation (Skivington *et al.* 2021). The overlaying of the intervention on top of the existing health system structure gives rise to an additional layer of complexity; the pathways through which the intervention influences system behaviour, dependent on programme design, implementation and context, will impact success or failure of the programme (Skivington *et al.* 2021).

Conventional methods for evaluation isolate and quantify individual effects of programmes on health system behaviour, assuming linear cause effect mechanisms for impact, disregarding the inherent complex nature of health systems in evaluations (Borghi and Chalabi 2017). Discounting system complexity when considering programme implementation can also give rise to unexpected or paradoxical behaviour and suboptimal service quality and delivery outcomes (Adam *et al.* 2012; Paina and Peters 2012). Health systems and implementation research require approaches that can retain system complexity in analysis, such as those that derive from systems thinking (de Savigny *et al.* 2017a).

Systems thinking is an approach that encourages exploration of relationships and interactions between different system 'agents' (stakeholders) and system sectors (e.g. financial, political, social), defining the boundaries and drivers for system behaviour (Gates 2016; McGill *et al.* 2021). There are a variety of tools that allow researchers to take a systems thinking approach to research, with use depending on the research study question (de Savigny *et al.* 2017b). For example, if the goal for research is to explore, describe and quantify stakeholder relationships and engagement, network analysis would be an appropriate tool for the study. If the study objectives detail identification of system processes to accomplish certain prioritised tasks, process mapping would be a fitting tool for analysis. If understanding of drivers for suboptimal or problematic system behaviour and testing solutions to promote desired behaviour is the research need, system dynamics methodology would be a suitable approach.

1.3.1 System dynamics methodology

When reflecting on the gaps in research identified by the systematic reviews of P4P programme impact in LMICs, there is clear potential for a systems thinking approach to evaluation, specifically the system dynamics methodology. System dynamics methodology originated in business management (Sterman 2000; Shepherd 2014), but now has widespread application from environment science to engineering, economics to recycling and waste research (Fiddaman 2002; Ford 2007; Radzicki 2009; Popli *et al.* 2017; Kunc *et al.* 2018; Lobontiu 2018). Two types of macro-level systems mapping and modelling methods are used in system dynamics; causal loop diagrams (CLDs) and system dynamics modelling (SDM).

Causal loop diagrams

CLDs are static maps that describe relationships and interactions between different system elements and sectors (Tomoaia-Cotisel *et al.* 2017). They help to visualise structural drivers for system behaviour; they can be used ex-ante (e.g. policy design) or retrospectively (e.g. policy evaluation) to aid identification of system delays and bottlenecks, and facilitators and barriers to optimal health system behaviour. This holistic systems perspective can yield knowledge on possible spill over effects to wider parts of the system, and any unexpected or unintended consequences for policy implementation, preventing suboptimal outcomes. System leverage points, targets for policies that are expected to induce optimal system behaviour, can be identified in the CLD and incorporated into the design of policies (Kwamie *et al.* 2014; Rwashana *et al.* 2014). CLDs have also been used outside of policy evaluation to identify mechanisms for health system resilience (Ozawa *et al.* 2016; Jamal *et al.* 2020), drivers for suboptimal childhood vaccination (Rwashana *et al.* 2009; Varghese *et al.* 2014; Kanniyan *et al.* 2021) and drivers for refugee and host community demand for healthcare (Noubani *et al.* 2020; Zablith *et al.* 2021).

System dynamics modelling

Where study design calls for testing changes to policy design and evaluation of subsequent impact on health system behaviour, researchers often combine CLD development with SDM. These quantitative simulation models are used to investigate macro-level system behaviour, using differential equations to simulate changes to system variables over time (Pruyt 2017). They can be used for policy evaluation before costly implementation or service restructuring, providing a relatively risk-free and low budget route for exploration of health system response to policies. In a similar vein to CLD methodology, they can help pinpoint system parameters that are causing bottlenecks or impeding the success of initiatives, unintended or unexpected outcomes and potential spill over effects to the wider system, with a crucial addition (Ansah *et al.* 2014; Rashwan *et al.* 2015; Mahmoudian-Dehkordi and Sadat 2017). System behaviour is likely to fluctuate over time as a result of policy design, implementation, context and other such stimuli; using simulation modelling allows researchers to not only pinpoint 'where' bottlenecks or delays are occurring, but 'when' and in response to 'what' changes in the wider system and at various time points. Production of a user-friendly model interface is also possible, facilitating communication of the model and results to stakeholders responsible for decision-making and policy implementation (Semwanga *et al.* 2016). SDMs have also been used outside of health system policy evaluation, to explore drivers for undesirable outcomes in emergency departments (Lane 1998, 2000), estimating future demand for care (Mielczarek and Zabawa 2016) and reasoning for variation in physician decision-making (Ghaffarzadegan *et al.* 2013).

1.4 Rationale for PhD

Despite the clear value of using a system dynamics approach for health systems research, there has been little application of it within LMIC settings, compared to research in high-income countries. System maps and models can be developed with primary and/or secondary data sources (including data collected for previous programme evaluations), enabling use in settings where access to stakeholders is not always guaranteed (e.g. humanitarian settings). It is vital that further research using a systems thinking approach is undertaken, to further our understanding of health system functioning in LMICs, and given the greater resource constraints, shed light on optimal programme designs before costly implementation.

Four studies have used a system dynamics approach to model P4P programmes in LMIC settings (Meker and Barlas 2015; Alonge *et al.* 2017; Renmans *et al.* 2017; Singh *et al.* 2021). Given the existing issues with conventional evaluations of P4P, indicative of low quality of evidence and mixed effects on health system functioning (Das *et al.* 2016; Patel 2018; Diaconu *et al.* 2021), there is great potential here for further exploration of programme pathways to effect and generation of recommendations for implementation using system dynamics.

1.5 Goal and objectives

The goal of this thesis is to use a systems thinking approach to further understand pathways to impact for P4P programmes and recommendations for more effective implementation in low-income settings. Specific objectives are to:

1. Determine current use and application of systems thinking methods (CLDs and SDMs) for health systems research, with reflection on use in LMIC health system settings.
2. Identify health system factors and feedback loops that facilitate or hinder the implementation of P4P programmes and its overall effectiveness.
3. Identify system leverage points which should be considered in the design of P4P programmes.
4. Explore how variations in the implementation, design and context of P4P could result in different outcomes to inform future design of P4P programmes.
5. Provide guidance on future health systems research using systems thinking to encourage uptake in LMIC settings.

1.6 Thesis outline

The thesis comprises of nine chapters:

Chapter 1 reviews the background information and presents the rationale and structure for the thesis. This includes the overall goal, objectives, ethical clearances and role of the candidate in the wider study that forms the basis for the thesis.

Chapter 2 is a systematic literature review (paper) identifying and describing application of SDM for health systems research, with reflection on current application in low-income settings and avenues for future research (**Objective 1**).

Chapter 3 provides background on the study setting (Tanzania), healthcare system (MCH services) and P4P programme that is the focus of this thesis.

Chapters 4 and 5 provide details on the methodologies employed in this thesis (CLD and SDM), including data use, model development and validation.

Chapters 6, 7 and 8 are a series of papers that illustrate study results. **Chapter 6** is an application of CLD methodology to further understand pathways to impact for P4P programmes, identify system delays and bottlenecks that affect service delivery and success of the programme and provide recommendations for the design of the programme based on key system leverage points **(Objectives 2 and 3)**. **Chapter 7** is a guidance piece for study design utilising CLD methodology for health systems research in low-income settings, based on the current literature and the candidates own experience of application (**Objectives 1 and 5**). **Chapter 8** is an application of SDM methodology, building on the foundation and knowledge gained from CLD analysis, to explore how changes in the design, implementation and context of the P4P programme can affect programme success, providing recommendations for future implementation and design (**Objective 4**).

Chapter 9 provides a summary of thesis findings and strengths and limitations of the work, closing with research and policy implications and recommendations.

1.7 Role of the candidate and declaration of funding

For the majority of PhD enrolment, the candidate was a Research Fellow at the London School of Hygiene and Tropical Medicine on the COSMIC Project (London School of Hygiene and Tropical Medicine 2022). The project aims to utilise computer modelling to optimise the design of health system programmes (P4P), to improve MCH service delivery and utilisation in low-resource settings. It is a 4.5 year project which started in 2018 and is a partnership between the London School of Hygiene and Tropical Medicine, Ifakara Health Institute in Tanzania, University of Zambia, Makerere University, Geneva Centre of Humanitarian Studies and Kuwait University. The project is funded by the Health Systems Research Initiative grant (MR/R013454/1), a joint initiative by Foreign, Commonwealth and Development Office, the Medical Research Council and Wellcome Trust, in collaboration with the Economic and Social Research Council.

The project objectives pertain to using various modelling approaches, namely CLD, SDM and agent-based modelling (ABM) to explore pathways to impact for P4P programmes and provide recommendations for future programme implementation, using Tanzania's programme as a case study. The candidate led the systems thinking (CLD and SDM) workstream of the project and developed the outline of work presented in this thesis, in discussion with the co-principal investigators for the COSMIC project who were also the candidate's PhD supervisors. The project contains an additional workstream, investigating the external validity of the models originally developed for Tanzania to another comparable setting (Zambia), which the candidate is also supporting. In relation to this thesis, the candidate:

- Developed the study design used in this thesis, in discussion with the supervision team and advisory committee.
- Developed the data collection tools and conducted data collection, in discussion with the supervision team and advisory committee, supported in data collection by a fellow project researcher.
- Developed study models and led analysis of study results.
- Drafted four first-author research papers, with paper revisions based on co-author feedback (Chapters 2, 6, 7 and 8).
- Wrote the remaining chapters of the thesis (Chapters 1, 3, 4, 5 and 9).

The candidate did not receive funding for PhD study and was self-funded for the entire duration.

1.8 Ethics approval

Ethics approval was sought for this study from London School of Hygiene and Tropical Medicine Research Ethics Committee (UK), Ifakara Health Institute Review Board (Tanzania), and Tanzania National Institute for Medical Research. Ethical clearance documentation from each institute can be found in Appendix 1, with study tools, participant information sheets and consent forms included in Appendix 2.

References

Adam T, Hsu J, de Savigny D *et al.* 2012. Evaluating health systems strengthening interventions in low-income and middle-income countries: are we asking the right questions? *Health Policy Plan* **27**: iv9–19.

Alkema L, Chou D, Hogan D *et al.* 2016. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Maternal Mortality Estimation Inter-Agency Group. *Lancet* **387**: 462–74.

Alonge O, Lin S, Igusa T *et al.* 2017. Improving health systems performance in low- and middle-income countries: a system dynamics model of the pay-for-performance initiative in Afghanistan. *Health Policy Plan* **32**: 1417–26.

Ansah JP, Eberlein RL, Love SR *et al.* 2014. Implications of long-term care capacity response policies for an aging population: A simulation analysis. *Health Policy (New York)* **116**: 105–13.

Bhatnagar A, George AS. 2016. Motivating health workers up to a limit: partial effects of performance-based financing on working environments in Nigeria. *Health Policy Plan* **31**: 868–77.

Binyaruka P, Patouillard E, Powell-Jackson T *et al.* 2015. Effect of Paying for Performance on Utilisation, Quality, and User Costs of Health Services in Tanzania: A Controlled Before and After Study. Ostermann J (ed.). *PLoS One* **10**: e0135013.

Borghi J, Chalabi Z. 2017. Square peg in a round hole: re-thinking our approach to evaluating health system strengthening in low-income and middle-income countries. *BMJ Glob Heal* **2**: e000406.

Das A, Gopalan SS, Chandramohan D. 2016. Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review. *BMC Public Health* **16**: 321.

Diaconu K, Falconer J, Verbel A *et al.* 2021. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev* **2021**, DOI: 10.1002/14651858.CD007899.pub3.

Engineer CY, Dale E, Agarwal A *et al.* 2016. Effectiveness of a pay-for-performance intervention to improve maternal and child health services in Afghanistan: a cluster-randomized trial. *Int J Epidemiol* **45**: 451–9.

Fiddaman TS. 2002. Exploring policy options with a behavioral climate-economy model. *Syst Dyn Rev* **18**: 243–67.

Ford A. 2007. Global Warming and System Dynamics. *International Conference of the System Dynamics Society*

Gates EF. 2016. Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Eval Program Plann* **59**: 62–73.

Gertler P, Vermeersch C. 2013. *Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes*. Date of access: December 12, 2022. URL: https://www.nber.org/system/files/working_papers/w19046/w19046.pdf

Ghaffarzadegan N, Epstein AJ, Martin EG. 2013. Practice Variation, Bias, and Experiential Learning in Cesarean Delivery: A Data-Based System Dynamics Approach. *Health Serv Res* **48**: 713–34.

Jamal Z, Alameddine M, Diaconu K *et al.* 2020. Health system resilience in the face of crisis: analysing the challenges, strategies and capacities for UNRWA in Syria. *Health Policy Plan* **35**: 26–35.

Kanniyan B, Govindarajan P K, Sheela P Haveri *et al.* 2021. Enactment of System Thinking Framework for Convolution of Immunization Service Governance. *Int J Res Pharm Sci* **12**: 849–56.

Kitson A, Brook A, Harvey G *et al.* 2017. Using Complexity and Network Concepts to Inform Healthcare Knowledge Translation. *Int J Heal policy Manag* **7**: 231–43.

Kliner M, Canaan M, Ndwandwe SZ *et al.* 2015. Effects of financial incentives for treatment supporters on tuberculosis treatment outcomes in Swaziland: a pragmatic interventional study. *Infect Dis Poverty* **4**: 29.

Kunc M, Mortenson MJ, Vidgen R. 2018. A computational literature review of the field of System Dynamics from 1974 to 2017. *J Simul* **12**: 115–27.

Kwamie A, Dijk H van, Agyepong IA. 2014. Advancing the application of systems thinking in health: realist evaluation of the Leadership Development Programme for district manager decision-making in Ghana. *Heal Res Policy Syst* **12**: 29.

Lagarde M, Burn S, Lawin L *et al.* 2021. *Exploring the Impact of Performance-Based Financing on Health Workers' Performance in Benin*. Date of access: December 12, 2022. URL: https://www.rbfhealth.org/sites/rbf/files/Benin RBFHRH report.pdf

Lane D. 1998. Emergency but no accident: a system dynamics study of casualty waiting times in the British NHS. *Eurohealth (Lond)* **4**.

Lane DC. 2000. Looking in the wrong place for healthcare improvements: a system dynamics study of an accident and emergency department. *J Oper Res Soc* **51**.

Lipsitz LA. 2012. Understanding health care as a complex system: the foundation for unintended consequences. *JAMA* **308**: 243–4.

Lobontiu N. 2018. *System Dynamics for Engineering Students*. Academic Press (ed.). Elsevier

London School of Hygiene and Tropical Medicine. 2022. COSMIC: Using Computer Modelling to Optimise the Design of Health System Programmes. Date of access: November 20, 2022. URL: https://www.lshtm.ac.uk/research/centres-projects-groups/cosmic

Mahmoudian-Dehkordi A, Sadat S. 2017. Sustaining critical care: using evidence-based simulation to evaluate ICU management policies. *Health Care Manag Sci* **20**: 532–47.

Mannion R, Davies HTO. 2008. Payment for performance in health care. *BMJ* **336**: 306–8.

Mayumana I, Borghi J, Anselmi L *et al.* 2017. Effects of Payment for Performance on accountability mechanisms: Evidence from Pwani, Tanzania. *Soc Sci Med* **179**: 61–73.

McGill E, Er V, Penney T *et al.* 2021. Evaluation of public health interventions from a complex systems perspective: A research methods review. *Soc Sci Med* **272**: 113697.

Meessen B, Soucat A, Sekabaraga C. 2011. Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? *Bull World Health Organ* **89**: 153–6.

Meker T, Barlas Y. 2015. Dynamic Consequences of Performance-Based Payment Systems in Public Hospitals. *Syst Res Behav Sci* **32**: 459–80.

Mielczarek B, Zabawa J. 2016. Modeling Healthcare Demand Using a Hybrid Simulation Approach. *Proceedings of the 2016 Winter Simulation Conference*. IEEE Press, 1535–1546.

Njuki R, Okal J, Warren CE *et al.* 2012. Exploring the effectiveness of the output-based aid voucher program to increase uptake of gender-based violence recovery services in Kenya: A qualitative evaluation. *BMC Public Health* **12**: 426.

Noubani A, Diaconu K, Ghandour L *et al.* 2020. A community–based system dynamics approach for understanding factors affecting mental Health and Health seeking behaviors in Beirut and Beqaa regions of Lebanon. *Global Health* **16**: 28.

Ozawa S, Paina L, Qiu M. 2016. Exploring pathways for building trust in vaccination and strengthening health system resilience. *BMC Health Serv Res* **16**: 639.

Paina L, Peters DH. 2012. Understanding pathways for scaling up health services through the lens of complex adaptive systems. *Health Policy Plan* **27**: 365–73.

Patel S. 2018. Structural, institutional and organizational factors associated with successful pay for performance programmes in improving quality of maternal and child health care in low and middle income countries: a systematic literature review. *J Glob Health* **8**, DOI: 10.7189/jogh.08.021001.

Paul E, Sossouhounto N, Eclou DS. 2014. Local stakeholders' perceptions about the introduction of performance-based financing in Benin: a case study in two health districts. *Int J Heal Policy Manag* **3**: 207–14.

Popli K, Sudibya GL, Kim S. 2017. A Review of Solid Waste Management using System Dynamics Modeling. *J Environ Sci Int* **26**: 1185–200.

Pruyt E. 2017. System dynamics: a tool for modelling and testing solutions. In: de Savigny D, Blanchet K, Adam T. *Applied Systems Thinking for Health Systems Research : A Methodological Handbook*. McGraw-Hill Education, 173–94.

Radzicki MJ. 2009. System Dynamics and Its Contribution to Economics and Economic Modeling. *Encyclopedia of Complexity and Systems Science*. New York, NY: Springer New York, 8990–9000.

Rashwan W, Abo-Hamad W, Arisha A. 2015. A system dynamics view of the acute bed blockage problem in the Irish healthcare system. *Eur J Oper Res* **247**: 276–93.

Renmans D, Holvoet N, Criel B. 2017. Combining Theory-Driven Evaluation and Causal Loop Diagramming for Opening the 'Black Box' of an Intervention in the Health Sector: A Case of Performance-Based Financing in Western Uganda. *Int J Environ Res Public Health* **14**: 1007.

Rwashana AS, Nakubulwa S, Nakakeeto-Kijjambu M *et al.* 2014. Advancing the application of systems thinking in health: understanding the dynamics of neonatal mortality in Uganda. *Heal*

Res Policy Syst **12**: 36.

Rwashana AS, Williams DW, Neema S. 2009. System dynamics approach to immunization healthcare issues in developing countries: A case study of Uganda. *Health Informatics J* **15**: 95–107.

de Savigny D, Blanchet K, Adam T. 2017a. *Applied Systems Thinking for Health Systems Research : A Methodological Handbook*. McGraw-Hill Education

de Savigny D, Blanchet K, Adam T. 2017b. Conclusion: mapping methods to research challenges. In: de Savigny D, Blanchet K, Adam T. *Applied Systems Thinking for Health Systems Research : A Methodological Handbook*. McGraw-Hill Education, 226–7.

Semwanga AR, Nakubulwa S, Adam T. 2016. Applying a system dynamics modelling approach to explore policy options for improving neonatal health in Uganda. *Heal Res Policy Syst* **14**: 35.

Shen GC, Nguyen HTH, Das A *et al.* 2017. Incentives to change: effects of performance-based financing on health workers in Zambia. *Hum Resour Health* **15**: 20.

Shepherd SP. 2014. A review of system dynamics models applied in transportation. *Transp B Transp Dyn* **2**: 83–105.

Singh NS, Kovacs RJ, Cassidy R *et al.* 2021. A realist review to assess for whom, under what conditions and how pay for performance programmes work in low- and middle-income countries. *Soc Sci Med* **270**: 113624.

Skivington K, Matthews L, Simpson SA *et al.* 2021. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ*: n2061.

Sterman JD. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill Companies Inc

Tomoaia-Cotisel A, Hyunjung K, Allen S *et al.* 2017. Causal loop diagrams: a tool for visualizing emergent system behaviour. In: de Savigny D, Blanchet K, Adam T (eds.). *Applied Systems Thinking for Health Systems Research : A Methodological Handbook*. McGraw-Hill Education.

UNICEF. 2021a. Maternal mortality. *Data.unicef.org*: 1. Date of access: October 19, 2022. URL: https://data.unicef.org/topic/maternal-health/maternal-mortality/

UNICEF. 2021b. Under five mortality. *Data.unicef.org*: 1. Date of access: October 19, 2022.

URL: https://data.unicef.org/topic/child-survival/under-five-mortality/

Varghese J, Kutty VR, Paina L *et al.* 2014. Advancing the application of systems thinking in health: understanding the growing complexity governing immunization services in Kerala, India. *Heal Res Policy Syst* **12**.

Witter S, Fretheim A, Kessy FL *et al.* 2012. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev*, DOI: 10.1002/14651858.CD007899.pub2.

Witvorapong N, Foshanji AI. 2016. The impact of a conditional cash transfer program on the utilization of non-targeted services: Evidence from Afghanistan. *Soc Sci Med* **152**: 87–95.

Zablith N, Diaconu K, Naja F *et al.* 2021. Dynamics of non-communicable disease prevention, diagnosis and control in Lebanon, a fragile setting. *Confl Health* **15**: 4.

2 Systematic review on the application of system dynamics to model health systems

2.1 Introduction

Abbreviations: Accident and Emergency Department (A&E); Accountable care organisation (ACO); Agent-based model (ABM); Caesarean delivery (CD); Cardiac catheterization (CC); Cardiothoracic surgery (CTS); Catharina Hospital Eindhoven (CHE); Community health system (CHS); Congestive heart failure (COHF); Discrete-event simulation (DES); Emergency department (ED); General internal medicine (GIM); Hampshire County Council (HCC); Integrated care system (IC); Intensive Care Unit (ICU); Long-term care services (LTC); Low- and middle-income country (LMIC); Maternal, newborn and child health (MNCH); Medication administration process (MAP); Mobile Stroke Unit (MSU); Payment for performance (P4P); Performance-based payment system (PBPS); Percutaneous Coronary Intervention Centre (PCI); Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA); Prospective Health Technology Assessment (ProHTA); ST-segment elevation myocardial infarction (STEMI); Supportive living (SL); System dynamics model (SDM); Total knee arthroplasty (TKA); Vaginal delivery (VD); Venous thromboembolism (VTE).

Chapter 2 of this thesis is a systematic review of the literature, describing current application of SDM (and ABM) for health systems research, with reflection on use for research in LMIC settings fulfilling **Objective 1** of this thesis:

1. Determine current use and application of systems thinking methods (CLDs and SDMs) for health systems research, with reflection on use in LMIC health system settings.

The results of the review are presented in a paper, ‘Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models’, published in BMC Health Services Research in November 2019. The chapter appendix contains supplementary material, including review search terms and further details on selected articles. Evidence of retention of copyright or use of published materials in this thesis can be found in Appendix 3.

2.2 Research paper 1: Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models

(Cover sheet on next page)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed <u>for each</u> research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1401642	**Title**	Ms
First Name(s)	Rachel		
Surname/Family Name	Cassidy		
Thesis Title	Using systems thinking to optimise health system interventions for improved maternal and child health in low-resource settings		
Primary Supervisor	Prof. Josephine Borghi		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMC Health Services Research		
When was the work published?	19th November 2019		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

Open Access CC-BY license

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Conceptualised study with PhD supervisors, led search, screened titles and full texts, responsible for data extraction and analysis. Wrote first draft of the manuscript and led revisions with input from co-authors.

SECTION E

Student Signature	[redacted]
Date	05/12/22

Supervisor Signature	[redacted]
Date	05/12/22

Cassidy et al. BMC Health Services Research (2019) 19:845
https://doi.org/10.1186/s12913-019-4627-7

BMC Health Services Research

RESEARCH ARTICLE

Open Access

Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models

Rachel Cassidy[1*], Neha S. Singh[1], Pierre-Raphaël Schiratti[2,3], Agnes Semwanga[4], Peter Binyaruka[5], Nkenda Sachingongu[6], Chitalu Miriam Chama-Chiliba[7], Zaid Chalabi[8], Josephine Borghi[1] and Karl Blanchet[1]

Abstract

Background: Mathematical modelling has been a vital research tool for exploring complex systems, most recently to aid understanding of health system functioning and optimisation. System dynamics models (SDM) and agent-based models (ABM) are two popular complementary methods, used to simulate macro- and micro-level health system behaviour. This systematic review aims to collate, compare and summarise the application of both methods in this field and to identify common healthcare settings and problems that have been modelled using SDM and ABM.

Methods: We searched MEDLINE, EMBASE, Cochrane Library, MathSciNet, ACM Digital Library, HMIC, Econlit and Global Health databases to identify literature for this review. We described papers meeting the inclusion criteria using descriptive statistics and narrative synthesis, and made comparisons between the identified SDM and ABM literature.

Results: We identified 28 papers using SDM methods and 11 papers using ABM methods, one of which used hybrid SDM-ABM to simulate health system behaviour. The majority of SDM, ABM and hybrid modelling papers simulated health systems based in high income countries. Emergency and acute care, and elderly care and long-term care services were the most frequently simulated health system settings, modelling the impact of health policies and interventions such as those targeting stretched and under resourced healthcare services, patient length of stay in healthcare facilities and undesirable patient outcomes.

Conclusions: Future work should now turn to modelling health systems in low- and middle-income countries to aid our understanding of health system functioning in these settings and allow stakeholders and researchers to assess the impact of policies or interventions before implementation. Hybrid modelling of health systems is still relatively novel but with increasing software developments and a growing demand to account for both complex system feedback and heterogeneous behaviour exhibited by those who access or deliver healthcare, we expect a boost in their use to model health systems.

Keywords: System dynamics, Agent-based, Hybrid, Health systems, Systematic review, Modelling

* Correspondence: rachel.cassidy@lshtm.ac.uk
[1]Department of Global Health and Development, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK
Full list of author information is available at the end of the article

Introduction

Health systems are complex adaptive systems [1]. As such, they are characterised by extraordinary complexity in relationships among highly heterogeneous groups of stakeholders and the processes they create [2]. Systems phenomena of massive interdependencies, self-organising and emergent behaviour, non-linearity, time lags, feedback loops, path dependence and tipping points make health system behaviour difficult and sometimes impossible to predict or manage [3]. Conventional reductionist approaches using epidemiological and implementation research methods are inadequate for tackling the problems health systems pose [4]. It is increasingly recognised that health systems and policy research need a special set of approaches, methods and tools that derive from systems thinking perspectives [5]. Health systems encompass a many tiered system providing services to local, district and national populations, from community health centres to tertiary hospitals. Attempting to evaluate the performance of such a multi-faceted organisation presents a daunting task. Mathematical modelling, capable of simulating the behaviour of complex systems, is therefore a vital research tool to aid our understanding of health system functioning and optimisation.

System dynamics model (SDM)

System dynamics models (SDM) and agent-based models (ABM) are the two most popular mathematical modelling methods for evaluating complex systems; while SDM are used to study macro-level system behaviour such as the movement of resources or quantities in a system over time, ABM capture micro-level system behaviour, such as human decision-making and heterogeneous interactions between humans.

While use of SDM began in business management [6, 7] it now has wide spread application from engineering to economics, from environmental science to waste and recycling research [8–13]. A SDM simulates the movement of entities in a system, using differential equations to model over time changes to system state variables. A stock and flow diagram can be used to provide a visual representation of a SDM, describing the relationships between system variables using stocks, rates and influencing factors. The diagram can be interpreted as mimicking the flow of water in and out of a bath tub [7]; the rates control how much 'water' (some quantifiable entity, resource) can leave or enter a 'bath tub' (a stock, system variable) which changes over time depending on what constraints or conditions (e.g. environmental or operational) are placed on the system. Often before the formulation of a stock and flow diagram, a causal loop diagram is constructed which can be thought of as a 'mental model' of the system [14], representing key dynamic hypotheses.

Agent-based model (ABM)

Unlike SDM, ABM is a ground-up representation of a system, simulating the changing states of individual 'agents' in a system rather than the broad entities or aggregate behaviour modelled in SDM. Aggregate system behaviour can however be inferred from ABM. Use of ABM to model system behaviour has been transdisciplinary, with application in economics to ecology, from social sciences to engineering [15–19]. There can be multiple types of agent modelled, each assigned their own characteristics and pattern of behaviour [20, 21]. Agents can learn from their own experiences, make decisions and perform actions based on set rules (e.g. heuristics), informed by their interactions with other agents, their own assigned attributes or based on their interaction with the modelled environment [22]. The interactions between agents can result in three levels of communication between agents; one-to-one communication between agents, one-to-many communication between agents and one-to-location communication where an agent can influence other agents contained in a particular location [22].

Why use SDM and ABM to model health systems?

ABM and SDM, with their ability to simulate micro- and macro-level behaviour, are complementary instruments for examining the mechanisms in complex systems and are being recognised as crucial tools for exploratory analysis. Their use in mapping health systems, for example, has steadily risen over the last three decades. ABM is well-suited to explore systems with dynamic patient or health worker activity, a limitation of other differential equation or event-based simulation tools [23–25]. Unlike discrete-event simulation (DES) for example, which simulates a queue of events and agent attributes over time [26], the agents modelled in ABM are decision makers rather than passive individuals. Closer to the true system modelled, ABM can also incorporate ongoing learning from events whereby patients can be influenced by their interactions with other patients or health workers and by their own personal experience with the health system [21]. SDM has also been identified as a useful tool for simulating feedback and activity across the care continuum [27–30] and is highly adept at capturing changes to the system over time [31]. This is not possible with certain 'snapshot in time' modelling approaches such as DES [32]. SDM is best implemented where the aim of the simulation is to examine aggregate flows, trends and sub-system behaviour as opposed to intricate individual flows of activity which are more suited to ABM or DES [33].

There are also models that can accommodate two or more types of simulation, known as hybrid models. Hybrid models produce results closer to true system behaviour by drawing on the strengths of one or more modelling methods while reducing the limitations associated with using a single simulation type [27]. The activity captured in such models emulates the individual variability of patients and health professionals while retaining the complex, aggregate behaviour exhibited in health systems.

Health scientists and policy makers alike have recognised the potential of using SDM and ABM to model all aspects of health systems in support of decision making from emergency department (ED) optimisation [34] to policies that support prevention or health promotion [35]. Before implementing or evaluating costly health policy interventions or health service re-structuring in the real world, modelling provides a relatively risk-free and low budget method of examining the likely impact of potential health system policy changes. They allow the simulation of 'what if' scenarios to optimise an intervention [36]. They can help identify sensitive parameters in the system that can impede the success of initiatives and point to possible spill-over effects of these initiatives to other departments, health workers or patients. Perhaps most important of all, these modelling methods allow researchers to produce simulations, results and a graphical-user interface in relation to alternative policy options that are communicable to stakeholders in the health system [37], those responsible for implementing system-wide initiatives and changes.

Study aim and objectives

Given the increasing amount of literature in this field, the main aim of the study was to examine and describe the use of SDM and ABM to model health systems. The specific objectives were as follows: (1) Determine the geographical, and healthcare settings in which these methods have been used (2) Identify the purpose of the research, particularly the health policies or interventions tested (3) Evaluate the limitations of these methods and study validation, and (4) Compare the use of SDM and ABM in health system research.

Although microsimulation, DES and Markov models have been widely used in disease health modelling and health economic evaluation, our aim in this study was to review the literature on mathematical methods which are used to model complex dynamic systems, SDM and ABM. These models represent two tenants of modelling: macroscopic (top-level) and microscopic (individual-level) approaches. Although microsimulation and DES are individual-based models like ABM, individuals in ABM are "active agents" i.e. decision-makers rather than "passive agents" which are the norm in microsimulation and DES models. Unlike Markov models which are essentially one-dimensional, unidirectional and linear, SDM are multi-dimensional, nonlinear with feedback mechanisms. We have therefore focussed our review on SDM and ABM because they are better suited to characterise the complexity of health systems. This study reviews the literature on the use of SDM and ABM in modelling health systems, and identifies and compares the key characteristics of both modelling approaches in unwrapping the complexity of health systems. In identifying and summarising this literature, this review will shed light on the types of health system research questions that these methods can be used to explore, and what they add to more traditional methods of health system research. By providing an over overview of how these models can be used within health system research, this paper is also expected to encourage wider use and uptake of these methods by health system researchers and policy makers.

Methods

The review was conducted in compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement [38].

Search strategy and information sources

The literature on ABM and SDM of health systems has not been confined to a single research discipline, making it necessary to widen the systematic review to capture peer-reviewed articles found in mathematical, computing, medicine and health databases. Accordingly, we searched MEDLINE, EMBASE, Cochrane Library, MathSciNet, ACM Digital Library, HMIC, Econlit and Global Health databases for literature. The search of health system literature was narrowed to identify articles that were concerned with modelling facility-based healthcare, services and related healthcare financing agreements which had been excluded or were not the focus of previous reviews [34, 35, 39–41]. The search criteria used for MEDLINE was as follows, with full search terms for each database and search terms used to locate SDM and ABM literature found in Additional file 1:

> *(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti. AND (agent-based OR agent based).ab,ti. AND (model*).ab,ti.*

In addition, the reference list of papers retained in the final stage of the screening process, and systematic reviews identified in the search, were reviewed for relevant literature.

Data extraction and synthesis

The screening process for the review is given in Fig. 1 (adapted from [38]). All search results were uploaded to Mendeley reference software where duplicate entries were removed. The remaining records were screened using their titles and abstracts, removing entries based on eligibility criteria given in Table 1. Post-abstract review, the full text of remaining articles was screened.

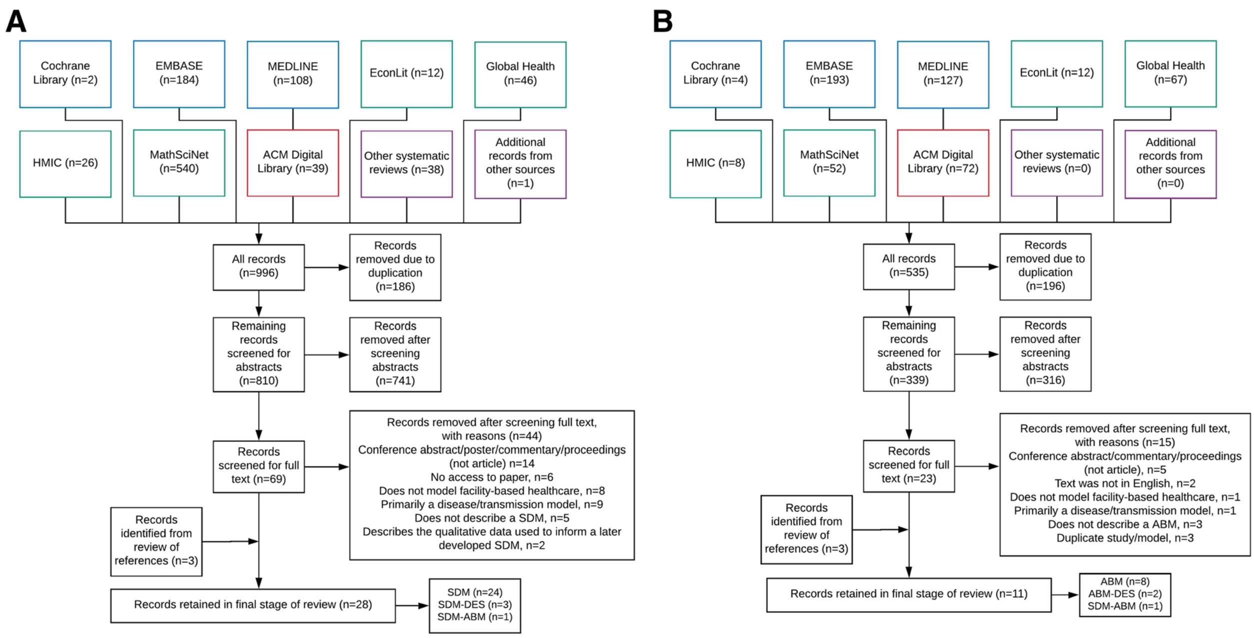


Fig. 1 a Flow-chart for systematic review of SDMs and **b** ABMs of health systems (Database research discipline is identified by colour; mathematical and computing (red), medicine (blue) and health (green) databases). Adapted from PRISMA [38]

Papers retained in final stage of screening were scrutinised, with data imported to Excel based on the following categories; publication date, geographical and healthcare setting modelled, purpose of research in addition to any policies or interventions tested, rationale for modelling method and software platform, validation and limitations of model. The results were synthesised using descriptive statistics and analysis of paper content that were used to answer the objectives.

The studies were first described by three characteristics: publication date, geographical setting, and what aspect of the health system was modelled and why. These characteristics were chosen for the following reasons. Publication date (Fig. 2) allows us to examine the quantity of SDM and ABM studies over time. Geographical settings (Fig. 2, top) allows us to see which health systems have been studied, as health systems in LMIC are very different from those in developed countries. Studies are classified as modelling health systems in high, upper middle, lower middle and low income countries as classified by The World Bank based on economy, July 2018 [42]. Finally, we examined which aspects of the health system have been modelled and the types of research/policy questions that the models were designed to address, to shed light on the range of potential applications of these models, and also potential gaps in their application to date.

The analysis of paper content was split into three sections; SDM use in health system research (including hybrid SDM-DES), ABM use in health system research (including hybrid ABM-DES) and hybrid SDM-ABM use in health system research. The quality of selected studies will not be presented as our aim was to compare and summarise the application of SDM and ABM in modelling health systems rather than a quality appraisal of studies.

Table 1 Eligibility criteria for review

Criteria	Inclusion	Exclusion
Type of study/model	Studies that describe the development and presentation of SDM or ABM or hybrid model.	Poster presentations, conference abstracts, review papers (reference list reviewed), commentaries, debate papers, papers that describe the qualitative data used to inform a later developed model, papers that only present conceptual SDM or ABM model, papers that present exclusively a DES model or other modelling method.
Setting	Facility-based healthcare or related policies/ financing arrangements	Papers that primarily describe a disease/transmission model or delivery of non-facility-based healthcare
Publication date	Up to May 2019	
Language	English	Other languages

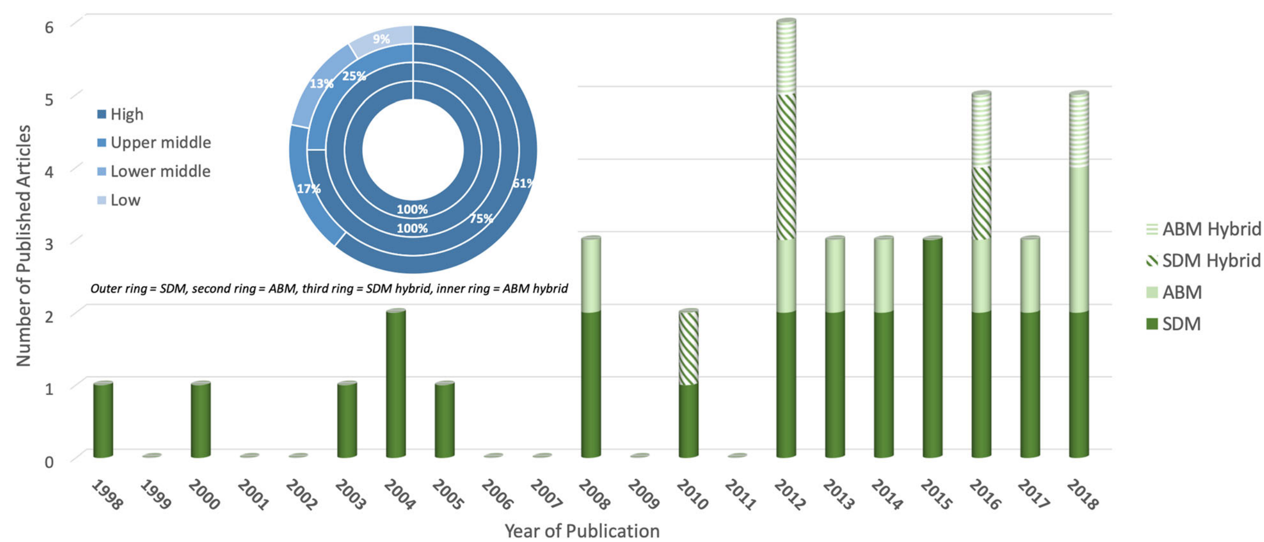


Fig. 2 Number of articles in the final review by year of publication and economic classification

Results

Study selection

The search initially yielded 535 citations for ABM and 996 citations for SDM of facility-based healthcare and services (see Fig. 1). Post-full text screening 11 ABM and 28 SDM papers were retained for analysis, six of which utilised hybrid modelling methods. Three of the hybrid modelling papers integrated SDM with DES [43–45], two integrated ABM with DES [24, 46] and one integrated SDM with ABM [47]. A summary table of selected papers is given in Table 2.

Descriptive statistics

Publication date

The first SDM paper to model health systems was published in 1998 [56] whilst the first publication [66] utilising ABM came almost a decade later (Fig. 2). We found an increasing trend in publications for both modelling approaches, with 90.9% (10/11) and 71.4% (20/28) of all ABM and SDM articles, respectively, having been published in the last decade. The first hybrid modelling article was published in 2010 [43], using SDM and DES to model the impact of an intervention to aid access to social care services for elderly patients in Hampshire, England.

Geographical setting

The proportion of papers that modelled health systems in high, upper middle, lower middle and low income countries is presented in Fig. 2. Eighteen (18/28) papers that employed SDM simulated health systems in high income countries including England [33, 36, 43, 45, 50, 54, 56, 57] and Canada [28, 51, 62]. Four SDM papers simulated upper middle income country health systems, including Turkey [52, 59] and China [64], with a nominal number of papers (5/28) focussing on lower middle or low income countries (West Bank and Gaza [48, 55], Indonesia [37], Afghanistan [30] and Uganda [60]). Almost all ABM papers (9/11) modelled a high income country health system, including the US [20, 23, 25] and Austria [65]. Two (2/11) ABM papers described an upper-middle income based health system (Brazil [22, 67]). All six articles that implemented a hybrid SDM or ABM simulated health systems based in high income countries, including Germany [44] and Poland [47].

Healthcare setting and purpose of research

The healthcare settings modelled in the SDM, ABM and hybrid simulation papers are presented in Fig. 3. Healthcare settings modelled using SDM included systems that were concerned with delivering emergency or acute care (11/28) [28, 31, 36, 45, 47, 50, 56–58, 61, 62], elderly or long-term care services (LTC)(12/28) [28, 31, 36, 43–45, 49–51, 54, 61, 62] and hospital waste management (4/28) [37, 48, 52, 55]. Twenty of the SDM papers selected in this review assessed the impact of health policy or interventions on the modelled system. Common policy targets included finding robust methods to relieve stretched healthcare services, ward occupancy and patient length of stay [28, 31, 36, 43, 49, 50, 54, 58, 62], reducing the time to patient admission [33, 53, 61], targeting undesirable patient health outcomes [47, 58, 60, 63], optimising performance-based incentive health system policies [30, 59] and reducing the total cost of care [33, 54, 61]. The remaining eight papers explored factors leading to undesirable emergency care system behaviour [56, 57], simulating hospital waste management systems and predicting future waste generation [37, 48, 55],

Table 2 Summary of studies included at full paper review (SDM) and studies included at full paper review (ABM)

Paper/Year/Ref	Purpose	Sector of health system modelled	Key results	Software platform
System dynamics models (SDMs)				
Al-Khatib (2016) [48]	Assess the impact of key factors on the hospital waste management system and compare the future total waste output between private, charitable and government hospitals.	• Model simulates hospital waste management in Nablus, Palestine. • Focus on three different types of hospital (private, charitable and government hospitals).	• The amount of waste generated heavily dependent on the number of beds. • Waste treatment was dependent on staff training and the enforcement of legislation.	• iThink.
Alonge (2017) [30]	Explore effective implementation structure for improving health system performance through pay-for-performance (P4P) initiative.	• The model is a generic representation of the pay for performance initiative in primary health facilities in Afghanistan.	• P4P initiative would likely have a beneficial impact on the volume and quality of health services if correctly implemented. • May prove ineffective if the impact of gaming is not mitigated or if the method for distributing financial rewards are inadequate.	• MATLAB and Simulink.
Ansah (2014) [49]	Assess the impact of different long-term care (LTC) capacity policies on uptake of acute care, demand for and utilisation of LTC services.	• Generic representation of LTC utilisation and resources for care and is not based or set in a particular health facility.	• Proactive adjustment of LTC capacity stemmed the number of acute care visits but required a modest increase in staff. • Movement of health staff (through delayed training or from LTC to the acute care sector) will impede the success of this policy.	• Does not state.
Brailsford (2004) [50]	To determine how emergency and on demand care is currently configured and what policies could alleviate pressure on the health system.	• Entire healthcare system that provides emergency centres etc) in Nottingham, England.	• Significant impact on elective hospital admissions as emergency cases are currently prioritised. • Redirecting certain elderly patients to appropriate services relieved pressure on emergency services.	• STELLA.
Brailsford (2010)[a] [43]	Investigate how local authorities such as Hampshire County Council (HCC) can improve access to services and support for older people, in particular assess the long-term impact of a new contact centre for patients.	• HCC system for long-term care, including a call centre that older patients can access to receive advice or be directed to appropriate care.	• The number of patients who contact the call centre on a second occasion (having failed to make contact the first time) where the health status of the patient has now deteriorated, fell drastically after the introduction of two additional call handlers.	• SDM is Vensim, DES model is Simul8.
Cepoiu-Martin (2018) [51]	To examine patient transition from home to supportive living (SL) or long term care (LTC) in persons with dementia and discern policy impact on the deficit of nurses and health care assistants.	• The Alberta Continuing Care System comprising of home living, SL or LTC services.	• Introducing benchmarks for hiring nurses and health care assistants in SL and LTC facilities will result initially in a greater deficit of staff but will stabilise the ratio of health professionals to patients in the long term.	• Does not state.
Chaerul (2008) [37]	To determine key factors that impact the management of hospital waste and predict future waste output.	• The model describes hospital waste management in the City of Jakarta, Indonesia.	• Hospital waste disposal is impacted by the reluctance of a densely populated cityto allow further waste to be dumped in landfill sites. • The simulation indicated that existing and new landfill sites will be at full capacity by 2011 and 2020, respectively.	• STELLA.

Table 2 Summary of studies included at full paper review (SDM) and studies included at full paper review (ABM) *(Continued)*

Paper/Year/Ref	Purpose	Sector of health system modelled	Key results	Software platform
Ciplak (2012) [52]	To predict future healthcare waste production and optimise the management of healthcare waste.	• Healthcare waste generation from healthcare facilities, the single healthcare waste treatment facility and alternative waste treatment facilities in Istanbul, Turkey.	• Employing stringent waste separation strategies would relieve the pressure on already at capacity waste treatment facility in Istanbul. • Up to 77% of healthcare waste could be diverted to alternative treatment technologies that do not require treatment at the incineration facility.	• Vensim.
De Andrade (2014) [53]	To examine the reasons for delayed ST-segment elevation myocardial infarction (STEMI) treatment and explore interventions that can speed up wait time in primary care facilities.	• A primary care hospital and a Percutaneous Coronary Intervention Centre (PCI) in Brazil.	• It was observed that 50% reduction in waiting time for patients is possible under a combination of interventions targeting ECG transmission and PCI centre team feedback time and patient transfer waiting time.	• Vensim.
Desai (2008) [54]	To forecast demand for older people's services and explore the future impact of challenges that accompany an ageing population.	• Adult Services Department of Hampshire County Council including 13 different types of care package that can be offered by the funding and assessment body.	• Providing care packages only to critical patients reduced the overall number of patients receiving acute care. • Savings can be made by increasing the number of unqualified care workers which can be fed back into care funding.	• STELLA.
Djanatliev (2012)[b] [47]	Presenting the functionality of the Prospective Health Technology Assessment (ProHTA) tool, which can simulate the impact of optimised technology prospectively before physical development.	• Mobile Stroke Unit (MSU) case study was simulated for Berlin, includes a generic hospital with emergency services where patients are taken by the MSU.	• In the simulation implementing MSU, 18.2% of patients received thrombolysis treatment compared with 10.6% in the simulation without MSU. • Fewer patients were also found to have developed severe disability in the simulation with MSU as a consequence of faster implemented treatment, reducing the long term costs for rehabilitation and care.	• AnyLogic.
Eleyan (2013) [55]	To predict general and medical waste generation for a complex hospital waste management system.	• Model simulates hospital waste management in three hospitals based in Jenin, Palestine.	• Increases in the amount of hospital waste are consistent with bed occupancy. Over the next 20 years, the total amount of waste generated will rise as will the total cost of treating hazardous waste.	• iThink.
Esensoy (2018) [28]	Transformation of stroke care to implement best practice.	• The model describes six sectors of Ontario health care system and the patient flow between them.	• When stroke best practice policy has been implemented (compared to the base case scenario), there is a reduction in length of stay across all sectors. • A reduction in bed utilisation was also observed with a 10 and 11.1% reduction in acute care and rehab sectors, respectively.	• Vensim.
Ghaffarzad. (2013) [32]	To explore physician decision making behind scheduled caesarean delivery (CD), unplanned CD and vaginal delivery (VD) and examine factors that influence procedure variation.	• The model does not reflect a particular hospital but is parameterised using patient information from hospital discharge databases in Florida.	• The biggest impact on physician delivery decision is from the delayed effect of colleague past experience. • Turning off all learning experiences reduces physician delivery variation for scheduled CD delivery from 6.5 to 4.7%.	• Vensim.

Table 2 Summary of studies included at full paper review (SDM) and studies included at full paper review (ABM) *(Continued)*

Paper/Year/Ref	Purpose	Sector of health system modelled	Key results	Software platform
Lane (1998) [56]	Explore the factors that lead to delays in Accident and Emergency Departments (A&E) and to elective admissions.	• A&E department at major inner-London teaching hospital coded in the study as 'St Dane's'.	• Reduction in bed numbers increases emergency admission waiting times and delays and cancellations to elective surgery admissions. • Increases in demand push the system to breaking point, with patients waiting hours to be admitted and health workers at full capacity.	• Does not state.
Lane (2000) [57]	The model depicts the performance of Accident and Emergency (A&E) at acute hospitals, investigating the sensitivity of waiting times to hospital bed numbers.	• A&E department at Inner-London teaching hospital coded in the study as 'St Dane's'.	• Reducing bed capacity increased the % of elective cancellations, negating the impact on other performance measures. • Deterioration of services is not attributed to lack of bed capacity but insufficient provision of A&E doctors who reach 100% utilisation.	• iThink.
Lattimer (2004) [36]	To evaluate 'front door' services of local emergency and urgent care facilities and test proposals for system change.	• Entire healthcare system that provides emergency or on demand care (GP, NHS Direct, Walk in centres etc) in Nottingham.	• Reducing emergency admissions from GP by 4% showed successive reduction in occupancy levels in A&E. • Interventions to lower admissions of patients over 60 resulted in a 1% reduction per annum in bed occupancy over 5 years.	• STELLA.
Mahmoudia. (2017) [58]	To explore the intended and unintended consequences of Intensive Care Unit (ICU) resource and bed management policies on patient mortality, emergency departments (ED) and general wards.	• Generic model of ICU, ED and general hospital wards.	• Whilst general ward admission control is not as effective at reducing ICU and ED occupancy rates, it outperforms other policies with regards to reducing patient mortality, arguably the more important ICU management performance measure.	• Does not state.
Meker (2015) [59]	To describe performance-based payment systems (PBPS) in second-step public hospitals and the impact on process measures in hospitals.	• Second-step public hospitals in Turkey.	• With reduced performance payments, physicians move to the private sector decreasing staff levels, reducing time spent with patients leading to a dramatic decrease of correct diagnosis and treatment.	• Does not state.
Mielczarek (2016)[a] [44]	To estimate the future demand for healthcare from patients with cardiac disease.	• Future demand for cardiac disease care in Wroclaw Region, Poland.	• Older population (over 60) will generate increasing demands for care, specifically the growth of cardiac patients was observed as more intense in men than women (increases of 34.4 and 30.15% respectively).	• Does not state.
Rashwan (2015) [31]	To explore the flow of elderly patients through the Irish healthcare system and anticipate the growing demand for services over the next 5 years.	• Generic emergency care facility in Ireland and six possible discharge locations.	• Under increasing demand, a combination of all three policies was necessary to significantly reduce elderly frail patients' length of stay in acute hospitals and reduce delayed discharge numbers.	• Does not state.
Semwanga (2016) [60]	To capture the dynamics of the Ugandan health system and evaluate what impact interventions might have on neonatal care.	• Does not focus on one type of health facility but incorporates different services and levels of care offered to this group.	• Integrating community health education, free delivery kits and motorcycle coupons has the biggest impact on reducing neonatal death. • Interventions targeting socioeconomic status had a greater impact on reducing neonatal mortality than those targeting service delivery.	• STELLA.

Table 2 Summary of studies included at full paper review (SDM) and studies included at full paper review (ABM) *(Continued)*

Paper/Year/Ref	Purpose	Sector of health system modelled	Key results	Software platform
Taylor (2005) [33]	To examine the impact of shifting cardiac catheterization (CC) services from tertiary to secondary level for low risk investigations and explore how improvements could be made to services.	• The CC service pathways at two English district general hospitals, referred to using the pseudonyms 'Veinbridge Hospital' and 'Ribsley Hospital'.	• Significant and stable improvements in service (reducing waiting list time and overall costs of service) were achieved with the implementation of strict referral guidelines for patients.	• STELLA.
Walker (2003) [61]	To model patient flow from feeder hospitals to a sub acute extended care hospital to show the impact of local rules used by the medical registrar (medical admitting officer).	• A single extended care facility in Victoria (Australia) and patient flow from feeder hospitals.	• Using the local rule, the cost of care exceeds the budget by 6%. Without the local rule, costs were 3% under budget. • The unprioritized list maintains waiting lists at a level that effectively short-circuits the feeder hospital second local rule of moving high acuity patients on to the wait list of the sub-acute hospital.	• iThink.
Wong (2010) [62]	To evaluate if smoothing the number of discharges over the week relieves the pressure on emergency departments (ED).	• Model describes a general internal medicine (GIM) program at a single tertiary care teaching hospital in Toronto, Canada.	• Both scenarios for 'smoothed average case' were similar, resulting in reduction of GIM in ED by 27% and GIM in ED length of stay by 31%. • For 'every day is a week day case', larger reductions observed.	• Vensim.
Worni (2012) [63]	To estimate what impact a policy to deny reimbursement of total knee arthroplasty (TKA) patient fees will have on venous thromboembolism (VTE) rates and any unintentional consequences.	• The model simulates all patients (9.7 million) in the US who have symptomatic osteoarthritis, over 65 and have Medicare insurance.	• Model output indicates new policy will result in 3-fold decrease in VTE rates. Fraction of those (in simulation with new policy) with bleeding complications is 6-fold higher and 6-fold more patients ineligible for TKA per year.	• Vensim.
Yu (2015) [64]	To explore the driving factors for a high proportion of patients in China not seeking medical care (also known as potential medical demand) and examine possible interventions.	• Three main sub-systems; medical demand of patients, outpatients in hospitals and outpatients in community health systems (CHS). It does not describe a specific hospital or CHS.	• An increase in the number of CHS and decrease in the number of hospitals was found to induce the biggest decrease in the number of patients not seeking care. • Varying the price of outpatient care in hospitals and CHS had minimal impact on increasing the number of patients who seek care.	• Vensim.
Zulkepli (2012)[a] [45]	Present a case study using hybrid modelling (SDM-DES), explore patient flow in an integrated care system (IC) and the impact of patient admission on health professional stress level.	• Three main sub-systems; patient flow through critical care facility, patient flow through intermediate care assessment and motivation and stress levels of health professionals.	• Due to high demand of intermediate care services but limited spaces bed blocking may occur, with an increase in patient admissions leading to an increase to health professional stress level.	• SDM is Vensim, DES model is Simul8.
Agent-based models (ABMs)				
Alibrahim (2018) [23]	To explore the effect of patient choice on the healthcare market, specifically providers that form accountable care organisations (ACO).	• A generalised simulation of patient (Medicare beneficiary, over 65 years old who has or can develop congestive heart failure) choice of medical provider (hospital or primary care physician facility) in the United States.	• Where providers were allowed to opt out of ACO network, they were able to optimise their own profits by not implementing a disease management programme - this led to a reduction in the overall quality of care, driving patients to attend	• AnyLogic.

Table 2 Summary of studies included at full paper review (SDM) and studies included at full paper review (ABM) *(Continued)*

Paper/Year/Ref	Purpose	Sector of health system modelled	Key results	Software platform
			alterative care facilities reducing the utilisation of that facility.	
Djanatliev (2012)[b] [47]	Presenting the functionality of the Prospective Health Technology Assessment (ProHTA) tool, which can simulate the impact of optimised technology prospectively before physical development.	• Mobile Stroke Unit (MSU) case study was simulated for Berlin, includes a generic hospital with emergency services where patients are taken by the MSU.	• In the simulation implementing MSU, 18.2% of patients received thrombolysis treatment compared with 10.6% in the simulation without MSU. • Fewer patients were also found to have developed severe disability in the simulation with MSU as a consequence of faster implemented treatment, reducing the long term costs for rehabilitation and care.	• AnyLogic.
Einzinger (2013) [65]	To create a tool capable of comparing reimbursement schemes in outpatient care.	• Compared different reimbursement schemes for Austrian outpatient health sector simulating the vast majority of health insured persons in Austria.	• Creation of a tool that can be used to compare health care reimbursement schemes in Austria.	• AnyLogic.
Hutzsch. (2008) [66]	To determine which mix of patients should be admitted to specialised hospitals to optimise resource utility and to consider the impact of unplanned patient arrivals on this process.	• Cardiothoracic surgery (CTS) and intensive care unit (ICU) at Catharina Hospital Eindhoven (CHE) in the Netherlands. CTS and ICU are broken down into their respective units such as the high care unit of CTS etc.	• An additional ward bed on the CTS ward decreased the frequency of sending pre- and post- operative admissions to other wards by a factor of 3 with minimal cost. • The brute force optimiser indicated that the number of IC high care beds should be increased and number of IC beds decreased to gain optimum throughput of patients in simulation.	• Java.
Huynh (2012) [20]	To assess the impact of redesigning medication administration process (MAP) workflow for registered nurses to improve medication administration safety.	• A local (anonymous) medical centre where nurses are administering medication to patients.	• Implementing a protocol for the order of MAP tasks to be performed improved the amount of time spent performing tasks. • When registered nurses performed tasks in the most frequently observed order (in the pilot study) this improved MAP task times.	• Netlogo.
Kittipitta. (2016)[c] [24]	To examine patient flow in an outpatient clinic of an orthopedic department and explore interventions that can improve clinical services to reduce patient waiting times.	• Orthopedic department at unidentified community hospital.	• Average waiting time for outpatient appointments fell by 32.03% under the new management policy.	• AnyLogic.
Liu (2014) [21]	To develop a tool that can be used as a decision support system for managers of emergency departments (ED) to assess risk, allocation of resources and identify weakness in emergency care service.	• ED at Hospital of Sabadell (University tertiary level hospital in Barcelona, Spain). The Department is split into sections A (critical patients) and B (least critical patients).	• A tool that can be used simulate the behaviour of agents in ED.	• Netlogo.
Liu (2016) [25]	To explore how accountable care organisations (ACO) can impact payers, healthcare providers and patients under a shared savings payment model for congestive heart failure (CHF) and achieve optimal outcomes.	• A generalised simulation of patients (Medicare beneficiary, over 65 years old who has or can develop congestive heart failure) seeking care (hospital or primary care physician facility) in Unites States.	• Quality orientated providers yielded higher financial returns to the payer agent (which were then shared between providers) than those that were profit-orientated.	• AnyLogic.

Table 2 Summary of studies included at full paper review (SDM) and studies included at full paper review (ABM) *(Continued)*

Paper/Year/Ref	Purpose	Sector of health system modelled	Key results	Software platform
Viana (2018)[c] [46]	To examine and improve patient flow through a pregnancy outpatient clinic in light of the uncertainty in demand for services from overdue patients.	• Overdue pregnancy outpatient clinic, pregnancy clinic and postnatal clinic at Akershus University Hospital, Norway.	• As expected increasing the number of midwives in the clinic reduces resource utilisation but combined with an increase in demand led to an increase in doctor utilisation. • Midwives act as a buffer (or bottleneck) to patients seeing doctors.	• AnyLogic.
Yousefi (2017) [67]	To apply group decision-making techniques for emergency department (ED) resource allocation and determine whether this approach improves performance indicators.	• A generic ED informed from the literature.	• Group-decision making between agents in the ED resulted in on average a 12.7% decrease in total waiting time and 14.4% decrease in the number of patients who left without being seen.	• Netlogo.
Yousefi (2018) [22]	To examine the behaviour of patients who leave public hospital emergency departments (ED) without being seen and the impact of preventative policies.	• ED at Hospital Risoleta Tolentino Neves, a tertiary hospital in Minas Gerais, Brazil.	• After applying preventative policies, average 42.14% reduction in the number of patients leaving without being seen in the ED and average 6.05% reduction in patient length of stay in ED was observed, with most effective policy to fast-track less critical patients after triage.	• NetLogo .

Note: [a]Articles implemented SDM-DES hybrid modelling
[b]Articles implemented SDM-ABM hybrid modelling
[c]Articles implemented ABM-DES hybrid modelling

estimating future demand for cardiac care [44], exploring the impact of patient admission on health professionals stress level in an integrated care system [45], and variation in physician decision-making [32].

ABM papers modelled systems focussed on delivering emergency or acute care (4/11) [21, 22, 47, 67] and accountable care organisations (ACO) or health insurance reimbursement schemes (3/11) [23, 25, 65]. Nine of the ABM papers assessed the impact of health policy or interventions on the modelled system. Common policy targets included decreasing the time agents spent performing tasks, waiting for a service or residing in parts

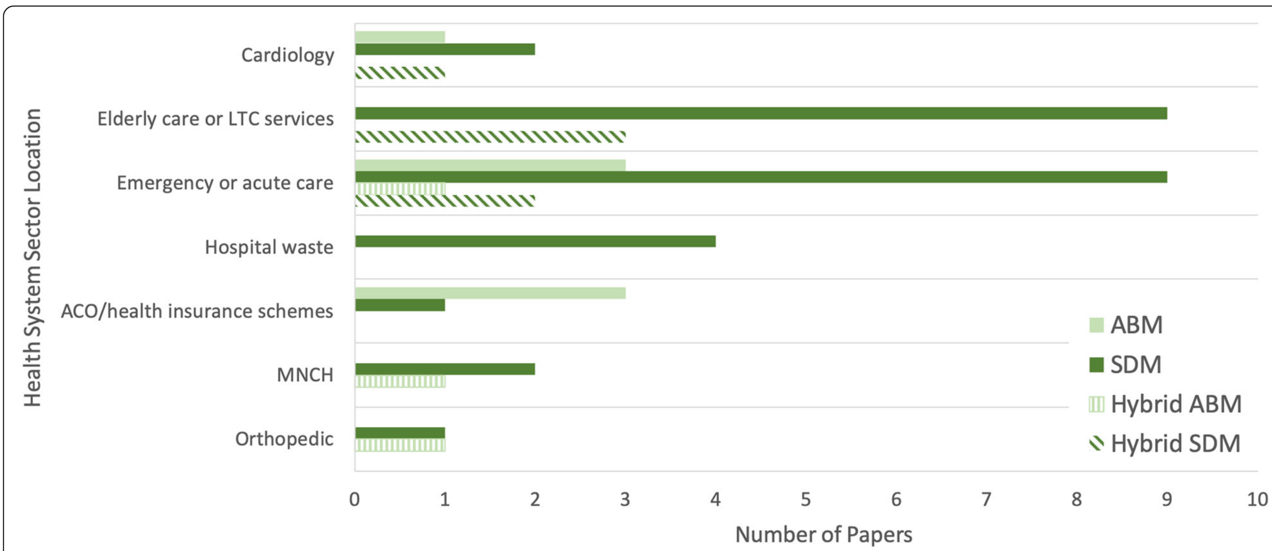


Fig. 3 The health system sector locations modelled in the SDM, ABM and hybrid modelling literature. Long-term care (LTC); Accountable care organisation (ACO); Maternal, newborn and child health (MNCH)

of the system [20, 22, 24, 67], reducing undesirable patient outcomes [23, 25, 47, 67], reducing the number of patients who left a health facility without being seen by a physician [22, 67] and optimising resource utility (beds and healthcare staff) [46, 66, 67]. The remaining two papers described simulation tools capable of comparing health insurance reimbursement schemes [65] and assessing risk, allocation of resources and identifying weaknesses in emergency care services [21].

Papers that utilised hybrid simulation, combining the strengths of two modelling approaches to capture detailed individual variability, agent-decision making and patient flow, modelled systems focussed on delivering elderly care or LTC services [43–45] and emergency or acute care [45, 47]. Four of the hybrid simulation papers assessed the impact of policy or intervention on the modelled system. Policy targets included improving access to social support and care services [43], reducing undesirable patient outcomes [47], decreasing patient waiting time to be seen by a physician [24] and improving patient flow through the system by optimising resource allocation [46]. The remaining two papers used hybrid simulation to estimate the future demand for health care from patients with cardiac disease [44] and model patient flow through an integrated care system to estimate impact of patient admission on health care professionals wellbeing [45].

SDM use in health systems research (including hybrid SDM-DES)

Rationale for using model

Gaining a holistic system perspective to facilitate the investigation of delays and bottlenecks in health facility processes, exploring counter-intuitive behaviour and monitoring inter-connected processes between subsystems was cited frequently as reasons for using SDM to model health systems [28, 36, 37, 48, 56]. SDM was also described as a useful tool for predicting future health system behaviour and demand for care services, essential for health resource and capacity planning [48, 60]. Configuration of the model was not limited by data availability [28, 52, 64] and could integrate data from various sources when required [51].

SDM was described as a tool for health policy exploration and optimising system interventions [33, 36, 51, 54, 58, 64], useful for establishing clinical and financial ramifications on multiple groups (such as patients and health care providers) [63], identifying policy resistance or unintended system consequences [59, 61] and quantifying the impact of change to the health system before real world implementation [62]. The modelling platform also provided health professionals, stakeholders and decision makers with an accessible visual learning environment that enabled engagement with experts necessary for model conception and validation [48, 50, 55, 57]. The model interface could be utilised by decision makers to develop and test alternative policies in a 'real-world' framework that strengthened their understanding of system-wide policy impact [31, 49, 58, 61].

SDM-DES hybrid models enabled retention of deterministic and stochastic system variability and preservation of unique and valuable features of both methods [44], capable of describing the flow of entities through a system and rapid insight without the need for large data collection [43], while simulating individual variability and detailed interactions that influence system behaviour [43]. SDM-DES offered dual model functionality [44] vital for simulating human-centric activity [45], reducing the practical limitations that come with using either SDM or DES to model health systems such as attempting to use SDM to model elements which have non-aggregated values (e.g. patient arrival time) [45] which is better suited for DES.

Healthcare setting

Sixteen papers that utilised SDM modelled systems that were concerned with the delivery of emergency or acute care, or elderly care or LTC services.

Ten of the reviewed papers primarily modelled sectors of the health system that delivered emergency or acute care[1,2]. Brailsford et al. [50], Lane et al. [56], Lane et al. [57] and Lattimer et al. [36] simulated the delivery of emergency care in English cities, specifically in Nottingham and London. Brailsford et al. [50] and Lattimer et al. [36] created models that replicated the entire emergency care system for the city of Nottingham, from primary care (i.e. General Practice surgeries) to secondary care (i.e. hospital admissions wards), to aid understanding of how emergency care was delivered and how the system would need to adapt to increasing demand. Lane et al. [56] and Lane et al. [57] modelled the behaviour of an ED in an inner-London teaching hospital, exploring the knock on effects of ED performance to hospital ward occupancy and elective admissions. Esensoy et al. [28] and Wong et al. [62] both modelled emergency care in Canada, Esensoy et al. [28] focussing on six sectors of the Ontario health system that cared for stroke patients while Wong et al. [62] simulated the impact of delayed transfer of General Internal Medicine patients on ED occupancy. Rashwan et al. [31], Walker

[1]One of the elderly or LTC services papers also modelled emergency or acute care but it was not the primary focus and is therefore not discussed here.

[2]The single SDM-ABM paper that modelled the delivery of emergency or acute care is discussed in section 'SDM-ABM use in health system research'.

et al. [61] and Mahmoudian-Dehkordi et al. [58] modelled patient flow through a generic emergency care facility with six possible discharge locations in Ireland, a sub-acute extended care hospital with patient flow from feeder facilities in Australia and an intensive care unit, ED and general wards in a generic facility.

Five of the SDM papers primarily simulated the behaviour of LTC facilities or care services for elderly patients[3]. Ansah et al. [49] modelled the demand and supply of general LTC services in Singapore with specific focus on the need for LTC and acute health care professionals. Desai et al. [54] developed a SDM that investigated future demand of care services for older people in Hampshire, England which simulated patient flow through adult social care services offering 13 different care packages. In modelling complex care service demand, Cepoiu-Martin et al. [51] explored patient flow within the Alberta continuing care system in Canada which offered supportive living and LTC services for patients with dementia. Brailsford et al. [43] used a hybrid SDM-DES model to investigate how local authorities could improve access to services and support for older people, in particular the long term impact of a new contact centre for patients. The SDM replicated the whole system for long term care, simulating the future demography and demand for care services and the nested DES model simulated the operational issues and staffing of the call centre in anticipation of growing demand for services. Zulkepli et al. [45] also used SDM-DES to model the behaviour of an integrated care system in the UK, modelling patient flow (DES) and intangible variables (SDM) related to health professionals such as motivation and stress levels.

Policy impact evaluation/testing

Twenty papers that utilised SDM tested the impact of policy or interventions on key health system performance or service indicators. The intended target of these policies ranged from relieving strained and under resourced healthcare services, decreasing healthcare costs to reducing patient mortality rates.

Ansah et al. [49], Brailsford et al. [50] and Desai et al. [54] aimed to reduce occupancy in acute or emergency care departments through policies that targeted elderly utilisation of these services. While demand for LTC services is expected to exponentially increase in Singapore, focus has been placed on expanding the acute care sector. Ansah et al. [49] simulated various LTC service expansion policies (static 'current' policy, slow adjustment, quick adjustment, proactive adjustment) and identified that proactive expansion of LTC services stemmed the number of acute care visits by elderly patients over time and required only a modest increase in the number of health professionals when compared with other policies. In Brailsford et al. [50] simulation of the entire emergency care system for Nottingham, England, policy testing indicated that while the emergency care system is operating near full capacity, yearly total occupancy of hospital beds could be reduced by re-directing emergency admissions from patients over 60 years of age (who make up around half of all admissions) to more appropriate services, such as those offered by community care facilities. To explore challenges that accompany providing care for an ageing population subject to budget restraints, Desai et al. [54] simulated the delivery and demand for social care services in Hampshire over a projected 5 year period. In offering care packages to only critical need clients and encouraging extra care services at home rather than offering residential care, the number of patients accessing acute care services reduced over the observed period.

Desai et al. [54], in addition to Taylor et al. [33] and Walker et al. [61], also examined policies that could reduce the total cost of care. Increasing the proportion of hired unqualified care workers (over qualified care workers who are employed at a higher cost rate) resulted in savings which could be fed back into care funding, although Desai et al. [54] remarked on the legal and practical limitations to this policy. Taylor et al. [33] examined the impact of shifting cardiac catheterization services from tertiary to secondary level hospitals for low risk investigations and explored how improvements could be made to services. Significant and stable improvements in service, including reduced waiting list and overall cost of service, were achieved with the implementation of strict (appropriate referral) guidelines for admitting patients. Walker et al. [61] modelled patient flow from feeder hospitals to a single sub-acute extended care facility in Victoria, Australia, to assess the impact of local rules used by the medical registrar for admission. The local admission policy which prioritised admissions from patients under the care of private doctors pushed the total cost of care over the facility budget by 6% whereas employing no prioritisation rule reduced the total cost of care to 3% under budget.

Semwanga et al. [60], Mahmoudian-Dehkordi et al. [58] and Worni et al. [63] evaluated the impact of health policy on undesirable patient outcomes (mortality and post-treatment complication rates). Semwanga et al. [60] tested the effectiveness of policies designed to promote maternal and neonatal care in Uganda, established from the literature. Policies that enabled service uptake, such as community health education, free delivery kits and motorcycle coupons were significant in reducing neonatal death over the simulated period. Mahmoudian-Dehkordi et al. [58]

[3]Six of the emergency or acute care review papers and one of the cardiology care papers also modelled elderly or LTC services but it was not the primary focus and are therefore not discussed here.

explored the intended and unintended consequences of intensive care unit resource and bed management policies on system performance indicators, including patient mortality. During a simulated crisis scenario, prioritising intensive care unit patient admission to general wards over emergency admissions was found to be the most effective policy in reducing total hospital mortality. Worni et al. [63] estimated the impact of a policy to reduce venous thromboembolism rates post-total knee arthroplasty surgery and identified unintentional consequences of the strategy. The policy prevented the reimbursement of patient care fees in the event that a patient was not taking the recommended prophylaxis medication and consequently develops venous thromboembolism. Simulation results indicated a positive 3-fold decrease in venous thromboembolism rates but an unintended 6-fold increase in the number of patients who develop bleeding complications as a result of compulsory prophylaxis treatment.

Validation (including sensitivity analysis)

Statistically-based models are usually used in quantitative data rich environments where model parameters are estimated through maximum likelihood or least-squares estimation methods. Bayesian methods can also be used to compare alternative statistical model structures. SDMs and ABMs on the other hand are not fitted to data observations in the traditional statistical sense. The data are used to inform model development. Both quantitative data and qualitative data (e.g. from interviews) can be used to inform the structure of the model and the parameters of the model. Furthermore, model structure and parameter values can also be elicited from expert opinion. This means that the nature of validation of ABMs and SDMs requires more scrutiny than that of other types of models.

With increasing complexity of such models, and to strengthen confidence in their use particularly for decision support, models are often subjected to sensitivity analysis and validation tests. Twenty-two papers that utilised SDM undertook model validation, the majority having performed behavioural validity tests (see Additional file 2 for details of validation methods for each model). Key model output such as bed occupancy [36, 50], department length of stay [62] and number of department discharges [31] were compared with real system performance data from hospitals [32, 33, 36, 48, 50, 54, 58, 59, 61, 62], local councils [54], nationally reported figs [31, 64]. as well being reviewed by experts [57, 60] as realistic. Others performed more structure orientated validity tests. Model conception [28, 60], development [30, 36, 50, 53, 54, 57, 62] and formulation [54, 56, 59] were validated by a variety of experts including health professionals [47, 53, 54, 57, 59, 62], community groups [56] and leaders [60], steering committees [36], hospital and care representatives [50, 56, 59], patient groups [60] and healthcare policy makers [60]. Further tests for structural validity included checking model behaviour when subjected to extreme conditions or extreme values of parameters [30, 31, 52, 57, 59, 60, 64], model dimensional consistency [31, 52, 57, 59, 60], model boundary adequacy [31] and mass balance [54] and integration error checks [31, 52]. Sensitivity analysis was performed to assess how sensitive model output was to changes in key parameters [49, 51, 57, 60, 64], to test the impact of parameters that had been based on expert opinion on model output [28] and varying key system parameters to test the robustness and effectiveness of policies [28, 30, 52, 53, 58] (on the assumption of imperfect policy implementation [28]).

Limitations of research

Most of the model limitations reported were concerned with missing parameters, feedback or inability to simulate all possible future health system innovations. Mielczarek et al. [44], Cepoiu-Martin et al. [51], Ansah et al. [49] and Rashwan et al. [31] did not take into account how future improvements in technology or service delivery may have impacted results, such as the possibility of new treatment improving patient health outcomes [51] and how this could impact the future utilisation of acute care services [49]. Walker et al. [61] and Alonge et al. [30] described how the models may not simulate all possible actions or interactions that occurred in the real system, such as all proactive actions taken by hospital managers to achieve budget targets [61] or all unintended consequences of a policy on the system [30]. De Andrade et al. [53] and Rashwan et al. [31] discussed the reality of model boundaries, that SDMs cannot encapsulate all health sub-sector behaviour and spill-over effects. Although these have been listed here as limitations, not accounting for possible future improvements in healthcare service or not simulating all possible actions in the modelled system did not prevent authors from fulfilling study objectives. When developing a SDM, it is not possible to account for all possible spill-over effects to other healthcare departments and this should not be attempted; model boundaries are set to only include variables and feedback that are pertinent to exploring the defined problem.

Simplification of model parameters was another common limitation. Wong et al. [62] stated that this would result in some model behaviour not holding in the real system, such as using weekly hospital admission and discharge averages in place of hourly rates due to the hospital recording aggregated data. This aggregation of model parameters may not have reflected real system complexity; Eleyan et al. [55] did not differentiate between service level and type of hospital when modelling health care waste production (described as future work) and Worni et al. [63] refrained from stratifying post-surgery complications by severity, potentially combining lethal and less harmful complications within the same stock (although this did not detract from the study conclusion that the rate of complications would increase as a result of the tested policy).

Data availability, lack of costing analysis and short time horizons were also considered credible limitations. Models that had been calibrated with real data were at risk of using datasets that contained measurement errors or incomplete datasets lacking information required to inform model structure or feedback [32]. Routine facility data required for model conception and formulation was unavailable which restricted the replication of facility behaviour in the model [36] and restricted validation of model behaviour [59], although it should be noted that this is only one method among many for SDM validation and the author was able to use other sources of data for this purpose. Lack of costing or cost effectiveness analysis when testing policies [60], particularly policies that required significant investment or capacity expansion [58], limited discussion on their feasibility in the real system. Models that simulated events over short time scales did not evaluate long term patient outcomes [33] or the long term effects of facility policies on certain groups of patient [57].

ABM use in health system research (including hybrid ABM-DES)

Rationale for using model

The model's ability to closely replicate human behaviour that exists in the real system was frequently cited [20–22, 25, 66], providing a deeper understanding of multiple agent decision-making [23, 67], agent networks [25] and interactions [21, 22]. The modelling method was described as providing a flexible framework capable of conveying intricate system structures [20], where simulations captured agent capacity for learning and adaptive behaviour [20, 25] and could incorporate stochastic processes that mimicked agent transition between states [25]. ABM took advantage of key individual level agent data [25] and integrated information from various sources including demographic, epidemiological and health service data [65]. The visualisation of systems and interface available with ABM software packages facilitated stakeholder understanding of how tested policies could impact financial and patient health outcomes [23], particularly those experts in the health industry with minimal modelling experience [67].

Integrating DES and ABM within a single model ensured an intelligent and flexible approach for simulating complex systems, such as the outpatient clinic described in Kittipittayakorn et al. [24]. The hybrid model captured both orthopaedic patient flow and agent decision-making that enabled identification of health care bottlenecks and optimum resource allocation [24].

Healthcare setting

Seven papers that utilised ABM modelled systems that were either concerned with delivering emergency or acute care[2], ACOs or health insurance reimbursement schemes.

Liu et al. [21] and Yousefi et al. [22] modelled behaviour in EDs in Spanish and Brazilian tertiary hospitals. Liu et al. [21] simulated the behaviour of eleven key agents in the ED including patients, admission staff, doctors, triage nurses and auxiliary staff. Patients were admitted to the ED and triaged before tests were requested and a diagnosis issued. Over time, agent states changed based on their interaction with other agents such as when a doctor decided upon a course of action for a patient (sending the patient home, to another ward, or continue with diagnosis and treatment). For further details of agent type and model rules for each paper, see Additional file 3.

Yousefi et al. [22] modelled the activities of patients, doctors, nurses and receptionists in a ED. Agents could communicate with each other, to a group of other agents or could send a message to an area of the ED where other agents reside. They made decisions based on these interactions and the information available to them at the time. The main focus of the simulation was on patients who left the ED without being seen by a physician; patients decided whether to leave the ED based on a 'tolerance' time extracted from the literature, which changed based on their interaction with other agents. In an additional paper, Yousefi et al. [67] simulated decision-making by patients, doctors, nurses and lab technicians within a generic ED informed from the literature. Group decision-making was employed, whereby facility staff could interact with each other and reach a common solution for improving the efficacy of the department such as re-allocating staff where needed. Yousefi et al. [67], Yousefi et al. [22] and Liu et al. [21] each used a finite state machine (a computational model which describes an entity that can be in one of a finite number of states) to model interactions between agents and their states.

Liu et al. [25] and Alibrahim et al. [23] modelled the behaviour of patients, health providers and payers using series of conditional probabilities, where health providers had participated in an ACO in the United States. Liu et al. [25] presented a model where health providers within an ACO network worked together to reduce congestive heart failure patient healthcare costs and were consequently rewarded a portion of the savings from the payer agent (hypothetically, the Centers for Medicare and Medicaid Services). Patients were Medicare beneficiaries over the age of 65 who developed diabetes, hypertension and/or congestive heart failure and sought care within the network of health providers formed of three hospitals and 15 primary care physician clinics. Alibrahim et al. [23] adapted Liu et al. [25] ACO network model to allow patients to bypass their nearest medical provider in favour of an alternative provider. The decision for a patient to bypass their nearest health centre was influenced by patient characteristics, provider characteristics and the geographical distance between health providers. Providers were also given a choice on whether to participate in an ACO network, where they would then need to implement a comprehensive congestive heart failure disease management programme.

Einzinger et al. [65] created a tool that could be used to compare different health insurance reimbursement schemes in the Austrian health sector. The ABM utilised anonymous routine data from practically all persons with health insurance in Austria, pertaining to medical services accessed in the outpatient sector. In the simulation, patients developed a chronic medical issue (such as coronary heart disease) that required medical care and led to the patient conducting a search of medical providers through the health market. The patient then accessed care at their chosen provider where the reimbursement system, notified of the event via a generic interface, reimbursed the medical provider for patients care.

Policy impact evaluation/testing

Nine papers tested the impact of policy on key health system performance or service indicators. The intended target of these policies ranged from decreasing patient length of stay, to reducing the number of patients who leave without being seen by a physician to reducing patient mortality and hospitalisation rates.

Huynh et al. [20], Yousefi et al. [22], Yousefi et al. [67] and Kittipittayakorn et al. [24] tested policies to reduce the time agents spent performing tasks, waiting for a service or residing in parts of the system. Huynh et al. [20] modelled the medication administration workflow for registered nurses at an anonymous medical centre in the United States and simulated changes to the workflow to improve medication administration safety. Two policies were tested; establishing a rigid order for tasks to be performed and for registered nurses to perform tasks in the most frequently observed order (observed in a real medical centre) to see if this improved the average amount of time spent on tasks. Yousefi et al. [67] modelled the effects of group decision-making in ED compared with the standard approach for resource allocation (where a single supervisor allocates resources) to assess which policy resulted in improved ED performance. Turning ‘on’ group decision-making and starting the simulation with a higher number of triage staff and receptionists resulted in the largest reduction of average patient length of stay and number of patients who left without being seen. This last performance indicator was the subject of an additional paper [22], with focus on patient-to-patient interactions and how this impacted their decision to leave the ED before being seen by a physician. Four policies adapted from case studies were simulated to reduce the number of patients leaving the ED without being seen and average patient length of stay. The policy of fast-tracking patients who were not acutely unwell during triage performed well as opposed to baseline, where acutely ill patients were always given priority. Kittipittayakorn et al. [24] used ABM-DES to identify optimal scheduling for appointments in an orthopaedic outpatient clinic, with average patient waiting time falling by 32% under the tested policy.

Liu et al. [25], Alibrahim et al. [23] and Yousefi et al. [67] tested the impact of health policy on undesirable patient outcomes (patient mortality and hospitalisation rates). Liu et al. [25] modelled health care providers who operated within an ACO network and outside of the network and compared patient outcomes. Providers who operated within the ACO network worked together to reduce congestive heart failure patient healthcare costs and were then rewarded with a portion of the savings. As part of their membership, providers implemented evidence-based interventions for patients, including comprehensive discharge planning with post-discharge follow-up; this intervention was identified in the literature as key to reducing congestive heart failure patient hospitalisation and mortality, leading to a reduction in patient care fees without compromising the quality of care. The ACO network performed well, with a 10% reduction observed in hospitalisation compared with the standard care network. In another study [23] six scenarios were simulated with combinations of patient bypass capability (turned “on” or “off”) and provider participation in the ACO network (no ACO present, optional participation in ACO or compulsory participation in ACO). Provider participation in the ACO, in agreement with Liu et al. [25], led to reduced mortality and congestive heart failure patient hospitalisation, with patient bypass capability marginally increasing provider ACO participation. Yousefi et al. [67] also modelled the impact of group decision-making in ED on the number of patient deaths and number of wrong discharges i.e. patients sent to the wrong sector for care after triage and are then discharged before receiving correct treatment.

Validation (including sensitivity analysis)

Nine of the 11 papers that utilised ABM undertook model validation, consisting almost exclusively of behavioural validity tests. Model output, such as patient length of stay and mortality rates, was reviewed by health professionals [46, 66] and compared with data extracted from pilot studies [20], health facilities (historical) [22, 24, 46, 65, 66], national health surveys [65] and relevant literature [23, 25]. Papers presented the results of tests to determine the equivalence of variance [20] and difference in mean [20, 24] between model output and real data. Structural validity tests included extreme condition testing [23, 46] and engaging health care experts to ensure the accuracy of model framework [22, 47]. Sensitivity analysis was performed to determine how variations or uncertainty in key parameters (particularly where they had not been derived from historical or care data [65]) affected model outcomes [23, 25].

Limitations of research

The majority of model limitations reported were concerned the use or availability of real system or case data. Huynh et al. [20], Yousefi et al. [67] and Liu et al. [25] formulated their models using data that was obtainable, such as limited sample data extracted from a pilot study [20], national average trends [25] and data from previous studies [67]. Yousefi et al. [22] case study dataset did not contain key system feedback, such as the tolerance time of patients waiting to be seen by a physician in the ED, although authors were able to extract this data from a comparable study identified in the literature.

Missing model feedback or parameters, strict model boundaries and simplification of system elements were also considered limitations. Huynh et al. [20], Hutzschenreuter et al. [66] and Einzinger et al. [65] did not model all the realistic complexities of their system, such as all possible interruptions to tasks that occur in patient care units [20], patient satisfaction of admission processes [66] (which will be addressed in future work), how treatment influences the course of disease or that morbid patients are at higher risk of developing co-morbidity than healthier patients, which would affect the service needs and consumption needs of the patient [65]. To improve the accuracy of the model, Huynh et al. stated that further research is taking place to obtain real, clinical data (as opposed to clinical simulation lab results) to assess the impact of interruptions on workflow. Liu et al.'s [21] model boundary did not include other hospital units that may have been affected by ED behaviour and they identify this as future work, for example to include hospital wards that are affected by ED behaviour. Alibrahim et al. [23] and Einzinger et al. [65] made simplifications to the health providers and networks that were modelled, such as assuming equal geographical distances and identical care services between health providers in observed networks [23], limiting the number of factors that influenced a patients decision to bypass their nearest health provider [65] and not simulating changes to health provider behaviour based on service utilisation or reimbursement scheme in place [23]. Alibrahim et al. [23] noted that although the model was constrained by such assumptions, the focus of future work would be to improve the capability of the model to accurately study the impact of patient choice on economic, health and health provider outcomes.

SDM-ABM use in health system research

A single paper used hybrid SDM-ABM to model health system behaviour. Djanatliev et al. [47] developed a tool that could be used to assess the impact of new health technology on performance indicators such as patient health and projected cost of care. A modelling method that could reproduce detailed, high granularity system elements in addition to abstract, aggregate health system variables was sought and a hybrid SDM-ABM was selected. The tool nested an agent-based human decision-making module (regarding healthcare choices) within a system dynamics environment, simulating macro-level behaviour such as health care financing and population dynamics. A case study was presented to show the potential impact of Mobile Stroke Units (MSU) on patient morbidity in Berlin, where stroke diagnosis and therapy could be initiated quickly as opposed to standard care. The model structure was deemed credible after evaluation by experts, including doctors and health economists.

Comparison of SDM and ABM papers

The similarities and differences among the SDM and ABM body of literature are described in this section and shown in Table 3. A high proportion of papers across both modelling methods simulated systems that were concerned with emergency or acute care. A high number of SDM papers (11/28) simulated patient flow and pathways through emergency care [28, 31, 36, 45, 47, 50, 56–58, 61, 62] with a subset evaluating the impact of policies that relieved pressure on at capacity ED's [28, 36, 50, 58, 62]. ABM papers simulated micro-level behaviour associated with emergency care, such as health professional and patient behaviour in EDs and what impact agent interactions have on actions taken over time [21, 22, 47, 67]. ACOs and health insurance reimbursement schemes, a common modelled healthcare setting among the ABM papers [23, 25, 65] was the focus of a single SDM paper [63] while health care waste management, a popular healthcare setting for SDM application [37, 48, 52, 55] was entirely absent among the selected ABM literature. SDM and ABM were both used to test the impact of policy on undesirable patient outcomes, including patient mortality [23, 25, 58, 60, 67] and hospitalisation rates [23, 25]. Interventions for reducing patient waiting time for services [24, 33, 53, 61, 67] and patient length of stay [22, 31, 67] were also tested using these methods, while policy exploration to reduce the total cost of care was more frequent among SDM studies [33, 54, 61].

SDM and ABM software platforms provide accessible, user-friendly visualisations of systems that enable engagement with health experts necessary for model validation [48, 50, 55, 57] and facilitate stakeholder understanding of how alternative policies can impact health system performance under a range conditions [31, 49, 58, 61]. The

Table 3 Comparison of content between SDM, ABM and hybrid models of health systems literature

	SDM papers	ABM papers	Hybrid papers
Purpose of research	*Testing policies or interventions:* • to relieve at-capacity healthcare services, reduce ward occupancy and patient length of stay [28, 31, 36, 43, 49, 50, 54, 58, 62]. • to reduce time to patient admission and treatment [33, 53, 61] • to reduce delayed discharges [31] • to increase the uptake of healthcare services and level of healthcare provision [60] • to target undesirable patient health outcomes (morbidity, mortality, post-treatment complications) [47, 58, 60, 63]. • to optimise performance-based incentive policies against health professional productivity, quality of care and volume of services [30, 59]. • to reduce the total cost of care [33, 47, 58, 60, 61, 63]. • to reduce deficit of health professionals [51] • to reduce generation of incineration-only health care waste [52] • to increase the number of patients who currently do not seek medical care [64] *Other:* • explore factors leading to undesirable emergency care system behaviour [56, 57] • simulating hospital waste management systems and predicting future waste generation [37, 48, 55]. • estimating future demand for cardiac care [44]. • exploring the impact of patient admission on health professionals stress level in an integrated care system (IC) [45]. • exploring variation in physician decision-making [32].	*Testing policies or interventions:* • to decrease the time agents spent performing tasks, waiting for a service or residing in parts of the system [20, 22, 24, 67]. • to reduce undesirable patient outcomes (mortality and hospitalisation) [23, 25, 47, 67]. • to reduce the number of patients who left a health facility without being seen by a physician [22, 67]. • to reduce number of patients who are wrongly discharged [67] • to optimise utility of resources (staff, beds) [46, 66, 67]. • on bypass rate of patients accessing care at alternative facilities [23] • to reduce total cost of care [25] *Other:* • Create tools capable of comparing health insurance reimbursement schemes [65]. • Assessing risk, allocation of resources and identifying weaknesses in emergency care services [21].	*Testing policies or interventions:* *SDM-DES* • to improve access to social support and care services [43]. *ABM-DES* • to decrease patient waiting time to be seen by a physician [24]. • to improve patient flow and length of stay through the system by optimising resource allocation [46]. *SDM-ABM* • to reduce undesirable patient outcomes (morbidity) [47]. *Other:* *SDM-DES* • Estimate the future demand for health care from patients with cardiac disease [44]. • Model patient flow through an integrated care system to estimate impact of patient admission on health care professional's wellbeing [45].
Healthcare setting modelled	• Cardiology care [33, 53] • Elderly care or LTC services [28, 31, 36, 49–51, 54, 61, 62] • Emergency or acute care [28, 31, 36, 50, 56–58, 61, 62] • Hospital waste management [37, 48, 52, 55] • ACO or health insurance schemes [63] • MNCH [32, 60] • Orthopaedic care [63]	• Cardiology care [66] • Emergency or acute care [21, 22, 67] • ACO or health insurance schemes [23, 25, 65]	*SDM-DES* • Cardiology care [44] • Elderly care or LTC services [43–45] • Emergency or acute care [45] *ABM-DES* • MNCH [46] • Orthopaedic care [24] *SDM-ABM* • Emergency or acute care [47]
Rationale for using model	• Gain holistic perspective of system to investigate delays and bottlenecks in health facility processes, exploring counter-intuitive behaviour and monitoring interconnected processes between sub-systems over time [28, 30, 31, 36, 37, 48, 56, 58]. • Useful tool for predicting future health system behaviour and demand for care services, essential for health resource and capacity planning [48, 60].	• Ability to closely replicate human behaviour that exists in the real system [20–22, 25, 66]. • Provides deeper understanding of multiple agent decision-making [23, 67], agent networks [25] and interactions [21, 22]. • Provides flexible framework capable of conveying intricate system structures [20], where simulations captured agent	*SDM-DES* • Enabled retention of deterministic and stochastic system variability and preservation of unique and valuable features of both methods [44]. • Capable of simulating flow of entities through system and provides rapid insight without need for large data collection [43]. • Can simulate individual variability

Table 3 Comparison of content between SDM, ABM and hybrid models of health systems literature *(Continued)*

	SDM papers	ABM papers	Hybrid papers
	• Configuration of model was not limited by data availability [28, 52, 64] and could integrate data from various sources when required [51]. • Used as a tool for health policy exploration and optimising health system interventions [33, 36, 51, 54, 57, 58, 64]. • Useful for establishing clinical and financial ramifications on multiple groups (such as patients and health care providers) [63]. • Identifying and simulating feedback, policy resistance or unintended system consequences [59, 61]. • Quantifying the impact of change to the health system before real world implementation [62]. • Visual learning environment enabled engagement with stakeholders necessary for model conception and validation [48, 50, 55, 57]. • Utilised by decision makers to develop and test alternative policies in a 'real-world' framework [31, 49, 58, 61]. • Suitable for quantitative analyses [53]. • Fast running simulation [54].	capacity for learning and adaptive behaviour [20, 25]. • Could incorporate stochastic processes that mimicked agent transition between states [25]. • Took advantage of key individual level agent data [25] and integrated information from various sources [65]. • Simulation allows patients to have multiple medical problems at the same time [65]. • Model can be made generalisable to other settings [65]. • Visualization of system facilitated stakeholder understanding of tested policy impact [23], particularly those in the health industry with minimal modelling experience [67].	and detailed interactions that influence system behaviour [43]. • Offered dual model functionality [44] vital for simulating human-centric activity [45], reducing the practical limitations that come with using a single simulation method to model health systems [45]. *ABM-DES* • Captured both patient flow through system and agent decision-making that enabled identification of health care bottlenecks and optimum resource allocation [24]. *SDM-ABM* • Could reproduce detailed, high granularity system elements in addition to abstract, aggregate health system variables [47].
Methods of validation	*Behavioural validity tests:* • Model output reviewed by experts [57, 60]. • Model output compared with historical data and relevant literature [31–33, 36, 48, 50, 54, 58, 59, 61, 62, 64]. *Structural validity tests:* • Model conception [28, 60], development [30, 36, 50, 53, 54, 57, 62] and formulation [54, 56, 59] validated by experts. • Extreme condition or value testing [30, 31, 52, 57, 59, 60, 64]. • Dimensional consistency checks [31, 52, 57, 59, 60]. • Model boundary accuracy checks [31]. • Mass balance checks [54]. • Integration error checks [31, 52]. *Sensitivity analysis* • to assess how sensitive model output was to changes in key parameters [49, 51, 57, 60, 64]. • to test the impact of parameters that had been based on expert opinion on model output [28]. • to test the robustness and effectiveness of policies [28, 30, 52, 53, 58, 63] (on the assumption of imperfect policy implementation [28]).	*Behavioural validity tests:* • Model output reviewed by experts [46, 66]. • Model output compared with historical data and relevant literature [20, 22–25, 46, 65, 66]. • F-test [20] and T-test [20, 24] (equivalence of variance and difference in mean tests). *Structural validity tests:* • Extreme condition or value testing [23, 46]. • Model framework reviewed by experts [22, 47]. *Sensitivity analysis:* • to determine how variations or uncertainty in key parameters (particularly where they had not been derived from historical or care data [65]) affected model outcomes [23, 25].	*Behavioural validity tests:* *ABM-DES* • Model output reviewed by experts [46]. • Model output compared with historical data [24, 46]. • T-test (difference in mean tests) [24]. *Structural validity tests:* *ABM-DES* • Extreme condition or value testing [46]. *SDM-ABM* • Model framework reviewed by experts [47]. *Sensitivity analysis:* *SDM-DES* • To assess how sensitive model output was to changes in key parameters [44].
Study limitations	• Did not consider how future improvements in technology or service delivery may impact results [31, 44, 49, 51]. • May not have simulated all possible actions or interactions that occurred in real system [30, 61]. • Model cannot encapsulate all health sub-sector behaviour and spill-over effects [31, 53].	• Model parameterised with best information available, sometimes missing key data [20, 22, 25, 67]. • Did not model all real system complexity, simplifications made to agents and their attributes [20, 23, 65, 66]. • Did not consider all hospital units affected by possible spill-over effects [21].	*SDM-DES* • Did not consider how future improvements in technology may impact results [44]. • Did not model all real system complexity, stable number of patients with disease per age group [44]. • Lack of technology support led

Table 3 Comparison of content between SDM, ABM and hybrid models of health systems literature *(Continued)*

	SDM papers	ABM papers	Hybrid papers
	• Simplification of real system in model [55, 62, 63]. • Lack of facility data required for model conception, formulation and validation [32, 36, 59]. • Lack of costing or cost effectiveness analysis [58, 60]. • Simulation was over a short time scale and did not evaluate long term patient outcomes [33, 57]. • Assumptions made in model development may not be generalisable to other settings [36, 63]. • Discussion with stakeholders that contributed to model development was not performed systematically [51]. • Quantifying model uncertainty was limited [64].		to simplifications in configuration of model (how information was passed between two distinct models) [45]. *ABM-DES* • Need more case studies to externally validate model [24].
Software platform	• iThink or STELLA (same software) [33, 36, 37, 48, 50, 54, 55, 57, 60, 61]. • MATLAB and Simulink [30]. • Vensim [28, 32, 52, 53, 62–64]. • Did not state [31, 49, 51, 56, 58, 59].	• AnyLogic [23, 25, 65]. • Java [66]. • Netlogo [20–22, 67].	*SDM-DES* • Vensim and Simul8 [43, 45]. • Does not state [44]. *ABM-DES* • AnyLogic [24, 46] *SDM-ABM* • AnyLogic [47].

ability to integrate information and data from various sources was also cited as rationale for using SDM and ABM [51]. Reasons for using SDM to model health systems, as opposed to other methods, included gaining a whole-system perspective crucial for investigating undesirable or counter-intuitive system behaviour across subsystems [28, 36, 37, 48, 56] and identifying unintended consequences or policy resistance with tested health policies [59, 61]. The ability to replicate human behaviour [20–22, 25, 66] and capacity for learning and adaptive behaviour [20, 25] was frequently cited as rationale for using ABM to simulate health systems.

Validation of SDMs and ABMs consisted mostly of behavioural validity tests where model output was reviewed by experts and compared to real system performance data or to relevant literature. Structural validity tests were uncommon among ABM papers while expert consultation on model development [30, 36, 50, 53, 54, 57, 62, 63], extreme condition [30, 31, 52, 57, 59, 60, 64] and dimensional consistency tests [31, 52, 57, 59, 60] were frequently reported in the SDM literature. The inability to simulate all actions or interactions that occur in the real system [20, 30, 61, 65, 66] and simplification of model parameters [23, 55, 62, 63, 65] were described as limitations in both SDM and ABM papers. Data availability for model conception and formulation [20, 22, 25, 32, 36, 67] and the impact of model boundaries (restricting exploration of interconnected sub-system behaviour [21, 31, 53]) were also cited limitations common to both sets of literature. Lack of costing analysis [58, 60], short time horizons [33, 57] and an inability to model future improvements in technology or service delivery [31, 44, 49, 51] were additionally cited among the SDM papers.

Discussion

Statement of principal findings

Our review has confirmed that there is a growing body of research demonstrating the use of SDM and ABM to model health care systems to inform policy in a range of settings. While the application of SDM has been more widespread (with 28 papers identified) there are also a growing number of ABM being used (11), just over half of which used hybrid simulation. A single paper used hybrid SDM-ABM to model health system behaviour. To our knowledge this is the first review to identify and compare the application of both SDM and ABM to model health systems. The first ABM article identified in this review was published almost a decade after the first SDM paper; this reflects to a certain extent the increasing availability of SDM and ABM dedicated software tools with the developments in ABM software lagging behind their SDM modelling counterparts.

Emergency and acute care, and elderly care and LTC services were the most frequently simulated health system setting. Both sets of services are facing exponential increases in demand with constraints on resources, presenting complex issues ideal for evaluation through simulation. Models were used to explore the impact and potential spill over effects of alternative policy options, prior to implementation, on patient outcomes, service use and efficiency under various structural and financial constraints.

Strengths and weaknesses of the study

To ensure key papers were identified, eight databases across four research areas were screened for relevant literature. Unlike other reviews in the field [39, 40], there was no restriction placed on publication date. The framework for this review was built to provide a general overview of the SDM and ABM of healthcare literature, capturing papers excluded in other published reviews as a result of strict inclusion criteria. These include reviews that have focussed specifically on compiling examples of modelled health policy application in the literature [35] or have searched for papers with a particular health system setting, such as those that solely simulate the behaviour of emergency departments [34]. One particularly comprehensive review of the literature had excluded papers that simulated hospital systems, which we have explicitly included as part of our search framework [39].

The papers presented in this review, with selection restricted by search criteria, provide a broad picture of the current health system modelling landscape. The focus of this review was to identify models of facility-based healthcare, purposely excluding literature where the primary focus is on modelling disease progression, disease transmission or physiological disorders which can be found in other reviews such as Chang et al. [39] and Long et al. [41]. The data sources or details of how data was used to conceptualise and formulate models are not presented in this paper; this could on its own be the focus of another study and we hope to publish these results as future work. This information would be useful for researchers who want to gain an understanding of the type and format of data used to model health systems and best practice for developing and validating such models.

Literature that was not reported in English was excluded from the review which may have resulted in a small proportion of relevant papers being missed. Papers that described DES models, the other popular modelling method for simulating health system processes, were not included in this review (unless DES methods are presented as part of a hybrid model integrated with SDM or ABM) but have been compiled elsewhere [68–70]. Finally, the quality of the papers was not assessed.

Implications for future research

A nominal number of SDM papers (9/28), an even lower proportion of ABM papers (2/11) and none of the hybrid methods papers simulated health systems based in low- or middle-income countries (LMICs). The lower number of counterpart models in LMICs can be attributed to a lack of capacity in modelling methods and perhaps the perceived scarcity of suitable data; however, the rich quantitative and qualitative primary data collected in these countries for other types of evaluation could be used to develop such models. Building capacity for using these modelling methods in LMICs should be a priority and generating knowledge of how and which secondary data to use in these settings for this purpose. In this review, we observed that it is feasible to use SDM to model low-income country health systems, including those in Uganda [60] and Afghanistan [30]. The need to increase the use of these methods within LMICs is paramount; even in cases where there is an absence of sufficient data, models can be formulated for LMICs and used to inform on key data requirements through sensitivity analysis, considering the resource and healthcare delivery constraints experienced by facilities in these settings. This research is vital for our understanding of health system functioning in LMICs, and given the greater resource constraints, to allow stakeholders and researchers to assess the likely impact of policies or interventions before their costly implementation, and to shed light on optimised programme design.

Health system professionals can learn greatly from using modelling tools, such as ABM, SDM and hybrid models, developed originally in non-health disciplines to understand complex dynamic systems. Understanding the complexity of health systems therefore require collaboration between health scientists and scientists from other disciplines such as engineering, mathematics and computer science. Discussion and application of hybrid models is not a new phenomenon in other fields but their utilisation in exploring health systems is still novel; the earliest article documenting their use in this review was published in 2010 [43]. Five of the six hybrid modelling papers [43–47] were published as conference proceedings (the exception Kittipittayakorn et al. [24]), demonstrating the need to include conference articles in systematic reviews of the literature in order to capture new and evolving applications of modelling for health systems research.

The configuration and extent to which two distinct types of models are combined has been described in the literature [71–75]. The hybrid modelling papers selected in this review follow what is described as 'hierarchical' or 'process environment' model structures, the former where two distinct models pass information to each other and the latter where one model simulates system processes within the environment of another model [72]. Truly 'integrated' models, considered the 'holy grail' [43] of hybrid simulation, where elements of the system are simulated by both methods of modelling with no clear distinction, were not identified in this review and in the wider literature remain an elusive target. In a recent review of hybrid modelling in operational research only four papers were identified to have implemented truly integrated hybrid simulation and all used bespoke software, unrestricted by the current hybrid modelling environments [76].

Of the six hybrid modelling papers, only Djanatliev et al. [47] presented a model capable of both ABM and SDM simulation. The crucial macro- and micro- level activity captured in such models represent feedback in the wider, complex system while retaining the variable behaviour exhibited by those who access or deliver healthcare. With increasing software innovation and growing demand for multi-method modelling in not only in healthcare research but in the wider research community, we need to increase their application to modelling health systems and progress towards the 'holy grail' of hybrid modelling.

Conclusions

We identified 28 papers using SDM methods and 11 papers using ABM methods to model health system behaviour, six of which implemented hybrid model structures with only a single paper using SDM-ABM. Emergency and acute care, and elderly care and LTC services were the most frequently simulated health system settings, modelling the impact of health policies and interventions targeting at-capacity healthcare services, patient length of stay in healthcare facilities and undesirable patient outcomes. A high proportion of articles modelled health systems in high income countries; future work should now turn to modelling healthcare settings in LMIC to support policy makers and health system researchers alike. The utilisation of hybrid models in healthcare is still relatively new but with an increasing demand to develop models that can simulate the macro- and micro-level activity exhibited by health systems, we will see an increase in their use in the future.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12913-019-4627-7.

Additional file 1. Search criteria used for each database.

Additional file 2. Descriptive table of validation methods used in SDM and ABM literature.

Additional file 3. Descriptive table of ABM model rules.

Abbreviations

ACO: Accountable care organisation; ABM: Agent-based model; DES: Discrete-event simulation; ED: Emergency Department; LTC: Long-term care; LMIC: Low- and middle-income countries; SDM: System dynamics model

Acknowledgements

Not applicable.

Authors' contributions

RC, KB, ZC specified the search criteria and selection of databases. RC screened titles, evaluated full text articles and was responsible for full text extraction. RC is lead author, NSS, KB, ZC, PS, JB, AS, PB, CC, NS provided guidance on the draft and final manuscript. All authors read and approved the final manuscript.

Funding

The work described in this paper was funded by the Health Systems Research Initiative (HSRI). MRC Grant Reference Number: MR/R013454/1.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Global Health and Development, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK. [2]Department of Mathematics, University College London, London, UK. [3]Sia Partners UK, London, UK. [4]Information Systems Department, College of Computing and Information Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda. [5]Ifakara Health Institute, PO Box 78373, Dar es Salaam, Tanzania. [6]Department of Gender Studies, School of Humanities and Social Sciences, University of Zambia, 10101 Lusaka, Zambia. [7]Economic and Business Research Programme, University of Zambia, Institute of Economic and Social Research, P O Box 30900, 10101 Lusaka, Zambia. [8]Department of Public Health, Environments and Society, London School of Hygiene and Tropical, London, UK.

Received: 14 June 2019 Accepted: 11 October 2019
[redacted]

References

1. Kitson A, Brook A, Harvey G, Jordan Z, Marshall R, O'Shea R, et al. Using Complexity and Network Concepts to Inform Healthcare Knowledge Translation. Int J Heal Policy Manag. 2017;7:231–43. https://doi.org/10.15171/ijhpm.2017.79.
2. Paina L, Peters DH. Understanding pathways for scaling up health services through the lens of complex adaptive systems. Health Policy Plan. 2012;27:365–73. https://doi.org/10.1093/heapol/czr054.
3. Lipsitz LA. Understanding health care as a complex system: the foundation for unintended consequences. JAMA. 2012;308:243–4. https://doi.org/10.1001/jama.2012.7551.
4. Adam T, de Savigny D. Systems thinking for strengthening health systems in LMICs: need for a paradigm shift. Health Policy Plan. 2012;27:iv1–3. https://doi.org/10.1093/heapol/czs084.
5. de Savigny D, Blanchet K, Adam T. Applied systems thinking for health systems research : a methodological handbook: McGraw-Hill Education; 2017.
6. Shepherd SP. A review of system dynamics models applied in transportation. Transp B Transp Dyn. 2014;2:83–105. https://doi.org/10.1080/21680566.2014.916236.
7. Sterman JD. Business dynamics: systems thinking and modeling for a complex world: McGraw-Hill Companies Inc; 2000.
8. Kunc M, Mortenson MJ, Vidgen R. A computational literature review of the field of system dynamics from 1974 to 2017. J Simul. 2018;12:115–27. https://doi.org/10.1080/17477778.2018.1468950.
9. System Dynamics for Engineering Students. Elsevier; 2018. doi:https://doi.org/10.1016/C2011-0-05346-2.
10. Radzicki MJ. System Dynamics and Its Contribution to Economics and Economic Modeling. Encycl. Complex. Syst. Sci. New York: Springer New York; 2009. p. 8990–9000. https://doi.org/10.1007/978-0-387-30440-3_539.
11. Ford A. Global warming and system dynamics. Int Conf Syst Dyn Soc. 2007.

12. Fiddaman TS. Exploring policy options with a behavioral climate-economy model. Syst Dyn Rev. 2002;18:243–67. https://doi.org/10.1002/sdr.241.
13. Popli K, Sudibya GL, Kim S. A review of solid waste management using system dynamics modeling. J Environ Sci Int. 2017;26:1185–200. https://doi.org/10.5322/JESI.2017.26.10.1185.
14. Groesser SN, Schaffernicht M. Mental models of dynamic systems: taking stock and looking ahead. Syst Dyn Rev. 2012;28:46–68. https://doi.org/10.1002/sdr.476.
15. Epstein JM. Generative social science: studies in agent-based computational modeling: STU-Stud. Princeton University Press; 2006.
16. Wilensky U, Rand W. An Introduction to Agent-Based Modeling: Mit Press; 2015.
17. Miller JH, Page SE. Complex Adaptive Systems: An Introduction to Computational Models of Social Life: STU-Stud. Princeton University Press; 2007.
18. Railsback SF, Grimm V. Agent-Based and Individual-Based Modeling: Princeton University Press; 2011.
19. Abar S, Theodoropoulos G, Lemarinier P, O'Hare G. Agent based Modelling and simulation tools: a review of the state-of-art software. Comput Sci Rev. 2017;24:13–33. https://doi.org/10.1016/j.cosrev.2017.03.001.
20. Huynh N, Snyder R, Vidal J, Tavakoli A, Cai B. Application of computer simulation modeling to medication administration process redesign. J Healthc Eng. 2012;3:649–62. https://doi.org/10.1260/2040-2295.3.4.649.
21. Liu Z, Cabrera E, Rexachs D, Luque E. A generalized agent-based model to simulate emergency departments. Sixth Int Conf Adv Syst Simul. 2014:65–70.
22. Yousefi M, Yousefi M, Fogliatto FS, Ferreira RPM, Kim JH. Simulating the behavior of patients who leave a public hospital emergency department without being seen by a physician: a cellular automaton and agent-based framework. Brazilian J Med Biol Res. 2018;51:e6961. https://doi.org/10.1590/1414-431X20176961.
23. Alibrahim A, Wu S. An agent-based simulation model of patient choice of health care providers in accountable care organizations. Health Care Manag Sci. 2018;21:131–43. https://doi.org/10.1007/s10729-016-9383-1.
24. Kittipittayakorn C, Ying K-C. Using the integration of discrete event and agent-based simulation to enhance outpatient service quality in an orthopedic department. J Healthc Eng. 2016;2016:4189206. https://doi.org/10.1155/2016/4189206.
25. Liu P, Wu S. An agent-based simulation model to study accountable care organizations. Health Care Manag Sci. 2016;19:89–101. https://doi.org/10.1007/s10729-014-9279-x.
26. Allen M, Spencer A, Gibson A, Matthews J, Allwood A, Prosser S, et al. Right cot, right place, right time: improving the design and organisation of neonatal care networks – a computer simulation study. Heal Serv Deliv Res. 2015;3:9. https://doi.org/10.3310/hsdr03200.
27. Brailsford SC. Tutorial: Advances and challenges in healthcare simulation modeling: 2007 Winter Simul. Conf., IEEE; 2007. p. 1436–48. https://doi.org/10.1109/WSC.2007.4419754.
28. Esensoy AV, Carter MW. High-Fidelity whole-system patient flow modeling to assess health care transformation policies. Eur J Oper Res. 2018;266:221–37. https://doi.org/10.1016/j.ejor.2017.09.019.
29. Homer JB, Hirsch GB. System dynamics modeling for public health: background and opportunities. Am J Public Health. 2006;96:452–8. https://doi.org/10.2105/AJPH.2005.062059.
30. Alonge O, Lin S, Igusa T, Peters DH. Improving health systems performance in low- and middle-income countries: a system dynamics model of the pay-for-performance initiative in Afghanistan. Health Policy Plan. 2017;32:1417–26. https://doi.org/10.1093/heapol/czx122.
31. Rashwan W, Abo-Hamad W, Arisha A. A system dynamics view of the acute bed blockage problem in the Irish healthcare system. Eur J Oper Res. 2015;247:276–93. https://doi.org/10.1016/j.ejor.2015.05.043.
32. Ghaffarzadegan N, Epstein AJ, Martin EG. Practice variation, bias, and experiential learning in cesarean delivery: A data-based system dynamics approach. Health Serv Res. 2013;48:713–34. https://doi.org/10.1111/1475-6773.12040.
33. Taylor K, Dangerfield B. Modelling the feedback effects of reconfiguring health services. J Oper Res Soc. 2005;56:659–75. https://doi.org/10.1057/palgrave.jors.2601862.
34. Mohiuddin S, Busby J, Savovic J, Richards A, Northstone K, Hollingworth W, et al. Patient flow within UK emergency departments: a systematic review of the use of computer simulation modelling methods. BMJ Open. 2017;7:e015007. https://doi.org/10.1136/bmjopen-2016-015007.
35. Atkinson J-A, Wells R, Page A, Dominello A, Haines M, Wilson A. Applications of system dynamics modelling to support health policy. Public Heal Res Pract. 2015;25. https://doi.org/10.17061/phrp2531531.
36. Lattimer V, Brailsford S, Turnbull J, Tarnaras P, Smith H, George S, et al. Reviewing emergency care systems I: insights from system dynamics modelling. Emerg Med J. 2004;21:685–91. https://doi.org/10.1136/emj.2002.003673.
37. Chaerul M, Tanaka M, Shekdar AV. A system dynamics approach for hospital waste management. Waste Manag. 2008;28:442–9. https://doi.org/10.1016/j.wasman.2007.01.007.
38. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535. https://doi.org/10.1136/bmj.b2535.
39. Chang AY, Ogbuoji O, Atun R, Verguet S. Dynamic modeling approaches to characterize the functioning of health systems: A systematic review of the literature. Soc Sci Med. 2017;194:160–7. https://doi.org/10.1016/j.socscimed.2017.09.005.
40. Rusoja E, Haynie D, Sievers J, Mustafee N, Nelson F, Reynolds M, et al. Thinking about complexity in health: a systematic review of the key systems thinking and complexity ideas in health. J Eval Clin Pract. 2018;24:600–6. https://doi.org/10.1111/jep.12856.
41. Long KM, Meadows GN. Simulation modelling in mental health: a systematic review. J Simul. 2017. https://doi.org/10.1057/s41273-017-0062-0.
42. Banco Mundial. World Bank: country and lending groups - DataBank. World Bank Gr. 2018.
43. Brailsford SC, Desai SM, Viana J. Towards the holy grail: Combining system dynamics and discrete-event simulation in healthcare: Proc. 2010 Winter Simul. Conf., IEEE; 2010. p. 2293–303. https://doi.org/10.1109/WSC.2010.5678927.
44. Mielczarek B, Zabawa J. Modeling Healthcare Demand Using a Hybrid Simulation Approach, vol. 2016: Proc. 2016 Winter Simul. Conf., IEEE Press. p. 1535–46.
45. Zulkepli J. Hybrid simulation for modelling large systems: an example of integrated care model: Proc. 2012 Winter Simul. Conf; 2012.
46. Viana J, Simonsen TB, Dahl FA, Flo K. A Hybrid Discrete Event Agent Based Overdue Pregnancy Outpatient Clinic Simulation Model. Proc. 2018 Winter Simul. Conf. Piscataway: IEEE Press; 2018. p. 1488–99.
47. Djanatliev A, German R, Kolominsky-Rabas P, Hofmann BM. Hybrid simulation with loosely coupled system dynamics and agent-based models for prospective health technology assessments: Proc. Winter Simul. Conf., Winter Simulation Conference; 2012. p. 69:1–69:12.
48. Al-Khatib IA, Eleyan D, Garfield J. A system dynamics approach for hospital waste management in a city in a developing country: the case of Nablus, Palestine. Environ Monit Assess. 2016;188:503. https://doi.org/10.1007/s10661-016-5487-9.
49. Ansah JP, Eberlein RL, Love SR, Bautista MA, Thompson JP, Malhotra R, et al. Implications of long-term care capacity response policies for an aging population: a simulation analysis. Health Policy. 2014;116:105–13. https://doi.org/10.1016/j.healthpol.2014.01.006.
50. Brailsford SC. Emergency and on-demand health care: modelling a large complex system. J Oper Res Soc. 2004;55.
51. Cepoiu-Martin M, Bischak DP. Policy choices in dementia care - an exploratory analysis of the Alberta Continuing Care System (ACCS) using system dynamics. Spec Issue Complex Forum "real" World Heal Syst Implic Complex. Theor Sci. 2018;24:278–84.
52. Ciplak N, Barton JR. A system dynamics approach for healthcare waste management: A case study in Istanbul Metropolitan City, Turkey. Waste Manag Res. 2012;30:576–86. https://doi.org/10.1177/0734242X12443405.
53. de Andrade L, Lynch C, Carvalho E, Rodrigues CG, Vissoci JRN, Passos GF, et al. System dynamics modeling in the evaluation of delays of care in ST-segment elevation myocardial infarction patients within a tiered health system. PLoS One. 2014;9:e103577. https://doi.org/10.1371/journal.pone.0103577.
54. Desai MS, Penn ML, Brailsford S, Chipulu M. Modelling of Hampshire adult services—gearing up for future demands. Health Care Manag Sci. 2008;11:167–76. https://doi.org/10.1007/s10729-007-9049-0.
55. Eleyan D, Al-Khatib IA, Garfield J. System dynamics model for hospital waste characterization and generation in developing countries. Waste Manag Res. 2013;31:986–95. https://doi.org/10.1177/0734242X13490981.
56. Lane D. Emergency but no accident: a system dynamics study of casualty waiting times in the British NHS. Eurohealth (Lond). 1998;4.
57. Lane DC. Looking in the wrong place for healthcare improvements: a system dynamics study of an accident and emergency department. J Oper Res Soc. 2000;51.
58. Mahmoudian-Dehkordi A, Sadat S. Sustaining critical care: using evidence-based simulation to evaluate ICU management policies. Health Care Manag Sci. 2017;20:532–47. https://doi.org/10.1007/s10729-016-9369-z.
59. Meker T, Barlas Y. Dynamic consequences of performance-based payment Systems in Public Hospitals. Syst Res Behav Sci. 2015;32:459–80. https://doi.org/10.1002/sres.2338.

60. Semwanga AR, Nakubulwa S, Adam T. Applying a system dynamics modelling approach to explore policy options for improving neonatal health in Uganda. Heal Res Policy Syst. 2016;14:35. https://doi.org/10.1186/s12961-016-0101-8.
61. Walker BC, Haslett T. The dynamics of local rules in hospital admission processes. Aust Health Rev. 2003;26:98–106.
62. Wong HJ, Wu RC, Caesar M, Abrams H, Morra D. Smoothing inpatient discharges decreases emergency department congestion: a system dynamics simulation model. Emerg Med J. 2010;27:593–8. https://doi.org/10.1136/emj.2009.078543.
63. Worni M, Pietrobon R, Zammar GR, Shah J, Yoo B, Maldonato M, et al. System dynamics to model the unintended consequences of denying payment for venous thromboembolism after total knee arthroplasty. PLoS One. 2012;7:e30578. https://doi.org/10.1371/journal.pone.0030578.
64. Yu W, Li M, Ge Y, Li L, Zhang Y, Liu Y, et al. Transformation of potential medical demand in China: A system dynamics simulation model. J Biomed Inform. 2015;57:399–414. https://doi.org/10.1016/j.jbi.2015.08.015.
65. Einzinger P, Popper N, Breitenecker F, Pfeffer N, Jung R, Endel G. The GAP-DRG Model: Simulation of Outpatient Care for Comparison of Different Reimbursement Schemes. Proc. 2013 Winter Simul. Conf. Simul. Mak. Decis. A complex world. Piscataway: IEEE Press; 2013. p. 2299–308.
66. Hutzschenreuter AK, Bosman PAN, Blonk-Altena I, van Aarle J, La Poutré H. Agent-based patient admission scheduling in hospitals. Belgian/Netherlands Artif Intell Conf. 2008:315–6. https://doi.org/10.1007/3-540-32062-8.
67. Yousefi M, Ferreira R. An agent-based simulation combined with group decision-making technique for improving the performance of an emergency department. Brazilian J Med Biol Res. 2017;50:1–10. https://doi.org/10.1590/1414-431X20175955.
68. Günal MM, Pidd M. Discrete event simulation for performance modelling in health care: a review of the literature. J Simul. 2010;4:42–51. https://doi.org/10.1057/jos.2009.25.
69. Salleh S, Thokala P, Brennan A, Hughes R, Dixon S. Discrete event simulation-based resource Modelling in health technology assessment. Pharmacoeconomics. 2017;35:989–1006. https://doi.org/10.1007/s40273-017-0533-1.
70. Zhang X. Application of discrete event simulation in health care: a systematic review. BMC Health Serv Res. 2018;18:687. https://doi.org/10.1186/s12913-018-3456-4.
71. Bennett PG. On linking approaches to decision-aiding: issues and prospects. J Oper Res Soc. 1985;36:659. https://doi.org/10.2307/2582261.
72. Chahal K, Eldabi T. Applicability of hybrid simulation to different modes of governance in UK healthcare: 2008 Winter Simul. Conf., IEEE; 2008. p. 1469–77. https://doi.org/10.1109/WSC.2008.4736226.
73. Lättilä L, Hilletofth P, Lin B. Hybrid simulation models – when, why, how? Expert Syst Appl. 2010;37:7969–75. https://doi.org/10.1016/j.eswa.2010.04.039.
74. Brailsford SC. Hybrid simulation in healthcare: New concepts and new tools: 2015 Winter Simul. Conf., IEEE; 2015. p. 1645–53. https://doi.org/10.1109/WSC.2015.7408284.
75. Morgan JS, Howick S, Belton V. A toolkit of designs for mixing discrete event simulation and system dynamics. Eur J Oper Res. 2017;257:907–18. https://doi.org/10.1016/j.ejor.2016.08.016.
76. Brailsford SC, Eldabi T, Kunc M, Mustafee N, Osorio AF. Hybrid simulation modelling in operational research: a state-of-the-art review. Eur J Oper Res. 2018. https://doi.org/10.1016/j.ejor.2018.10.025.

Publisher's Note

2.3 Supplementary material to research paper 1

Table S1: Agent-based modelling (ABM) literature search strategy

Database	Agent-based modelling (ABM) literature search strategy
ACM Library	+(("agent-based" "agent based") +("model*") +("health system*" "health care" "healthcare" "health service*" "health polic*" "health facil*" "primary care" "secondary care" "tertiary care" "hospital*"))
Cochrane	(health system*):ti,ab,kw OR (health care*):ti,ab,kw OR (healthcare):ti,ab,kw OR (health service*):ti,ab,kw OR (health polic*):ti,ab,kw OR (health facil*):ti,ab,kw OR (primary care):ti,ab,kw OR (secondary care):ti,ab,kw OR (tertiary care):ti,ab,kw OR (hospital*):ti,ab,kw (agent based):ti,ab,kw OR (agent-based):ti,ab,kw (model*):ti,ab,kw (#1 AND #2 AND #3)
Econlit	(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti,kw. (agent-based OR agent based).ab,ti,kw. (model*).ab,ti,kw. 1 and 2 and 3
Embase Classic+Embase	(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti,kw. (agent-based OR agent based).ab,ti,kw. (model*).ab,ti,kw. 1 and 2 and 3
Global Health	(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti. (agent-based OR agent based).ab,ti. (model*).ab,ti. 1 and 2 and 3
HMIC Health Management Information Consortium	(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti. (agent-based OR agent based).ab,ti. (model*).ab,ti. 1 and 2 and 3
MathSciNet[1]	"(Anywhere=(health*) AND Anywhere=(agent-based)[1] AND Anywhere=(model*))"
Ovid MEDLINE(R)	(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti,kw. (agent-based OR agent based).ab,ti,kw. (model*).ab,ti,kw. 1 and 2 and 3

Note to Figure: [1]The search was sent twice to include results for 'agent based' and agent-based'.

Table S2: System dynamics modelling (SDM) literature search strategy

Database	System dynamics modelling (SDM) literature search strategy
ACM Library	+(("system dynamic*") +("model*") +("health system*" "health care" "healthcare" "health service*" "health polic*" "health facil*" "primary care" "secondary care" "tertiary care" "hospital*"))
Cochrane	(health system*):ti,ab,kw OR (health care*):ti,ab,kw OR (healthcare):ti,ab,kw OR (health service*):ti,ab,kw OR (health polic*):ti,ab,kw OR (health facil*):ti,ab,kw OR (primary care):ti,ab,kw OR (secondary care):ti,ab,kw OR (tertiary care):ti,ab,kw OR (hospital*):ti,ab,kw (system dynamic*):ti,ab,kw (model*):ti,ab,kw (#1 AND #2 AND #3)
Econlit	(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti,kw. (system dynamic*).ab,ti,kw. (model*).ab,ti,kw. 1 and 2 and 3
Embase Classic+Embase	(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti,kw. (system dynamic*).ab,ti,kw. (model*).ab,ti,kw. 1 and 2 and 3
Global Health	(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti,kw. (system dynamic*).ab,ti,kw. (model*).ab,ti,kw. 1 and 2 and 3
HMIC Health Management Information Consortium	(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti. (system dynamic*).ab,ti. (model*).ab,ti. 1 and 2 and 3
MathSciNet	"(Anywhere=(health*) AND Anywhere=(system dynamic*) AND Anywhere=(model*))"
Ovid MEDLINE(R)	(health system* OR health care OR healthcare OR health service* OR health polic* OR health facil* OR primary care OR secondary care OR tertiary care OR hospital*).ab,ti,kw. (system dynamic*).ab,ti,kw. (model*).ab,ti,kw. 1 and 2 and 3

Table S3: Descriptive table of validation methods used and outcome variables validated in SDM and ABM literature

System dynamics models (SDMs)			
Paper/Year/Ref	Purpose	Validation method(s) used	Any outcome variable(s) validated? (behaviour/parameter/structure confirmation tests)
Al-Khatib (2016) [1]	Assess the impact of key factors on the hospital waste management system and compare the future total waste output between private, charitable and government hospitals.	*Behaviour validity* Model output compared to real data.	Unclear which model output was validated with data collected from Nablus city hospitals.
Alonge (2017) [2]	Explore effective implementation structure for improving health system performance through pay-for-performance (P4P) initiative.	*Structure validity* Direct structure tests (structure-confirmation test, parameter-confirmation test). Structure-orientated behavioural tests (phase relationship test, extreme-condition test, behaviour sensitivity test).	Each element (and how they were parameterised) of the model was confirmed by healthcare and P4P experts in Afghanistan. 'Gaming' element of model reported in the literature.
Ansah (2014) [3]	Assess the impact of different long-term care (LTC) capacity policies on uptake of acute care, demand for and utilisation of LTC services.		
Brailsford (2004) [4]	To determine how emergency and on demand care is currently configured and what policies could alleviate pressure on the health system.	*Behaviour validity* Model output compared to real data. *Structure validity* Direct structure tests (dimensional consistency test, structure-confirmation test, parameter-confirmation test).	Total daily bed occupancy and individual ward occupancy are given as examples of model output that were validated with real health facility performance data from Nottingham hospitals. The model was developed in close collaboration with the Steering Committee (representatives from all healthcare providers in Nottingham).
Brailsford (2010)*1 [5]	Investigate how local authorities such as Hampshire County Council (HCC) can improve access to services and support for older people, in particular assess the long-term impact of a new contact centre for patients.		

Cepoiu-Martin (2018) [6]	To examine patient transition from home to supportive living (SL) or long term care (LTC) in persons with dementia and discern policy impact on the deficit of nurses and health care assistants.		
Chaerul (2008) [7]	To determine key factors that impact the management of hospital waste and predict future waste output.	*Behaviour validity* Model output compared to published data.	Unclear which model output was validated with published data on health and waste management in Jakarta, Indonesia.
Ciplak (2012) [8]	To predict future healthcare waste production and optimise the management of healthcare waste.	*Behaviour validity* Model output compared to published data. *Structure validity* Direct structure tests (dimensional consistency). Structure-orientated behavioural tests (extreme-condition test, integration error test).	Population and total healthcare waste materials generated are given as examples of model output that were validated with published data on health and waste management in Istanbul, Turkey.
De Andrade (2014) [9]	To examine the reasons for delayed ST-segment elevation myocardial infarction (STEMI) treatment and explore interventions that can speed up wait time in primary care facilities.	*Structure validity* Direct structure tests (structure confirmation test).	The model was developed in close collaboration ('mediated modelling') with health professionals at a primary hospital in Foz do Iguaçu, Brazil.
Desai (2008) [10]	To forecast demand for older people's services and explore the future impact of challenges that accompany an ageing population.	*Behaviour validity* Model output compared to real data. *Structure validity* Direct structure tests (structure-confirmation test, parameter-confirmation test). Structure-orientated behavioural tests (mass-balance check)	Researchers worked in close collaboration and consulted staff at Hampshire County Council during model conceptualisation and formulation. Model output (stocks) for initial contact, eligibility and initial assessment, care manager assessment and create a care package were compared with real data from Hampshire County Council.

Djanatliev (2012)*2 [11]	Presenting the functionality of the Prospective Health Technology Assessment (ProHTA) tool, which can simulate the impact of optimised technology prospectively before physical development.	*Structure validity* Direct structure tests (structure-confirmation test, parameter-confirmation test).	Model achieved credibility from review by experts (doctors, health economists, medical informaticians and knowledge management experts).
Eleyan (2013) [12]	To predict general and medical waste generation for a complex hospital waste management system.	*Behaviour validity* Model output compared to published data.	Model output, general waste generated and hazardous waste generated, were compared with published waste data from hospitals in Iran.
Esensoy (2018) [13]	Transformation of stroke care to implement best practice.	*Behaviour validity* Model output shared with experts. *Structure validity* Direct structure tests (structure-confirmation test, parameter-confirmation test).	Preliminary model results were shared with policy and stroke experts for validation. Model development conducted in close collaboration with Ministry of Health and Long-term Care and the Central East Local Integration Network in Ontario, Canada.
Ghaffarzad. (2013) [14]	To explore physician decision making behind scheduled caesarean delivery (CD), unplanned CD and vaginal delivery (VD) and examine factors that influence procedure variation.	*Behaviour validity* Model output compared to out-of-sample real data.	Model output, total scheduled caesarean deliveries and total caesarean deliveries, was compared with empirical data from the Florida all-payer hospital discharge database (concerning deliveries at non-federal acute-care hospitals).
Lane (1998) [15]	Explore the factors that lead to delays in Accident and Emergency Departments (A&E) and to elective admissions.	*Structure validity* Direct structure tests (structure confirmation test, parameter confirmation test).	Researchers worked closely with the A&E department of a London hospital, pseudonym 'St Danes' during model development. Aspects of model formulation were reviewed by members of Casualty Watch, the London Ambulance Service and the Emergency Bed Service.
Lane (2000) [16]	The model depicts the performance of Accident and Emergency (A&E) at acute hospitals, investigating the sensitivity of waiting times to hospital bed numbers.	*Behaviour validity* Model output shared with experts. *Structure validity*	Model output (shown using graphs) and performance indicators were judged as realistic by staff from (pseudonym) St Danes hospital, London, England. Model structure was reviewed by experts (A&E system) and model parameters were confirmed using available hospital process information.

		Direct structure tests (dimensional consistency, structure confirmation test, parameter confirmation test). Structure-orientated behavioural tests (extreme-condition test).	Aspects of model formulation were reviewed by members of Southwark Community Health Council, Casualty Watch, the London Ambulance Service, the Emergency Bed Service and the Registrar at St Danes.
Lattimer (2004) [17]	To evaluate 'front door' services of local emergency and urgent care facilities and test proposals for system change.	*Behaviour validity* Model output compared to real data. *Structure validity* Direct structure tests (structure confirmation test, parameter confirmation test).	Model output compared to data from the emergency and urgent care system in Nottingham, England. Daily bed occupancy rates were given as an example of model output that was validated. The project team, steering group and health professionals from Nottingham emergency and urgent care system contributed to the development and refinement of the quantitative model.
Mahmoudia. (2017) [18]	To explore the intended and unintended consequences of Intensive Care Unit (ICU) resource and bed management policies on patient mortality, emergency departments (ED) and general wards.	*Behaviour validity* Model output compared to health facility data extracted from the literature.	Model output, specifically average occupancy of ED, average ICU occupancy and ward occupancy, are compared to relevant health facility statistics identified in the literature.
Meker (2015) [19]	To describe performance-based payment systems (PBPS) in second-step public hospitals and the impact on process measures in hospitals.	*Behaviour validity* Model output compared to real data. *Structure validity* Direct structure tests (dimensional consistency). Structure-orientated behavioural tests (extreme-condition test).	Model output, specifically number of patients examined per month, number of tests performed per month and number of surgeries performed per month, are compared to data from a second-step public hospital in Istanbul.
Mielczarek (2016)[*1] [20]	To estimate the future demand for healthcare from patients with cardiac disease.		
Rashwan (2015) [21]	To explore the flow of elderly patients through the Irish healthcare system and anticipate the growing demand for services over the next five years.	*Behaviour validity* Model output compared to reported data. *Structure validity*	The model structure was discussed and reviewed by Irish Health Service Executive (HSE) officials and domain experts. Model output, specifically number of delayed discharges for each possible destination (home, another hospital, rehabilitation, convalescence, long term care, death, other),

		Direct structure tests (dimensional consistency, structure-confirmation test). Structure-orientated behavioural tests (boundary adequacy, extreme-condition test, integration error test).	were compared to annual delayed discharge figures reported by the HSE.
Semwanga (2016) [22]	To capture the dynamics of the Ugandan health system and evaluate what impact interventions might have on neonatal care.	*Behaviour validity* Model output compared to historical and reported data. *Structure validity* Direct structure tests (dimensional consistency, parameter-confirmation, structure-confirmation test). Structure-orientated behavioural tests (mass-balance test, extreme-condition test).	Model output, including graphical behaviour and results, were compared to historical/national health surveys and reports. During brainstorming sessions, researchers and neonatal and maternal healthcare staff from hospitals in Uganda were asked to validate the model structure.
Taylor (2005) [23]	To examine the impact of shifting cardiac catheterization (CC) services from tertiary to secondary level for low risk investigations and explore how improvements could be made to services.	*Behaviour validity* Model output compared to historical data (Theil inequality statistic, R^2 statistic, MAPE statistic).	Simulated model output was compared with real data from Ribsley Hospital for CC services, specifically district-based elective CC investigation rate, elective CC investigation referral rate, average time spent on CC investigation list, CC investigation waiting list, elective CC investigation rate, tertiary-based elective CC investigation rate.
Walker (2003) [24]	To model patient flow from feeder hospitals to a sub acute extended care hospital to show the impact of local rules used by the medical registrar (medical admitting officer).	*Behaviour validity* Model output compared to historical data.	Model results were compared with historical data from a sub-acute extended care facility in Australia.
Wong (2010) [25]	To evaluate if smoothing the number of discharges over the week relieves the pressure on emergency departments (ED).	*Behaviour validity* Model output compared to historical data. *Structure validity* Direct structure tests (structure-confirmation test, parameter-confirmation test).	Model output, specifically those concerning ED and ward censuses and their respective length of stay, were compared with historical data from the Toronto General Hospital, Canada. Staff physicians at the hospital were engaged during the study and their feedback was used to validate model structure, assumptions and feedback.

Worni (2012) [26]	To estimate what impact a policy to deny reimbursement of total knee arthroplasty (TKA) patient fees will have on venous thromboembolism (VTE) rates and any unintentional consequences.		
Yu (2015) [27]	To explore the driving factors for a high proportion of patients in China not seeking medical care (also known as potential medical demand) and examine possible interventions.	*Behaviour validity* Model output compared to historical data.	Model output, specifically number of visits to hospital, were compared to historical data from the Chinese Health Statistical Yearbook.
Zulkepli (2012)*1 [28]	Present a case study using hybrid modelling (SDM-DES), explore patient flow in an integrated care system (IC) and the impact of patient admission on health professional stress level.		
Agent-based models (ABMs)			
Alibrahim (2018) [29]	To explore the effect of patient choice on the healthcare market, specifically providers that form accountable care organisations (ACO).	*Behaviour validity* Model output compared to published data. *Structure validity* Structure-orientated behavioural tests (extreme-condition test, parameter-confirmation tests).	Model results relating to financial and population outcomes were compared with those from ACO studies, identified in the literature. Patient bypass rates were comparable with previous studies of Medicare patients and per-patient payment figures were in line with estimates identified in the literature.
Djanatliev (2012)*2 [11]	Presenting the functionality of the Prospective Health Technology Assessment (ProHTA) tool, which can simulate the impact of optimised technology prospectively before physical development.	*Structure validity* Direct structure tests (conceptual confirmation test).	Model achieved credibility from review by experts (doctors, health economists, medical informaticians and knowledge management experts).
Einzinger (2013) [30]	To create a tool capable of comparing reimbursement schemes in outpatient care.	*Behaviour validity* Model output compared to reported data. *Structure validity*	The overall prevalence of disease in the model was comparable with national health survey data, confirming parameterisation of model with routine care data was valid.

		Structure-orientated behavioural tests (parameter-confirmation tests).	
Hutzsch. (2008) [31]	To determine which mix of patients should be admitted to specialised hospitals to optimise resource utility and to consider the impact of unplanned patient arrivals on this process.	*Behaviour validity* Model output compared to real data and shared with experts.	Model output, including the annual number of surgery patients, number of admission requests and back-up capacity usage in medium care, were comparable with real data from the Catharina Hospital Eindhoven, the Netherlands. Domain experts from the hospital were consulted and determined model output to be credible.
Huynh (2012) [32]	To assess the impact of redesigning medication administration process (MAP) workflow for registered nurses to improve medication administration safety.	*Behaviour validity* Model output compared to real data (F-test, T-test)	Model output, specifically amount of time spent performing tasks and variation in number of tasks performed, were compared to observed pilot study data using t-tests. An F-test was also used to determine equivalence of variance between simulated and observed pilot study data.
Kittipitta. (2016)*3 [33]	To examine patient flow in an outpatient clinic of an orthopedic department and explore interventions that can improve clinical services to reduce patient waiting times.	*Behaviour validity* Model output compared to real data (T-test)	Model output, specifically average waiting time, average throughput time and average utilisation, were compared to observed and recorded department operations using t-tests.
Liu (2014) [34]	To develop a tool that can be used as a decision support system for managers of emergency departments (ED) to assess risk, allocation of resources and identify weakness in emergency care service.		
Liu (2016) [35]	To explore how accountable care organisations (ACO) can impact payers, healthcare providers and patients under a shared savings payment model for congestive heart failure (COHF) and achieve optimal outcomes.	*Behaviour validity* Model output compared to published data.	Model output, specifically COHF-related hospitalisation rate and mortality rate, were compared to published results from clinical trials and national health reports.
Viana (2018)*3 [36]	To examine and improve patient flow through a pregnancy outpatient clinic in light of the uncertainty in demand for services from overdue patients.	*Behaviour validity* Model output compared to real data and reviewed by experts. *Structure validity*	Model output, specifically arrival patterns, was compared with historical data from the outpatient clinic at Akershus University Hospital, Norway and patient length of stay results were deemed credible by clinic staff. Clinic staff also reviewed the visualisation of the model.

		Structure-orientated behavioural tests (structure-confirmation test, extreme-condition test).	
Yousefi (2017) [37]	To apply group decision-making techniques for emergency department (ED) resource allocation and determine whether this approach improves performance indicators.		
Yousefi (2018) [38]	To examine the behaviour of patients who leave public hospital emergency departments (ED) without being seen and the impact of preventative policies.	*Behaviour validity* Model output compared to reported data (t-test). *Structure validity* Structure-orientated behavioural tests (structure-confirmation tests).	Model output, specifically total time patients spent in hospital and weekly number of discharged patients, was compared to historical data from the ED at Hospital Risoleta Tolentino Neves, Minas Gerais, Brazil using t-tests. The simulation model was reviewed by the hospital manager and district co-ordinator.

*Note: [*1] Articles implemented SDM-DES hybrid modelling. [*2] Articles implemented SDM-ABM hybrid modelling. [*3] Articles implemented ABM-DES hybrid modelling.*

Table S4: Descriptive table of ABM model rules

Agent-based models (ABMs)				
Paper/Year/Ref	Purpose	Type of agents	Key model rules	Type/source of rules
Alibrahim (2018) [29]	To explore the effect of patient choice on the healthcare market, specifically providers that form accountable care organisations (ACO).	Patient Provider (hospitals, primary care physician clinics) Payer (Medicare)	Patients could bypass their assigned healthcare provider based on patient characteristics (age, race, gender, income), provider characteristics (mortality rate, hospitalisation rate, presence of disease management programme in place by membership of ACO), and distance between providers. If patients chose to bypass, they must travel 60 mins to other facility. Providers delivered congestive heart failure care to patients and could chose to be part of the ACO network. If providers chose to be part of the ACO network they would share cost savings with the payer (Medicare) but have to provide comprehensive disease management programs. Payer agents calculate the savings made between providers participating in the ACO or not and return a percentage of savings to providers.	Literature (epidemiological studies, cost-effectiveness studies, national/public health surveys). Psychological theory (Theory of planned behaviour). Behaviour economics.
Einzinger (2013) [30]	To create a tool capable of comparing reimbursement schemes in outpatient care.	Patient Medical provider	A patient develops a chronic medical issue (such as coronary heart disease) that requires care, leading to the patient conducting a search of medical providers through the health market with preference for those who are closer. The health market (classified as a single object in model) returns a list of suitable providers using distance from the patient, suitable wait time and selecting providers who can provide the highest number of services required. The patient accesses care at their chosen provider. The reimbursement system is notified of this event via a generic interface and reimburses the medical provider for the patients care.	Utility/game theory (standard gamble utility assessment technique). Heuristics (greedy algorithm). Agent behaviour adapted from a prototype, universal model of a healthcare system.
Djanatliev (2012)*2 [11]	Presenting the functionality of the Prospective Health Technology Assessment (ProHTA) tool, which can	Patient Mobile Stroke Unit	Mobile Stroke Units (MSU), a new healthcare technology innovation, are presented as a case study. The ABM module simulates patient state (generally categorised as prevention, pre-treatment, treatment, post-treatment stages) which are reached by patients traversing a workflow, with decisions based on set	Observational studies. Determined by domain experts.

	simulate the impact of optimised technology prospectively before physical development.		probabilities when a patient has a stroke i.e. patient contacting emergency services, contacting GP, going directly to the hospital. If a MSU is available, this is dispatched to treat the patient which would lead to reduced long term complications.	
Hutzsch. (2008) [31]	To determine which mix of patients should be admitted to specialised hospitals to optimise resource utility and to consider the impact of unplanned patient arrivals on this process.	OR scheduling agent Resource agent	OR scheduling agent manages the use of cardiothoracic surgery (CTS) operating room. Resource agents are each of the units that form the CTS and intensive care unit (ICU) departments, such as the high care unit of CTS, the main ward for CTS, high care intensive care unit. Patient priority and care pathway is determined by selected patient characteristics. Each medical unit has their own preference for the type of patient they will admit. Hospital resources (such as use of operating rooms) are limited and with their availability in flux due to the need from other surgical disciplines and admission of emergency patients. Where emergency patients are admitted, bottlenecks can occur in ICU where beds are needed by elective CTS patients but priority is given to acute cases.	Routine health facility data. Determined by healthcare experts.
Huynh (2012) [32]	To assess the impact of redesigning medication administration process (MAP) workflow for registered nurses to improve medication administration safety.	Registered nurse	A registered nurse is engaged in a single task until its completion but can be interrupted by another health professional. Based on transition probabilities, the nurse then moves on to a new task until all tasks are completed.	Observational study.
Kittipitta. (2016)*3 [33]	To examine patient flow in an outpatient clinic of an orthopaedic department and explore interventions that can improve clinical services	Orthopaedic outpatient Doctor Nurse Healthcare assistant Radiologist Biomedical scientist Administration staff	Patients who are 85 years or older or have a particular condition requiring a fast track consultation are attended to by a doctor first, all other patients are sent to the waiting area. Scheduling of patients in the waiting area is dependent on the number of walk-in patients and scheduled patients are waiting. The doctor decides if the patient should be sent for an examination (to then return for another consultation), requires medication and when they are to be sent home.	Observational study. Determined by healthcare experts. Routine health facility data.

	to reduce patient waiting times.	Patient information system[p] Examination centre[p] Loudspeaker system[p]	The patient collects any medication required from the pharmacy.	
Liu (2014) [34]	To develop a tool that can be used as a decision support system for managers of emergency departments (ED) to assess risk, allocation of resources and identify weakness in emergency care service.	Patient Admission staff Triage Nurse Doctor Auxiliary staff Nurse Laboratory test Internal test External test Ambulance Carebox	Each agent has assigned behaviour such as waiting for the next task, arranging a test, providing treatment, moving a patient to a different area of the ward etc. Patients are admitted and triaged before tests are requested and a diagnosis issued. Over time a patient's status may change where the doctor will decide a new course of action (send the patient home, to another ward, or continue with diagnosis and treatment).	Determined by healthcare experts. Routine health facility data.
Liu (2016) [35]	To explore how accountable care organisations (ACO) can impact payers, healthcare providers and patients under a shared savings payment model for congestive heart failure (COHF) and achieve optimal outcomes.	Patient[p] Provider (hospitals, primary care physician clinics) Payer (Medicare)	Patients were passive and were not decision-makers. Providers considered whether to conduct the CHF intervention. Provider behaviour was dependent on the financial return of conducting the intervention, patient health outcomes, peer pressure from other providers and perceived difficulty in conducting intervention. Payer agents calculate the savings made between providers participating in the ACO or not and return a percentage of savings to providers.	Literature (epidemiological studies, cost-effectiveness studies, national/public health surveys). Psychological theory (Theory of planned behaviour). Behaviour economics.
Viana (2018)[*3] [36]	To examine and improve patient flow through a pregnancy outpatient clinic in light of the uncertainty in demand for services from overdue patients.	Patient	The assigned characteristics of the patient (particularly if the pregnancy is considered overdue) and utilisation of staff will determine where the patient is sent after arriving for her appointment, when the patient is attended to by a midwife or doctor and how long the entire process takes.	Determined by healthcare experts. Routine health facility data.

Yousefi (2017) [37]	To apply group decision-making techniques for emergency department (ED) resource allocation and determine whether this approach improves performance indicators.	Patient Doctor Technician Triage nurse Emergency room nurse Receptionist	Behaviour of agents are modelled by a finite state machine, where agent interactions result in a change of state. Each agent has a set list of possible tasks they may complete i.e. patients can wait for treatment, then receive treatment, then move to a different section of the ED etc. Agent communication also informed group decision-making whereby a group of agents could decide where to place resources (allocate a nurse to a different area of the ED) if an area of the ED was struggling.	Observational study. Multi-attribute decision making theory. Literature (modelling studies, observational studies, routine health facility data).
Yousefi (2018) [38]	To examine the behaviour of patients who leave public hospital emergency departments (ED) without being seen and the impact of preventative policies.	Patient Doctor Nurse Receptionist	Agents can communicate with each other, to a group of agents or agents can send a message to an area of the ED where other agents reside. Agents make decisions based on these interactions and information available to them at the time. Patients decide whether to leave the emergency department based on a tolerance time, which can change upon interaction with other agents.	Cellular automata. Observational study. Literature (modelling studies, observational studies, routine health facility data). Determined by healthcare experts. Routine health facility data.

*Note: *2 Articles implemented SDM-ABM hybrid modelling. *3 Articles implemented ABM-DES hybrid modelling. p Considered in the published model as a passive, non-decision-making agent.*

References

[1] Al-Khatib IA, Eleyan D, Garfield J. A system dynamics approach for hospital waste management in a city in a developing country: the case of Nablus, Palestine. Environ Monit Assess 2016;188:503. doi:http://dx.doi.org/10.1007/s10661-016-5487-9.

[2] Alonge O, Lin S, Igusa T, Peters DH. Improving health systems performance in low- and middle-income countries: a system dynamics model of the pay-for-performance initiative in Afghanistan. Health Policy Plan 2017;32:1417–26. doi:http://dx.doi.org/10.1093/heapol/czx122.

[3] Ansah JP, Eberlein RL, Love SR, Bautista MA, Thompson JP, Malhotra R, et al. Implications of long-term care capacity response policies for an aging population: a simulation analysis. Health Policy 2014;116:105–13. doi:https://dx.doi.org/10.1016/j.healthpol.2014.01.006.

[4] Brailsford SC. Emergency and on-demand health care: modelling a large complex system. J Oper Res Soc 2004;55.

[5] Brailsford SC, Desai SM, Viana J. Towards the holy grail: Combining system dynamics and discrete-event simulation in healthcare. Proc. 2010 Winter Simul. Conf., IEEE; 2010, p. 2293–303. doi:10.1109/WSC.2010.5678927.

[6] Cepoiu-Martin M, Bischak DP. Policy choices in dementia care - an exploratory analysis of the Alberta Continuing Care System (ACCS) using system dynamics. Spec Issue Complex Forum “real” World Heal Syst Implic Complex Theor Sci 2018;24:278–84.

[7] Chaerul M, Tanaka M, Shekdar A V. A system dynamics approach for hospital waste management. Waste Manag 2008;28:442–9. doi:http://dx.doi.org/10.1016/j.wasman.2007.01.007.

[8] Ciplak N, Barton JR. A system dynamics approach for healthcare waste management: A case study in Istanbul Metropolitan City, Turkey. Waste Manag {\&} Res J Int Solid Wastes Public Clean Assoc ISWA 2012;30:576–86. doi:https://dx.doi.org/10.1177/0734242X12443405.

[9] de Andrade L, Lynch C, Carvalho E, Rodrigues CG, Vissoci JRN, Passos GF, et al. System dynamics modeling in the evaluation of delays of care in ST-segment elevation myocardial infarction patients within a tiered health system. PLoS One 2014;9:e103577. doi:https://dx.doi.org/10.1371/journal.pone.0103577.

[10] Desai MS, Penn ML, Brailsford S, Chipulu M. Modelling of Hampshire Adult Services—gearing up for future demands. Health Care Manag Sci 2008;11:167–76. doi:10.1007/s10729-007-9049-0.

[11] Djanatliev A, German R, Kolominsky-Rabas P, Hofmann BM. Hybrid Simulation with Loosely Coupled System Dynamics and Agent-based Models for Prospective Health Technology Assessments. Proc. Winter Simul. Conf., Winter Simulation Conference; 2012, p. 69:1----69:12.

[12] Eleyan D, Al-Khatib IA, Garfield J. System dynamics model for hospital waste characterization and generation in developing countries. Waste Manag Res 2013;31:986–95. doi:http://dx.doi.org/10.1177/0734242X13490981.

[13] Esensoy AV, Carter MW. High-Fidelity Whole-System Patient Flow Modeling to Assess Health Care Transformation Policies. Eur J Oper Res 2018;266:221–37. doi:10.1016/j.ejor.2017.09.019.

[14] Ghaffarzadegan N, Epstein AJ, Martin EG. Practice variation, bias, and experiential learning in cesarean delivery: A data-based system dynamics approach. Health Serv Res 2013;48:713–34. doi:https://dx.doi.org/10.1111/1475-6773.12040.

[15] Lane D. Emergency but no accident: a system dynamics study of casualty waiting times in the British NHS. Eurohealth (Lond) 1998;4.

[16] Lane DC. Looking in the wrong place for healthcare improvements: a system dynamics study of an accident and emergency department. J Oper Res Soc 2000;51.

[17] Lattimer V, Brailsford S, Turnbull J, Tarnaras P, Smith H, George S, et al. Reviewing emergency care systems I: insights from system dynamics modelling. Emerg Med J 2004;21:685–91. doi:http://dx.doi.org/10.1136/emj.2002.003673.

[18] Mahmoudian-Dehkordi A, Sadat S. Sustaining critical care: using evidence-based simulation to evaluate ICU management policies. Health Care Manag Sci 2017;20:532–47. doi:http://dx.doi.org/10.1007/s10729-016-9369-z.

[19] Meker T, Barlas Y. Dynamic Consequences of Performance-Based Payment Systems in Public Hospitals. Syst Res Behav Sci 2015;32:459–80. doi:10.1002/sres.2338.

[20] Mielczarek B, Zabawa J. Modeling Healthcare Demand Using a Hybrid Simulation Approach. Proc. 2016 Winter Simul. Conf., IEEE Press; 2016, p. 1535–46.

[21] Rashwan W, Abo-Hamad W, Arisha A. A system dynamics view of the acute bed blockage problem in the Irish healthcare system. Eur J Oper Res 2015;247:276–93. doi:10.1016/j.ejor.2015.05.043.

[22] Semwanga AR, Nakubulwa S, Adam T. Applying a system dynamics modelling approach to explore policy options for improving neonatal health in Uganda. Heal Res Policy Syst 2016;14:35. doi:https://dx.doi.org/10.1186/s12961-016-0101-8.

[23] Taylor K, Dangerfield B. Modelling the feedback effects of reconfiguring health services. J Oper Res Soc 2005;56:659–75. doi:10.1057/palgrave.jors.2601862.

[24] Walker BC, Haslett T. The dynamics of local rules in hospital admission processes. Aust Health Rev 2003;26:98–106.

[25] Wong HJ, Wu RC, Caesar M, Abrams H, Morra D. Smoothing inpatient discharges decreases emergency department congestion: a system dynamics simulation model. Emerg Med J 2010;27:593–8. doi:https://dx.doi.org/10.1136/emj.2009.078543.

[26] Worni M, Pietrobon R, Zammar GR, Shah J, Yoo B, Maldonato M, et al. System dynamics to model the unintended consequences of denying payment for venous thromboembolism after total knee arthroplasty. PLoS One 2012;7:e30578. doi:http://dx.doi.org/10.1371/journal.pone.0030578.

[27] Yu W, Li M, Ge Y, Li L, Zhang Y, Liu Y, et al. Transformation of potential medical demand in China: A system dynamics simulation model. J Biomed Inform 2015;57:399–414. doi:http://dx.doi.org/10.1016/j.jbi.2015.08.015.

[28] Zulkepli J. HYBRID SIMULATION FOR MODELLING LARGE SYSTEMS: AN EXAMPLE OF INTEGRATED CARE MODEL. Proc. 2012 Winter Simul. Conf., 2012.

[29] Alibrahim A, Wu S. An agent-based simulation model of patient choice of health care providers in accountable care organizations. Health Care Manag Sci 2018;21:131–43. doi:http://dx.doi.org/10.1007/s10729-016-9383-1.

[30] Einzinger P, Popper N, Breitenecker F, Pfeffer N, Jung R, Endel G. The GAP-DRG Model: Simulation of Outpatient Care for Comparison of Different Reimbursement Schemes. Proc. 2013 Winter Simul. Conf. Simul. Mak. Decis. a Complex World, Piscataway, NJ, USA: IEEE Press; 2013, p. 2299–308.

[31] Hutzschenreuter AK, Bosman PAN, Blonk-Altena I, van Aarle J, La Poutré H. Agent-based patient admission scheduling in hospitals. Belgian/Netherlands Artif Intell Conf 2008:315–6. doi:10.1007/3-540-32062-8.

[32] Huynh N, Snyder R, Vidal J, Tavakoli A, Cai B. Application of computer simulation modeling to medication administration process redesign. J Healthc Eng 2012;3:649–62. doi:http://dx.doi.org/10.1260/2040-2295.3.4.649.

[33] Kittipittayakorn C, Ying K-C. Using the integration of discrete event and agent-based simulation to enhance outpatient service quality in an orthopedic department. J Healthc Eng 2016;2016:4189206. doi:http://dx.doi.org/10.1155/2016/4189206.

[34] Liu Z, Cabrera E, Rexachs D, Luque E. A Generalized Agent-Based Model to Simulate Emergency Departments. Sixth Int Conf Adv Syst Simul 2014:65–70.

[35] Liu P, Wu S. An Agent-Based Simulation Model to Study Accountable Care Organizations. Health Care Manag Sci 2016;19:89–101. doi:10.1007/s10729-014-9279-x.

[36] Viana J, Simonsen TB, Dahl FA, Flo K. A Hybrid Discrete Event Agent Based Overdue Pregnancy Outpatient Clinic Simulation Model. Proc. 2018 Winter Simul. Conf., Piscataway, NJ, USA: IEEE Press; 2018, p. 1488–99.

[37] Yousefi M, Ferreira R. An agent-based simulation combined with group decision-making technique for improving the performance of an emergency department. Brazilian J Med Biol Res 2017;50:1–10. doi:10.1590/1414-431X20175955.

[38] Yousefi M, Yousefi M, Fogliatto FS, Ferreira RPM, Kim JH. Simulating the behavior of patients who leave a public hospital emergency department without being seen by a physician: a cellular automaton and agent-based framework. Brazilian J Med Biol Res 2018;51:e6961. doi:http://dx.doi.org/10.1590/1414-431X20176961.

METHODOLOGY

3. Programme and study setting

3.1 Introduction

Abbreviations: 2 doses of intermittent preventative treatment (IPT2); Antenatal care (ANC); Antiretroviral therapy (ART); Council Health Management Team (CHMT); Focused Antenatal Care (FANC); Health Management Information System (HMIS); Human Immunodeficiency Virus (HIV); Low- and middle-income country (LMIC); Maternal and child health (MCH); Medical Stores Department (MSD); Ministry of Health and Social Welfare (MoHSW); Payment for performance (P4P); Sustainable Development Goal (SDG); Regional Health Management Team (RHMT); World Health Organisation (WHO).

In this chapter the programme theories underpinning use of P4P to improve health outcomes and service delivery in LMICs, landscape for MCH care in Tanzania and details on the Pwani P4P programme, the intervention of focus in this thesis, are presented.

3.2 Theories underpinning use of payment for performance to improve health system outcomes and potential for unintended consequences in low- and middle-income settings

A recent scoping review on the theories underpinning use of P4P within the health sector found economic and psychological theories were predominantly used to justify utilisation of the programme to improve health and service delivery outcomes, with a minority of articles also referencing theories from other disciplines such as political, social, management and organisation science (Paul *et al.* 2021). Economic theory was the most referenced theory base for use of P4P, specifically the principal agent-theory to enact positive system change. The agency theory stipulates that one organisation (the principal) issues work to another individual or collective (the agent), with compensation then paid to the agent for the work (Sekwat 2000; Lohmann *et al.* 2016). For P4P programmes in LMICs, the principal is normally the Ministry of Health or donor group, with health providers acting as the agent. Inherent goal conflict, where the agent and principal's interests and priorities are not aligned, is thought to be minimised by use of rewards or penalties (e.g. financial) to the agent. Specifically in reference to P4P, this is often managed through use of contracts agreed on by both parties where expectations for programme participation are clear, with compliance managed through supervision and verification activities (Lohmann *et al.* 2016).

Paul *et al.* (2021) described the economic theory base for utilisation of P4P to comprise of three broad categories of justification; theories related to neoclassical economics, new institutional economics and behavioural economics. Neoclassical economics (such as public choice theory) are concerned with uptake and supply of services, with efficiency to be encouraged through competition between different healthcare providers (Munger 2015). Institutional economic theories, such as contract theory, incentive theory and property rights theory, are generally concerned with the rules that influence stakeholder behaviour and interactions leading to altered economic performance, and what system structure might be optimal where individuals are pursuing multiple or conflicting goals (Preker *et al.* 2007; Paul *et al.* 2021). Lastly, behavioural economics, drawing from psychology to argue that rewards could increase healthcare provider motivation related to incentivised tasks but could also lead to undesirable outcomes (such as misreporting performance data) (Preker *et al.* 2007).

The psychological theory base for implementing P4P consists of theories that describe possible positive system change and caution against inducing perverse behaviour or system outcomes. The most highly referenced psychological theories related to P4P include self-determination theory (goals and regulatory processes through which goals are pursued, innate psychological needs as foundation for goal and process integration) (Deci and Ryan 2000), crowding-out theory (undermining of intrinsic motivation through appeal to extrinsic motivation with rewards system) (Lohmann *et al.* 2016), expectancy theory (motivation dependent on expectancy between effort and outcomes, and attractiveness of the outcome) (Ogundeji *et al.* 2018), cognitive evaluation theory (pathways through which extrinsic focussed rewards impact intrinsic motivation) (Deci and Ryan 1985; Bhatnagar and George 2016) and goal theory (motivation and performance increase with clear goals, difficult but achievable goals and presence of feedback to adjust future effort) (Preker *et al.* 2007).

Programme implementation theories specific to the Tanzania Pwani P4P programme are described in Chapter 8, with an overview of Pwani programme design given in section 3.4.

There are several potential unintended effects of programmes that use rewards-based systems to induce positive behaviour and outcomes (Miller and Singer Babiarz 2013). Positive unintended consequences of programme implementation include improvements to other services not directly targeted by the programme, ‘spill over effects’ (Sherry *et al.* 2017) and improvements in overall patient satisfaction of quality of care and patient-provider interactions (Diaconu *et al.* 2021).

‘Tunnel vision’ is an example of a negative unintended consequence of the programme, where health providers who are required to carry out multiple tasks may shift focus and effort away from non-incentivised activities (Holmstrom and Milgrom 1991; Aryankhesal *et al.* 2015). With programme designs that have different reward amounts for each targeted service, there is a risk that health providers might focus on performing tasks that have the highest marginal return

(Sherry *et al.* 2017). Where programmes stipulate the same reward for different services, this can also lead to focus of effort towards the easier to accomplish targets (Lagarde *et al.* 2013); both programme designs can lead to cherry picking of patients to improve health provider performance and boost incentive payments.

Crowding out of intrinsic motivation is also a concern, where a rewards-based system appeals to health providers extrinsic rather than intrinsic motivation to conduct service provision; externally controlled motivation (extrinsic) is thought to be more unstable and prone to changes in one's environment compared to internal (intrinsic) motivation (Deci and Ryan 2000; Lohmann *et al.* 2016). With a rewards system that requires submission of data, there is a risk that data may be misreported or distorted to conceal true performance (Kalk *et al.* 2010; Aryankhesal *et al.* 2015; Turcotte-Tremblay *et al.* 2020), although there is limited evidence within health on the incidence of gaming, with further uncertainty garnered by lack of knowledge on gaming in non-P4P health providers for comparison purposes (Van Herck *et al.* 2010). There are certain programme design features that aim to mitigate undesirable unintended consequences of P4P, for example, implementing gaming safeguarding measures such as auditing or introduction of penalties (Kovacs *et al.* 2020).

3.3 Tanzania and maternal and child health

Tanzania, located in East Africa, shares its land border with seven other sub-Saharan African countries, predominantly Kenya, Zambia and Mozambique. It has an estimated population of 59,734,213, with an annual population growth of 2.9% (The World Bank 2022a). In July 2020, Tanzania formally transitioned from a low-income country to lower-middle-income status, attributed to the county's continued macroeconomic stability, geographical location and natural endowments (The World Bank 2022b).

The healthcare system in Tanzania comprises broadly of six tiers of healthcare provider: (i) traditional or informal health providers; (ii) dispensaries; (iii) health centres; (iv) district hospitals; (v) regional hospitals and (vi) specialty hospitals. Each tier serves different catchment populations, offering varying levels of care, with different cadre numbers and composition (Maluka *et al.* 2018). There are approximately 6,640 dispensaries in Tanzania (Kapologwe *et al.* 2020), each have a catchment area of less than 10,000, with a minimum of one clinical assistant and nurse and offer fundamental outpatient care (Maluka *et al.* 2018). There are fewer health centres, approximately 695 (Kapologwe *et al.* 2020), with a catchment area of around 50,000; they offer a broader range of services by clinical officers and nurses, including inpatient care and some surgical services (Maluka *et al.* 2018). There are approximately 108 district hospitals that offer further inpatient, outpatient and surgical services not offered at primary care facilities

(Kapologwe *et al.* 2020). They accept referrals from primary care facilities (Ifeanyichi *et al.* 2021) and on average serve 250,000 people (Githendu *et al.* 2020). Each region in Tanzania has at least one regional level hospital (28 in total) (Ishijima *et al.* 2021), with further specialised care facilities.

In relation to provision of MCH services, primary health care facilities (dispensaries and health centres) in Tanzania usually provide antenatal care (ANC) (Chamani *et al.* 2021) and basic emergency obstetric and newborn care, including services such as facility-based deliveries and administration of antibiotics and labour medications (Ministry of Health and Social Welfare 2012). Where more comprehensive emergency obstetric and newborn care is required, patients may be referred to larger health centres or district, regional or tertiary hospitals, which offer advanced care such as caesarean section deliveries and blood transfusion services (Ministry of Health and Social Welfare 2012).

Tanzania has had varied progress in MCH over the last three decades (Afnan-Holmes *et al.* 2015; World Health Organisation 2015). Whilst under 2 years old immunisation against measles has fluctuated (84%, 2020) (The World Bank 2022a), indicators for the Sustainable Development Goal (SDG) 3, maternal mortality (524 per 100,000 live births, 2017) and under-5 mortality (49 per 1,000 live births, 2020) have seen steady improvement (The World Bank 2022c), although substantial continued progress is needed to achieve SDG targets (UNDP 2022).

A key strategy to reduce maternal and perinatal mortality, used by countries like Tanzania, has been to try and increase the number of women who deliver in a health facility (Kohi *et al.* 2018). In 2016, less than two-thirds of women in Tanzania chose to have a facility-based delivery (62.6%) (National Bureau of Statistics 2016); in 2022, the statistic is approximately 81.9% (Ministry of Health, Community Development, Gender, Elderly and Children 2022), on track for the 85% target set by the Tanzania Government for 2025 (Ministry of Health, Community Development, Gender, Elderly and Children 2021). There are a number of factors thought to influence a woman's decision to seek a facility-based delivery, including socio-demographic factors (like education and wealth) (Bishanga *et al.* 2018), geography (patients reside in urban vs. rural area) (Dewau *et al.* 2021), perceived risk of complications and birth preparedness (Bishanga *et al.* 2018; Konje *et al.* 2020), community, social and cultural factors (Mahiti *et al.* 2015; Konje *et al.* 2020), perceived quality of the health facility (including health provider attitude) (Mahiti *et al.* 2015; Armstrong *et al.* 2016), timing of first ANC visit (Mageda and Mmbaga 2015) and number of ANC visits (Feyissa and Genemo 2014; Bishanga *et al.* 2018).

In line with World Health Organisation (WHO) recommendations, the Ministry of Health and Social Welfare (MoHSW) in Tanzania implemented a Focused Antenatal Care (FANC) programme in 2002. The programme shifted focus from quantity of ANC visits, to strengthening the quality of patient visits (Women and Health Initiative & Maternal Health Task Force 2014). Health providers received training under FANC, with new guidelines and a checklist of actions to be performed in each ANC visit, including drug administration for prevention of malaria and history taking. The guidelines recommended four ANC visits, with an initial visit taking place before week 16 (Kasagama *et al.* 2022). The reduction in visits and focus on quality services was expected to improve continuity of care and negate some barriers to service uptake for patients (Women and Health Initiative & Maternal Health Task Force 2014). The results from the Tanzania Demographic Health Survey in 2016 (next survey is currently ongoing in 2022) indicate the percentage of women who have at least one ANC visit is high (98%), the percentage of women who receive 4 or more visits is relatively low (50.6%) (National Bureau of Statistics 2016). Factors attributed to patient ANC attendance include timely attendance of first ANC visit (Mrisho *et al.* 2009; Kasagama *et al.* 2022), presence of education and communication campaigns (Kasagama *et al.* 2022; Laisser *et al.* 2022), perceived quality and content of care (including health provider attitude) (Mrisho *et al.* 2009; Camacho *et al.* 2022; Laisser *et al.* 2022), and wealth status (Moshi 2021).

There exist broad critical health system-related barriers impeding delivery and coverage of MCH services in Tanzania. Health worker shortages and cadre imbalance at facilities is an issue in many regions, attributed in part to low recruitment rates, inconsistent quality of training and inadequate human resource management (MoHSW 2014; Futures Group 2015). The expected shortage of healthcare workforce is estimated to be 52% (Ministry of Health, Community Development, Gender, Elderly and Children 2021).

Provision of medical commodities is also a widespread issue. The Medical Stores Department (MSD) is an autonomous government body responsible for supply of medical commodities to public health facilities in Tanzania (MoHSW 2008). Government funding is deposited directly in MSD accounts and used for procurement (Githendu *et al.* 2020). Provision operates on a ‘pull system’, where facilities place orders on a quarterly basis to the MSD (Binyaruka and Borghi 2017). Dispensing of medicines at public health facilities is managed on a cost-sharing basis in Tanzania; for certain groups of patients (including pregnant women), services and medicines are supposed to be free, but in reality a large proportion of patients face out of pocket medical costs (Perkins *et al.* 2009; Neke *et al.* 2018). Historically, the MSD has faced issues in delivering its mandate, resulting in commodity stock outs at the facility level (although recent reforms have indicated some performance improvements) (Githendu *et al.* 2020). The 2014 Service Delivery Indicators survey (World Bank Group 2014) found that 49.1% of maternal priority medications were available across facilities in Tanzania, with 58.8% of child priority medications in stock.

When drugs are out of stock, patients resort to purchasing drugs at private pharmacies, increasing the cost of care seeking.

3.4 Pwani region's payment for performance programme

As part of a concerted effort to make progress towards Millennium Development Goals 4 and 5 (preceding the Sustainable Development Goals) (Borghi *et al.* 2013) concerning MCH, Tanzania introduced a P4P programme in Pwani region in 2011. The programme was implemented by the MoHSW in Tanzania, with funding from the Norwegian Ministry of Foreign Affairs. The programme encouraged health workers, district and regional level managers to reach certain MCH service and administration targets through provision of incentives for performance. Incentives were to be split between health workers and used to strengthen facility operations (e.g. purchasing additional medicines where needed). All facilities (hospitals, health centres and dispensaries) across the seven districts in the Pwani region (Figure 1) could participate if they opened or already had a facility bank account, delivered MCH services and provided service data from the previous year (2010).

The work presented in this thesis will focus on modelling the maternal care system and its response to the Pwani P4P programme in Tanzania. Tanzania was selected as a case study for this project due to the wealth of data already available on the programme (Borghi *et al.* 2013, 2021; Olafsdottir *et al.* 2014; Binyaruka *et al.* 2015, 2018a, 2018b; Anselmi *et al.* 2017; Binyaruka and Borghi 2017; Mayumana *et al.* 2017; Binyaruka and Anselmi 2020) and because the programme was shown to improve certain health service indicators and uptake of healthcare (Binyaruka *et al.* 2015), specifically percentage of institutional deliveries and percentage of ANC clients receiving at least 2 doses of intermittent preventative treatment (IPT2) for malaria.

During P4P, health providers would report Health Management Information System (HMIS) data on service delivery to district managers every cycle (6 months), where performance was then verified to ensure data quality. District and regional managers, and the national health management team would physically attend facilities and verify data (Anselmi *et al.* 2017). Data from the year preceding the programme (2010) was used to inform the first cycle (January – June 2011) of P4P targets. Targets for health providers were either to improve by a certain percentage each cycle ('percentage point increase') or to reach a certain level of absolute performance ('overall result') (Ministry of Health and Social Welfare 2012; Cassidy *et al.* 2021) (see Table 1).

Figure 1: Map of Tanzania by region, and map of Pwani region by district. Map created in ArcGIS using files downloaded from DIVA-GIS (Hijmans et al. 2004) and Map Library (Map Library).

Notes to Figure: The original GIS shapefile for Pwani region did not show new district Kibaha Town (National Bureau of Statistics 2016). The shapefile was updated to include this new district, therefore the district boundary between Kibaha and Kibaha Town may not be spatially accurate.

An example of a 'percentage point increase' style target is percentage of institutional deliveries; if the previous cycle performance was 70%, the facility would need to improve by at least 5% to receive an incentive payment. An example of an 'overall result' style target is percentage of ANC clients receiving 2 doses of intermittent preventative treatment (IPT2); facilities needed to achieve 80% or more each cycle to receive payment.

Payment was expected to be made within three months of the end of a cycle; however, payments were often delayed (Borghi *et al.* 2021). Payments received by providers were to be split in two; the larger portion (75%) was to be distributed between health workers at the facility and the smaller portion (25%) was to be used to improve facility operations (i.e. buying additional medicine or renovating facilities). District managers, otherwise known as the Council Health Management Team (CHMT) and regional managers, otherwise known as the Regional Health Management Team (RHMT), were also eligible to receive payments. Examples of district and regional targets include improving provider performance within the district/region, improving availability of medicines and timeliness of reporting.

Table 1: Facility coverage, content of care and HMIS strengthening indicators and performance targets set during the Pwani P4P programme in Tanzania for health facilities. Source: (Ministry of Health and Social Welfare 2012; Cassidy *et al.* 2021).

Indicator	**Measure**	**Baseline coverage (previous cycle)**				
		0-20%	**21-40%**	**41-70%**	**71-85%**	**85%+***
Coverage indicators						
% of institutional deliveries	Percentage point increase	15%	10%	5%	5%	Maintain
% of mothers attending a facility within 7 days of delivery	Percentage point increase	15%	10%	5%	5%	Maintain
% of women using long term contraceptives	Percentage point increase	20%	15%	10%	Maintain above 71%	Maintain
% children under 1 year receiving Penta3 vaccine	Overall result	50%	65%	75%	80% +	Maintain
% children under 1 year receiving measles vaccine	Overall result	50%	65%	75%	80% +	Maintain
Content of care indicators						
% ANC clients receiving IPT2	Overall result	80%	80%	80%	80%+	Maintain above 80%
% HIV+ ANC clients on ART	Overall result	40%	60%	75%	75%+	Maintain
% of newborns receiving polio vaccine (OPV0)	Overall result	60%	75%	80%	80%+	Maintain
HMIS strengthening						
HMIS monthly reports correctly filled and submitted on time to CHMT	Overall result	100%	100%	100%	100%	100%

Notes to Table: +*85% or more. Antiretroviral therapy (ART), Antenatal care (ANC), Council Health Management Team (CHMT), Health Management Information System (HMIS), Human Immunodeficiency Virus (HIV), Intermittent Preventative Treatment (IPT2).

References

Afnan-Holmes H, Magoma M, John T *et al.* 2015. Tanzania's Countdown to 2015: an analysis of two decades of progress and gaps for reproductive, maternal, newborn, and child health, to inform priorities for post-2015. *Lancet Glob Heal* **3**: e396–409.

Anselmi L, Binyaruka P, Borghi J. 2017. Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania. *Implement Sci* **12**: 10.

Armstrong CE, Martínez-Álvarez M, Singh NS *et al.* 2016. Subnational variation for care at birth in Tanzania: is this explained by place, people, money or drugs? *BMC Public Health* **16**: 795.

Aryankhesal A, Sheldon TA, Mannion R *et al.* 2015. The dysfunctional consequences of a performance measurement system: the case of the Iranian national hospital grading programme. *J Health Serv Res Policy* **20**: 138–45.

Bhatnagar A, George AS. 2016. Motivating health workers up to a limit: partial effects of performance-based financing on working environments in Nigeria. *Health Policy Plan* **31**: 868–77.

Binyaruka P, Anselmi L. 2020. Understanding efficiency and the effect of pay-for-performance across health facilities in Tanzania. *BMJ Glob Heal* **5**: e002326.

Binyaruka P, Borghi J. 2017. Improving quality of care through payment for performance: examining effects on the availability and stock-out of essential medical commodities in Tanzania. *Trop Med Int Heal* **22**: 92–102.

Binyaruka P, Patouillard E, Powell-Jackson T *et al.* 2015. Effect of Paying for Performance on Utilisation, Quality, and User Costs of Health Services in Tanzania: A Controlled Before and After Study. Ostermann J (ed.). *PLoS One* **10**: e0135013.

Binyaruka P, Robberstad B, Torsvik G *et al.* 2018a. Who benefits from increased service utilisation? Examining the distributional effects of payment for performance in Tanzania. *Int J Equity Health* **17**: 14.

Binyaruka P, Robberstad B, Torsvik G *et al.* 2018b. Does payment for performance increase performance inequalities across health providers? A case study of Tanzania. *Health Policy Plan* **33**: 1026–36.

Bishanga DR, Drake M, Kim Y-M *et al.* 2018. Factors associated with institutional delivery:

Findings from a cross-sectional study in Mara and Kagera regions in Tanzania. Zeng W (ed.). *PLoS One* **13**: e0209672.

Borghi J, Binyaruka P, Mayumana I *et al.* 2021. Long-term effects of payment for performance on maternal and child health outcomes: evidence from Tanzania. *BMJ Glob Heal* **6**: e006409.

Borghi J, Mayumana I, Mashasi I *et al.* 2013. Protocol for the evaluation of a pay for performance programme in Pwani region in Tanzania: A controlled before and after study. *Implement Sci* **8**: 80.

Camacho EM, Smyth R, Danna VA *et al.* 2022. Women's preferences for antenatal care in Tanzania: a discrete choice experiment. *BMC Pregnancy Childbirth* **22**: 296.

Cassidy R, Tomoaia-Cotisel A, Semwanga AR *et al.* 2021. Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach. *Soc Sci Med* **285**: 114277.

Chamani AT, Mori AT, Robberstad B. 2021. Implementing standard antenatal care interventions: health system cost at primary health facilities in Tanzania. *Cost Eff Resour Alloc* **19**: 79.

Deci EL, Ryan RM. 1985. Cognitive Evaluation Theory. *Intrinsic Motivation and Self-Determination in Human Behavior*. Boston, MA: Springer US.

Deci EL, Ryan RM. 2000. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychol Inq* **11**: 227–68.

Dewau R, Angaw DA, Kassa GM *et al.* 2021. Urban-rural disparities in institutional delivery among women in East Africa: A decomposition analysis. Wehrmeister FC (ed.). *PLoS One* **16**: e0255094.

Diaconu K, Falconer J, Verbel A *et al.* 2021. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev* **2021**.

Feyissa TR, Genemo GA. 2014. Determinants of institutional delivery among childbearing age women in Western Ethiopia, 2013: unmatched case control study. *PLoS One* **9**: e97194.

Futures Group. 2015. SNAPSHOT: TANZANIA'S HEALTH SYSTEM. Date of access: July 1, 2022. URL: https://www.healthpolicyproject.com/pubs/803_TanzaniaHealthsystembriefFINAL.pdf

Githendu P, Morrison L, Silaa R *et al.* 2020. Transformation of the Tanzania medical store department through global fund support: an impact assessment study. *BMJ Open* **10**: e040276.

Van Herck P, De Smedt D, Annemans L *et al.* 2010. Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res* **10**: 247.

Hijmans R, Guarino L, Bussink C *et al.* 2004. A geo-graphic information system for the analysis of biodiversity data: Manual for DIVA-GIS. Date of access: July 19, 2019. URL: http://www.diva-gis.org/

Holmstrom B, Milgrom P. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *J Law, Econ Organ* **7**: 24–52.

Ifeanyichi M, Broekhuizen H, Juma A *et al.* 2021. Economic Costs of Providing District- and Regional-Level Surgeries in Tanzania. *Int J Heal Policy Manag*, DOI: 10.34172/ijhpm.2021.09.

Ishijima H, Suzuki S, Masaule F *et al.* 2021. Measuring hospital performances of regional referral hospitals in Tanzania. *J Hosp Adm* **10**: 1.

Kalk A, Paul FA, Grabosch E. 2010. "Paying for performance" in Rwanda: does it pay off? *Trop Med Int Heal* **15**: 182–90.

Kapologwe NA, Meara JG, Kengia JT *et al.* 2020. Development and upgrading of public primary healthcare facilities with essential surgical services infrastructure: a strategy towards achieving universal health coverage in Tanzania. *BMC Health Serv Res* **20**: 218.

Kasagama E, Todd J, Renju J. 2022. Factors associated with changes in adequate antenatal care visits among pregnant women aged 15-49 years in Tanzania from 2004 to 2016. *BMC Pregnancy Childbirth* **22**: 18.

Kohi TW, Mselle LT, Dol J *et al.* 2018. When, where and who? Accessing health facility delivery care from the perspective of women and men in Tanzania: a qualitative study. *BMC Health Serv Res* **18**: 564.

Konje ET, Hatfield J, Kuhn S *et al.* 2020. Is it home delivery or health facility? Community perceptions on place of childbirth in rural Northwest Tanzania using a qualitative approach. *BMC Pregnancy Childbirth* **20**: 270.

Kovacs RJ, Powell-Jackson T, Kristensen SR *et al.* 2020. How are pay-for-performance schemes in healthcare designed in low- and middle-income countries? Typology and systematic literature review. *BMC Health Serv Res* **20**: 291.

Lagarde M, Wright M, Nossiter J *et al.* 2013. *Challenges of Payment-for-Performance in Health Care and Other Public Services – Design, Implementation and Evaluation*. Date of access: June 28, 2023. URL: https://piru.ac.uk/assets/files/Challenges%20of%20payment%20for%20performance%20in%20healthcare%20and%20other%20public%20services%20final.pdf

Laisser R, Woods R, Bedwell C *et al.* 2022. The tipping point of antenatal engagement: A qualitative grounded theory in Tanzania and Zambia. *Sex Reprod Healthc* **31**: 100673.

Lohmann J, Houlfort N, De Allegri M. 2016. Crowding out or no crowding out? A Self-Determination Theory approach to health worker motivation in performance-based financing. *Soc Sci Med* **169**: 1–8.

Mageda K, Mmbaga EJ. 2015. Prevalence and predictors of institutional delivery among pregnant mothers in Biharamulo district, Tanzania: a cross-sectional study. *Pan Afr Med J* **21**: 51.

Mahiti GR, Mkoka DA, Kiwara AD *et al.* 2015. Women's perceptions of antenatal, delivery, and postpartum services in rural Tanzania. *Glob Health Action* **8**: 28567.

Maluka S, Chitama D, Dungumaro E *et al.* 2018. Contracting-out primary health care services in Tanzania towards UHC: how policy processes and context influence policy design and implementation. *Int J Equity Health* **17**: 118.

Map Library. 2019. *Map Mak Ltd*. Date of access: July 19, 2019. URL: http://www.maplibrary.org

Mayumana I, Borghi J, Anselmi L *et al.* 2017. Effects of Payment for Performance on accountability mechanisms: Evidence from Pwani, Tanzania. *Soc Sci Med* **179**: 61–73.

Miller G, Singer Babiarz K. 2013. *Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs*. Date of access: June 28, 2023. URL: https://ssrn.com/abstract=2245441

Ministry of Health, Community Development, Gender, Elderly and Children. 2021. *Health Sector Strategic Plan July 2021 June 2026 (HSSP V)*. Date of access: December 14, 2022. URL: https://mitu.or.tz/wp-content/uploads/2021/07/Tanzania-Health-Sector-Strategic-Plan-V-17-06-2021-Final-signed.pdf

Ministry of Health, Community Development, Gender, Elderly and Children. 2022. RCH: Labour and delivery data. *HMIS*. Date of access: December 14, 2022. URL:

https://hmisportal.moh.go.tz/hmisportal/#/dashboards/routine/hmisportal_Y2vo6aovxhj

Ministry of Health and Social Welfare. 2012. *The Pwani Region Pay-for-Performance (P4P) Pilot Design Document*.

MoHSW. 2008. *Mapping of the Medicines Procurement and Supply Management System in Tanzania*. Date of access: December 14, 2022. URL: https://docplayer.net/17837515-Mapping-of-the-medicines-procurement-and-supply-management-system-in-tanzania.html

MoHSW. 2014. *Human Resource For Health and Social Welfare Strategic Plan 2014 - 2019*. Date of access: December 14, 2022. URL: https://www.jica.go.jp/project/tanzania/006/materials/ku57pq00001x6jyl-att/HRHSP_2014-2019.pdf

Moshi F V. 2021. Prevalence and factors which influence early antenatal booking among women of reproductive age in Tanzania: An analysis of data from the 2015-16 Tanzania Demographic Health Survey and Malaria Indicators Survey. Marotta C (ed.). *PLoS One* **16**: e0249337.

Mrisho M, Obrist B, Schellenberg JA *et al.* 2009. The use of antenatal and postnatal care: perspectives and experiences of women and health care providers in rural southern Tanzania. *BMC Pregnancy Childbirth* **9**: 10.

Munger MC. 2015. Public Choice Economics. *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, 534–9.

National Bureau of Statistics. 2016. *Tanzania Demographic and Health Survey and Malaria Indicator Survey*.

Neke N, Reifferscheid A, Buchberger B *et al.* 2018. Time and cost associated with utilization of services at mobile health clinics among pregnant women. *BMC Health Serv Res* **18**: 920.

Ogundeji YK, Sheldon TA, Maynard A. 2018. A reporting framework for describing and a typology for categorizing and analyzing the designs of health care pay for performance schemes. *BMC Health Serv Res* **18**: 686.

Olafsdottir AE, Mayumana I, Mashasi I *et al.* 2014. Pay for performance: an analysis of the context of implementation in a pilot project in Tanzania. *BMC Health Serv Res* **14**: 392.

Paul E, Bodson O, Ridde V. 2021. What theories underpin performance-based financing? A scoping review. *J Health Organ Manag* **35**: 344–81.

Perkins M, Brazier E, Themmen E *et al.* 2009. Out-of-pocket costs for facility-based maternity care in three African countries. *Health Policy Plan* **24**: 289–300.

Preker AS, Liu X, Velenyi EV *et al.* 2007. *Public Ends, Private Means : Strategic Purchasing of Health Services.* World Bank (ed.). Washington DC.

Sekwat A. 2000. Principal-agent theory: a framework for improving health care reform in Tennessee. *J Health Hum Serv Adm* **22**: 277–91.

Sherry TB, Bauhoff S, Mohanan M. 2017. Multitasking and Heterogeneous Treatment Effects in Pay-for-Performance in Health Care: Evidence from Rwanda. *Am J Heal Econ* **3**: 192–226.

The World Bank. 2022a. Tanzania, Overview. Date of access: June 30, 2022. URL: https://data.worldbank.org/country/tanzania

The World Bank. 2022b. The World Bank in Tanzania. Date of access: June 30, 2022. URL: https://www.worldbank.org/en/country/tanzania/overview

The World Bank. 2022c. Tanzania, SDG Goal. Date of access: June 30, 2022. URL: https://data.worldbank.org/country/tanzania

Turcotte-Tremblay A-M, Gali Gali IA, Ridde V. 2020. An Exploration of the Unintended Consequences of Performance-Based Financing in 6 Primary Healthcare Facilities in Burkina Faso. *Int J Heal Policy Manag*.

UNDP. 2022. Sustainable Development Goals, Good Health and Well-being. Date of access: June 30, 2022. URL: https://www.undp.org/sustainable-development-goals#good-health

Women and Health Initiative & Maternal Health Task Force. 2014. *Focused Antenatal Care in Tanzania*. Date of access: June 30, 2022. URL: https://cdn2.sph.harvard.edu/wp-content/uploads/sites/32/2014/09/HSPH-Tanzania5.pdf

World Bank Group. 2014. *Tanzania 2014 Service Delivery Indicators: Health Technical Report*. Date of access: June 30, 2022. URL: https://openknowledge.worldbank.org/bitstream/handle/10986/24796/Health0service0delivery0in0Tanzania.pdf?sequence=1&isAllowed=y

World Health Organisation. 2015. *Countdown to 2015. A Decade of Tracking Progress for Maternal, Newborn and Child Survival - The 2015 Report.* Date of access: June 30, 2022. URL: https://www.countdown2030.org/2015/2015-final-report

4. Causal loop diagram

4.1 Introduction

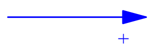
Abbreviations: Antenatal care (ANC); Causal loop diagram (CLD); Community Health Fund (CHF); Council Health Management Team (CHMT); Focus group discussion (FGD); Health Facility Governing Committee (HFGC); Maternal and child health (MCH); Medical Stores Department (MSD); Payment for performance (P4P).

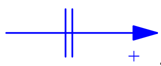
This chapter describes and expands on the methodology presented in the published papers (Cassidy *et al.* 2021) (see Chapter 6) and (Cassidy *et al.* 2022) (see Chapter 7) for CLD development and validation. The chapter opens with why a CLD approach is used in this thesis, an overview of the building blocks of CLDs and an outline of the method. Further details on data sources, creation of the CLD and presentation of the CLD then follow.

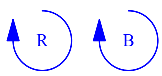
As previously mentioned, CLDs help us to visualise structural drives for system behaviour; systematic causes for suboptimal (and optimal) health system outcomes. For programme implementation research, this holistic system perspective can also yield evidence of possible spill over effects of policies to the wider system and unexpected or unintended consequences of policies, knowledge critical for policy design and implementation. A CLD approach is therefore used in this thesis to further understand pathways to impact for P4P programmes, identify bottlenecks that affect service delivery and success of the programme, and provide recommendations for the design of the programme based on findings, which fulfil **Objectives 2 and 3** of this thesis:

> 2. Identify health system factors and feedback loops that facilitate or hinder the implementation of P4P programmes and its overall effectiveness.
>
> 3. Identify system leverage points which should be considered in the design of P4P programmes.

4.2 Causal loop diagrams

CLD notation (Sterman 2000) consists of system elements Variable, arrows with attributed polarity that represent the relationships between system elements 

, some of which contain delay functions 

. Where two or more variables are connected, they can form reinforcing and balancing loops 

.

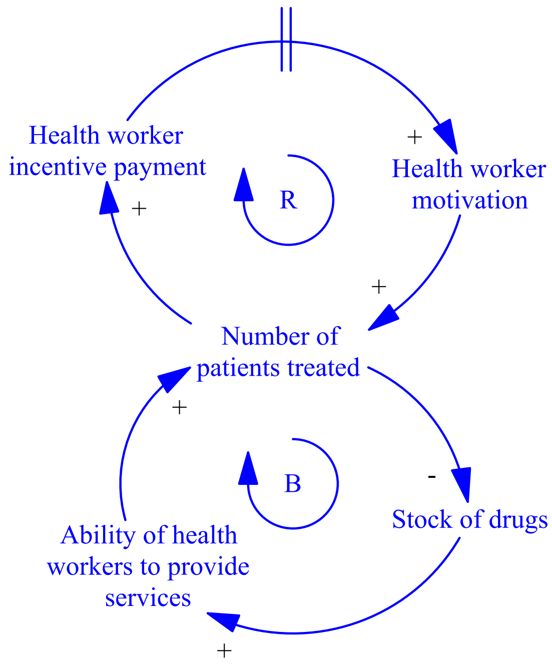


Figure 1: Simple example of a CLD. Source: Adapted from Cassidy et al. (2021).

A simple example of a CLD is given in Figure 1. The cause and effect relationship between 'number of patients treated' and 'health worker incentive payment' is represented with a positive arrow, which can be interpreted as an increase in the number of patients leads to an increase to health worker incentive payments. The relationship between health worker incentive payment and health worker motivation is represented with a positive delay arrow, which can be interpreted as health worker incentive payments lead to further motivated workforce (after a time delay). These three variables are connected and form a reinforcing, amplified cycle of behaviour referred to as a feedback loop; the more patients that are treated leads to an increase in health worker incentive payments, which leads to an increase in health worker motivation, which leads to an increase in the number of patients treated, and so it continues. The cycle forms a growing, positive action over time. Where cycles exhibit desirable system behaviour that should be promoted, they are referred to as 'virtuous' cycles of behaviour, where they exhibit undesirable behaviour they are labelled 'vicious' cycles.

Balancing loops form where cycles of behaviour are prevented from exhibiting growing action by the presence of one or more variables, creating a dampening effect on cycle behaviour. In the example, an increase in the stock of drugs leads to ability of health workers to provide services, which facilitates patient treatment, which in turn reduces the availability of medicines at the facility. The cycle is prevented from spiralling continuously by the reduction in availability of medicines as a result of service delivery.

In this thesis, both primary (stakeholder consultation) and secondary (previous programme evaluation data) sources were used to develop a CLD of the Pwani P4P programme in Tanzania. First, an initial CLD, describing shared understanding of programme effects on the health system was developed using previous evaluation data collected on the Pwani programme (detailing health worker and district manager experiences of the programme). This initial CLD was then validated using a portion of the programme evaluation data that had been set aside to be used for validation and stakeholder consultations (conducted at the time of this study). Analysis of the developed CLD revealed a number of catalytic variables (system variables that affect multiple outcomes and should be considered when designing P4P programmes) and system levers (system variables not targeted by the current programme design but should be too enhance the effect of the programme), leading to generation of recommendations for future programme implementation. Results from analysis of the CLD can be found in Chapter 6.

4.3 Data sources

4.3.1 Secondary data

Qualitative data collected as part of a process evaluation of the Pwani P4P programme (Borghi et al. 2013) were used to develop and validate the CLD (Table 1). The evaluation sought to assess how P4P had been received and implemented in health facilities, factors that facilitated or hindered the success of the programme and identify any unintended outcomes, with three round of data collection conducted between December 2011-March 2013 (Borghi et al. 2013). Data collection was conducted via one-on-one interviews or focus group discussions (FGDs), consisting of forty-three stakeholder interviews (with members of the CHMT, those in charge of facilities or MCH services and health workers) and eight FGDs with CHMTs, Health Facility Governing Committees (HFGCs) and health workers. Stakeholders were selected for interview from five (Bagamoyo, Kibaha Town, Kisarawe, Mafia island and Mkuranga) of the seven Pwani districts (Chapter 3, Figure 1), with ten primary care facilities purposively sampled to obtain data from providers with various levels of care and facility ownership. Data were collected via audio recording in Swahili by four local researchers working in pairs, with transcripts produced verbatim in English, in Word software.

Table 2: Description of Pwani P4P programme process evaluation data collected between December 2011-March 2013. Interviews and focus group discussions describing stakeholder experiences of the programme and programme implementation. Source: Adapted from Cassidy *et al.* 2021.

District	**Facility/CHMT**	**Stakeholder Type**	**No. of Interviews**	**No. of FGDs (no. of participants in each FGD)**
District A	Health Centre	Health worker	2	
	Dispensary	Health worker	3	
	CHMT	CHMT	5	1 (7)
District B	Health Centre	Health worker	4	
	Dispensary	Health worker	1	
		HFGC		1 (5) and 1 (4)
	CHMT	CHMT	3	
District C	Dispensary	Health worker	1	
	Dispensary	Health worker	1	
		HFGC		1 (5)
	CHMT	CHMT	3	
District D	Health Centre	Health worker	2	
	Dispensary	Health worker	1	
	CHMT	CHMT	4	1 (9)
District E	Health Centre	Health worker	4	1 (7)
		HFGC		1 (4)
	Dispensary	Health worker	4	
	CHMT	CHMT	5	1 (10)
Total			**43**	**8 (51 participants)**

Notes to Table: Council Health Management Team (CHMT) Focus group discussions (FGDs), Health Facility Governing Committee (HFGC).

4.3.2 Primary data

Further interviews were conducted at the time of this study to complement the above secondary dataset. The CLD that was created and validated using the described secondary data was presented to stakeholders during three rounds of stakeholder engagement cycles between March – November 2020 as an additional step for model validation. Twenty-one stakeholders who participated in the evaluation or implementation of the Pwani programme were approached via email to contribute to the study. For the last cycle of stakeholder discussions, invited stakeholders received a flyer (Appendix 2) and link to a short film co-developed by the PhD candidate, research team and Preston Street Films, describing the project and information to be requested from stakeholders. Consent was sought from all participants through a consent form, with verbal consent for participation and audio/written recording of consent taken again at the start of the interview. At the start of the interview, the participant information sheet was summarised and stakeholders were given time to ask questions. Stakeholders were interviewed individually and asked to comment on the structure of the developed CLD to ascertain if the

CLD resonated with their experience of the programme and knowledge of the health system or whether modifications needed to be made. The topic guide, participant information sheet and consent form can be found in Appendix 2 of this thesis.

4.4 Creation of CLD

There were three steps to developing the CLD (i) creation of individual CLDs; (ii) creation of one initial shared CLD and, (iii) validation of initial shared CLD. Each step is discussed in detail below, an overview is given here. In the first step, the secondary process evaluation data (Table 1) were used to develop individual CLDs (Kim and Andersen 2012; Tomoaia-Cotisel 2018) representative of stakeholder experience and understanding of how P4P affects their day to day activities. In the second step, these singular CLDs were combined in a step-wise process (Tomoaia-Cotisel 2018) to develop a single CLD, a initial shared mental model of the mechanism through which P4P impacts the health system. In the third step, the CLD underwent validation to check to what extent additional stakeholders interviewed at the time of the original data collection agreed on the structure of the system (using a subset of the Pwani programme evaluation data, set aside for validation) and to check to what extent stakeholders consulted at the time of this study agreed that the CLD represented their knowledge and experience of the programme and the health system (stakeholder consultations). The initial, shared CLD pre- and post-validation can be viewed in Appendix 4.

4.4.1 Step 1: Creation of individual CLDs

The secondary data described in the previous section were split into two datasets, used for development and validation of the CLD. Transcripts from districts A, C and E (Table 1) were used in Step 1 and 2 to develop a single, initial shared CLD (Figure 2). These districts were selected as they offered variation in terms of geographic location. Transcripts from the remaining two districts (B and D) were used for validation of the initial shared CLD (see Step 3: Validation of initial shared CLD)

To develop the initial individual CLDs, Purposive Text Analysis (Kim and Andersen 2012) adapted for CLDs (Tomoaia-Cotisel 2018) was used to extract information from the secondary data for CLD development. The approach involved systematically reading stakeholder transcripts and extracting cause and effect quotations that described how stakeholders received the intervention in their facility or district, or described factors that facilitated or hindered health provider success during the programme. Diagrams are then created to show these cause-and-effect relationships, which when drawn together form a CLD. Excel was used to store the extracted information, with Vensim software used to draw the CLD representative of each stakeholder transcript (Ventana Systems Inc. 2015)

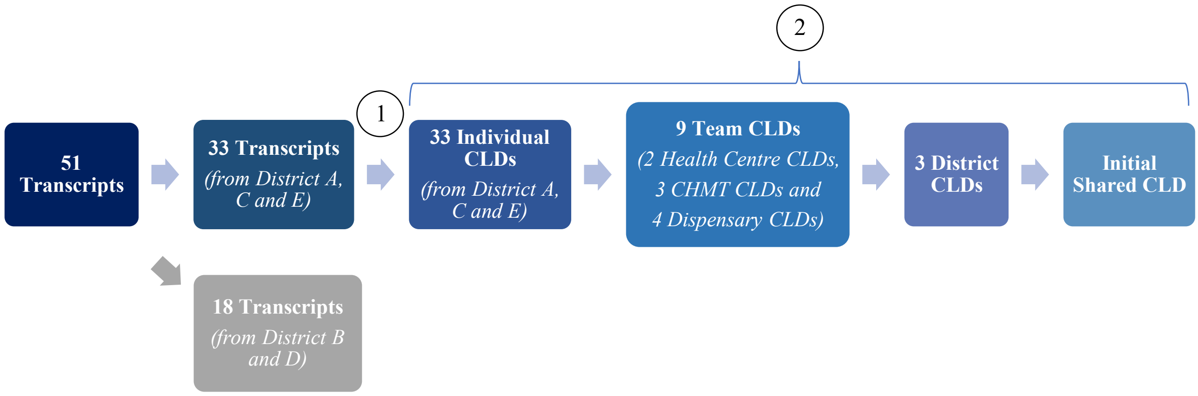


Figure 2: Process for creating (1) individual CLDs and (2) initial *shared* CLD. Source: Cassidy *et al.* 2021

Notes to Figure: Step (1) Development of individual CLDs based on districts A, C and E (Table 1) and Step (2) Merging individual causal loop diagrams to create a single shared causal loop diagram. The 51 transcripts comprise of the transcripts from 43 individual interviews and from 8 focus group discussions. A CLD was developed for each transcript, one FGD transcript was used to develop one CLD. Adapted from Tomoaia-Cotisel (2018). Causal Loop Diagram (CLD), Council Health Management Team (CHMT).

An example of using purposive text analysis to perform data extraction and development of a CLD is given in Box 1. In the example, the interviewer asked the stakeholder how health providers addressed challenges to the provision of quality health care in their facilities during a payment for performance programme (1a). Quotations were deemed relevant and extracted if they described events or scenarios that furthered understanding of how stakeholders responded to the programme or demonstrated health system behaviour that facilitated or hindered facilities delivering quality health care (1b). Isolated cause and effect statements, with their associated quotations were extracted from transcripts and stored in an Excel file. The direction of the relationship (positive or negative) was also noted; in the given example, an increase in the stock of drugs and equipment at facilities resulted in providers being able to deliver health services (1c). At the end of this data extraction process, all cause and effect statements were drawn as simple diagrams with a polarity indicating the direction of the relationship (1d). Each of these simple diagrams were then combined to form a single CLD representative of an individual's mental model of the system (1e).

Box 1: Example of applying Purposive Text Analysis to text. Source: Adapted from Cassidy *et al.* 2022.

(1a) Question: Are there any strategies being implemented that aim to address these challenges (to provision of quality health services)?

(1b) Quotation: 'Yes, there is strategy done in the district, which is community health fund. We realized that the shortage of equipments and drugs was becoming a common problem which resulted in poor health service delivery [1], the community health fund was established as alternative to solve those problems. So once the government supply insufficient medicine [2] the community health fund money are used to substitute [3/4].

Main argument: When the Medical Stores Department (autonomous government department that procures and distributes health commodities to facilities, MSD) cannot provide drugs and equipment, facilities must draw on other sources of funding like the community health fund to buy medical commodities.

(1c) Causal structure:

[1]

Causal variable	*Relationship*	*Effect variable*
Stock of drugs/equipment	Increase	Delivery of health services

[2]

Causal variable	*Relationship*	*Effect variable*
MSD issue facility resources	Increase	Stock of drugs/equipment

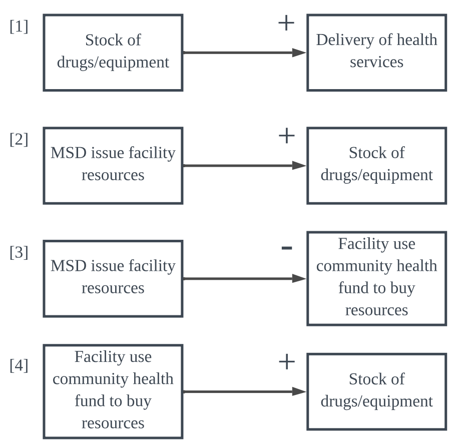
[3]

Causal variable	*Relationship*	*Effect variable*
MSD issue facility resources	Decrease	Facility use community health fund to buy resources

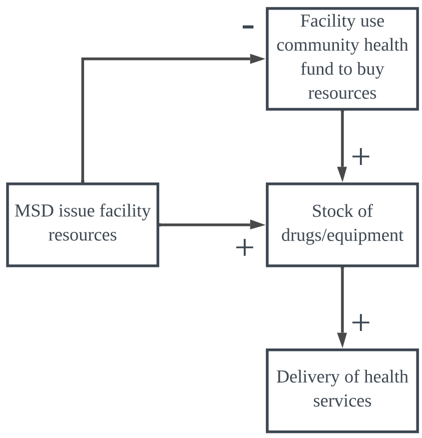
[4]

Causal variable	*Relationship*	*Effect variable*
Facility use community health fund to buy resources	Increase	Stock of drugs/equipment

(1d) Causal structure diagrams



(1e) Part of a larger causal loop diagram



Stakeholders may use different terminology to describe events; as coding progresses, it becomes easier to standardise variable names assigned to cause and effect statements. Each stakeholder CLD was then 'mildly pruned' (Tomoaia-Cotisel 2018; Milsom 2021) to retain feedback loops and those linear feedback linkages that contained delays. In pruning, any singular, linear, feedback linkage that did not contain a delay is removed from the CLD. Pruning is necessary so that the CLD contains that feedback which accounts for and describes most of the system behaviour over the time period of interest (the length of P4P programme duration in Pwani).

4.4.2 Step 2: Creation of one initial shared CLD

To create the initial shared CLD (Figure 2), individual CLDs were combined into team CLDs (representative of facility or district management) through a process called CLD Combination (Tomoaia-Cotisel 2018). Individual stakeholder CLDs within teams were ordered according to their level of 'complexity', in the sense of having the highest number of variables, links, loops and delays. The most complex CLD was labelled the 'anchor' CLD and compared to the second most complex CLD. This altered CLD was then compared to the third most complex CLD and so on until all individual stakeholder CLDs within that team had been combined into one team CLD. Next, team CLDs were combined into three district level CLDs using the same approach. Lastly, the three-district level CLDs were combined to create a shared (single) CLD. This initial shared CLD can be viewed in Appendix 4 (Figure A4.1).

There were three possible avenues for action when an anchor CLD was compared to the new CLD to make a combined CLD: (i) addition of new information from the new CLD; (ii) selection of further complexity from the new CLD and, (iii) merging of variables (Tomoaia-Cotisel 2018). The combined CLD may be developed as a result of multiple actions (e.g. addition of new information and selection of a further complex loop). An example of an 'addition' action is shown in Figure 3. The anchor CLD (Figure 3A) is altered to reflect information extracted from the new CLD (Figure 3B) to create a new combined CLD (Figure 3C).

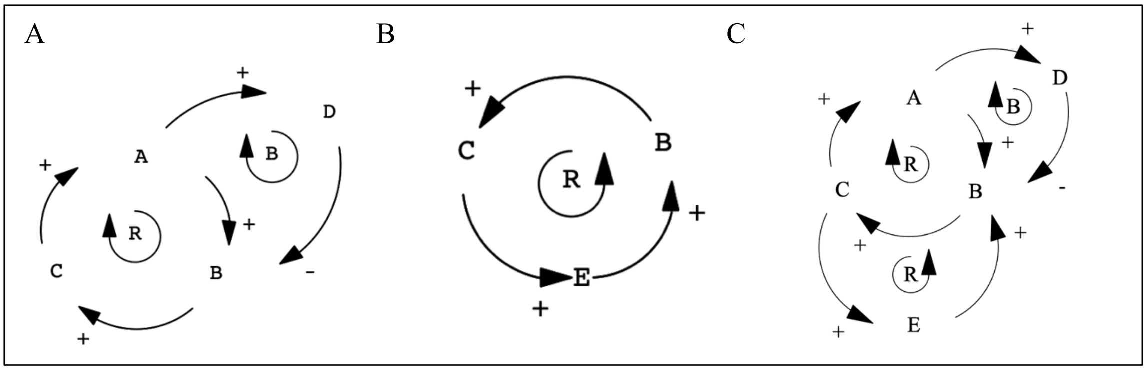


Figure 3: An example of an ‘additive’ action; the anchor CLD (A) is altered to reflect information extracted from the new CLD (B) to create a new combined CLD (C).

An example of a ‘selective’ action is shown in Figure 4. The anchor CLD (Figure 4A) is altered to reflect more complex information on a feedback loop or dynamic identified in the new CLD (Figure 4B) to create a new combined CLD (Figure 4C). The dynamic exists in a simpler loop in the anchor CLD (Figure 4A) but the additional information provided by the new CLD (Figure 4B) contributes to further understanding on the dynamic behaviour of that loop. The dynamic behaviour previously described may not have been adequately expressed, perhaps due to stakeholder knowledge or experience.

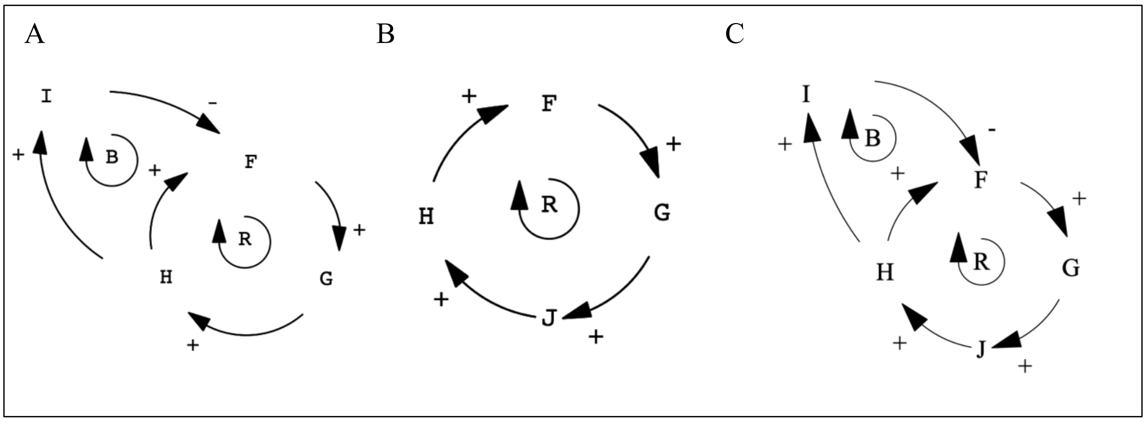


Figure 4: An example of an ‘selective’ action; the anchor CLD (A) is altered to reflect more complex information on a feedback loop or dynamic identified in the new CLD (B) to create a combined CLD (C).

An example of a ‘merging’ action is shown in Figure 5. The anchor CLD (Figure 5A) and new CLD (Figure 5B) contain feedback loops or dynamics that describe the same system behaviour that could be condensed into a simpler loop in the combined CLD (Figure 5C). For example, in the anchor CLD, a stakeholder describes availability of drugs leading to increased facility performance during P4P. In the new CLD, another stakeholder describes availability of medical supplies (i.e. gloves) leading to increased performance during the programme. This feedback could be condensed to show 'availability of medical commodities' leads to increased performance.

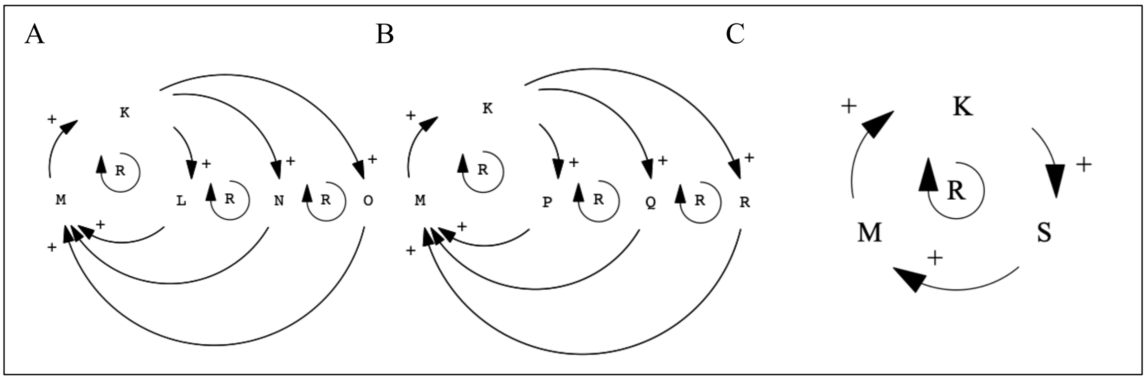


Figure 5: An example of an 'merging' action; the anchor CLD (A) and new CLD (B) contain feedback loops or dynamics that describe the same system behaviour and could be condensed into a simpler loop in the new combined CLD (C).

Although most individual stakeholder CLDs and combinations of CLDs proved to either be (i) shared experiences of the programme or (ii) not directly contradictory experiences (adding further complexity or segments to the diagram), there were occasions where stakeholders reported different information on system processes (such as facility financing mechanisms). When this happened, a review of the data extraction table and original transcript took place to check researcher interpretation of the reported process or feedback, and consideration of which stakeholder would have a more intimate experience of the process. This would also prompt further follow up of the system process in stakeholder consultations.

4.4.3 Step 3: Validation of initial shared CLD

Lastly, validation of the initial shared CLD was performed to ensure that critical input from each of the three stakeholder groups (health centres, dispensaries and CHMT) had not been lost or misinterpreted during the CLD development process (Figure 6). Validation comprised of two stages: first, the initial shared CLD was validated to check to what extent process evaluation data from additional districts agreed with the structure of the system (Tomoaia-Cotisel 2018). Second, the updated initial shared CLD was further validated to check to what extent additional stakeholders interviewed at the time of this study agreed that the CLD reflected their experience of the programme (Andersen *et al.* 2012; Rwashana *et al.* 2014). Changes made to the CLD at each validation stage can be viewed in Appendix 4.

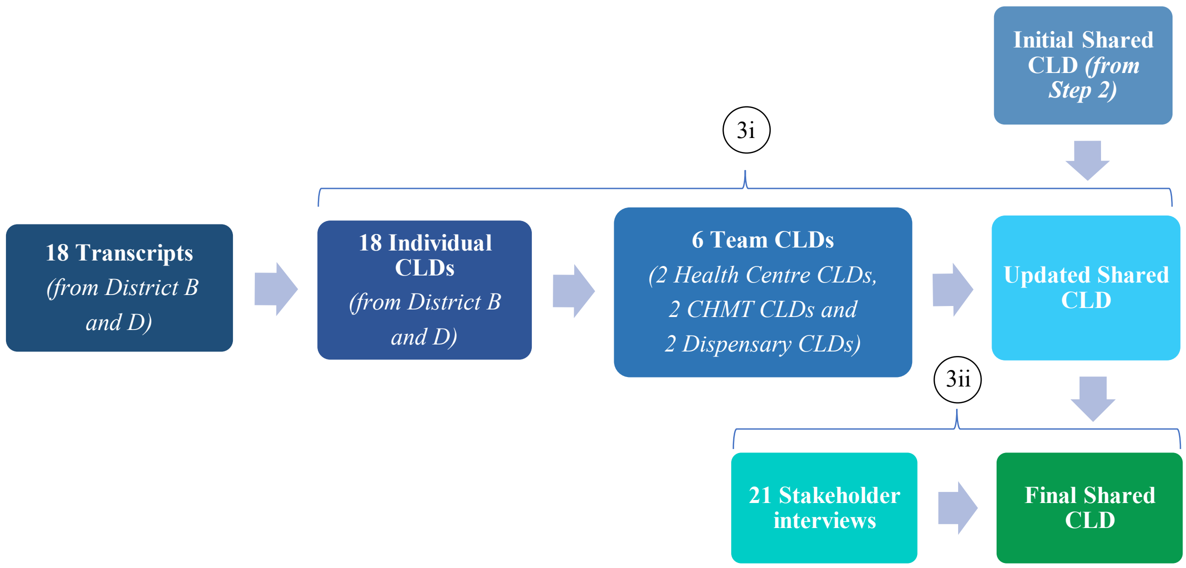


Figure 6: Validation of initial shared CLD. Source: Cassidy *et al.* 2021.

Notes to Figure: Step (3i) Comparison with team level CLDs that were not used to develop the shared CLD in the previous stage. Step (3ii) New stakeholder interviews to validate CLD structure. Adapted from Tomoaia-Cotisel (2018), Andersen *et al.* (2012) and Rwashana *et al.* (2014). Causal Loop Diagram (CLD), Council Health Management Team (CHMT).

In the first stage, transcripts from districts B and D were used to generate 18 individual CLDs which were then combined into six team CLDs. Each team level CLD was compared to the initial shared CLD from Step 2, to see to what extent the team level CLDs confirmed the structure of the initial shared CLD, or if any changes were required to the diagram (Tomoaia-Cotisel 2018). The aim of this validation test was to explore whether saturation has been reached in the diagram based on the addition of further team level CLDs; does the comparison to unused team CLDs result in changes to the initial shared CLD, or does the diagram suitably represent feedback and processes as reported by key stakeholders?

With each comparison and subsequent combination, the number of new links and variables were noted and used to create diagram saturation curves (Tomoaia-Cotisel 2018). The variable saturation curve (Figure 7) and link saturation curve diagrams (Figure 8) illustrate this information and show if saturation of information has been reached with each new combination. The variable saturation curve reports an initial increase in the number of variables with the first CLD team comparison (four new variables), no new variables when compared to the second, third and fourth team CLD, one further addition made by comparison to the fifth team CLD, with no further additions through comparison with the sixth team CLD. The link saturation curve also shows an increase in the number of new links as a result of CLD combination; four new links through comparison with the first team CLD, no new links from the second and third team CLD, two new links from the fourth CLD, one new link from the fifth CLD and no new links from the sixth team CLD.

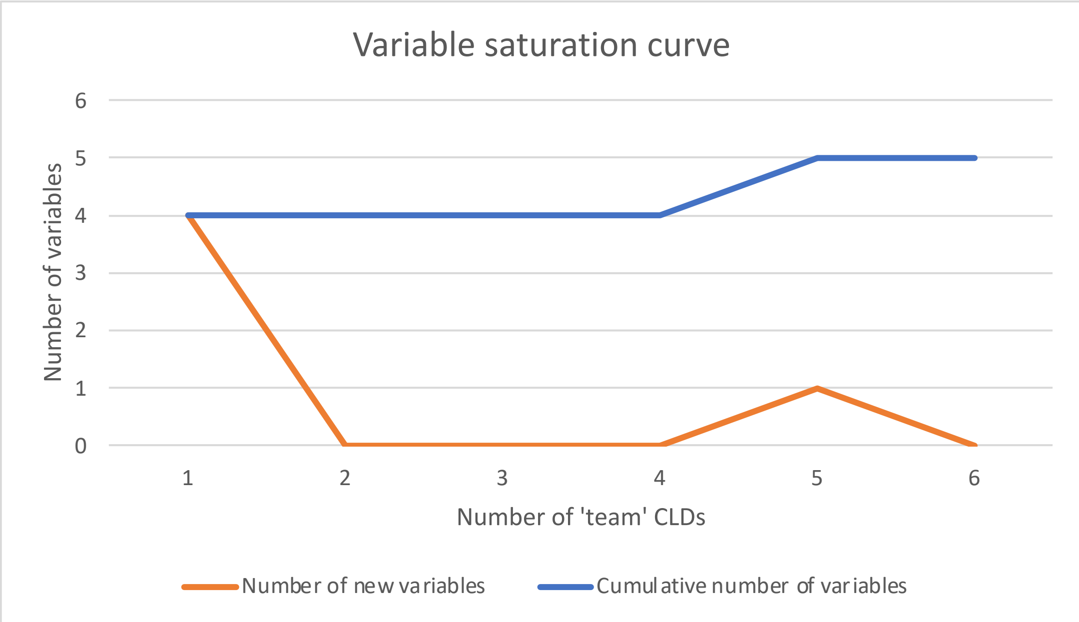


Figure 7: Variable saturation curve graph.

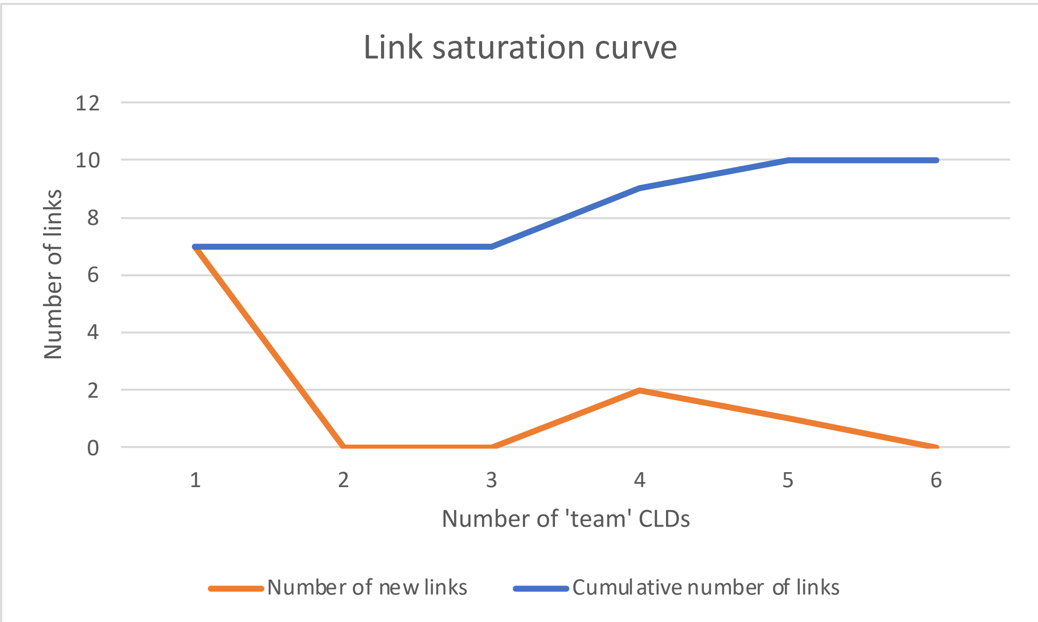


Figure 8: Link saturation curve graph.

Minor structural changes were made to the shared CLD to reflect the new elements identified in the six ‘validation’ set of team level CLDs. This updated shared CLD can be viewed in Appendix 4 (Figure A4.2). Modifications include an expansion of the existing medical commodities procurement process, additional complexity incorporated into other existing loops (facility data reporting processes and service demand creation) and new links between perceived service quality and demand creation.

Some further minor refinements were made to the shared CLD to condense or otherwise adjust repetitive or lengthy, inefficient loops (comparable to the ‘merging’ action described in Figure 4). Inefficient loops are loops that contain repetitive information or additional information that is superfluous to the loop dynamic. This updated, shared CLD can be viewed in Appendix 4 (Figure A4.3). Simplifications include combining the previously separated utilisation of ANC and utilisation of child health services into one variable, combining ANC service delivery and child health service delivery into one variable, condensing a loop on health worker efficiency and condensing loops that contain the variable ‘mother and child health’; the variable was adjacent to ‘facility performance’ in all feedback loops and proved an unnecessary addition to the loop.

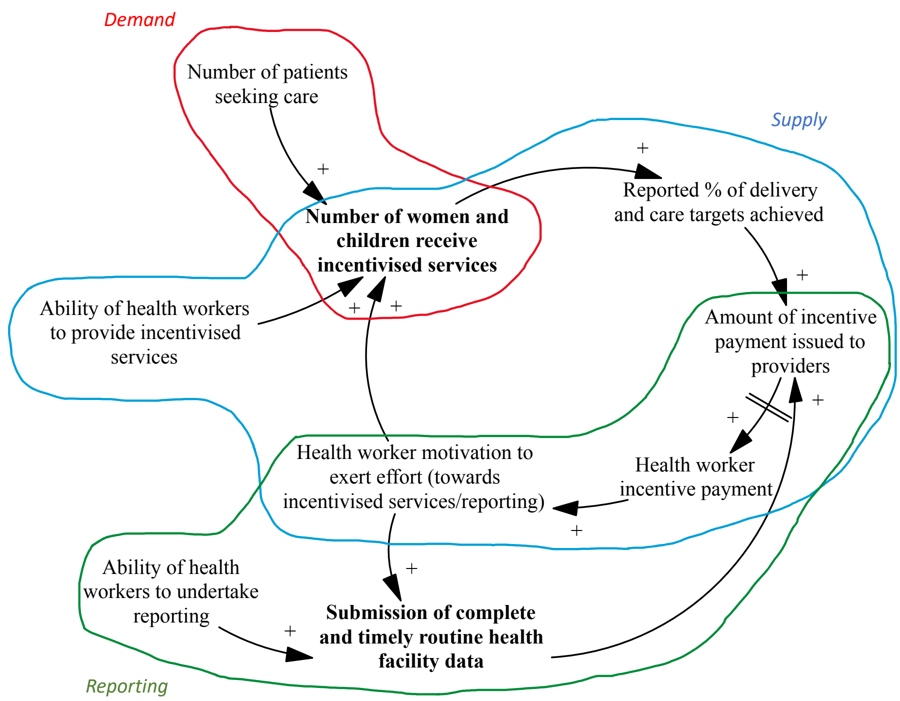
In the second stage of validation, the CLD was presented to twenty-one stakeholders closely involved with the evaluation and implementation of the P4P programme (Andersen *et al.* 2012; Rwashana *et al.* 2014). Stakeholders were asked in individual interviews if they recognised the structure and elements in the CLD and if any changes needed to be made to reflect their own experience of the health system and the P4P programme. This stage of validation was in place to minimise unconscious bias that may have been introduced by the researcher during CLD development and to identify and amend any misinterpretation of previous interview data in the shared CLD, including refinement of variable names. The validation interviews also provided an opportunity to elicit any further information that may have been missing from the interview data and to resolve any conflicts in the data. Minor adjustments were made to the CLD as a result of these interviews. This final CLD can be viewed in Appendix 4 of this thesis (Figure A4.4) and in the Results section of this thesis, Chapter 6 Appendix D (Figure D1).

Modifications include refinement of existing variable names, strengthened understanding on the use of facility and CHMT funding, additional complexity incorporated into hiring of staff and its effect on key outcomes (data reporting, health worker motivation and service delivery), new drivers for health worker motivation and perceived quality of services, and further condensing of inefficient feedback loops.

4.5 Presentation of CLD

Within the CLD two categories of performance targets were identified: 'Number of women and children who receive incentivised services' and 'Submission of routine health facility data by providers' (shown in **bold** in a high-level snapshot of CLD, Figure 9). During CLD development and validation, it became clear that there were three core mechanisms responsible for provider achievement of (or failure to reach) targets during the programme: (1) mechanisms that result in changes in the supply of services, (2) mechanisms that result in changes to facility reporting, and (3) mechanisms that result in changes in demand for services. The results section (Chapter 6) presents an overview of each of these mechanisms and the corresponding sections of the CLD (with the overall CLD shown in Chapter 6 Appendix D).

Figure 9: High level snapshot of causal loop diagram. Source: Cassidy *et al.* 2021



Notes to Figure: Three main mechanisms responsible for provider achievement of (or failure to reach) targets during P4P are shown in different colours. Changes in the supply of services (blue), changes to facility reporting (green), and changes in demand for services (red).

References

Andersen DL, Luna-Reyes LF, Diker VG *et al.* 2012. The disconfirmatory interview as a strategy for the assessment of system dynamics models. *Syst Dyn Rev* **28**: 255–75.

Borghi J, Mayumana I, Mashasi I *et al.* 2013. Protocol for the evaluation of a pay for performance programme in Pwani region in Tanzania: A controlled before and after study. *Implement Sci* **8**: 80.

Cassidy R, Borghi J, Rwashana Semwanga A *et al.* 2022. How to do (or not to do)…Using Causal Loop Diagrams for Health System Research in Low- and Middle-Income Settings. *Health Policy Plan*: czac064.

Cassidy R, Tomoaia-Cotisel A, Semwanga AR *et al.* 2021. Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach. *Soc Sci Med* **285**: 114277.

Kim H, Andersen DF. 2012. Building confidence in causal maps generated from purposive text data: mapping transcripts of the Federal Reserve. *Syst Dyn Rev* **28**: 311–28.

Milsom P. 2021. International trade and investment liberalization, corporate power and noncommunicable disease prevention policy: A case study of nutrition and alcohol policy non-decisions in South Africa. , DOI: 10.17037/PUBS.04662735.

Rwashana AS, Nakubulwa S, Nakakeeto-Kijjambu M *et al.* 2014. Advancing the application of systems thinking in health: understanding the dynamics of neonatal mortality in Uganda. *Heal Res Policy Syst* **12**: 36.

Sterman J. 2000. Part 1: Perspective and Process. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill Companies Inc, 22.

Tomoaia-Cotisel A. 2018. The Journey toward the Patient-Centered Medical Home: A Grounded, Dynamic Theory of Primary Care Transformation. *PhD Thesis, LSHTM*, DOI: 10.17037/PUBS.04647856.

Ventana Systems Inc. 2015. Vensim PLE Plus. Date of access: July 2, 2022. URL: https://vensim.com

5. System dynamics model

5.1 Introduction

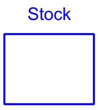
Abbreviations: 2 doses of intermittent preventative treatment (IPT2); Antenatal care (ANC); Causal loop diagram (CLD); Community health fund (CHF); Community health worker (CHW); Council Health Management Team (CHMT); District Executive Director (DED); Health Facility Governing Committee (HFGC); Low- and middle-income country (LMIC); Maternal and child health (MCH); Medical Stores Department (MSD); Payment for performance (P4P); System dynamics model (SDM).

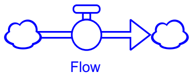
This chapter describes and expands on the methodology presented in the ready for publication paper (see Chapter 8) on SDM development and validation. In this chapter, details on why a SDM approach is used in this thesis, an overview of the building blocks of SDM and an outline of the method are presented. Further details on model software, creation of the SDM and simulation and scenario testing follow.

Development and analysis of the CLD (methods Chapter 4, results Chapter 6) revealed bottlenecks to programme success and system elements that should be incorporated or targeted to enhance the effect of the programme (catalytic variables and system levers), leading to recommendations for future design and implementation of the programme. CLDs are static diagrams; they are not capable of simulating system behaviour over time in response to an intervention or testing the impact of design changes to interventions on key outcomes. This is important when we consider pathways to effect for programmes or recommendations for design, as programme impact is expected to fluctuate over time in response to the wider health system and contextual environment in which the programme is implemented. SDMs can simulate programme behaviour over time and allow the user to assess which programme design yields optimal system outcomes. Using an SDM approach fulfils **Objective 4** of this thesis:

> 4. Explore how variations in the implementation, design and context of P4P could result in different outcomes to inform future design of P4P programmes.

5.2 System dynamics models

Stock and flow diagram notation are often used to construct SDMs (Pruyt 2017) which consists of stocks 

, flows 

, auxiliary variables 

and constants 

.

A simple example of a stock and flow diagram is presented in Figure 1, demonstrating replenishment and depletion of medicine at a health facility on a monthly basis. ‘Stock of medicine’ is a stock, a container where the quantity can be observed to change over time in response to the inflow ‘replenishment of medicine’ and outflow ‘depletion of medicine’. Stock of medicine depletes depending on the static, constant variable ‘medicine used’. Stock of medicine is replenished through the dynamic auxiliary variable ‘medicine procured’, which fluctuates in response to constant variables ‘availability of medicine from supplier’ and ‘medicine requested’ from the health facility. In real procurement and supply processes, these constant variables are likely to be influenced by other system elements and fluctuate over time but for simplicity are given constant variable status here. The numbers attached to the constant variables indicate their value; 100 items of medicine are used every month at the facility, facilities request 300 items of medication (supplies for a three-month service period), with the supplier fulfilling only 75% of the requested order. Availability of medicine is observed to fluctuate because of this dynamic in the accompanying illustration (shown as a graph in the stock).

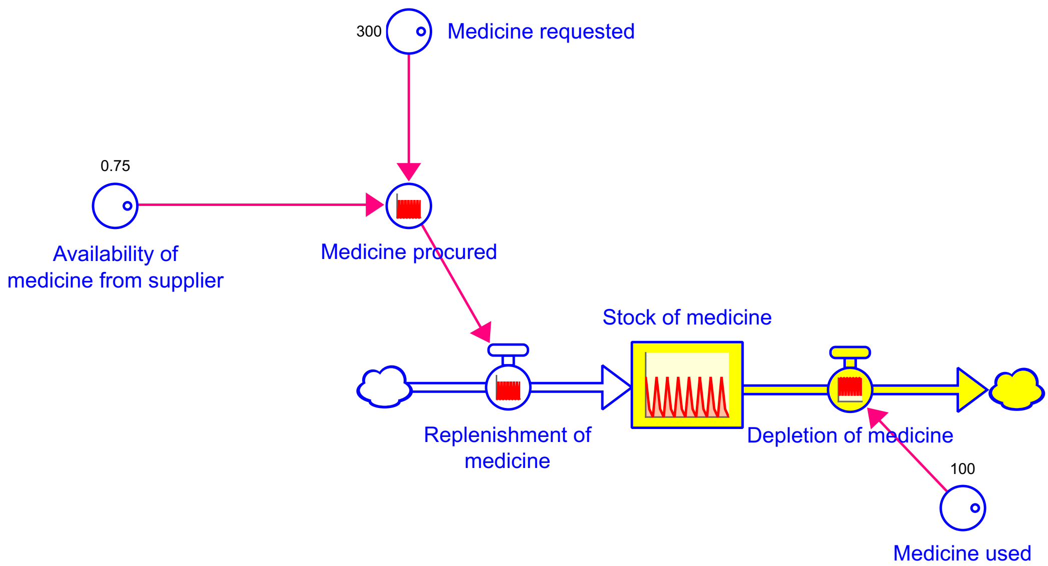


Figure 1: Simple example of a SDM.

In this thesis, SDM was used to explore the mechanisms through which a P4P programme affected MCH service delivery outcomes in a primary care facility setting and the effect of changes in the design, implementation, and context of a P4P programme in a low-income setting, Tanzania. A previously developed CLD (Chapter 4) was used as a framework to inform development of the model; the model consisted of six sectors that managed activity related to supply and demand side mechanisms responsible for provider achievement of targets during the P4P programme. Primary and secondary data sources (including data from a previous impact evaluation and evidence drawn from the literature) were used to populate the model, which later underwent validation and verification tests to build confidence in the model. A series of analyses were performed to determine sensitivity of key outcomes to changes in parameters, which shed light on the effect of programme context on key content of care (percentage of women who received two doses of IPT during ANC) and coverage (percentage of women who had a facility-based delivery) outcomes.

5.3 Model software

STELLA Architect (version 2.1.4) (isee systems inc 2021) is a widely used software for SDM development. Although Vensim (Ventana Systems Inc. 2015) was used to develop the CLD in the previous chapter of this thesis (Chapter 4) and can be used to develop SDMs, STELLA was chosen as the preferred modelling software. STELLA has extensive modelling documentation available on their support website, easy to navigate for those developing a model for the first time and a useful resource when trying to decode errors and bugs. The software also has capacity for development of user-friendly model interfaces; a frontend for the model that allows navigation and experimentation without the backend of the model being visible. This particularly appealed to the candidate as it was anticipated the model would be presented to stakeholders for their feedback, where an interface could be used to demonstrate the functionality and results of the model.

5.4 Creation of the SDM

Development and validation of the model can be broadly summarised as following four stages; (i) defining the purpose and goal of the model (ii) creation of model sectors (iii) validation of the model (iv) sensitivity analysis.

5.4.1 Step 1: Model purpose

The first step for SDM development was to (i) define the problem/health system behaviour to be investigated and (ii) define the goals of the model. The CLD (details of development in Chapter 4), which identified pathways to impact of P4P on delivery and coverage of MCH using the

Pwani programme as a case study, was used as a blueprint for determining model purpose, sector selection and creation (Figure 2). The health system behaviour explored with SDM was the performance of facilities during the P4P programme in Pwani. The goal of the SDM model was to (i) explore this phenomenon and (ii) test whether changes to implementation of the programme can result in further improved health provider performance in a 'typical' primary care provider. In the model, the performance of providers for two incentivised services were monitored; a content of care indicator (percentage of women who received two doses of IPT during ANC) and a coverage indicator (percentage of women who had a facility-based delivery), as these indicators showed some improvement during the P4P programme in Pwani and were the primary outcomes in the CLD. Using SDM allowed exploration of the mechanisms underpinning these improvements. The model time step is the smallest unit of activity in the model, the performance reporting unit (months), with simulation start time January 2011 and a time horizon of 54 months. The simulation period covers programme commencement (January 2011), the period of programme evaluation (January 2012 – March 2015) and a short period post-evaluation (up to July 2015), to consider both the short- and long-term effects of the programme which may fluctuate over time (Borghi *et al.* 2021).

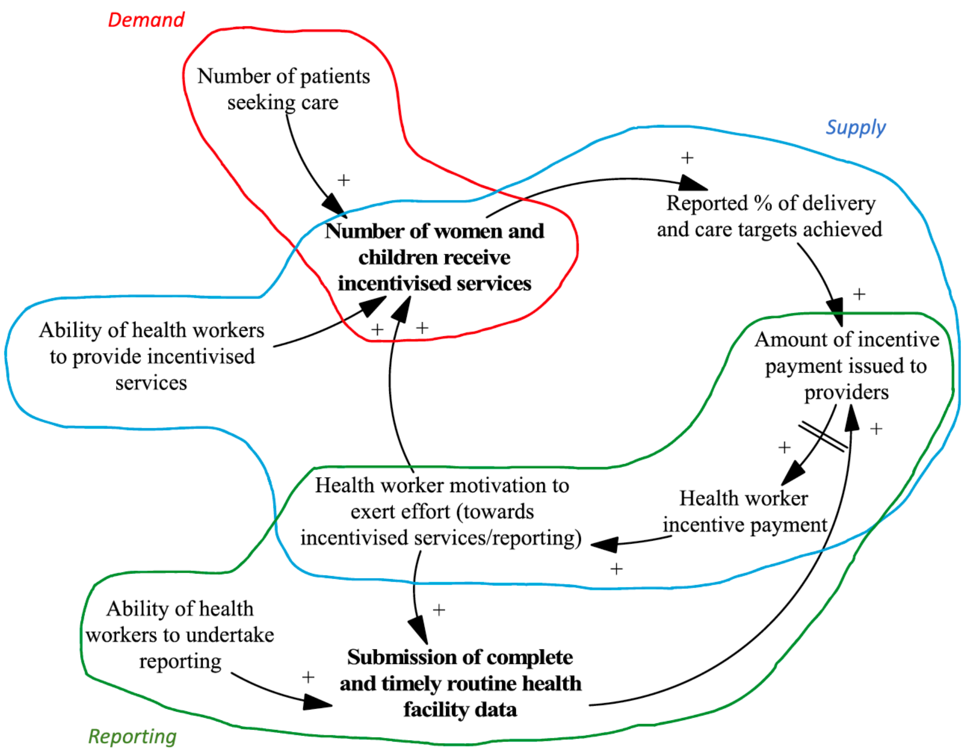


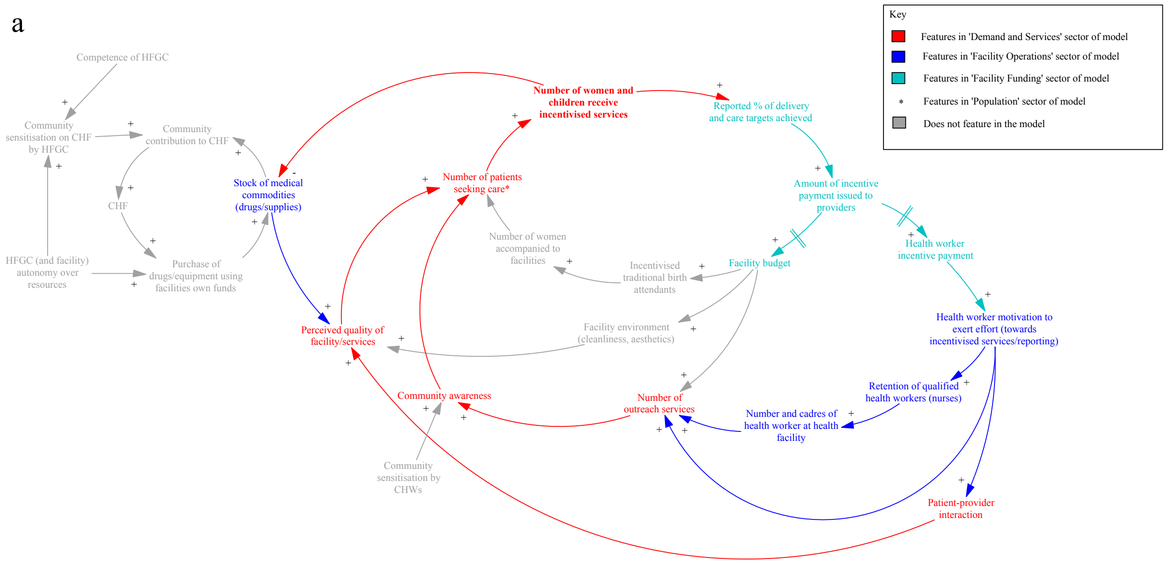
Figure 2: High level snapshot of causal loop diagram used as a blueprint to develop SDM. Notes to Figure: Three main mechanisms responsible for provider achievement of (or failure to reach) targets during P4P are shown in different colours. Changes in the supply (blue), changes to facility reporting (green), and changes in demand for services (red). Source: Cassidy *et al.* 2021.

5.4.2 Step 2: Model sectors

The second step for development was the creation of model sectors that drive behaviour in different compartments of the model. The CLD was used as a framework to inform development of the simulation model. Six model sectors were generated from the structures identified in the CLD (structure taken from the CLD and used to develop the SDM are shown in different colours in Figure 3, mapped to the different sectors of the SDM). Structure identified in the 'demand' component of the CLD (Figure 3a) fed into development of the '**Demand and Services**', '**Facility Operations'**, **'Facility Funding'** and **'Population'** sectors (Figure 4). Structure identified in the 'supply' component of the CLD (Figure 3b) fed into development of the '**Demand and Services**', '**Facility Operations'**, **'Facility Funding'**, '**District Manager Operations**' and '**Facility Commodities**' sectors (Figure 4).

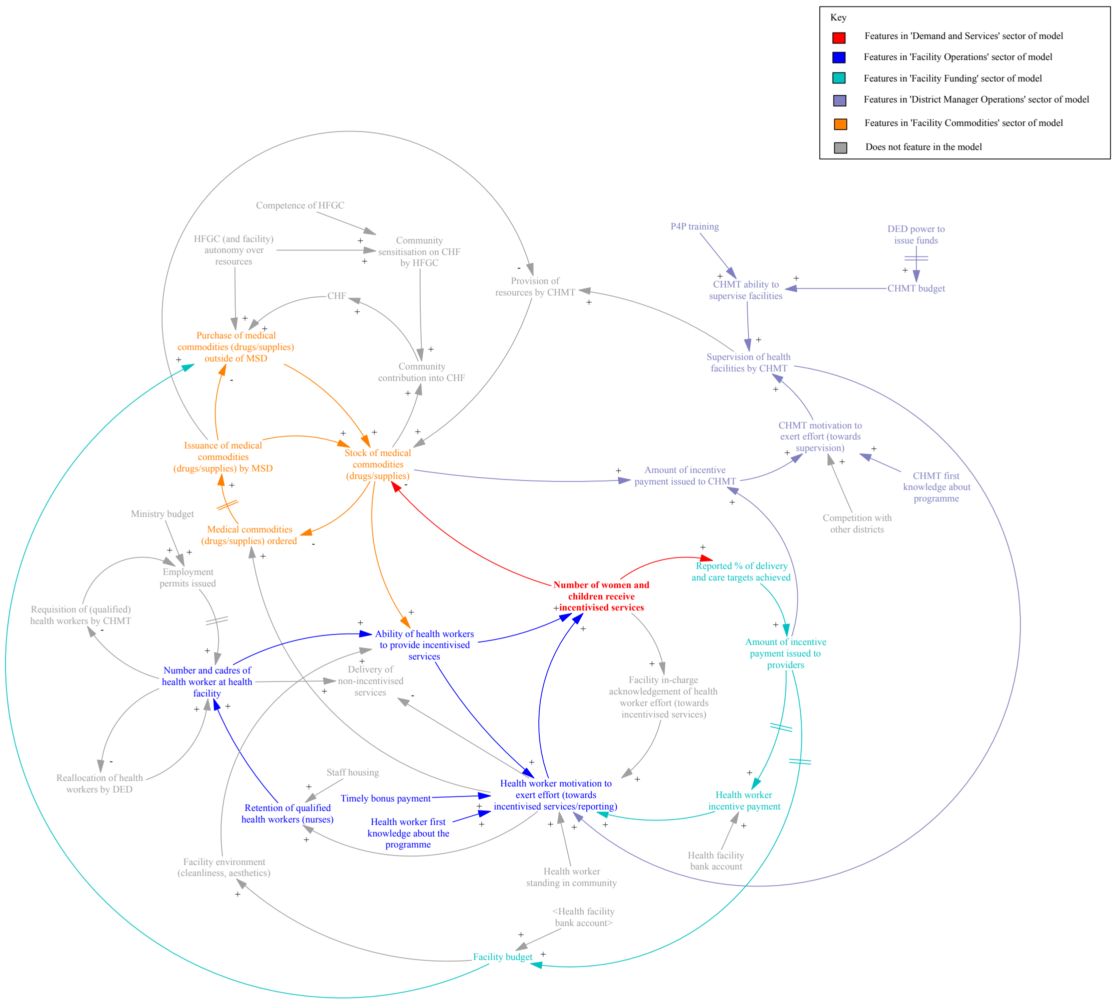
In the SDM, dynamics related to facility reporting were not included (Figure 3c). The main focus was on capturing the facility level supply side dynamics related to facility performance as this was the primary target of P4P. As a result, the SDM does not capture a number of demand side elements, shown in 'grey' in Figure 3 including: the dynamics around payment into a community health fund (voluntary community health insurance fund that was used to support provision of services at the facility), mechanisms for employing health workers and the activities of community health workers and traditional birth attendants in service demand creation. An agent-based model is currently under development which will explore the effect of community and service demand dynamics on facility-based deliveries.

a



Key
Features in 'Demand and Services' sector of model
Features in 'Facility Operations' sector of model
Features in 'Facility Funding' sector of model
Features in 'Population' sector of model
Does not feature in the model
Competence of HFGC
Community sensitisation on CHF by HFGC
Community contribution to CHF
CHF
HFGC (and facility) autonomy over resources
Purchase of drugs/equipment using facilities own funds
Stock of medical commodities (drugs/supplies)
Number of women and children receive incentivised services
Reported % of delivery and care targets achieved
Number of patients seeking care*
Amount of incentive payment issued to providers
Health worker incentive payment
Number of women accompanied to facilities
Incentivised traditional birth attendants
Facility budget
Health worker motivation to exert effort (towards incentivised services/reporting)
Perceived quality of facility/services
Facility environment (cleanliness, aesthetics)
Community awareness
Number of outreach services
Retention of qualified health workers (nurses)
Number and cadres of health worker at health facility
Community sensitisation by CHWs
Patient-provider interaction

b



Key
Features in 'Demand and Services' sector of model
Features in 'Facility Operations' sector of model
Features in 'Facility Funding' sector of model
Features in 'District Manager Operations' sector of model
Features in 'Facility Commodities' sector of model
Does not feature in the model
Competence of HFGC
HFGC (and facility) autonomy over resources
Community sensitisation on CHF by HFGC
P4P training
DED power to issue funds
CHMT ability to supervise facilities
CHMT budget
Provision of resources by CHMT
CHF
Purchase of medical commodities (drugs/supplies) outside of MSD
Community contribution into CHF
Supervision of health facilities by CHMT
CHMT motivation to exert effort (towards supervision)
Issuance of medical commodities (drugs/supplies) by MSD
Stock of medical commodities (drugs/supplies)
Amount of incentive payment issued to CHMT
CHMT first knowledge about programme
Competition with other districts
Ministry budget
Medical commodities (drugs/supplies) ordered
Employment permits issued
Number of women and children receive incentivised services
Reported % of delivery and care targets achieved
Requisition of (qualified) health workers by CHMT
Ability of health workers to provide incentivised services
Amount of incentive payment issued to providers
Number and cadres of health worker at health facility
Delivery of non-incentivised services
Facility in-charge acknowledgement of health worker effort (towards incentivised services)
Reallocation of health workers by DED
Staff housing
Retention of qualified health workers (nurses)
Timely bonus payment
Health worker motivation to exert effort (towards incentivised services/reporting)
Health worker incentive payment
Health worker first knowledge about the programme
Facility environment (cleanliness, aesthetics)
Health worker standing in community
Health facility bank account
<Health facility bank account>
Facility budget

c

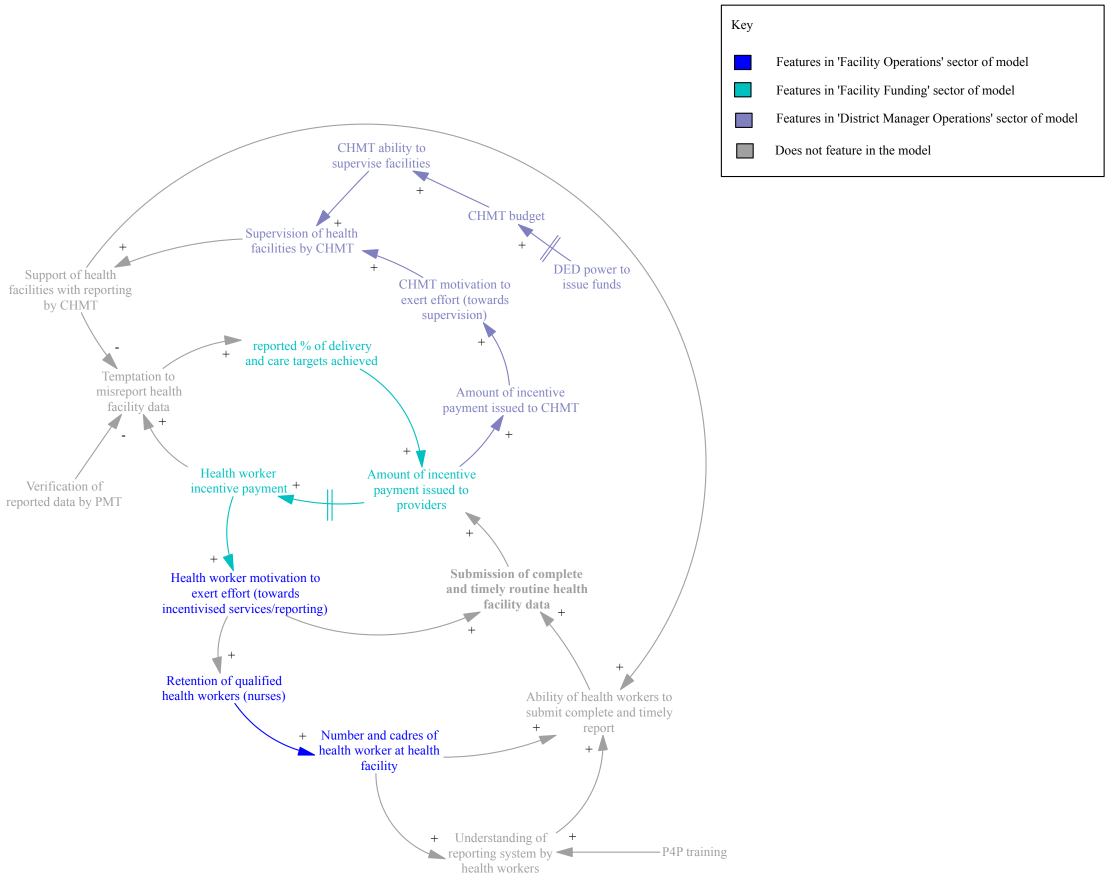


Figure 3: Structure identified in the previously developed CLD, specifically drawn from the (a) ‘demand’ (b) ‘supply’ and (c) ‘reporting’ components, that feature in the simulation model. Abbreviations: Community health fund (CHF); Community health worker (CHW); Council Health Management Team (CHMT); District Executive Director (DED); Health Facility Governing Committee (HFGC); Medical Stores Department (MSD); Payment for performance (P4P).

Simulation model

The purpose of each **sector**, *key sector outputs* used as input to other sectors and description of how sectors pass information is given here (Figure 4), with detailed individual model sector diagrams and description of model equations given in Chapter 8 Supplementary Files 2 and 3, respectively. This section outlines the model functioning in the absence of P4P.

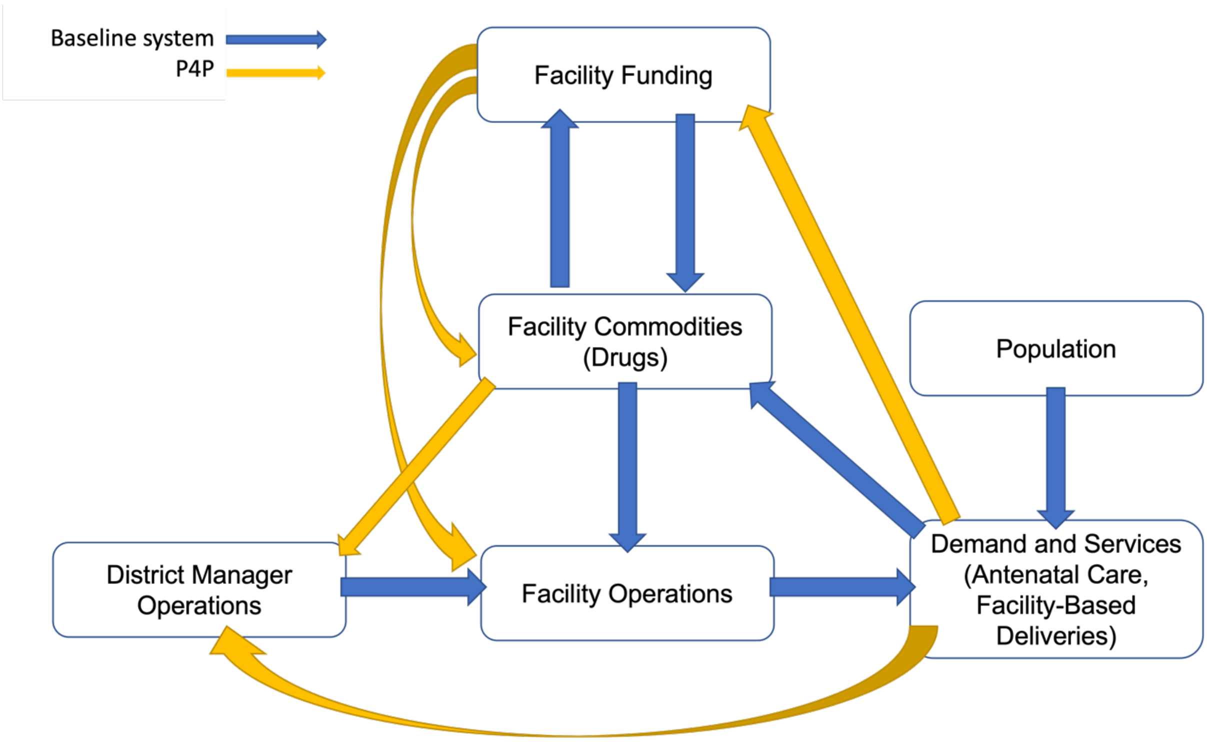


Figure 4: High level overview of simulation model. The model contains six subsectors which pass information to each other (arrows). The user can run the model with payment for performance switched 'on' (P4P, yellow) and 'off' (baseline system, blue).

The **Population** sector controls population dynamics that feed into the **Demand and Services** sector**.** It controls ageing in the population (neonates, infants, pre-schoolers, children, reproductive age adults and adults above 50) over time, which is driven by: (i) the respective age mortality rates; (ii) fertility rate. The sector generates the following key output and population group of interest *number of newly pregnant women,* which contributes to the flow of patients seeking care in the **Demand and Services** sector. The population sector has been structured so that the model can be later adapted to focus on other types of service provision (e.g. neonatal vaccination).

The **Demand and Services** sector controls the number of ANC patients that receive services and facility-based deliveries. Patients can attend up to four ANC appointments, with three possible pathways for each ANC visit (i) dropping out and not attending ANC visit, (ii) receiving treatment (up to two doses of IPT across all ANC visits, with goal of two doses for each patient) or (iii) do not receive treatment. Treatment receipt is dependent on (i) provider readiness to deliver care (controlled in the **Facility Operations** sector) and (ii) attendance rates for each antenatal care visit.

The percentage of facility-based deliveries is determined by (i) the number of antenatal care visits (ii) distance to facility; (iii) awareness of maternal and child health and healthcare in the community (in part estimated from ability to perform outreach controlled in **Facility Operations** sector and fraction of women attending antenatal care); (iv) perceived quality of facility/services (estimated from availability of drugs in **Facility Commodities** sector and patient-provider interaction from **Facility Operations** sector). For each patient who receives a service (ANC or facility-based delivery), a single unit of drug is 'used' with drug availability depleting in the **Facility Commodities** sector. The **Demand and Services** sector generates two primary service delivery outputs of the model, *percentage of women who receive IPT2* and *percentage of women who seek facility-based delivery.*

The **Facility Commodities** sector controls the replenishment and depletion of malaria (IPT) and labour drugs at the facility level. The expected number of ANC and facility-based delivery patients is fed in from the **Demand and Services** sector and used to place orders for drugs on a quarterly basis to the Medical Stores Department (autonomous government department responsible for provision of medical commodities). Depending on the availability of drugs at Medical Stores, facilities may need to try and address the deficit of drugs. Facilities can use funds (facility held funds, managed in the **Facility Funding** sector) where available to purchase additional drugs. Key outputs in the **Facility Commodities** sector are the *availability of IPT drugs* and *availability of labour drugs,* which deplete depending on the number of patients treated in the **Demand and Services** sector.

The **Facility Operations** sector manages facility-level dynamics including provider readiness (related to administration of IPT during ANC). Provider readiness is dependent on (i) knowledge of health workers (IPT); (ii) number of health workers at health facility (percentage of positions filled); (iii) availability of IPT drugs fed in from the **Facility Commodities** sector; (iv) health worker motivation to exert effort towards incentivised services. Health worker motivation is dependent on availability of drugs (IPT and labour) fed in from **Facility Commodities** sectors, district manager supervision (quality) fed in from **District Manager Operations** sector and number of health workers at health facility (percentage of positions filled). Key outputs in this sector are *provider readiness (related to delivery of IPT)* and *health worker motivation to exert effort towards incentivised services.*

The **Facility Funding** sector manages the funding that is held and used at the facility level and can be used to purchase additional drugs where needed. The key output from this sector is *facility held funds available.*

The **District Manager Operations** sector manages supervision visits by members of the CHMT to facilities. The district manager supervision (quality) is dependent on district level resources, management team motivation and the skill level of district managers. Supervision visits affect knowledge of health workers related to IPT and health worker motivation. The key output for this sector is *district manager supervision (quality).*

Introduction of P4P intervention

Health providers have set targets they need to reach each cycle (6 months) to receive P4P incentive payments. Payment was expected to be made within three months of the conclusion of the six month performance cycle, however payments were often delayed. In the model, the performance targets are for specific services monitored in the **Demand and Services** sector. These are (i) percentage of women who receive IPT2 and (ii) percentage of women who seek facility-based delivery. Depending on performance against these targets, providers may receive incentive payments which are deposited in the **Facility Funding** sector.

The payment is split 75:25, with the larger portion allocated for health worker incentive payments and the remaining portion to be used to improve facility operations (e.g. purchasing additional medical commodities where needed). The health worker incentive payment is fed from **Facility Funding** to the **Facility Operations** sector. Incentive payments (specifically timeliness of payments) influence health worker trust in the programme and health worker motivation to exert effort towards incentivised services. The remainder of the incentive payment, in the model, supplements facility held funds (**Facility Funding sector**) and can be used to purchase drugs (malaria and labour drugs) where needed in the **Facility Commodities** sector. A new key output from the **Facility Funding** sector is *staff incentives*.

The CHMT are also eligible for incentive payments, which are processed in the **District Manager Operations** sector, with payments influencing district manager motivation to support facilities. In the simulation model, the district management targets (and determinants of incentive payment issued) are to reduce stockouts of medicine (observed in the **Facility Commodities** sector) and overall performance of health providers (observed in **Demand and Services** sector).

Data

The model was populated with both primary and secondary data sources, see Chapter 8 Supplementary File 4 for details. Examples of secondary data include population and housing census reports, country and district-level health surveys, data from the impact evaluation conducted on the Pwani P4P programme and evidence drawn from the literature. The previous evaluation conducted on the Pwani P4P is described elsewhere (Borghi *et al.* 2013, 2021; Binyaruka *et al.* 2015), with a summary provided here. The impact evaluation investigated the effect of the P4P programme on all targeted MCH services (including percentage of women who receive IPT2 and percentage of women who seek facility-based delivery) through a controlled before and after study design. Surveys were conducted in all six districts of Pwani region (where P4P had been implemented) and in five control districts in neighbouring regions. The evaluation consisted of a health facility survey, health worker survey, exit patient survey and survey of women who had delivered in the last 12 months. Data collection took place at three time points to observe programme impact: 'baseline' (January 2012), 'short term' (February 2013) and 'long term' (February and March 2015).

During model development, two members of the original programme evaluation team were consulted to provide insight into model dynamics related to impact of district manager supervision on health worker skill level. Model equations reflect this discussion, where effect of district manager supervision on health worker knowledge is dependent on the 'base level' of knowledge at the facility. Where this is lower, it will take a few supervision visits to raise the health worker knowledge (specifically related to provision of IPT during ANC).

5.4.3 Step 3: Model validation

The third step for model development was subjecting the model to a series of verification and validation tests to build confidence in the structure, behaviour, and robustness of the model. To check for internal validity, every equation in the model was reviewed for dimensional consistency, that model units were appropriate for the given variable i.e. population parameters are measured in units of 'persons', and that units used for outputs were appropriate based on variable input units. STELLA Architect does have functionality which returns an error message when dimensional consistency is violated but every equation was verified on the happenstance that an error had not been picked up by the software. The model was subjected to extreme condition testing, whereby model parameters were adjusted to extreme values and model output was evaluated to ensure expected results. For example, when the dropout rate for attending a first ANC visit is 0.999, only a handful of patients are expected to attend this first visit and move through the ANC part of the demand and services sector; or when the provision of medicine by the Medical Stores is severely impacted, a drastic depletion of medicine available at facilities should be observed. The model performed well when subjected to testing, producing expected behaviour under extreme conditions. Model equations and structure were also independently reviewed by a team member.

To check for external validity, selected model output projections were also compared to real data where available, with equation and parameter adjustments made where required so that model outputs were aligned with data (model calibration). The model was adequately able to replicate known trends, see Chapter 8 Supplementary File 5 for further details.

To check model face validity, the model was presented to nine key stakeholders involved in the implementation or evaluation of the Pwani programme during virtual interviews (conducted via Zoom) as a final validation step. A model interface was developed in STELLA Architect to assist with presentation of model outputs and key assumptions, see Chapter 8 Supplementary File 6 for the interview guide and details on model interface. The interview consisted of two segments; (i) stakeholders were shown key model output and dynamics and asked to comment on whether model behaviour was realistic and aligned with their experience of the P4P programme and the health system, and (ii) stakeholders were shown model assumptions and asked to provide their feedback on their validity.

The feedback received during these interviews resulted in some changes to the model's structure. Key changes consisted of: (i) inclusion of an 'alternative facility held funding' variable (ii) adjustment of effect of amount of incentive on trust in programme and motivation, and (iii) adjustment of effect of payment delays on trust in programme and motivation. For change (i), stakeholders suggested inclusion of other types of facility held funding (other than P4P), as this would affect the purchasing power of facilities for buying additional drugs during the programme. In the current version of the model, this variable 'alternative facility held funding available' is static, but in future iterations of the model this will be dynamic. For change (ii), stakeholders commented that health workers would want to improve their performance, so a lower incentive payment (reflective of performance) would not be demotivating but would spur health workers on to try and improve their performance.

For change (iii), stakeholders were presented with three scenarios (possible assumptions) related to effect of payment delays in the model; payment delays do not affect trust and motivation, any delay affects trust and motivation and only severe delays (4+ months) affect trust and motivation. The consensus from stakeholders was that the third scenario, severe delays affect trust and motivation, was most likely, with communication of delays and expected payment dates sustaining trust and motivation up to a point. If payments are not made after a certain period (assumed 4+ months in the model), trust and motivation decrease until a payment is made. This relationship between payment delays and trust and motivation is retained in the model when payment delays are enacted. Stakeholders also reflected on the presentation of the model, commenting that a high-level diagram showing how the model worked would be useful to them (see Figure 4).

Stakeholders remarked on the importance of community health workers and traditional birth attendants in increasing community awareness of services and escorting women to facilities for facility-based deliveries. These dynamics are not included in this current version of the model for the reasons set out above.

5.4.4 Step 4: Sensitivity analysis

The final step for model development was subjecting the model to sensitivity analysis to determine the sensitivity of key outcomes (percentage of women who receive at least two doses of IPT during ANC, percentage of women who seek facility-based delivery) to changes in model parameters. Model parameters deemed appropriate for analysis (see Chapter 8 Supplementary File 8) were adjusted by 10%, with key outcome results recorded. Initial stock values, constant variable values (including table and graphical function values) were adjusted and the effect on model outputs simulated. Equation based parameters (flows and auxiliary variables) and constant variables where the value was not appropriate for adjustment (such as the 'on' and 'off' switch for turning the intervention on and off in the model, represented by '0' and '1' in the model) were not subjected to sensitivity analysis.

The following scale was used to determine sensitivity to changes in model variables; sensitive (5% ≤ change in outcome < 15%), very sensitive (15% ≤ change in outcome < 25%) and highly sensitive (25% ≥ change in outcome). The scale is adapted from Semwanga *et al.* (2016) and presented with smaller intervals for higher sensitivity categories, to further distinguish 'very sensitive' from 'highly sensitive' results. As the variables included in the sensitivity analysis reflect health system characteristics, this analysis also shed light on the likely effect of changes to the health system context in which P4P is implemented on key outcomes.

5.5 Simulation and scenario testing

Model scenarios were selected to contribute evidence towards the knowledge gap identified by reviews of P4P effects in LMIC settings (Das *et al.* 2016; Patel 2018; Diaconu *et al.* 2021); to further understanding on pathways to effect for P4P, acknowledging the influence of programme design, implementation and context.

The model was first used to explore how health system performance changed under P4P, to examine the effects of the programme as it was implemented (as mentioned, there were delays in issuing programme incentive payments) on pathways to effect for the programme (e.g. health worker motivation, availability of medicine) and targeted services, percentage of women who receive at least two doses of IPT during ANC and percentage of women who seek facility-based delivery. The effect of changes in programme implementation (payments made on time vs. with delays) and design (adjusting the share of funds between staff incentives and funds to strengthen facility operations) on these outcomes were then tested in the model. Finally, the sensitivity analysis results were used to explore the effect of changes to programme and health system contextual factors (including provision of medicine from Medical Stores, amount of alternative facility held funding and staffing levels) on targeted services, percentage of women who receive at least two doses of IPT during ANC and percentage of women who seek facility-based delivery.

References

Binyaruka P, Patouillard E, Powell-Jackson T *et al.* 2015. Effect of Paying for Performance on Utilisation, Quality, and User Costs of Health Services in Tanzania: A Controlled Before and After Study. Ostermann J (ed.). *PLoS One* **10**: e0135013.

Borghi J, Binyaruka P, Mayumana I *et al.* 2021. Long-term effects of payment for performance on maternal and child health outcomes: evidence from Tanzania. *BMJ Glob Heal* **6**: e006409.

Borghi J, Mayumana I, Mashasi I *et al.* 2013. Protocol for the evaluation of a pay for performance programme in Pwani region in Tanzania: A controlled before and after study. *Implement Sci* **8**: 80.

Cassidy R, Borghi J, Semwanga AR *et al.* 2022. How to do (or not to do)…using causal loop diagrams for health system research in low and middle-income settings. *Health Policy Plan*, DOI: 10.1093/heapol/czac064.

Cassidy R, Tomoaia-Cotisel A, Semwanga AR *et al.* 2021. Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach. *Soc Sci Med* **285**: 114277.

Das A, Gopalan SS, Chandramohan D. 2016. Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review. *BMC Public Health* **16**: 321.

Diaconu K, Falconer J, Verbel A *et al.* 2021. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev* **2021**, DOI: 10.1002/14651858.CD007899.pub3.

isee systems inc. 2021. STELLA Architect.

Patel S. 2018. Structural, institutional and organizational factors associated with successful pay for performance programmes in improving quality of maternal and child health care in low and middle income countries: a systematic literature review. *J Glob Health* **8**, DOI: 10.7189/jogh.08.021001.

Pruyt E. 2017. System dynamics: a tool for modelling and testing solutions. In: de Savigny D, Blanchet K, Adam T. *Applied Systems Thinking for Health Systems Research : A Methodological Handbook*. McGraw-Hill Education, 173–94.

Semwanga AR, Nakubulwa S, Adam T. 2016. Applying a system dynamics modelling approach to explore policy options for improving neonatal health in Uganda. *Heal Res Policy Syst* **14**: 35.

RESULTS

6. Causal loop diagram

6.1 Introduction

Abbreviations: 2 doses of intermittent preventative treatment (IPT2); Causal loop diagram (CLD); Community health fund (CHF); Council Health Management Team (CHMT); Community health worker (CHW); District Executive Director (DED); Focus group discussion (FGD); Health Facility Governing Committee (HFGC); Health Management Information System (HMIS); Human Immunodeficiency Virus (HIV); Low- and middle-income country (LMIC); Maternal and child health (MCH); Medical Stores Department (MSD); Ministry of Health and Social Welfare (MoHSW); Payment for performance (P4P); Pilot Management Team (PMT); Regional Health Management Team (RHMT).

In Chapter 6, the results from analysis of the CLD are presented, fulfilling **Objectives 2 and 3** of this thesis):

> 2. Identify health system factors and feedback loops that facilitate or hinder the implementation of P4P programmes and its overall effectiveness.
>
> 3. Identify system leverage points which should be considered in the design of P4P programmes.

The results of the study are presented in a paper, ‘Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach’, published in Social Science & Medicine in September 2021. Evidence of retention of copyright or use of published materials in this thesis can be found in Appendix 3.

The chapter appendix contains supplementary material, including information on interpretation of CLDs, additional information of P4P programme targets, the study interview tool and additional results.

6.2 Research paper 2: Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach

(Cover sheet on next page)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1401642	**Title**	Miss
First Name(s)	Rachel		
Surname/Family Name	Cassidy		
Thesis Title	Using systems thinking to optimise health system interventions for improved maternal and child health in low-resource settings		
Primary Supervisor	Prof. Josephine Borghi		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Social Science & Medicine		
When was the work published?	28th July 2021		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

Open Access CC-BY license

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Conceptualised study with PhD supervisors, collected new data to validate model, led analysis of secondary and newly collected data, developed conceptual model. Wrote first draft of the manuscript and led revisions with input from co-authors.

SECTION E

Student Signature	[redacted]
Date	05/12/22

Supervisor Signature	[redacted]
Date	05/12/22

Social Science & Medicine 285 (2021) 114277

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach

Rachel Cassidy [a,*], Andrada Tomoaia-Cotisel [b], Agnes Rwashana Semwanga [c], Peter Binyaruka [d], Zaid Chalabi [e,f], Karl Blanchet [g], Neha S. Singh [a], John Maiba [d], Josephine Borghi [a]

[a] Department of Global Health and Development, London School of Hygiene and Tropical Medicine, 15-17, Tavistock Place, London, WC1H 9SH, UK
[b] RAND Corporation, Boston, MA, 02116, USA
[c] Information Systems Department, College of Computing and Information Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda
[d] Ifakara Health Institute, PO Box 78373, Dar Es Salaam, Tanzania
[e] Department of Public Health, Environments and Society, London School of Hygiene and Tropical Medicine, 15-17, Tavistock Place, London, WC1H 9SH, UK
[f] Bartlett School of Environment, Energy and Resources, University College London, Central House, 14 Upper Woburn Place, London, WC1H 0NN, UK
[g] Geneva Centre of Humanitarian Studies, University of Geneva and the Graduate Institute, Rue Rothschild 22, 1211, Genève, Switzerland

ARTICLE INFO

Keywords:
Health systems
Causal loop diagram
Payment for performance
Maternal and child health
Primary care
Evaluation
Tanzania

ABSTRACT

Payment for performance (P4P) has been employed in low and middle-income (LMIC) countries to improve quality and coverage of maternal and child health (MCH) services. However, there is a lack of consensus on how P4P affects health systems. There is a need to evaluate P4P effects on health systems using methods suitable for evaluating complex systems. We developed a causal loop diagram (CLD) to further understand the pathways to impact of P4P on delivery and uptake of MCH services in Tanzania. The CLD was developed and validated using qualitative data from a process evaluation of a P4P scheme in Tanzania, with additional stakeholder dialogue sought to strengthen confidence in the diagram. The CLD maps the interacting mechanisms involved in provider achievement of targets, reporting of health information, and population care seeking, and identifies those mechanisms affected by P4P. For example, the availability of drugs and medical commodities impacts not only provider achievement of P4P targets but also demand of services and is impacted by P4P through the availability of additional facility resources and the incentivisation of district managers to reduce drug stock outs. The CLD also identifies mechanisms key to facility achievement of targets but are not within the scope of the programme; the activities of health facility governing committees and community health workers, for example, are key to demand stimulation and effective resource use at the facility level but both groups were omitted from the incentive system. P4P design considerations generated from this work include appropriately incentivising the availability of drugs and staffing in facilities and those responsible for demand creation in communities. Further research using CLDs to study heath systems in LMIC is urgently needed to further our understanding of how systems respond to interventions and how to strengthen systems to deliver better coverage and quality of care.

1. Introduction

Payment for performance (P4P) programmes have been employed in many low- and middle-income countries (LMICs) to improve the quality and coverage of maternal and child health (MCH) services. Under P4P, health care providers, managers and/or organisations receive bonus payments that are tied to the delivery of pre-determined services or quality improvements (Mannion and Davies, 2008). The theoretical rationale for using financial incentives is to align incentives and behaviours of stakeholders within the health system in light of the agency relationships between managers, health care providers and patients, together with asymmetric information in these relationships (Fichera et al., 2014). Financial incentives are expected to motivate health workers to adhere to clinical care guidelines and increase the availability and quality of care delivered to patients (Gagné and Deci, 2005; Das et al., 2016). Many evaluations of P4P in LMIC have focused on estimating effects on elements within the health system, such as health worker job satisfaction (Shen et al., 2017; Engineer et al., 2016), health worker motivation (Shen et al., 2017; Engineer et al., 2016; Bhatnagar and George, 2016), availability of medical commodities (Das et al.,

* Corresponding author.
E-mail address: rachel.cassidy@lshtm.ac.uk (R. Cassidy).

https://doi.org/10.1016/j.socscimed.2021.114277
Received 9 February 2021; Received in revised form 13 July 2021; Accepted 27 July 2021
Available online 28 July 2021

2016; Engineer et al., 2016; Bhatnagar and George, 2016), patient perceived quality of care (Das et al., 2016; Engineer et al., 2016; Paul et al., 2014) and accountability mechanisms (supervision of providers by managers (Bhatnagar and George, 2016; Paul et al., 2014; Mayumana et al., 2017) and community engagement in provision of services (Engineer et al., 2016; Mayumana et al., 2017)).

There has been less attention to the causal mechanisms through which P4P improves service delivery or coverage of health services. Causal mediation analysis was recently employed to unpack the mechanisms through which P4P improves service indicators in two low-income settings, isolating potential mediators of programme effect (Anselmi et al., 2017; Ngo et al., 2017). However, such analyses examine one-directional static single chains of causality, ignoring feedback mechanisms, overlooking dynamics in the health system as a whole, and disregard intrinsic time delays. We must consider the *holistic* impact of interventions on the health system, not just acknowledging that connections and mediators exist in isolation but how they affect each other over time. This knowledge is critical to understanding which design elements of P4P work and promote optimal health system behaviour (as intended) and which lead to suboptimal behaviour or negative unintended consequences, undermining programme success.

A recent realist review (Singh et al., 2021) identified pathways underpinning P4P effectiveness, including outreach activities to generate demand for services, greater availability of drugs and medical supplies and provider adherence to clinical guidelines. The review also pointed to relevant contextual factors underpinning programme effectiveness, including facility staffing levels and facility autonomy. Whilst informative, few of the studies included in this review were designed to evaluate pathways to P4P effectiveness or provide evidence of a link between a given mechanism and outcome.

Tools that derive from systems thinking methodologies can be used to better understand complex systems, such as health systems, and unpack the pathways to impact of interventions such as P4P (Borghi and Chalabi, 2017; Peters, 2014; Atun, 2012). Causal loop diagrams (CLDs) can identify and explore system problems and support decision making within health systems. They can also be used as a complementary tool to enhance other evaluation methods, such as realist evaluations, where there is a need to identify (and visualise) health system programme mechanisms and outcomes, and the context in which they are implemented (Singh et al., 2021; Renmans et al., 2020). CLDs are not a suitable choice for testing and modelling potential solutions to problems. Instead, system dynamics models, which often utilise CLDs in their development, are a better fit for this research need (de Savigny et al., 2017).

CLDs depict cause and effect relationships between variables in a system and provide a visual representation of system structure, capturing cyclic 'looping' feedback (Tomoaia-Cotisel et al., 2017). CLDs use arrows, where arrow polarity signifies the effect of changes in one variable on another. Delays in influence of one variable on another can be shown in CLDs using the symbol of two lines through an arrow. Reinforcing (R) and balancing (B) loops are identified in a CLD using numbered, circular arrows; reinforcing loops describe positive/amplified behaviour and balancing loops describe negative/stabilising behaviour. For more information on interpretation of CLDs, please see Appendix A. There has been a steady rise in the application of CLDs to evaluate the impact of policies on health systems in high income settings (Rashwan et al., 2015; Schoenenberger et al., 2016), most recently during the COVID-19 pandemic (Bradley et al., 2020; Sahin et al., 2020). To our knowledge, only four studies have used these methods to examine the effect of P4P interventions on health systems (Singh et al., 2021; Alonge et al., 2017; Meker and Barlas, 2015; Renmans et al., 2017), three in low-income settings (Singh et al., 2021; Alonge et al., 2017; Renmans et al., 2017).

The aim of this study was to develop a CLD to further understand the pathways to impact of P4P on delivery and uptake of MCH services in Tanzania, a low-income setting, and reflect on the insights gained from using this approach as compared to conventional evaluation methods. Tanzania was selected as a case study as it had implemented a P4P programme which was known to be effective in improving service uptake (Binyaruka et al., 2015), and resulted in health system improvements (provider kindness and greater drug availability) which mediated programme effects (Anselmi et al., 2017). There was also a wealth of evaluation data on the health system effects of the programme (Mayumana et al., 2017; Binyaruka et al., 2015, 2018a; Binyaruka and Borghi, 2017; Olafsdottir et al., 2014; Borghi et al., 2013; Binyaruka and Anselmi, 2020) to inform the CLD.

1.1. Study setting

Tanzania has experienced mixed progress in MCH over the last three decades (Afnan-Holmes et al., 2015) and implemented a P4P programme in 2011 as part of a concerted effort to make progress towards Millennium Development Goals 4 and 5 (Borghi et al., 2013). The design of the programme has been described extensively elsewhere (Binyaruka et al., 2015; Borghi et al., 2013), but a summary follows. The Ministry of Health and Social Welfare in Tanzania, with funding from the Norwegian Ministry of Foreign Affairs, introduced a P4P initiative in 2011 in the region of Pwani. To be eligible to participate in the programme, facilities had to provide MCH services, hold or open a bank account and provide facility performance data from the previous year (2010–2011), which was used to set initial MCH service coverage targets. Facilities were eligible for incentive payments if they met targets for each 6 month cycle; either a percentage increase on the previous cycle's performance or an absolute performance target (MoHSW Ministry of Health and Social Welfare, 2012; Binyaruka et al., 2018b) (see Appendix B). For primary health care facilities (dispensaries and health centres), 75 % of this payment was to be distributed among health workers at the facility and the remaining funds were to be spent on facility improvements/demand creation (25 %). Managers at the district and regional level who were responsible for supporting facilities and verifying facility performance data, the Council Health Management Team (CHMT) and Regional Health Management Team (RHMT), were also eligible for incentives (Appendix B).

2. Methods

2.1. Secondary data

We used qualitative data collected through a process evaluation during the Tanzania P4P programme (Borghi et al., 2013) to develop and validate a CLD (Table 1). These data describe how P4P was implemented in different facilities, factors that affected the success of the programme and potential unintended consequences (Borghi et al., 2013). Although secondary care facilities participated in the programme and consequent evaluation, due to programme design differences between providers, we focussed our evaluation on primary care facilities. Three rounds of data collection took place between December 2011–March 2013. Interviews were conducted in five of the seven districts in Pwani (Kibaha Town, Bagamoyo, Mkuranga, Kisarawe and Mafia island). Ten primary care health facilities were purposively sampled to reflect differences in level of care and ownership. Forty-three interviews were conducted with health workers, those in-charge of MCH care, those in-charge of facilities and members of the CHMT. Eight focus groups discussions (FGDs) were conducted with Health Facility Governing Committees (HFGC), CHMTs and health workers. Interviews were conducted in Swahili by four local social scientists working in pairs. All interviews were audio recorded and verbatim transcripts produced in Word, with transcripts translated to English.

2.2. Primary data

The CLD that was developed and validated using the secondary data

Table 1
Description of secondary data used to develop and validate causal loop diagram, collected between December 2011–March 2013.

District	Facility/ CHMT	Stakeholder Type	No. Of Interviews	No. Of FGDs
District A	Health Centre	Health worker	2	
	Dispensary	Health worker	3	
	CHMT	CHMT	5	1
District B	Health Centre	Health worker	4	
	Dispensary	Health worker	1	
		HFGC		2
	CHMT	CHMT	3	
District C	Dispensary	Health worker	1	
	Dispensary	Health worker	1	
		HFGC		1
	CHMT	CHMT	3	
District D	Health Centre	Health worker	2	
	Dispensary	Health worker	1	
	CHMT	CHMT	4	1
District E	Health Centre	Health worker	4	1
		HFGC		1
	Dispensary	Health worker	4	
	CHMT	CHMT	5	1
Total			**43**	**8**

Notes to Table: Council Health Management Team (CHMT) Focus group discussions (FGDs), Health Facility Governing Committee (HFGC).

described in the previous section was also validated by additional stakeholders in three rounds of data collection between March and December 2020. Twenty-one stakeholders who were closely involved with the evaluation and implementation of P4P in Tanzania were invited to interview via email communication. Interviews were conducted over Zoom due to COVID-19 travel restrictions. In the final round of data collection, stakeholders were also sent a flyer and link to a short film introducing the research and purpose of interviews. Stakeholders were asked to confirm the structure of the CLD or indicate if changes needed to be made to reflect their experience of P4P (see Appendix C for interview tool).

2.3. Creation of CLD

There were three steps to developing the CLD. First we used secondary data (Table 1) to develop individual CLDs (Kim and Andersen, 2012; Tomoaia-Cotisel, 2018) representing stakeholder understanding of how P4P affects their local health system. Second, individual CLDs were combined in a step-wise process (Tomoaia-Cotisel, 2018) resulting in a single CLD, an initial shared mental model of P4P's impact on the health system. Third, the combined CLD structure was validated to check to what extent additional stakeholders interviewed at the time of the original data collection agree on the structure of the system (Tomoaia-Cotisel, 2018) and to check to what extent additional stakeholders interviewed at the time of this study agreed that the CLD reflected their experience of the programme (Rwashana et al., 2014; Andersen et al., 2012).

2.4. Step 1: creation of individual CLDs

Interview and FGD transcripts were split into two groups; transcripts from districts A, C and E (Table 1) were used in Step 1 and 2 to develop an initial shared CLD (Fig. 1). These three districts (A, C and E) were selected to develop the initial shared CLD to represent variation in stakeholder group and geographical location. Transcripts from the remaining two districts (B and D) were used in Step 3 for initial validation of the CLD (Fig. 2).

To develop individual stakeholder-specific CLDs, cause and effect relationships from each transcript were elicited using Purposive Text Analysis (Kim and Andersen, 2012) adapted for CLDs (Tomoaia-Cotisel, 2018). Quotations were coded if they described events or scenarios that furthered understanding of how providers or health managers responded to the intervention in their facility or district, or demonstrated health system behaviour that facilitated or hindered facilities achieving P4P targets. Using this transformative process (Kim and Andersen, 2012; Tomoaia-Cotisel, 2018), coding was used to develop a single CLD for each stakeholder interview (using Excel to store this information and Vensim software (Ventana Systems Inc. Vens, 2015) to develop the CLD).

2.5. Step 2: creation of initial shared CLD

To create the initial shared CLD (Fig. 1), we combined individual CLDs into team CLDs (representative of facility or district management) through a process called CLD Combination (Tomoaia-Cotisel, 2018). Individual stakeholder CLDs within teams were ordered according to their level of 'complexity', in terms of the number of variables, links, loops and delays. The most complex CLD, the 'anchor' CLD, was compared to the second most complex CLD. The anchor CLD was altered to reflect new information in the second CLD, through a new segment of the CLD or refinement of existing content. This altered CLD was then compared to the third most complex CLD and so on until all individual stakeholder CLDs within that team had been combined into one team

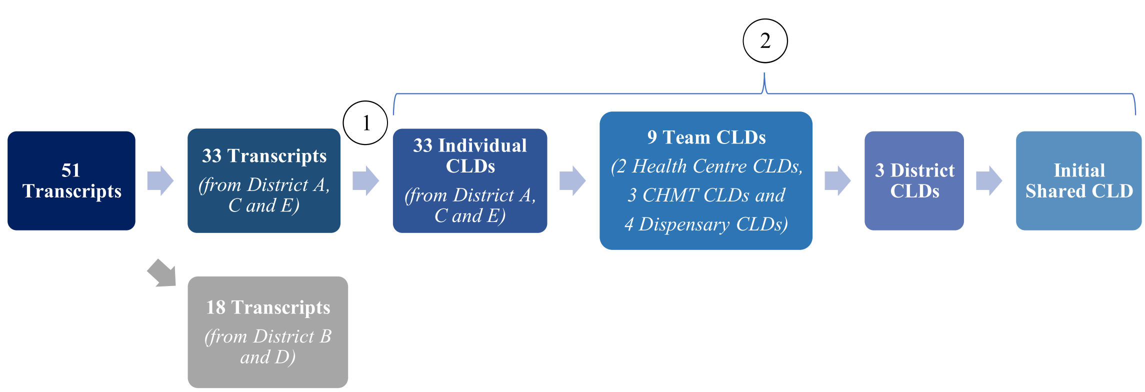


Fig. 1. Process for creating (1) individual CLDs and (2) initial shared CLD.
Notes to Figure: Step (1) Development of individual CLDs based on districts A, C and E (Table 1) and Step (2) Merging individual causal loop diagrams to create a single shared causal loop diagram. The 51 transcripts comprise of the transcripts from 43 individual interviews and from 8 focus group discussions. A CLD was developed for each transcript, one FGD transcript was used to develop one CLD. Adapted from (Tomoaia-Cotisel, 2018). Causal Loop Diagram (CLD), Council Health Management Team (CHMT).

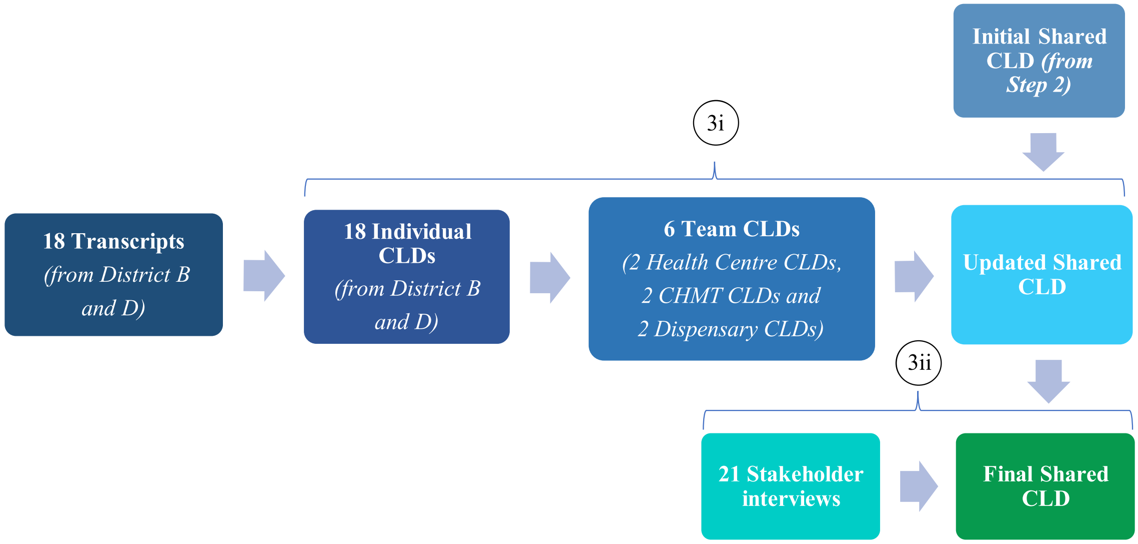


Fig. 2. Validation of initial shared CLD.
Notes to Figure: Step (3i) Comparison with team level CLDs that were not used to develop the shared CLD in the previous stage. Step (3ii) New stakeholder interviews to validate CLD structure. Adapted from (Tomoaia-Cotisel, 2018; Rwashana et al., 2014; Andersen et al., 2012). Causal Loop Diagram (CLD), Council Health Management Team (CHMT).

CLD. Where stakeholders described the same event but one CLD contained more information (a more complex loop), the complex loop was retained if the additional information was deemed necessary to understanding the behaviour of that particular part of the system. Next, we combined team CLDs into three district level CLDs using the same approach. Lastly, we combined the three-district level CLDs to create a shared (single) CLD.

2.6. Step 3: validation of initial shared CLD

Lastly, validation of the initial shared CLD was performed to ensure that critical input from each of the three stakeholder groups (health centres, dispensaries and CHMT) had not been lost or misinterpreted during the CLD development process (Fig. 2). Validation comprised of two stages: first, the initial shared CLD was validated to check to what extent additional teams interviewed at the time of the original data collection agree on the structure of the system (Tomoaia-Cotisel, 2018) and second, the updated shared CLD was validated to check to what extent additional stakeholders interviewed at the time of this study agreed that the CLD reflected their experience of the programme (Rwashana et al., 2014; Andersen et al., 2012).

In the first stage, we used interview and FGD data from districts B and D to generate 18 individual CLDs and then combined these individual CLDs into six team CLDs. We then compared each team level CLD to the initial shared CLD from Step 2, to see to what extent the team level CLDs confirmed the structure of the shared CLD or if any changes were required to the diagram (Tomoaia-Cotisel, 2018). Structural changes were made to the shared CLD to reflect the new elements identified in the team level CLDs (additional links and variables to expand concepts/loops already present in the CLD, modifications that increased understanding of supply of medical commodities at the facility).

In the second stage of validation, the CLD resulting from the first phase of validation was presented to twenty-one stakeholders closely involved with the evaluation and implementation of the P4P programme (Rwashana et al., 2014; Andersen et al., 2012). Stakeholders were asked if they recognised the structure and elements in the CLD and if any changes were needed to reflect their own experience of the health system and the P4P programme. This process aimed to minimise unconscious bias, to identify and amend any misinterpretation of data and elicit any further missing content. Structural changes were made to the CLD as a result of these interviews (strengthened understanding on use of facility and CHMT funding, additional drivers for health worker motivation, additional complexity included on pathways for addressing staffing levels at facilities).

2.7. Presentation of CLD

We identify two categories of performance targets: 'Number of women and children who receive incentivised services' and 'Submission of routine health facility data by providers' (shown in **bold** in a high-level snapshot of CLD, Fig. 3). We identified three core mechanisms responsible for provider achievement of (or failure to reach) targets during the programme: (1) changes in the supply of services, (2) changes to facility reporting, and (3) changes in demand for services. We present an overview of each mechanism and the corresponding sections of the CLD (with the overall CLD shown in Appendix D), including stakeholder quotes from the qualitative data the CLD was developed from.

3. Results

3.1. Changes in the supply of services

The mechanisms that result in changes in the supply of services during the P4P programme are presented in Fig. 4, with individual loops shown in Fig. 5. The 'motivation – service delivery' loop (Fig. 5A, R1) is a virtuous cycle of growing action where incentive payments to providers increase health worker motivation to exert effort towards incentivised services. At the start of the programme before any payments are made, the promise of future bonus payments motivates health workers to achieve targets. On receiving the P4P incentive payment, health workers feel further motivated to reach targets. This initial boost and then sustained level of motivation is dependent on bonus payments being made on time; where payments to facilities are delayed (a common issue during the first year of the scheme) staff become frustrated and apathetic about the programme. Health workers also feel motivated to continually exert effort where their exertion is recognised by those in senior roles at the facility (Fig. 5A, R2) and where supervision visits by the CHMT are taking place (Fig. 5A, R3), as this makes health workers feel valued. However, the CHMT can only perform supportive supervision where funds for per diems and transport are available.

Health worker motivation to deliver incentivised services leads to timely requisition of medical commodities as seen in Fig. 5B, R4. As

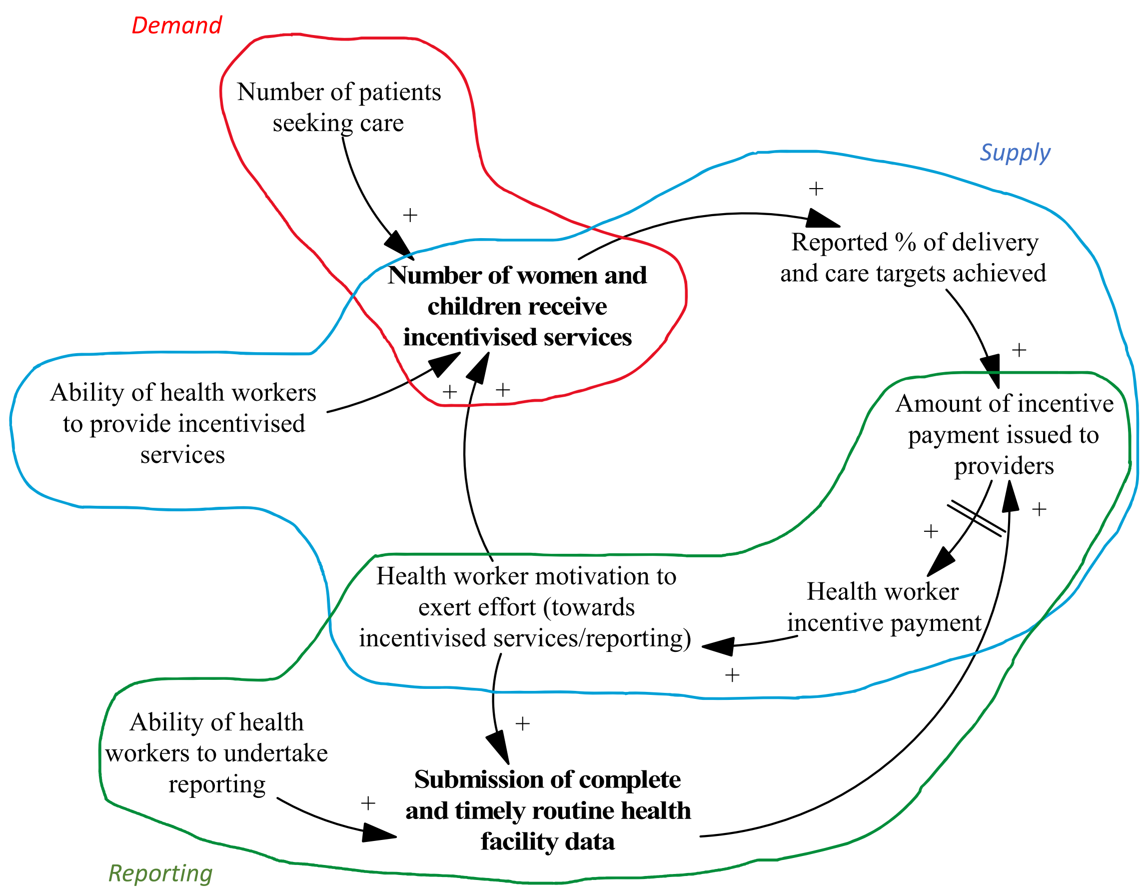


Fig. 3. High level snapshot of causal loop diagram.
Notes to Figure: Three main mechanisms responsible for provider achievement of (or failure to reach) targets during P4P are shown in different colours. Changes in the supply of services (blue), changes to facility reporting (green), and changes in demand for services (red).

shown in the 'ability' loop (Fig. 5B, B1), providers can only deliver incentivised services where there are adequate levels of medical commodities. Providers become proactive in their requisition of these items in an effort to reduce stockouts; to ensure drugs and supplies are sent to the facility, requests need to be submitted within a strict timeframe. Requests are sent to the district pharmacist who liaises with the Medical Stores Department (MSD) (an autonomous government department) to procure drugs and medical supplies (Fig. 5B, B2). Delays and missing drugs in orders received from the MSD leads to failure of P4P service delivery targets if facilities are unable to procure from another source:

> "One of the indicators was vaccines they are supposed to be given, but there are no vaccines to offer, at the end the facility will not score but that is not the facility's fault, it is somebody's fault. You may find (…) that the MSD does not supply all the drugs requested (….)".
>
> District level stakeholder, January 2012.

When the MSD are unable to fulfil an order, two virtuous cycles of growing action become dominant; CHMT provision of resources and facility purchase of drugs/medical supplies outside of MSD. The reduction of drug and medical supply stockouts at the facility level is an incentivised indicator for CHMT, through which CHMT members are motivated to support facilities (Fig. 5C, R7) and provide medicine and medical equipment where needed (Fig. 5C, R5). Facilities also use their own funds to purchase medicine and medical supplies from other sources (outside of the MSD) where needed (Fig. 5C, R6):

> "There are changes, we used to get few drugs but since P4P started there is an improvement, if we get problems, we face our doctor (in charge) we use (…) P4P money to buy drugs. We take this opportunity to ask him to identify unavailable drugs in the facility then we buy them".
>
> Facility level stakeholder, July 2012.

However, this is dependent on health providers having funding available (achieving P4P targets and receiving bonus payments) (Fig. 5C, R8), facilities setting up and having access to a bank account and an active HFGC. The HFGC, comprised of community members and health workers, support provider decision-making on use of funds at the facility and approve the release of funds. An additional source of funding outside of P4P that can be used by facilities to purchase medicine and medical supplies is the Community Health Fund (CHF) (Fig. 5C, R9). Providers saw this voluntary health insurance scheme as an opportunity to raise additional funds for service delivery (as premium revenue is kept by the facility) and increase the likelihood of achieving P4P targets (see 'mechanisms that result in changes in the demand for services' section for further details on the Fund).

Health worker motivation is tied to worker ability to deliver services. Where there are shortages in medicine, medical supplies (Fig. 5D, B3) and inadequacies in the facility environment (old mattresses, lack of cleaning equipment) (Fig. 5D, R10) impeding health worker ability to deliver services, health workers feel frustrated and demotivated, affecting staffing levels at facilities:

> "They are frustrated by this (…) they had a medical doctor there, but he was not happy that he was sent to a facility that did not have a lot of equipment. He could not practice the skills he received during his training … so he was frustrated to the extent that he was planning to leave".

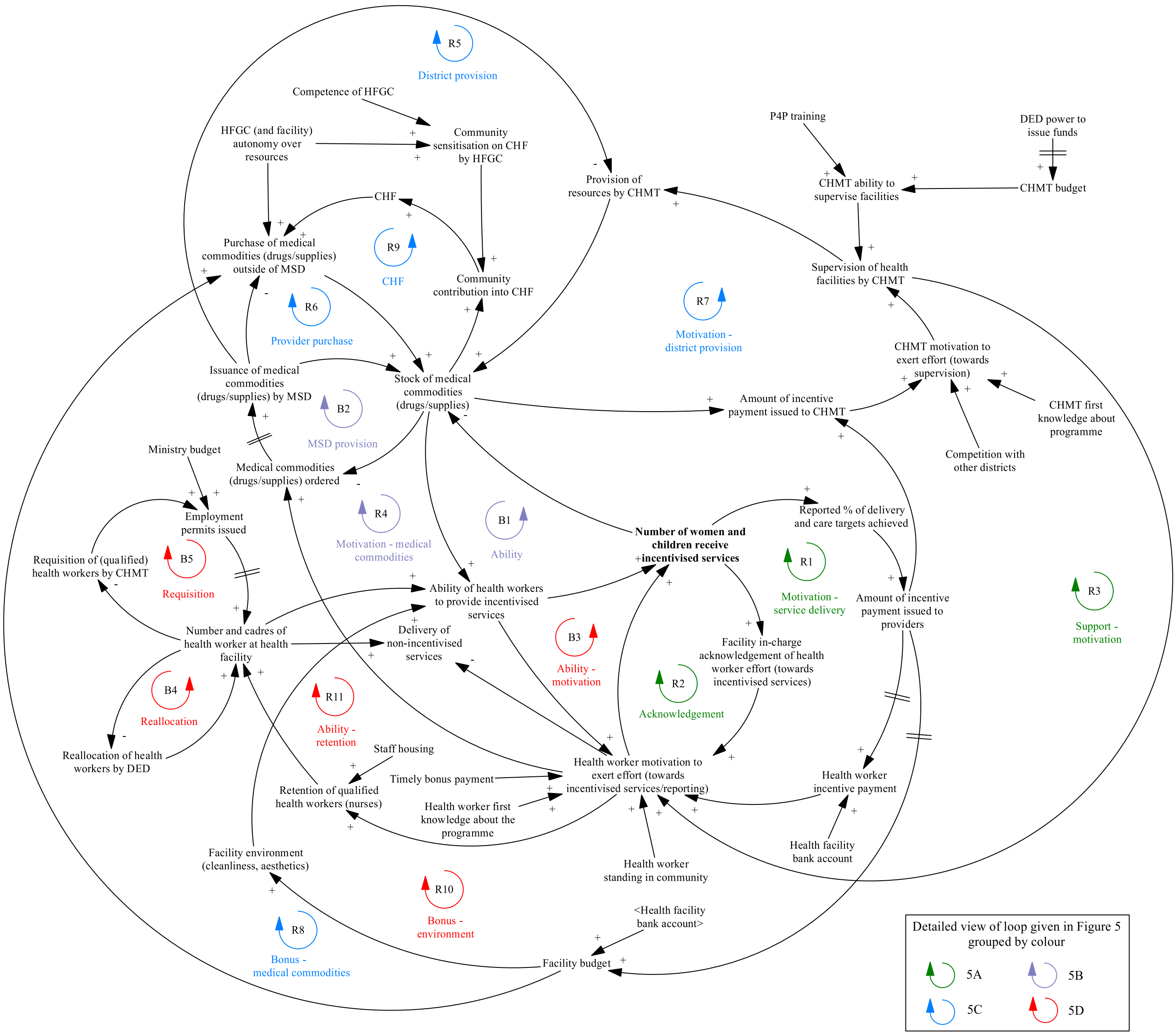


Fig. 4. Changes in the supply of services during the programme that impact facility achievement of targets.
Notes to Figure: Council Health Management Team (CHMT), Community Health Fund (CHF), District Executive Director (DED), Health Facility Governing Committee (HGFC), Medical Stores Department (MSD), Payment for performance (P4P).

Programme evaluation researcher, November 2020.

Adequate staffing levels and variety in the cadre of staff ensure an appropriate skill mix at the facility, to deliver more specialised services such as delivery care (Fig. 5D, R11). There were concerns that in facilities with depleted staffing levels, health worker motivation to deliver incentivised services and achieve reporting targets would result in task-shifting away from non-incentivised services (illustrated in Fig. 5D). To address vacancies the District Executive Director would reallocate staff to facilities in need (Fig. 5D, B4) and the CHMT request funding/permits for new staff (Fig. 5D, B5).

3.2. Changes to facility reporting

The mechanisms that result in changes to facility reporting during the P4P programme are presented in Fig. 6, with individual loops shown in Fig. 7. The 'motivation – reporting' loop (Fig. 7A, R12) is a virtuous cycle of growing action where incentive payments to providers increase health worker reporting of facility activity to the CHMT. This task can take considerable time, and facilities need adequate staffing to achieve this target alongside service delivery (Fig. 7A, R13) (with mechanisms for addressing staffing levels discussed in the previous section):

> "Effort is done, we are expecting to get money in the third round, what was causing us not to get the money was the failure of submitting reports, the facility had one nurse. She said that she was

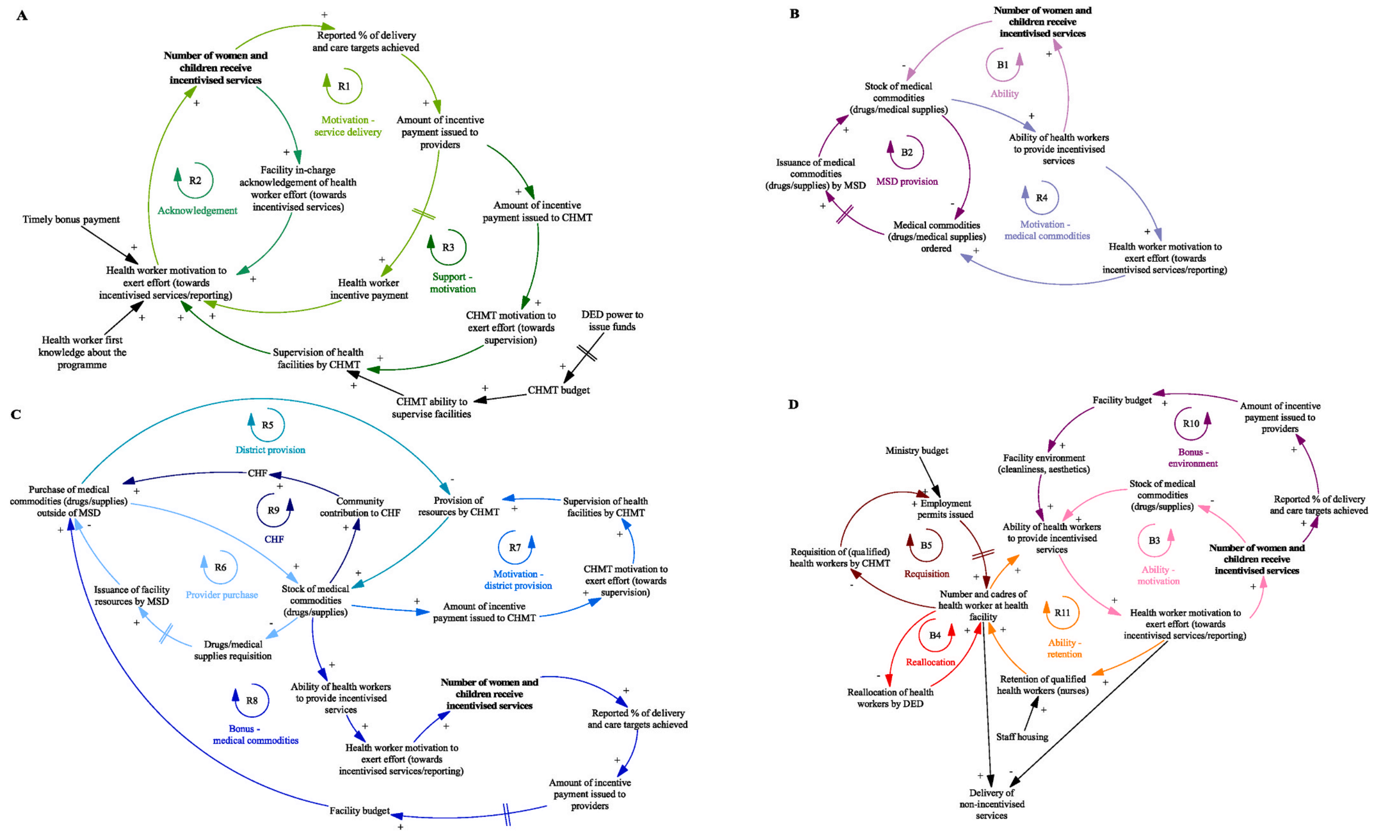


Fig. 5. Detailed views of the mechanisms that result in changes in the supply of services during the programme and impact facility achievement of targets. Notes to Figure: Council Health Management Team (CHMT), Community Health Fund (CHF), District Executive Director (DED), Health Facility Governing Committee (HGFC), Medical Stores Department (MSD), Payment for performance (P4P).

overloaded, but since I arrived here the first thing I did was to make sure we submit reports".

Facility level stakeholder, July 2012.

The 'understanding' loop (Fig. 7A, R14) is another virtuous cycle of growing action where health worker ability to undertake reporting is dependent on their knowledge of the reporting system. Health workers are sent for training at the start of the programme including on what routine health facility data should be reported to the CHMT. In facilities with high staff turnover, the training knowledge was lost with providers unable to achieve this target.

District manager (CHMT) incentives are partly driven by completeness of provider reporting (Fig. 7B, R15). CHMT members advise providers on record keeping and reporting during supervision visits. This offsets lack of provider knowledge (Fig. 7A, R14). In cases where facilities are unable to physically submit reports (due to lack of funds, transport or staff) the CHMT collect reports to support timely submission. The bonus payments encouraged district managers to make supervision visits:

"The bonus is like a carrot we have to run for it (…) we are trying to improve our systems as time goes on (…). So, we decided to start collecting report(s) because we discovered this will be very helpful to us. Though we face transport problem(s), I remember the last trip I went for supervision (I) was not paid, I spent my own money from my pocket because the budget for supervision was very minimal (…)".

District level stakeholder, January 2012.

The 'temptation to misreport' loop (Fig. 7B, R16) is a vicious cycle illustrating the temptation to game the system and record higher levels of service delivery than actually provided to achieve higher incentive payments. Where mis-reporting is suspected during verification visits, an investigation and potential suspension of facility and CHMT incentive payments is implemented. CHMT supervision visits act as a deterrent for misreporting (Fig. 7B, B6); district managers compare reported data with facility records to ensure reported performance is accurate.

3.3. *Changes in demand for services*

The mechanisms that result in changes in demand for services during the P4P programme are presented in Fig. 8, with individual loops shown in Fig. 9. Improved patient-provider interaction (perceived kindness and respect from health workers) observed during P4P leads to an increase in the patient perceived quality of services (Fig. 9A, R17) and facility reputation, affecting the care seeking of other women.

"… workers are very polit(e) and kind to patients not like before. This surprise[s] the pregnant mothers, it is not like before when the workers were abusing them. Through P4P the pregnant mothers get good serve [health services] so she may tell her fellow [women] to come to the facility too [so] finally many of them will come to deliver [their babies] in the facility".

Facility level stakeholder, December 2011.

A key mechanism to boost demand for services involves additional outreach activities carried out by providers in the community (Fig. 9A, R18). In some settings, this includes both community sensitisation activities and use of a mobile clinic to offer immunization services. The additional interaction between patients and providers in the community provides the opportunity to raise awareness of services offered at the facility and build trust between women in the community and facility

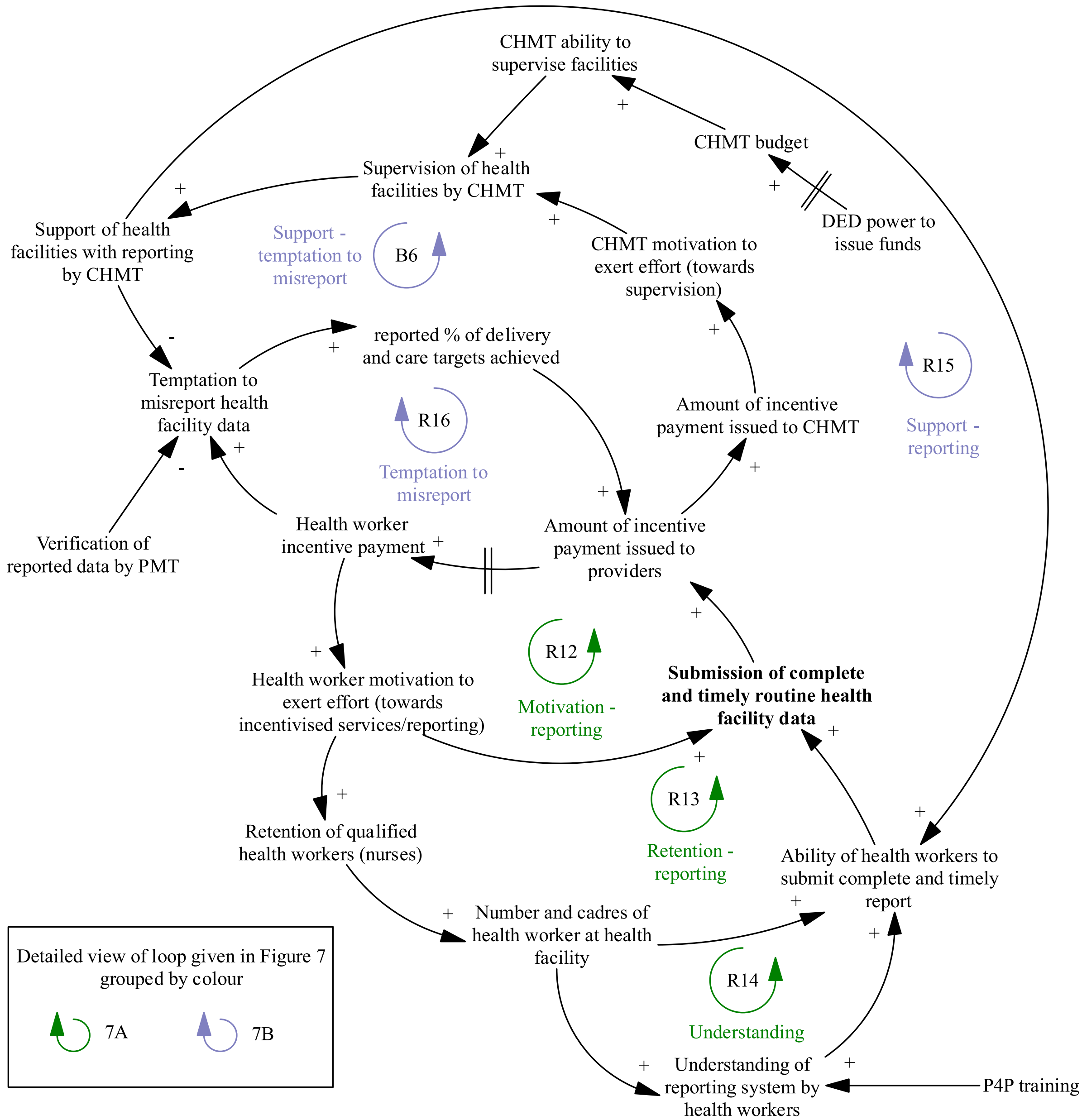


Fig. 6. Changes to facility reporting during the programme that impact facility achievement of targets.
Notes to Figure: Council Health Management Team (CHMT), District Executive Director (DED), Payment for performance (P4P), Pilot Management Team (PMT).

workers, resulting in greater service uptake. Although health workers feel motivated to perform outreach services, their ability to do so is heavily dependent on funding for transport/fuel (Fig. 9B, R19) and adequate staffing levels (Fig. 9B, R20). The work of community health workers (CHWs) in engaging community members and promoting health education is critical to raising awareness of health issues and available services (Fig. 9A):

> "You know the CHW, normally they come from the same community, and (…) are trusted people in that community. The community people are the one(s) who have chosen that person to be a community health worker (…) so if anything, if that CHW tells the community about maybe malaria, they trust it through the CHW more than anybody else (…) sensitisation becomes easier because it is their own people who tell the story".
>
> National level stakeholder, November 2020.

Other mechanisms for increasing demand include improving the facility environment (Fig. 9B, R22) and incentivising traditional birth attendants (Fig. 9B, R21). Providers used their bonus payments to purchase cleaning products, mattresses and other items to improve facility cleanliness and aesthetics. The facility environment is expected to impact patient perceived quality of services and decision to seek care, with improvements due to P4P, likely to increase demand from patients. Another innovative method employed by some facilities is to incentivise traditional birth attendants to boost the number of patients seeking care; incentive payments awarded to facilities who had increased their service performance are partially redistributed to traditional birth attendants who accompany women to attend facilities for institutional deliveries.

A key element that feeds into patient perceived quality of services and decision to seek care is availability of medicine and medical supplies (Fig. 9C, B7); by increasing the availability of drugs and supplies (Fig. 5C, R8), P4P reduces the likelihood of patients paying out of pocket, which increases demand. Availability of medical commodities

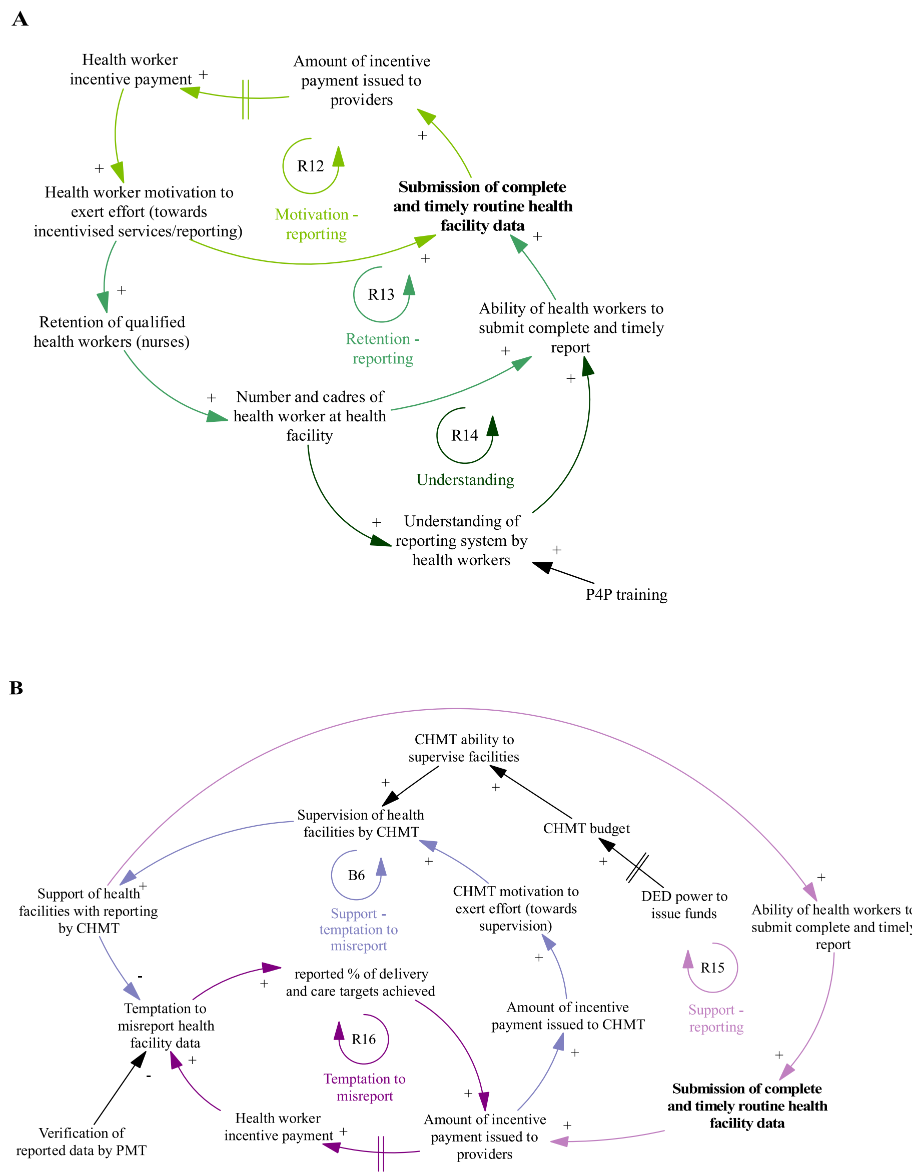


Fig. 7. Detailed views of the mechanisms that result in changes to facility reporting during the programme and impact facility achievement of targets.
Notes to Figure: Council Health Management Team (CHMT), District Executive Director (DED), Payment for performance (P4P), Pilot Management Team (PMT).

also influences patient decision-making on membership of the community health insurance scheme, CHF (Fig. 9C, R9). When drugs and supplies are in stock, patients are more likely to register with the CHF, as they perceive services to be of better quality. Membership of the CHF also reduces the likelihood of paying out of pocket for care. The additional revenue from the CHF increases resource availability at the facility level which further increases demand for services. This cycle produces optimal behaviour when stocks of drugs and supplies are already at satisfactory levels; where they are diminished due to supply chain issues (Fig. 5B, B2) or funding (Fig. 5C, R8), this creates a vicious cycle. Lack of medicine leads to reduced payment into CHF, where community members anticipate their contributions will not guarantee availability of medicine, leading to lack of funding for medicine and supplies. An important measure to prevent this downward spiral is providers having an active, competent HFGC. In addition to advising providers on use of facility funds, the HFGC also promote community contribution into the Fund through community mobilisation and education:

"The health facility governing committee (HFGC) were not active, I remember it was in May when they were told about their roles as HFGC members. They were told why facilities run out of drugs, it just because people do not want to join CHF, I am telling you that HFGC members came up with action plan, they planed that when they go back to their villages, they are going to join CHF as well as to sensitize other village members to join CHF. This will let the community know that they must contribute for drugs."

District level stakeholder, July 2012.

4. Discussion

We used CLDs to provide insight into how facilities and district managers responded to P4P and shed light on mechanisms involved in provider achievement of MCH and facility reporting targets, and contextual factors supporting or impeding these. On the supply side, we observed how health worker motivation and ability of health workers to provide services were critical to achievement of P4P targets. Health

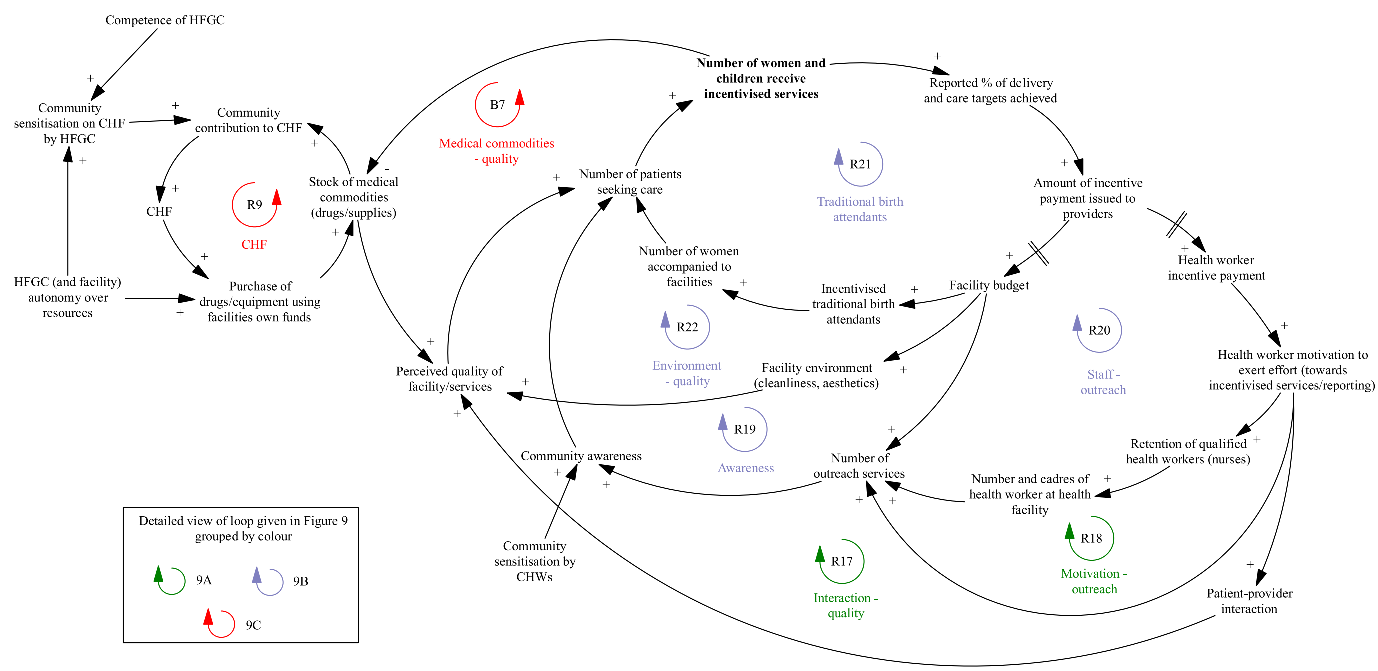


Fig. 8. Changes in demand for services during the programme that impact facility achievement of targets.
Notes to Figure: Council Health Management Team (CHMT), Community Health Fund (CHF), Community Health Workers (CHWs), District Executive Director (DED), Health Facility Governing Committee (HGFC).

worker motivation and ability to deliver services were dependant on factors directly affected by P4P (timely receipt of incentive payments, ability to purchase drugs and medical supplies using incentive payments) but importantly also on factors outside of programme influence (number of health workers, drugs and medical supplies supply chain, core facility funding). In the same vein, we observed that routine reporting of health facility data was heavily dependent on support given by the CHMT (directly influenced by P4P) but also by the composition of health workers at the facility (not directly affected by P4P). On the demand side, we observed the importance of patient perceived quality of services and community awareness of facility services (both partly influenced by P4P) in leading to a higher number of patients seeking care and facility achievement of P4P targets.

These three overarching mechanisms that resulted in provider achievement of targets are closely interconnected (Fig. 3), with changes in one part of the system leading to knock-on effects in other parts of the system. Using the CLD, it is possible to identify *catalytic variables* in the system; variables that affect multiple outcomes or mechanisms and therefore deserve careful consideration in the design of P4P schemes. Facility readiness, and especially the availability of drugs and medical supplies, is critical to service delivery. Not only in the direct sense of availability of drugs enabling health workers to deliver services but it is also critical to health worker motivation to deliver services. Facility readiness also influences patient perceived quality of facility services and feeds into the decision to seek care at the facility and decision to financially support the facility by enrolling in the community-based health insurance scheme. This variable was key to facilitating the supply and demand side mechanisms that led to facility achievement of P4P targets.

Staffing levels and supervision of facilities by district level managers are also catalytic variables. Supply of services at the facility, outreach activities (impacting demand for services) and facility reporting mechanisms (timely completion and submission of reports) only exhibited optimal behaviour where there were adequate levels of staffing at facilities. Supply of services and facility reporting mechanisms were also influenced by district management team supervision, with support leading to a more motivated workforce and facilitating provider ability to undertake routine data reporting.

The CLD also unearths potential *system levers* which are not targeted by P4P but could be incorporated to enhance the effect of the programme. CHF, the community-based health insurance scheme, was an additional source of revenue for facilities that could be used to purchase medical commodities and enhance their ability to achieve targets. We found that facilities often drew on the CHF as a lever to enhance performance. Community contributions into the CHF were dependent on community sensitisation on CHF by the HFGC, and the availability of drugs and supplies (as their absence led to out-of-pocket payments). HFGC members were not incentivised as part of P4P and yet were integral to facility success during the programme through their role in mobilising community contributions to the CHF and as signatories on facility expenditures. CHWs were also a non-incentivised group that were instrumental in stimulating demand for services at facilities, leading to facility achievement of P4P targets. CHWs provided a crucial flow of information from providers to the wider community; they were seen as trusted members of the community, able to promote health education and spread awareness on facility services and operation. The lack of incentivisation of these stakeholders sometimes undermined their leverage by facilities to achieve performance goals, where this created bad feeling. Incentivising other key stakeholders who operate at the facility and community level seems an appropriate element in the design of P4P schemes in LMICs.

This study sheds light on those P4P design features which were most important in achieving outcomes, and how programme design could be improved to enhance effects. For example, the facility-level incentive and incentivisation of district managers based on drug availability was critical to the programme resulting in the reduction in stock outs of drugs and supplies – which was a catalytic variable key to service delivery and demand for care. In settings where the availability of drugs and supplies is limited, it is essential that a share of the P4P incentive payments go to facilities to enable their procurement of drugs and supplies, and that other stakeholders that can facilitate access to drugs and supplies be incentivised as well (in this case district managers). To

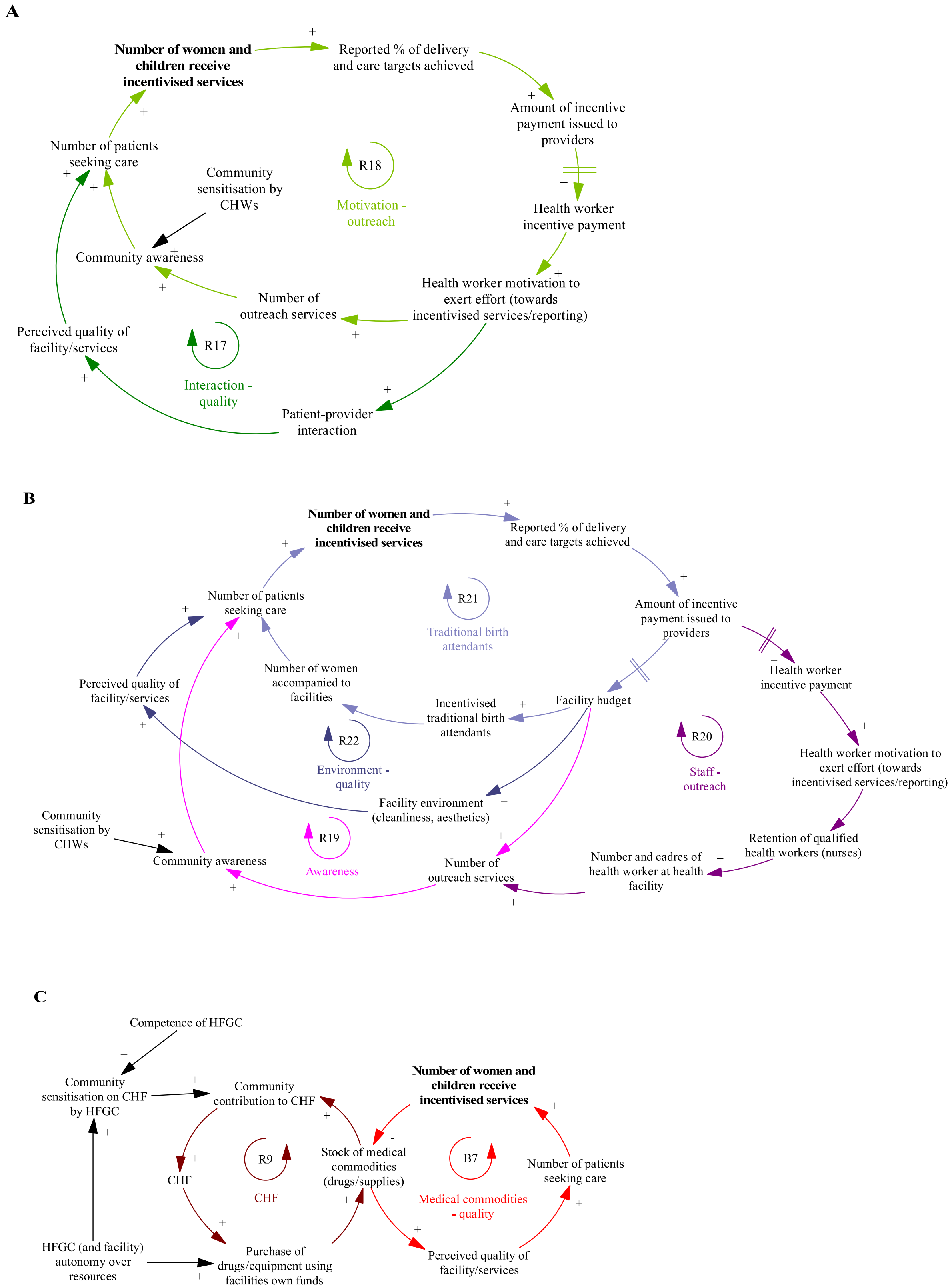


Fig. 9. Detailed views of the mechanisms that result in changes in demand for services during the programme and impact facility achievement of targets.
Notes to Figure: Council Health Management Team (CHMT), Community Health Fund (CHF), Community Health Workers (CHWs), District Executive Director (DED), Health Facility Governing Committee (HGFC).

further strengthen the effect of the programme on drugs and supplies, and align incentives across levels of the health system, the central MSD might have been incentivised (as has been observed in the design of the most recent, scaled-up version of P4P in Tanzania) (MoHSW Ministry of Health and Social Welfare, n.d.). Delays, due to late submission of stock orders by facilities or stockouts at the MSD, may have been mitigated if the MSD had been either incentivised or supported as part of the programme design. Strengthening the supply chain of medical commodities might be assisted through other system strengthening initiatives outside of P4P, such as a redesign of logistic systems and availability of vehicles for transport of medical commodities (strategic reforms recently tested by the Global Health Fund in discussion with the Tanzania government (Githendu et al., 2020)).

Our study also identifies health system features which were critical to the supply and demand of services but were not impacted by the programme, due to their omission from the design. Staffing levels were critical to achieving outcomes, but this was largely outside of the control of facilities and districts. While district managers could reallocate staff within the district from higher to lower staffed facilities, they could not recruit new staff, even for facilities that were understaffed relative to staffing norms. Although reallocation of staff could be an incentivised target for district managers, without the capacity to hire new staff (managed at the government level and subject to restrictions on budget) this may not be a suitable target for P4P in Tanzania. Lastly, HFGC and CHW were critical stakeholders to boost demand for services, and in the case of HFGC, ensure full and effective use of facility resources, yet they were omitted from the incentive system. To maximise programme effectiveness, all stakeholders that are critical to achieving supply and demand side goals should be identified and, if possible, integrated into the incentive system. Our findings suggest that P4P as currently designed, would work best in facilities with adequate drug and supply availability and staffing levels.

Our study adds to the existing evidence base examining the effects of the P4P pilot in Tanzania on the health system and population (Mayumana et al., 2017; Anselmi et al., 2017; Binyaruka et al., 2015, 2018a; Binyaruka and Borghi, 2017; Olafsdottir et al., 2014; Binyaruka and Anselmi, 2020), by identifying those variables which are really catalytic both in terms of achieving performance targets (e.g. drugs), and limiting their achievement (staffing). The CLD also identifies pathways to improvements and potential pathways to harm (unintended negative effects), and system levers which are outside the scope of the programme but can be leveraged by providers to help achieve programme goals.

Two other studies have used CLDs to evaluate the impact of P4P programmes on health systems in low-income countries (Alonge et al., 2017; Renmans et al., 2017). A CLD of P4P in Afghanistan (Alonge et al., 2017) also identified the effect of service utilisation on facility revenue, and of health worker motivation on uptake of services. However, the Afghan CLD includes a highly a composite quality variable (representing time spent with patients, drug availability, perception of care and other measures). Our study shows these measures of quality do not necessarily move in the same direction over time and are, therefore, better observed separately. The Afghan CLD also excludes supervision and staffing which we found to be important influences on outcomes. A CLD of P4P in Uganda (Renmans et al., 2017) identified the importance of district-level supervision on health worker motivation and knowledge, as in our study, and investments in the facility environment leading to increased care seeking. However, medicines and infrastructure are combined in a 'work environment' variable, and unlike our study, the medicines supply chain is not included in the CLD. Reporting of health facility data was not an incentivised target in the Ugandan P4P programme and was therefore excluded from the CLD.

Singh et al. (2021) used a CLD to synthesize evidence identified in a realist review. Like our study, the realist review identifies drug availability, health worker kindness and outreach services as key mechanisms underpinning P4P effects on utilisation outcomes. Our study contributes further evidence on availability of drugs as a critical factor in community demand for community based health insurance, and the positive relationship between insurance uptake and drug availability. We also identify pathways between facility readiness and health worker motivation, and between supervision and deterrence from misreporting data.

There are a number of limitations to this study. Data used to develop the CLD were not collected for this purpose, which may have limited the degree of causal statements. However, this approach is highly cost-effective by limiting the primary data collection that is needed. As the CLD was developed by one researcher, there is a risk that unconscious bias may have gone unchecked in the CLD. However, we found the CLD to be well supported during validation. On the methods front, we could have used more objective methods to compare pairs of CLDs prior to combining them; mathematical graph theory has been used previously to compare pairs of CLDs (Markóczy, 1995; Schaffernicht and Groesser, 2011). However, because of the large number of CLDs involved, it was more practical to use qualitative reasoning methods to compare the CLDs and then combine them. While we had planned to conduct face to face validation interviews, due to COVID-19 these were conducted via Zoom. However, this online format worked effectively. To reduce the risk of recall bias stakeholders were encouraged to say when they were not confident in their recollection of events. Stakeholders often offered anecdotes and reflections to support their confirmation of model structure (or recommendation for changes) which strengthened confidence in their ability to provide evidence on their experience of the programme.

Another limitation is the generalisability of the CLD to represent pathways to impact of P4P on delivery and uptake of MCH services in other types of facilities (secondary care providers). Study authors decided to exclude secondary data collected on secondary care facilities due to the substantial programme design differences between hospitals and lower-level facilities (health centres and dispensaries), and the much larger number of primary care facilities included in the programme. Given these facility operation and design differences, the current CLD would not be generalisable to secondary care facilities.

A further limitation of this work is that we did not have data from patients themselves which may have highlighted other variables of relevance to care seeking practices. We intend to develop a system dynamics simulation model in later work that will use survey data from patients to explore the dynamic hypothesis raised in the CLD. The CLD gave an indication of variables which were more or less frequently mentioned, but it does not allow us to quantify the relative impact of different variables or loops within the system. Without quantifying relative and combined effects, it is difficult to estimate how key outcomes would be impacted by P4P design changes and understand the reasons for the dynamic behaviour playing out over time. However, in the system dynamics modelling research we have planned, we will be able to identify the key/dominant loops in the CLD by quantifying how mechanisms/loops change over time in response to P4P using the developed simulation model. The model will allow us to quantify relationships between variables and measure the effect a given loop has on key outcomes.

The CLD identified key mechanisms underpinning facility achievement of P4P targets, catalytic mechanisms impacting multiple outcomes and potential levers, and design modifications to improve programme effectiveness. Further research using CLDs to study heath systems in LMIC is urgently needed to further our understanding of how systems respond to interventions and how to strengthen systems to deliver better coverage and quality of care.

Authors' contributions

RC, JB, KB and NSS conceived the study. All authors contributed to the final structure and content of the paper. RC, JB, ATC, ARS, ZB and KB contributed to the determination of methods applied in this study. PB and JB managed the provision of secondary data. RC led the analysis of secondary and newly collected data in the study. RC and JM collected new data, interviews with stakeholders, to validate the causal loop diagram. RC wrote the first draft of the paper with input from JB and ZC. All authors contributed to the development of the paper and reviewed and approved the final version.

Declaration of competing interest

None.

Acknowledgements

This research is funded by the Medical Research Council under the Health Systems Research Initiative grant (MR/R013454/1). This study received a favourable ethical opinion from the Observational/Interventions Research Ethics Committee at The London School of Hygiene and Tropical Medicine (LSHTM Ethics Ref: 16139 – 2), the Institutional Review Board at Ifakara Health Institute (IHI/IRB/No:15 -2019) and National Institute for Medical Research (NIMR/HQ/R.8a/Vol. IX/3154) in Tanzania.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2021.114277.

References

Afnan-Holmes, H., Magoma, M., John, T., Levira, F., Msemo, G., Armstrong, C.E., et al., 2015. Tanzania's Countdown to 2015: an analysis of two decades of progress and gaps for reproductive, maternal, newborn, and child health, to inform priorities for post-2015. Lancet Glob Heal 3, e396–409. https://doi.org/10.1016/S2214-109X(15)00059-5.

Alonge, O., Lin, S., Igusa, T., Peters, D.H., 2017. Improving health systems performance in low- and middle-income countries: a system dynamics model of the pay-for-performance initiative in Afghanistan. Health Pol. Plann. 32, 1417–1426. https://doi.org/10.1093/heapol/czx122.

Andersen, D.L., Luna-Reyes, L.F., Diker, V.G., Black, L., Rich, E., Andersen, D.F., 2012. The disconfirmatory interview as a strategy for the assessment of system dynamics models. Syst. Dynam. Rev. 28, 255–275. https://doi.org/10.1002/sdr.1479.

Anselmi, L., Binyaruka, P., Borghi, J., 2017. Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania. Implement. Sci. 12, 10. https://doi.org/10.1186/s13012-016-0540-1.

Atun, R., 2012. Health systems, systems thinking and innovation. Health Pol. Plann. 27 https://doi.org/10.1093/heapol/czs088 iv4–8.

Bhatnagar, A., George, A.S., 2016. Motivating health workers up to a limit: partial effects of performance-based financing on working environments in Nigeria. Health Pol. Plann. 31, 868–877. https://doi.org/10.1093/heapol/czw002.

Binyaruka, P., Anselmi, L., 2020. Understanding efficiency and the effect of pay-for-performance across health facilities in Tanzania. BMJ Glob Heal 5, e002326. https://doi.org/10.1136/bmjgh-2020-002326.

Binyaruka, P., Borghi, J., 2017. Improving quality of care through payment for performance: examining effects on the availability and stock-out of essential medical commodities in Tanzania. Trop. Med. Int. Health 22, 92–102. https://doi.org/10.1111/tmi.12809.

Binyaruka, P., Patouillard, E., Powell-Jackson, T., Greco, G., Maestad, O., Borghi, J., 2015. Effect of paying for performance on utilisation, quality, and user costs of health services in Tanzania: a controlled before and after study. PloS One 10, 1–16. https://doi.org/10.1371/journal.pone.0135013.

Binyaruka, P., Robberstad, B., Torsvik, G., Borghi, J., 2018a. Who benefits from increased service utilisation? Examining the distributional effects of payment for performance in Tanzania. Int. J. Equity Health 17, 14. https://doi.org/10.1186/s12939-018-0728-x.

Binyaruka, P., Robberstad, B., Torsvik, G., Borghi, J., 2018b. Does payment for performance increase performance inequalities across health providers? A case study of Tanzania. Health Pol. Plann. 33, 1026–1036. https://doi.org/10.1093/heapol/czy084.

Borghi, J., Chalabi, Z., 2017. Square peg in a round hole: re-thinking our approach to evaluating health system strengthening in low-income and middle-income countries. BMJ Glob Heal 2, e000406. https://doi.org/10.1136/bmjgh-2017-000406.

Borghi, J., Mayumana, I., Mashasi, I., Binyaruka, P., Patouillard, E., Njau, I., et al., 2013. Protocol for the evaluation of a pay for performance programme in Pwani region in Tanzania: a controlled before and after study. Implement. Sci. 8, 80. https://doi.org/10.1186/1748-5908-8-80.

Bradley, D.T., Mansouri, M.A., Kee, F., Garcia, L.M.T., 2020. A systems approach to preventing and responding to COVID-19. EClinicalMedicine 21, 100325. https://doi.org/10.1016/j.eclinm.2020.100325.

Das, A., Gopalan, S.S., Chandramohan, D., 2016. Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review. BMC Publ. Health 16, 321. https://doi.org/10.1186/s12889-016-2982-4.

de Savigny, D., Blanchet, K., Adam, T., 2017. Conclusion: mapping methods to research challenges. In: de Savigny, D., Blanchet, K., Adam, T. (Eds.), Applied Systems Thinking for Health Systems Research : a Methodological Handbook. McGraw-Hill Education.

Engineer, C.Y., Dale, E., Agarwal, A., Agarwal, A., Alonge, O., Edward, A., et al., 2016. Effectiveness of a pay-for-performance intervention to improve maternal and child health services in Afghanistan: a cluster-randomized trial. Int. J. Epidemiol. 45, 451–459. https://doi.org/10.1093/ije/dyv362.

Fichera, E., Nikolova, S., Sutton, M., 2014. Comparative Performance Evaluation: Quality. Encyclopedia of Health Economics. Elsevier, pp. 111–116. https://doi.org/10.1016/B978-0-12-375678-7.01313-4.

Gagné, M., Deci, E.L., 2005. Self-determination theory and work motivation. J. Organ. Behav. 26, 331–362. https://doi.org/10.1002/job.322.

Githendu, P., Morrison, L., Silaa, R., Pothapregada, S., Asiimwe, S., Idris, R., et al., 2020. Transformation of the Tanzania medical store department through global fund support: an impact assessment study. BMJ Open 10, e040276. https://doi.org/10.1136/bmjopen-2020-040276.

Kim, H., Andersen, D.F., 2012. Building confidence in causal maps generated from purposive text data: mapping transcripts of the Federal Reserve. Syst. Dynam. Rev. 28, 311–328. https://doi.org/10.1002/sdr.1480.

Mannion, R., Davies, H.T.O., 2008. Payment for performance in health care. BMJ 336, 306–308. https://doi.org/10.1136/bmj.39463.454815.94.

Markóczy, L., 1995. A method for eliciting and comparing causal maps. J. Manag. 21, 305–333. https://doi.org/10.1016/0149-2063(95)90060-8.

Mayumana, I., Borghi, J., Anselmi, L., Mamdani, M., Lange, S., 2017. Effects of payment for performance on accountability mechanisms: evidence from Pwani, Tanzania. Soc. Sci. Med. 179, 61–73. https://doi.org/10.1016/j.socscimed.2017.02.022.

Meker, T., Barlas, Y., 2015. Dynamic consequences of performance-based payment systems in public hospitals. Syst. Res. Behav. Sci. 32, 459–480. https://doi.org/10.1002/sres.2338.

MoHSW (Ministry of Health and Social Welfare), 2012. The Pwani Region Pay-for-Performance (P4P) Pilot Design Document. Dar es Salaam. MoHSW, United Republic of Tanzania.

MoHSW (Ministry of Health and Social Welfare), n.d. Result Based Financing, Operational Manual. Dar es Salaam: MoHSW, United Republic of Tanzania.

Ngo, D.K.L., Sherry, T.B., Bauhoff, S., 2017. Health system changes under pay-for-performance: the effects of Rwanda's national programme on facility inputs. Health Pol. Plann. 32, 11–20. https://doi.org/10.1093/heapol/czw091.

Olafsdottir, A.E., Mayumana, I., Mashasi, I., Njau, I., Mamdani, M., Patouillard, E., et al., 2014. Pay for performance: an analysis of the context of implementation in a pilot project in Tanzania. BMC Health Serv. Res. 14, 392. https://doi.org/10.1186/1472-6963-14-392.

Paul, E., Sossouhounto, N., Eclou, D.S., 2014. Local stakeholders' perceptions about the introduction of performance-based financing in Benin: a case study in two health districts. Int. J. Health Pol. Manag. 3, 207–214. https://doi.org/10.15171/ijhpm.2014.93.

Peters, D.H., 2014. The application of systems thinking in health: why use systems thinking? Health Res. Pol. Syst. 12, 51. https://doi.org/10.1186/1478-4505-12-51.

Rashwan, W., Abo-Hamad, W., Arisha, A., 2015. A system dynamics view of the acute bed blockage problem in the Irish healthcare system. Eur. J. Oper. Res. 247, 276–293. https://doi.org/10.1016/j.ejor.2015.05.043.

Renmans, D., Holvoet, N., Criel, B., 2017. Combining theory-driven evaluation and causal loop diagramming for opening the "black box" of an intervention in the health sector: a case of performance-based financing in Western Uganda. Int. J. Environ. Res. Publ. Health 14, 1007. https://doi.org/10.3390/ijerph14091007.

Renmans, D., Holvoet, N., Criel, B., 2020. No mechanism without context: strengthening the analysis of context in realist evaluations using causal loop diagramming. N. Dir. Eval. 2020, 101–114. https://doi.org/10.1002/ev.20424.

Rwashana, A.S., Nakubulwa, S., Nakakeeto-Kijjambu, M., Adam, T., 2014. Advancing the application of systems thinking in health: understanding the dynamics of neonatal mortality in Uganda. Health Res. Pol. Syst. 12, 36. https://doi.org/10.1186/1478-4505-12-36.

Sahin, O., Salim, H., Suprun, E., Richards, R., MacAskill, S., Heilgeist, S., et al., 2020. Developing a preliminary causal loop diagram for understanding the wicked complexity of the COVID-19 pandemic. Systems 8, 20. https://doi.org/10.3390/systems8020020.

Schaffernicht, M., Groesser, S.N., 2011. A comprehensive method for comparing mental models of dynamic systems. Eur. J. Oper. Res. 210, 57–67. https://doi.org/10.1016/j.ejor.2010.09.003.

Schoenenberger, L.K., Bayer, S., Ansah, J.P., Matchar, D.B., Mohanavalli, R.L., Lam, S.S., et al., 2016. Emergency department crowding in Singapore: insights from a systems thinking approach. SAGE Open Med 4. https://doi.org/10.1177/2050312116671953, 205031211667195.

Shen, G.C., Nguyen, H.T.H., Das, A., Sachingongu, N., Chansa, C., Qamruddin, J., et al., 2017. Incentives to change: effects of performance-based financing on health workers in Zambia. Hum. Resour. Health 15, 20. https://doi.org/10.1186/s12960-017-0179-2.

Singh, N.S., Kovacs, R.J., Cassidy, R., Kristensen, S.R., Borghi, J., Brown, G.W., 2021. A realist review to assess for whom, under what conditions and how pay for performance programmes work in low- and middle-income countries. Soc. Sci. Med. 270, 113624. https://doi.org/10.1016/j.socscimed.2020.113624.

Tomoaia-Cotisel, A., 2018. The Journey toward the Patient-Centered Medical Home: A Grounded, Dynamic Theory of Primary Care Transformation. PhD Thesis. LSHTM. https://doi.org/10.17037/PUBS.04647856.

Tomoaia-Cotisel, A., Hyunjung, K., Allen, S., Blanchet, K., 2017. Causal loop diagrams: a tool for visualizing emergent system behaviour. In: de Savigny, D., Blanchet, K., Adam, T. (Eds.), Applied Systems Thinking for Health Systems Research : a Methodological Handbook. McGraw-Hill Education.

Ventana Systems Inc, 2015. Vensim PLE Plus.

6.3 Supplementary material to research paper 2

Appendix A: How to read a causal loop diagram

A CLD shows cause and effect relationships between variables in a system. The polarity indicated on each connecting arrow signifies the effect of changes in one variable on another. So, for example, if incentives result in more motivated health workers there would be a positive arrow linking ‘health worker incentive payment’ and ‘health worker motivation’ (Figure S1).

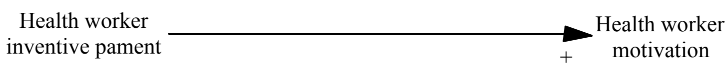


Figure S1: Example of a cause and effect relationship represented in CLD.

Delays in effect are indicated by two lines cutting through a connecting arrow (Figure S2). For example, this might apply if there were delays in payments reaching facilities.

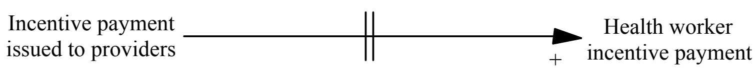


Figure S2: Example of a delay in effect represented in CLD.

Two or more connected variables can form reinforcing (Figure S3) or balancing loops (Figure S4). For example, if incentives result in motivated health workers who then treat more patients to achieve further incentive payments, there would be positive arrows linking ‘health worker incentive payment’ to ‘health worker motivation’, ‘health worker motivation’ to ‘number of patients treated’ and ‘number of patients treated’ to ‘incentives’ forming a reinforcing loop. There exists a reinforced or amplified behaviour between these variables, a knock-on effect leading to a

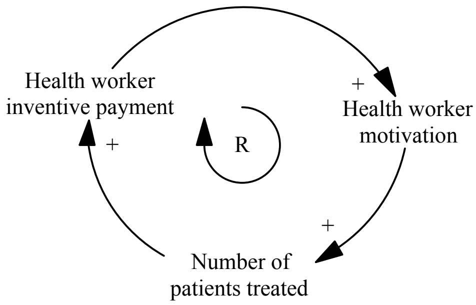


Figure S3: Example of a reinforcing loop.

growing action over time. Where this loop produces desirable behaviour, it is referred to as a virtuous cycle and where undesirable behaviour persists it is referred to as a vicious cycle.

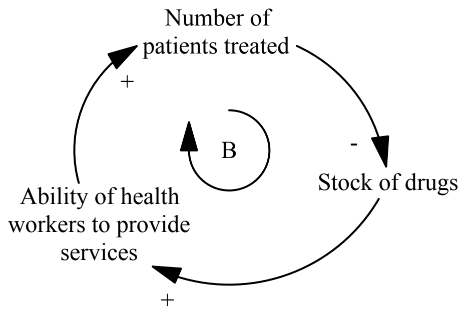


Figure S4: Example of a balancing loop.

In a balancing loop, the loop is prevented from exhibiting spiralling or amplified behaviour by the presence of one or more variables. For example, adequate stock of drugs enables health workers to treat patients leading to a reduction in the overall stock of drugs at the facility, is an example of a balancing loop. The loop is prevented from exhibiting amplified behaviour by one or more variables (volume of patients decreasing stock of drugs).

Please see [1,2] for more information on how to interpret CLDs.

Appendix B: Facility and district managers indicators and performance targets during the pilot P4P programme in Tanzania

Table S1: Facility coverage, content of care and HMIS strengthening indicators and performance targets set during the pilot P4P programme in Tanzania.

Indicator	**Measure**	**Baseline coverage (previous cycle)**				
		0-20%	**21-40%**	**41-70%**	**71-85%**	**85%+**
Coverage indicators						
% of institutional deliveries	Percentage point increase	15%	10%	5%	5%	Maintain
% of mothers attending a facility within 7 days of delivery	Percentage point increase	15%	10%	5%	5%	Maintain
% of women using long term contraceptives	Percentage point increase	20%	15%	10%	Maintain above 71%	Maintain
% children under 1 year receiving Penta3 vaccine	Overall result	50%	65%	75%	80% +*	Maintain
% children under 1 year receiving measles vaccine	Overall result	50%	65%	75%	80% +*	Maintain
Content of care indicators						
% ANC clients receiving IPT2	Overall result	80%	80%	80%	80%+*	Maintain above 80%
% HIV+ ANC clients on ART	Overall result	40%	60%	75%	75%+*	Maintain
% of newborns receiving polio vaccine (OPV0)	Overall result	60%	75%	80%	80%+*	Maintain
HMIS strengthening						

HMIS monthly reports correctly filled and submitted on time to CHMT	Overall result	100%	100%	100%	100%	100%

Notes to Table: +*80% or more. Antiretroviral therapy (ART), Antenatal care (ANC), Council Health Management Team (CHMT), Management Information System (HMIS), Human Immunodeficiency Virus (HIV), Intermittent Preventative Treatment (IPT2). Source: Binyaruka *et al.* [3] and MoHSW [4].

Table S2: Council Health Management Team and Regional Health Management Team performance indicators set during the pilot P4P programme in Tanzania.

CHMT/RHMT/Both	**Indicator**	**Measure**
	Coverage indicators	
Both	% of maternal and newborn deaths that are appropriately audited on time	Overall result
	Health system strengthening	
CHMT	% of facilities reporting stock-outs of either one or more of the tracer medicines in a specified period (< 8 days)	Overall result
	HMIS strengthening	
CHMT	% of facilities included in the HMIS monthly reports exported through DHIS to RHMT in timely manner	Overall result
	Management	
RHMT	Submission to MoHSW of a Semi-Annual Regional Health Profile report, based on DHIS	Overall result
CHMT	% of facilities receiving a copy of a Quarterly District Health Profile report, based on DHIS	Overall result
	Overall	
Both	Overall performance along P4P facility-based indicators	Overall result

Notes to Table: Council Health Management Team (CHMT), District Health Information Software (DHIS), Health Management Information System (HMIS), Ministry of Health and Social Welfare (MoHSW), Regional Health Management Team (RHMT). Source: MoHSW [4].

Appendix C: Stakeholder CLD validation interview tool

Interviewer: This series of interviews have been organised by researchers from Ifakara Health Institute and the London School of Hygiene and Tropical Medicine. We are hoping to conduct interviews with experts, such as yourself, to validate a map we have created of the Tanzania maternal and child health (MCH) system response to payment for performance (P4P). We developed the map using interview data that was collected during the pilot P4P programme in Tanzania (2011-2013); interviews were conducted with health workers, facility in-charges and district level managers on how the programme had been received by providers and managers, and what factors had facilitated or hindered effective implementation of the programme. We are currently focussing on the primary care facilities that offered MCH services and took part in the pilot programme (excluding up-graded health centres).

Interviewer: To ensure our system map accurately represents the real health system behaviour and processes that developed under the pilot we now require this map to be validated by experts with knowledge of the pilot programme.

Interviewer: During this interview, I will show you system maps that are representative of how we believe the health system functioned following the introduction of the pilot P4P programme. Using your knowledge, experience and feedback of health system operation we will then refine the structure of our maps to ensure they reflect the pilot P4P programme.

Interviewer: In the next phase of our project we are going to be looking at the differences between the pilot and other health system strengthening programmes that have taken place in the country, including the up-scaled Results-Based Financing programme (RBF, 2016-2019) and Direct Health Facility Financing programme (DHFF, 2019-Present). If you have time at the end of the interview, I would be very interested to hear your opinion on the core (intended and observed) differences in health system transformation and outcomes between the three programmes.

Interviewer: Just before we begin, I have received a copy of your consent form but I would just like to seek your verbal consent that you are happy to continue with the interview and you are happy for me to take written notes and an audio-recording of this session. This is only for our records and shared only with our research team. You can change your mind or stop the interview at any time.

Interviewer: *If no* That is okay I will take written notes instead.

Once participant has given consent, open the Vensim diagram that shows the system map and ask the interviewee if they can see the map on their screen

Interviewer: We have this large system map of the Tanzania health system response to P4P but to make the most use of the time we have today, I am going to focus the interview on one area of the map. The map has been split into three segments corresponding to the (i) demand, (ii) supply and (iii) reporting-side mechanisms underpinning achievement of targets during P4P, with targets represented in the diagram in bold labelled ‘Number of women and children receive incentivised services’ and ‘submission of routine health facility data by providers’.

Interviewer: In today’s interview we are going to focus on the part of the map that describes *refer to (i), (ii) or (iii)*. I have highlighted the portion of the system map that corresponds to *refer to (i), (ii) or (iii)* so that we can still see how this part of the map connects to other elements of the map (just to show it doesn’t operate in isolation). I will describe what we are seeing in the map then periodically stop to check, to your knowledge, that this process occurred during the pilot P4P programme. Your feedback will help us validate our diagram and make any necessary refinements.

Interviewer: Just a few comments on what we are seeing here. We have variables and arrows connecting each of the variables. This indicates some kind of causal relationship exists between pairs of variables. You will also notice that the arrows have polarity attached to them, plus and minus signs. These indicate the direction of causality. For example, as ‘Amount of inventive payment issued to providers' increases, so does 'Health worker salary top up' (i.e. health workers receive bonus payments for improved performance during P4P).

*Stop here and check if the interviewee understands what you have described – does this make sense? *

Interviewer: You will also notice there are two small dashes across the arrow; this indicates a delay in effect. Taking the same example, although facilities who improve their performance during P4P should receive a bonus payment, there were often delays between the incentive payment being issued and health workers receiving this money (particularly at the beginning of the programme). This delay in effect is represented by those two dashes across the arrow.

*Stop here and check if the interviewee understands what you have described – does this make sense? *

Interviewer then proceeds with taking the interviewee round the rest of this map segment, periodically stopping to check interviewee understanding and to ask if any modifications should be made to the map to reflect their experience of the programme

The interviewer does not have to explicitly run through these questions while discussing the map, can instead probe 'Does this make sense? Are we missing anything important in this section of the map? Is there anything that you feel should be removed in the map?'. When an interviewee gives their feedback on the map, it will generally fall into these compartments and help the modeller to go back and make modifications to the map:

- Does this part of the system exist to your knowledge?
- Are appropriate system variables represented? If not, what variables are missing or should be removed?

- Are appropriate in and out flows represented? If not, what flows are missing or should be removed?

- Is the polarity of in and out flows accurately represented? If not, what changes would you make?

- Are appropriate delays in the system represented? If not, what delays are missing or should be removed?

When interviewer has finished with validating the system map

Interviewer: We may have already touched on this during our discussion of the map but I would also be interested to hear your view on what you think could have been changed in the implementation of the pilot programme to help facilities achieve targets (and improve the delivery and coverage of MCH services?).

Interviewer: As I said earlier, in the next phase of our project we are going to be looking at the differences between the pilot and other health system strengthening programmes that have taken place in the country, including the up-scaled Results-Based Financing programme (RBF, 2016-2019) and Direct Health Facility Financing programme (DHFF, 2019-Present). If you have time now, I would be very interested to hear your opinion on:

What are the key similarities and differences between P4P pilot program and/or 1) RBF program and (2) DHFF program?

Interviewer: Thank you very much for your time today, this has been incredibly useful. If you feel comfortable doing so, is there anyone you would recommend for us to interview next?

Appendix D: Mechanisms that result in changes in the supply of services, facility reporting and demand for services, highlighted in the CLD

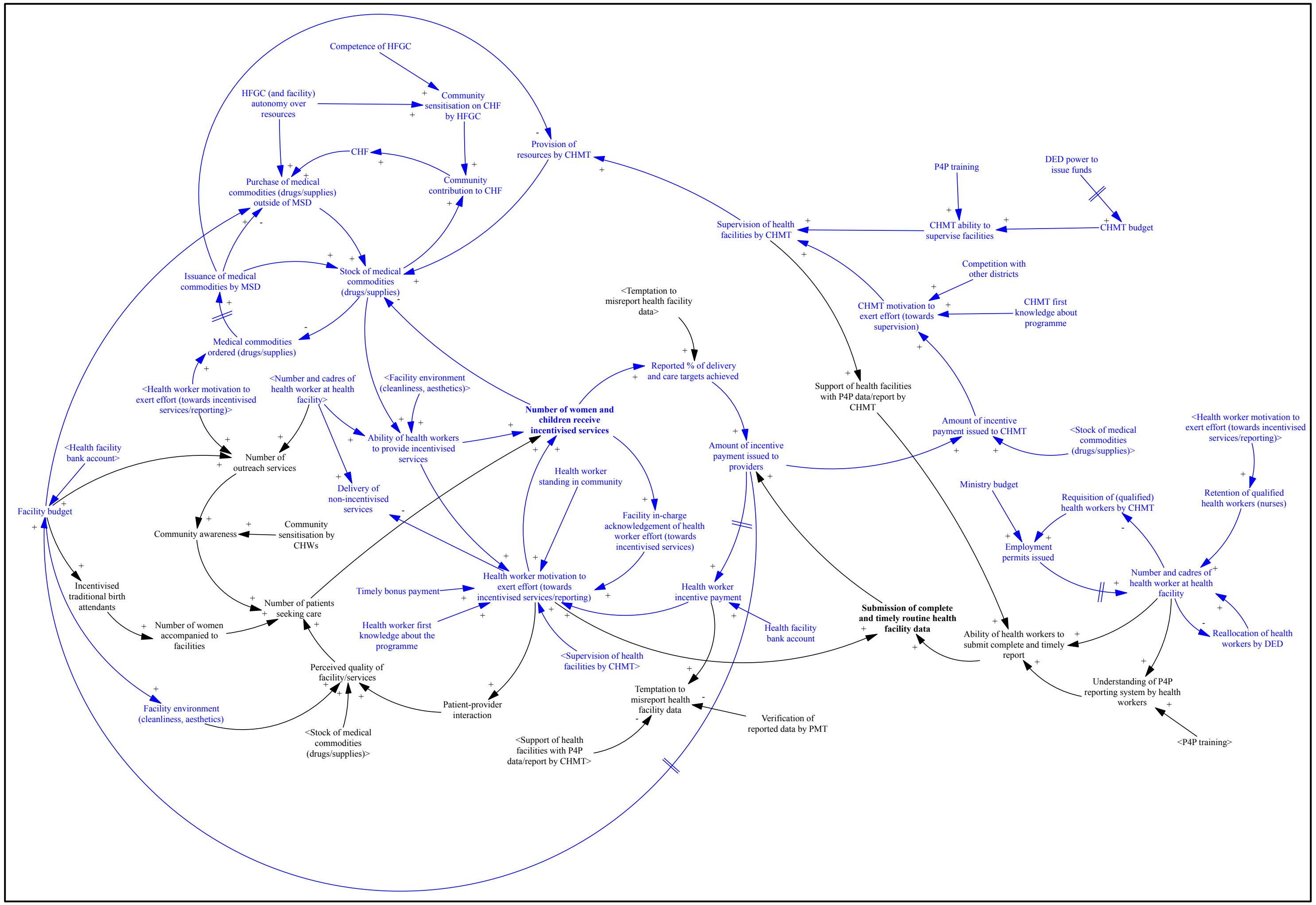


Figure S5: Mechanisms that result in changes in the supply of services highlighted in the CLD.

Notes to Figure: Health Management Team (CHMT), Community Health Fund (CHF), Community Health Workers (CHWs), District Executive Director (DED), Health Facility Governing Committee (HGFC), Medical Stores Department (MSD), Payment for performance (P4P), Pilot Management Team (PMT).

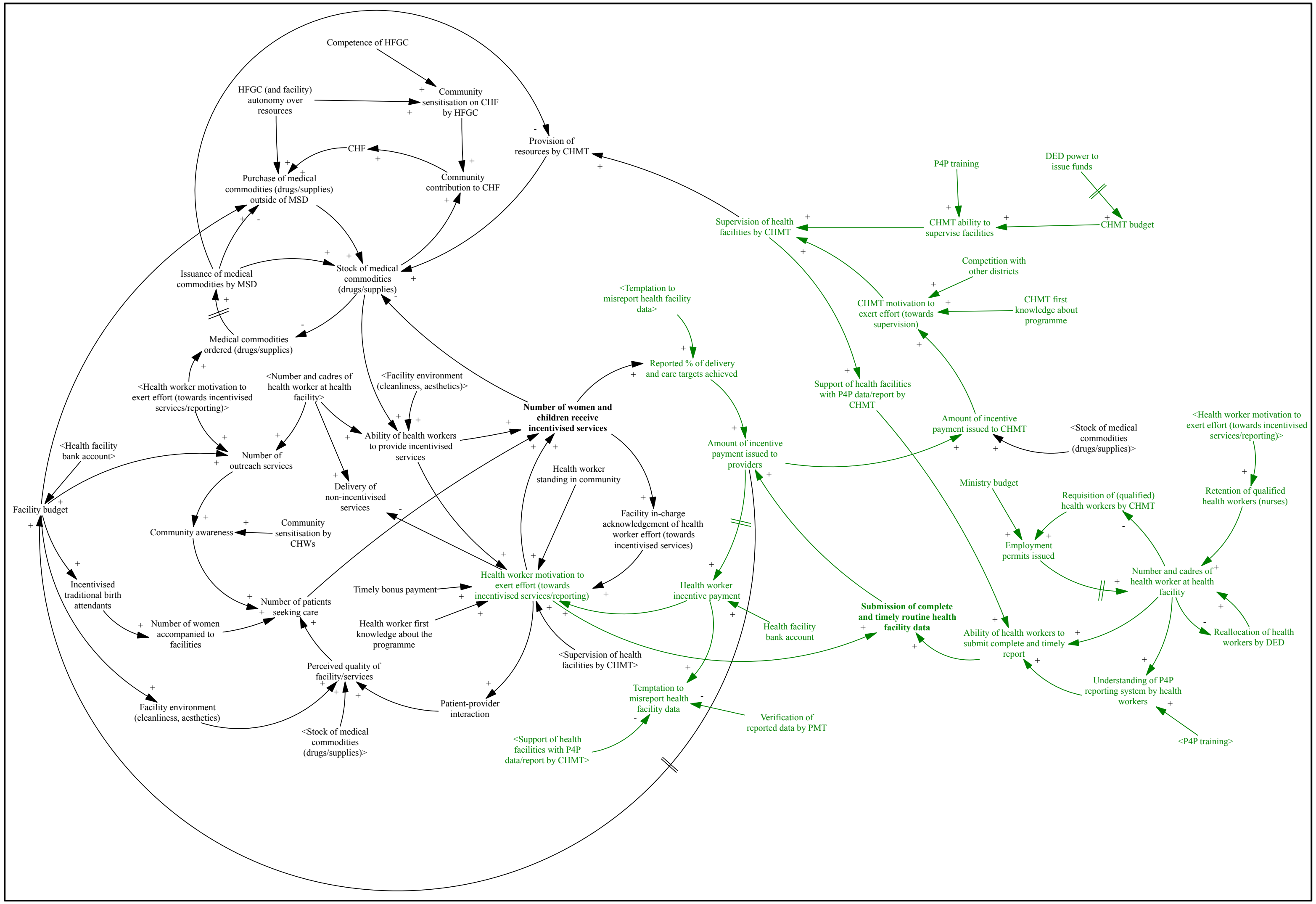


Figure S6: Mechanisms that result in changes to facility reporting highlighted in the CLD.

Notes to Figure: Health Management Team (CHMT), Community Health Fund (CHF), Community Health Workers (CHWs), District Executive Director (DED), Health Facility Governing Committee (HGFC), Medical Stores Department (MSD), Payment for performance (P4P), Pilot Management Team (PMT).

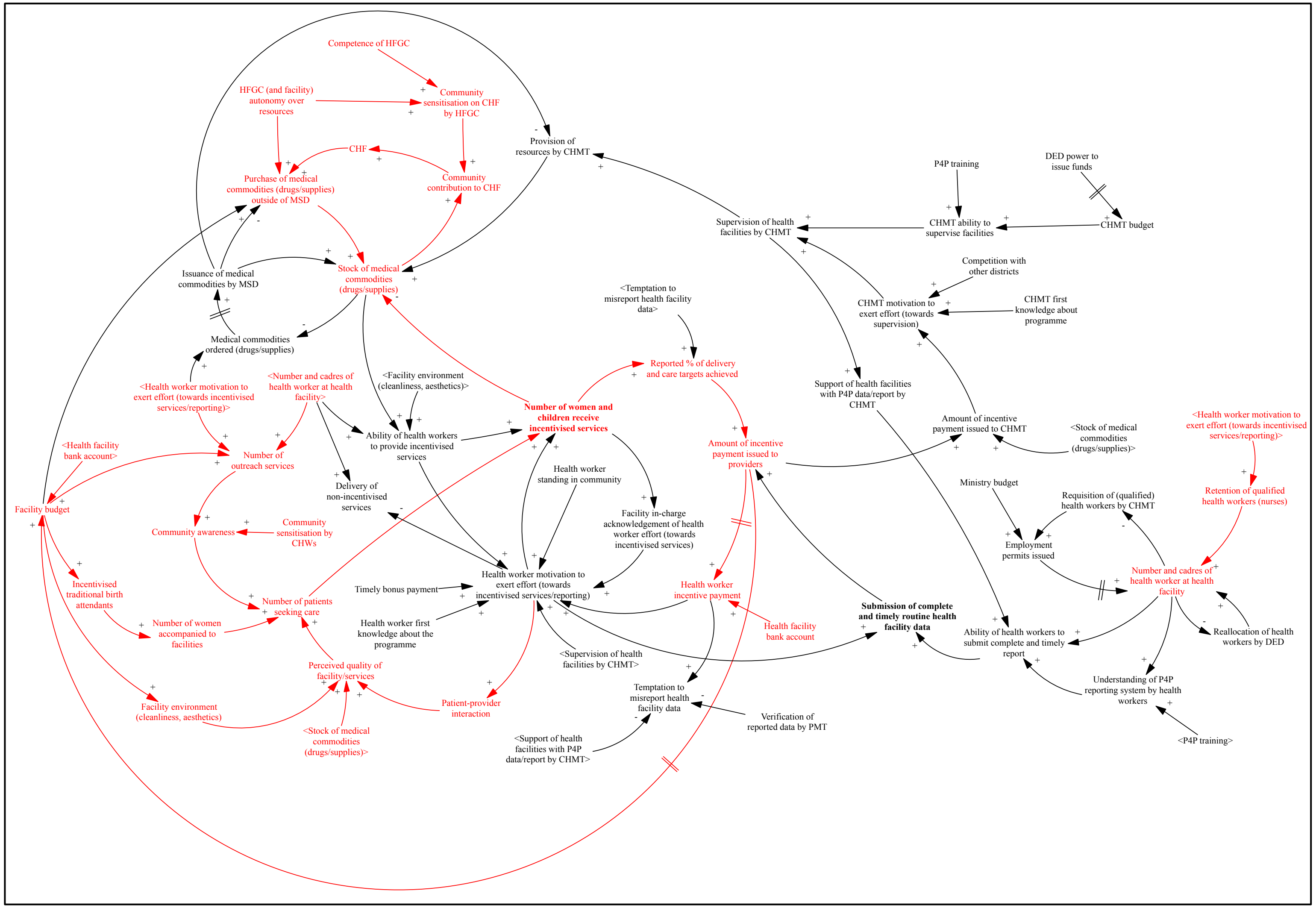


Figure S7: Mechanisms that result in changes in demand for services highlighted in the CLD.

Notes to Figure: Health Management Team (CHMT), Community Health Fund (CHF), Community Health Workers (CHWs), District Executive Director (DED), Health Facility Governing Committee (HGFC), Medical Stores Department (MSD), Payment for performance (P4P), Pilot Management Team (PMT).

References

[1] Tomoaia-Cotisel A, Hyunjung K, Allen S, Blanchet K. Causal loop diagrams: a tool for visualizing emergent system behaviour. In: de Savigny D, Blanchet K, Adam T, editors. Applied systems thinking for health systems research : a methodological handbook, McGraw-Hill Education; 2017.

[2] Sterman JD. Causal loop diagrams. Business Dynamics: Systems Thinking and Modeling for a Complex World, McGraw-Hill Companies Inc; 2000.

[3] Binyaruka P, Patouillard E, Powell-Jackson T, Greco G, Maestad O, Borghi J. Effect of Paying for Performance on Utilisation, Quality, and User Costs of Health Services in Tanzania: A Controlled Before and After Study. PLoS One 2015;10:1–16. doi:10.1371/journal.pone.0135013.

[4] MoHSW (Ministry of Health and Social Welfare), 2012. The Pwani Region Pay-for-Performance (P4P) Pilot Design Document. Dar es Salaam: MoHSW, United Republic of Tanzania.

7. Recommendations for using causal loop diagrams in health systems research

7.1 Introduction

Abbreviations: Causal loop diagram (CLD); Group Model Building (GMB); Low- and middle-income country (LMIC); Medical Stores Department (MSD); System dynamics model (SDM).

Chapter 7 is a guidance piece for study design utilising CLD methodology for health systems research in LMICs, based on the candidates own experience of application and evidence drawn from the literature, fulfilling **Objectives 1 and 5** of this thesis:

> 1. Determine current use and application of systems thinking methods (CLDs and SDMs) for health systems research, with reflection on use in LMIC health system settings.
>
> 5. Provide guidance on future health systems research using systems thinking to encourage uptake in LMIC settings.

The results of the study are presented in a paper, 'How to do (or not to do)…using causal loop diagrams for health system research in low and middle-income settings', published in Health Policy and Planning in August 2022. Evidence of retention of copyright or use of published materials in this thesis can be found in Appendix 3.

7.2 Research paper 3: How to do (or not to do)…using causal loop diagrams for health system research in low and middle-income settings

(Cover sheet on next page)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1401642	**Title**	Miss
First Name(s)	Rachel		
Surname/Family Name	Cassidy		
Thesis Title	Using systems thinking to optimise health system interventions for improved maternal and child health in low-resource settings		
Primary Supervisor	Prof. Josephine Borghi		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Health Policy and Planning		
When was the work published?	3rd August 2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

Open Access CC-BY license

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Conceptualised study with PhD supervisors, led search, screened titles and full texts, responsible for data extraction and analysis. Wrote first draft of the manuscript and led revisions with input from co-authors.

SECTION E

Student Signature	[redacted]
Date	05/12/22

Supervisor Signature	[redacted]
Date	05/12/22

Health Policy and Planning, **37**, 2022, 1328–1336
DOI: https://doi.org/10.1093/heapol/czac064
Advance access publication date: 3 August 2022
How to Do (Or Not to Do)

How to do (or not to do)…using causal loop diagrams for health system research in low and middle-income settings

Rachel Cassidy[1,*], Josephine Borghi[1], Agnes Rwashana Semwanga[2], Peter Binyaruka[3], Neha S Singh[1], and Karl Blanchet[4]

[1]Department of Global Health and Development, London School of Hygiene and Tropical Medicine, 15-17, Tavistock Place, London WC1H 9SH, UK
[2]Information Systems Department, College of Computing and Information Sciences, Makerere University, PO Box 7062, Kampala, Uganda
[3]Ifakara Health Institute, PO Box 78373, Dar Es Salaam, Tanzania
[4]Geneva Centre of Humanitarian Studies, University of Geneva and the Graduate Institute, 40, Boulevard du Pont-d'Arve, 1211 Genève, Switzerland

*Corresponding author. Department of Global Health and Development, London School of Hygiene and Tropical Medicine, 15-17, Tavistock Place, London WC1H9SH, UK. E-mail: rachel.cassidy@lshtm.ac.uk

Accepted on 2 August 2022

Abstract

Causal loop diagrams (CLDs) are a systems thinking method that can be used to visualize and unpack complex health system behaviour. They can be employed prospectively or retrospectively to identify the mechanisms and consequences of policies or interventions designed to strengthen health systems and inform discussion with policymakers and stakeholders on actions that may alleviate sub-optimal outcomes. Whilst the use of CLDs in health systems research has generally increased, there is still limited use in low- and middle-income settings. In addition to their suitability for evaluating complex systems, CLDs can be developed where opportunities for primary data collection may be limited (such as in humanitarian or conflict settings) and instead be formulated using secondary data, published or grey literature, health surveys/reports and policy documents. The purpose of this paper is to provide a step-by-step guide for designing a health system research study that uses CLDs as their chosen research method, with particular attention to issues of relevance to research in low- and middle-income countries (LMICs). The guidance draws on examples from the LMIC literature and authors' own experience of using CLDs in this research area. This paper guides researchers in addressing the following four questions in the study design process; (1) What is the scope of this research? (2) What data do I need to collect or source? (3) What is my chosen method for CLD development? (4) How will I validate the CLD? In providing supporting information to readers on avenues for addressing these key design questions, authors hope to promote CLDs for wider use by health system researchers working in LMICs.

Keywords: Health systems research, research methods, study design, complex systems, systems thinking, causal loop diagrams, low- and middle-income countries

Key messages

- Causal loop diagrams enable identification and visualization of drivers for complex system behaviour, including spill over effects or unintended consequences of policy and managerial decisions.
- They can be built and validated using different sources of data, depending on resource and health system setting constraints.
- It is vital that further research using a systems thinking lens be conducted in LMICs, taking into account the delivery and resource constraints experienced by facilities and actors, to further our understanding of health system functioning and optimization.

Introduction

Health systems are complex systems due to the large number of system elements (people, resources, processes), the varying and extensive relationships between them, and their responsiveness to their external environment (Lipsitz, 2012; Barasa *et al.*, 2017). They produce non-linear behaviour that evolves over time (Sterman, 2000a) and in response to relationships that exist between system elements (Lipsitz, 2012). Treating the health system as a static, linear system in evaluations results in oversight of potential unintended consequences, with health policies leading to suboptimal or undesirable outcomes due to focus on singular events and failure to observe the feedback and relationships between system elements (Adam and de Savigny, 2012).

For this reason, tools designed to manage and analyse complex behaviour need to be used to guide the design of health system interventions, and evaluate their effects (Skivington *et al.*, 2021). In taking a 'systems thinking' approach to research, emphasis is placed on connections and relationships between system elements as part of a larger, evolving system (Peters, 2014). Methods derived from systems thinking enable evaluation of interventions on the wider, interconnected dynamic system whilst observing the important underlying mechanisms and interactions that drive health system behaviour (Gates, 2016). Causal loop diagrams (CLDs) are one such method providing a visual representation of the relationships between system elements and their interactions, leading to understanding of what drives problematic system behaviour (Adam, 2014).

By helping to identify key health system constraints and/or evaluate potential health system improvements prior to implementation to guide programme design, CLDs can ensure investments are well targeted, which is especially useful in resource constrained health systems Furthermore, CLDs can be employed even where routine health information system data are limited, as literature, policy reports and stakeholder interviews can be used to support development of models. CLDs can be used to better understand the 'mechanisms for action' in the health system before interventions are implemented to inform their design (Borghi and Chalabi, 2017), or after their implementation to determine what worked, how and why.

However, to date, the use of CLDs has been limited in health systems research in low- and middle-income country (LMIC) studies (Borghi and Chalabi, 2017; Cassidy *et al.*, 2019). This paper introduces the reader to CLDs and their potential usages as a health systems research and policy tool, with particular attention to issues of relevance for LMIC studies. We then guide the reader through the stages of CLD development and validation (Box 1), using examples from the LMIC literature and authors' own experience of using CLDs in Tanzania and Uganda.

Box 1. Four guiding steps that underpin the design and conduct of CLDs for health systems research

(1) What is the scope of this research?

To define the phenomena or behaviour that you are trying to unpack, there are three key elements to consider:

- Time frame of interest
- Boundary of issue
- Level of system aggregation

(2) What data do I need to collect or source?

To further understanding on what is driving phenomena/behaviour, we can source and analyse:

- Primary data (e.g. key informant interviews and group model building)
- Secondary data (e.g. programme evaluation data, published literature, health surveys or reports, policy documents and systematic or realist review).
- Primary and secondary data

(3) What is my chosen method for CLD development?

Method for analysing and extracting data for CLD development:

- Ex post development (e.g. thematic analysis and purposive text analysis)
- Real-time development (e.g. group model building)

(4) How will I validate the CLD?

Method for confirming the CLD is still grounded in the experience of those with expert knowledge of the phenomena/behaviour:

- Stakeholder dialogue, including group model building activities
- Comparison to primary/secondary data sources

What are CLDs?

CLDs (Box 2) are diagrams that help us better understand what actions or mechanisms drive behaviour in a system (Tomoaia-Cotisel *et al.*, 2017); feedback (interactions between system elements, causing cycles of cause-and-effect behaviour) and loops (cycles of behaviour) emerge through development of these diagrams, illuminating desirable or undesirable behaviour (Sterman, 2000a). We can also identify spill-over effects of actions or interventions to wider parts of the system and unintended consequences that can lead to unexpected outcomes.

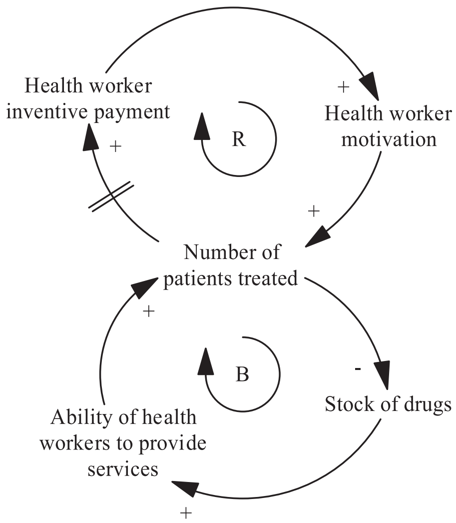
When can I use a CLD?

There are a variety of potential applications of CLDs of relevance to the health systems research and policy community. CLDs can be used ex-ante, to inform the design of a health systems intervention or policy, or to develop a theory of change to guide its evaluation (McGill *et al.*, 2021). Used in this way CLDs can determine the likely risks to a future programme that can be monitored during implementation to enable course correction (Sarriot *et al.*, 2015) and/or understand underlying mechanisms (drivers) for health system behaviour, and leverage points which can be targeted to produce optimum system behaviour (Kwamie *et al.*, 2014; Cassidy *et al.*, 2021). CLDs can also be used retrospectively to explore how policy implementation changes over time (Nigenda *et al.*, 2015), or to explore why health policies have succeeded or failed (Agyepong *et al.*, 2012; Paina *et al.*, 2014). They can be used in conjunction with existing health system frameworks, for example by identifying interconnections and/or dynamic behaviour between the WHO health system building blocks (Sharma *et al.*, 2020). Finally, CLDs can also support the synthesis of evidence regarding a health systems intervention, used to present the results of realist and systematic reviews (Namatovu and Semwanga, 2020; Singh *et al.*, 2021).

CLDs can also be used outside programme evaluation to explore how health systems respond to shocks or disruption (Ozawa *et al.*, 2016; Jamal *et al.*, 2020), and identify factors leading to system resilience, specifically the 'absorptive, adaptive and transformative capabilities' of the system. CLDs can highlight supply and/or demand side mechanisms related to a particular health condition, such as drivers for inadequate childhood immunization (Rwashana *et al.*, 2009; Varghese *et al.*, 2014; Kanniyan *et al.*, 2021), uptake and provision of mental health services (Trani *et al.*, 2016; Noubani *et al.*, 2020) and refugee and host community demand for healthcare (Noubani *et al.*, 2020; Zablith *et al.*, 2021).

Box 2. Origin, building blocks and interpretation of CLDs

System dynamics (the methodological field in which CLDs originate from) began as a tool for industrial and business management but now has widespread application across various research domains, including health system research (Pruyt 2017).



Notes to Box: Building blocks of CLD are presented; variables, arrows with polarity, reinforcing and balancing feedback loops and delays. Source: Adapted from Cassidy *et al.* (2021).

A simple CLD is presented in this box, showing the impact of a payment for performance intervention on the delivery of services at a health facility (Cassidy *et al.* 2021). Arrows with polarity indicate a causal relationship between two system variables and the direction of effect, for example, an increase in incentive payments during the intervention results in health workers feeling further motivated to deliver incentivised health services.

Delays in effect can also be represented, identified as a double line through an arrow. For example, we observe a delay in effect between the number of patients treated and an increase in the incentive payment then issued to health workers.

A series of arrows that close to form a 'loop' are labelled as either a reinforcing or balancing loop. A reinforcing loop exhibits amplified or spiralling behaviour (all arrows in the loop have the same polarity). An increase in health worker incentive payments leading to an increase in health worker motivation and the number of patients who are then treated leading to an increase in the incentive payments then issued to health workers is an example of a loop that shows reinforcing behaviour. A balancing loop is prevented from exhibiting spiralling behaviour by the presence of one or more variables and instead presents a dampened behaviour. An increase in the stock of drugs available at facilities results in an increase in health worker ability to provide services and the number of patients who are then treated. However, an increase in the number of patients treated results in a decrease in the stock of drugs at the facility.

For more information on interpretation and best practice for drawing CLDs (naming variables, identification of loops, etc.), please see Sterman (2000c) and Tomoaia-Cotisel *et al.* (2017)

How to design a causal loop diagram study for a LMIC health system setting

What is the scope of this research?

When defining the scope of the CLD, there are three elements that need to be considered: the time frame of interest, the boundary of the issue, and the level of system aggregation (Kim, 2000). For what period of time did the policy or behaviour of interest unfold and therefore what period of time will be reflected in the CLD? What is the boundary i.e. where do we draw the line for what should be included in the diagram and what is external to it? Will the focus be on capturing community and/or facility dynamics (Rwashana *et al.*, 2014), or is the focus on district or state (Cassidy *et al.*, 2021), national (Paina *et al.*, 2014) and/or global level dynamics (Glenn *et al.*, 2020)? Relatedly, what is the level of aggregation in the CLD or level of detail needed to understand patterns of behaviour? To model the behaviour of interest, do actions and outcomes that occur on a daily, weekly, monthly or yearly basis need to be captured? When determining the scope of a CLD, the goal should always be to use CLDs to map key structural drivers for a given behaviour or problem of interest, not to try and map the feedback that drives behaviour in the entire, wider health system (Sterman, 2000a). This is key to avoiding overly complex diagrams which may obscure key dynamics around the behaviour or phenomena of interest.

The decision regarding the scope of the CLD can evolve during the process of the research, in response to discussion with stakeholders, new findings or resource availability for the project. In Cassidy *et al.* (2021), the research sought to determine constraints to achieving key service delivery targets in primary care facilities during a results-based financing programme. The time frame of interest was the duration of the programme (two years). The boundary and aggregation were informed by the research question (primary care facilities) and stakeholder experiences; their description of key events that led to their achievement or failure of targets during the intervention (at the facility, community and district-level) guided CLD development.

What data do I need to collect or source?

CLDs can be generated using a variety of data sources, including primary and secondary data; often a combination of sources is used.

Primary data

Popular primary data sources include key informant interviews (Sharma *et al.*, 2020) and group model building (GMB) sessions (Noubani *et al.*, 2020). In GMB, development of the CLD takes place with direct real-time input from stakeholders present (more on GMB in the next section) whereas with key informant interviews CLDs are developed post-hoc. The purpose of data collection is to obtain causal information on drivers for a behaviour/phenomenon of interest; this information will then be mapped out in the CLD. Stakeholders can also be asked to comment on potential leverage points within the system and actions that could be taken to alleviate problematic behaviour which can be represented in the CLD.

A recent paper has compared CLDs developed from key informant interviews to those developed through GMB

(Valcourt *et al.*, 2020). Although the CLDs developed from individual interviews yielded more variables and causal links, the CLDs produced from GMB workshops contained more feedback loops and more information on dynamic system behaviour. This was thought to be attributed to the design of GMB workshops, where stakeholders are actively encouraged to focus on feedback effects and dynamic behaviour. The decision to opt for key informant interviews versus GMB will be driven by several factors, including the availability of stakeholders, the topic under investigation (suitability for group discussion) and experience of the team. Due to global restrictions on travel during the recent COVID-19 pandemic, primary data collection has also successfully taken place through online mediums (Wilkerson *et al.*, 2020; Cassidy *et al.*, 2021).

Selection of stakeholders can be driven by researchers' own knowledge of influential actors of the system under study or inferred from the literature. Who has expert knowledge of the problem we want to investigate? Those involved in funding, policy formulation, implementation and users/beneficiaries will have varying perspectives on the system and drivers for health system behaviour. Depending on the research question, different groups may need to be consulted to create a complete picture. Key informants can also be identified via snowballing during an initial round of interviews. Many studies in the current literature incorporate a provider perspective, with fewer including, or with the sole focus on the patient experience. Examples of study design and data collection tools for patients can be found in (Rwashana *et al.*, 2014; Zablith *et al.*, 2021).

Secondary data

Where a CLD is being used to understand causal pathways or programme mechanisms ex-post; programme evaluation data can be used to support the construction of a CLD (Varghese *et al.*, 2014; Sarriot *et al.*, 2015; Cassidy *et al.*, 2021). Other secondary data such as published or grey literature (Yu *et al.*, 2018; Kurnianingtyas *et al.*, 2020), health surveys or reports (Li *et al.*, 2019) and policy documents (Nigenda *et al.*, 2015) can also be used to develop CLDs. Data extracted through a systematic or realist review can be cleaned, integrated and categorized to generate cause and effect relationships that can be represented in a CLD (Namatovu and Semwanga, 2020; Singh *et al.*, 2021).

The decision to use secondary data to develop a CLD may be driven by difficulty in accessing stakeholders for primary data collection and/or a rich source of secondary data being available and suited for CLD development (Cassidy *et al.*, 2021). Whilst secondary data might be less resource intensive to obtain, care should be taken to ensure the data contributes causal information on what is driving behaviour in the system. CLDs developed using primary data can also be triangulated with evidence from the literature and other secondary sources (Alonge *et al.*, 2017; Ahmad *et al.*, 2019). For studies where repeated access to stakeholders for CLD development is not possible (e.g. humanitarian settings), a combination of primary and secondary data sources may be preferable.

The results from a CLD developed using secondary data can be presented to stakeholders for triangulation and validation to ensure key information has been retained in the diagram (Agyepong *et al.*, 2012; Cassidy *et al.*, 2021). Stakeholder engagement encourages buy-in to the research, with higher likelihood of uptake of findings by stakeholders and policy makers (Zimmerman *et al.*, 2016).

What is my chosen method for CLD development?

There are different approaches for developing a CLD. Depending on the purpose of the research and data requirements, researchers may choose ex-post development (developing CLD from data collected/sourced) or real time development (developing the CLD with stakeholders). For further information on presentation of CLDs, see Box 3.

Ex-post development

Thematic analysis is a popular choice for extracting information that can then be used for CLD development. Deductive, inductive and blended coding (Skjott Linneberg and Korsgaard, 2019) have been used to analyse primary and secondary sources of data in preparation for CLD development. With the former method, codebooks can be developed using relevant literature, conceptual frameworks and middle range theories (Kwamie *et al.*, 2014; Xu and Mills, 2017) and used to traverse and extract variables, their relationships and linkages to be represented in the CLD. Codebooks can be updated where the researcher identifies new themes during data analysis. Deductive coding provides structure for traversing data from the outset but there is a possibility that new themes and concepts that emerge from the data might be missed.

With inductive coding, codes are derived directly from the data (Renmans *et al.*, 2017; Lembani *et al.*, 2018); codes or categories can be iteratively refined, and data reanalysed. Inductive coding is a suitable choice, where there is a lack of theoretical background to the research topic (Skjott Linneberg and Korsgaard, 2019). In practice, blended coding is often used to harness the strengths of each approach (Elliott, 2018).

Purposive Text Analysis is another option for analysing data and extracting information for CLD development (Kim and Andersen, 2012). This approach involves systematically reviewing key informant transcripts, extracting quotations that describe drivers for behaviour of interest, and extraction of cause-and-effect statements, with diagrams that represent these relationships. Cassidy *et al.* (2021) used this approach to develop their CLD (Box 4) and a method called CLD Combination (Tomoaia-Cotisel, 2018) to systematically merge together key informant CLDs into a single CLD. This approach involves ordering key informant CLDs in order of their 'complexity' (number of links, variables and delays). The most complex and second most complex CLD are compared. Additions are made to the most complex CLD where new information about system behaviour is revealed. Key informant CLDs are continually compared to this 'anchor' CLD until information from all CLDs are represented in one CLD.

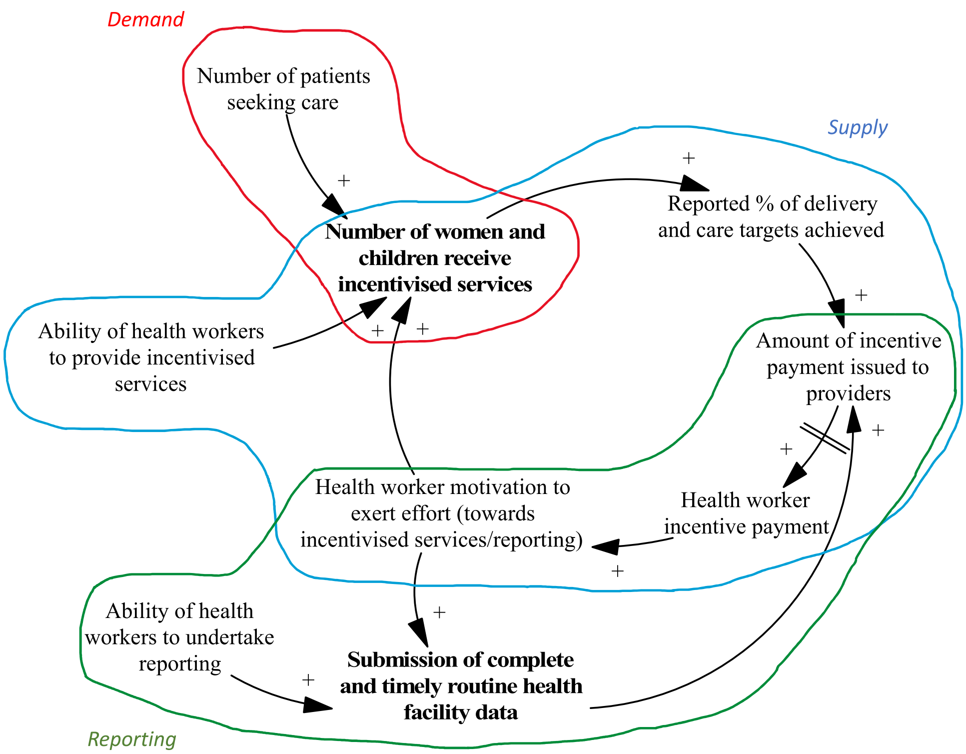
Real time development

For real time development of CLDs, GMB is a popular choice. Scripts are freely available that can help researchers guide GMB sessions (Hovmand *et al.*, 2011), hosted on the Scriptapedia website (Wikibooks Contributors, 2022). There are a range of activities that can be undertaken in GMB sessions depending on the purpose of the workshop; examples include encouraging stakeholders to discuss and list variables they think are driving a system process 'Variable Elicitation script'

Box 3. Tips for presentation of CLD

Software There are software packages specifically developed for creating CLDs such as Vensim (Ventana Systems Inc. 2015) and STELLA (Isee Systems Inc. 2021). Each have different licenses available to purchase depending on the functionality needed by the user. At the time of writing, Vensim offers a free personal learning edition for educational use.

Presentation For ease of viewing, analysis and validation, it can be helpful to develop multiple CLDs or present the CLD in smaller segments (Sterman 2000b). Multiple CLDs can be used to demonstrate the policy effect and emergent behaviour (Paina *et al.* 2014; Renmans *et al.* 2017) and shifting community perspective on vaccine acceptance (Varghese *et al.* 2014) in the system at different time steps. A single CLD can also be presented in smaller segments, for example, key mechanisms related to the supply, demand and reporting of healthcare services (Rwashana *et al.* 2014; Cassidy *et al.* 2021); perception, causes and health seeking practices related to mental health (Noubani *et al.* 2020) can be presented in segments (that are connected in the wider, whole CLD). An example of how to illustrate where these segments connect in the wider CLD is given here (Cassidy *et al.* 2021).



Notes to Box: High-level snapshot of how three smaller diagrams presented in the paper fit together in the larger CLD is given. Three main mechanisms responsible for provider achievement of (or failure to reach) targets during payment for performance programmes are shown here. Mechanisms that result in changes in the supply of services (blue), mechanisms that result in changes in facility reporting (green) and mechanisms that result in changes in demand for services (red). Source: Cassidy *et al.* (2021).

and developing a CLD as a group exercise 'Initiating and Elaborating a Causal Loop Diagram script' (Trani *et al.*, 2016; Noubani *et al.*, 2020; Wikibooks contributors, 2022).

A combination of methods can also be used to develop CLDs. For example, researchers may start by developing an initial CLD from secondary data or prospective interviews and then use a GMB workshop to develop a final CLD (Lembani *et al.*, 2018; Jamal *et al.*, 2020). Alternatively, CLDs can be initially developed through GMB sessions before triangulating the results with thematic analysis of subsequent key informant interviews (Zablith *et al.*, 2021). Triangulating the results with data sources or presentation and discussion of the CLD with stakeholders lends weight to the validity of the CLD to represent real health system behaviour (see validation).

How will I validate the CLD?

The developed CLD needs to be validated to minimize any unconscious bias that may have been introduced by the researcher during development or misinterpretation of data. Stakeholder dialogue is the most popular method to validate CLDs in the LMIC health literature, with illustrative tools provided in Cassidy *et al.* (2021) and Rwashana *et al.* (2014). The decision to approach stakeholders will be dependent on accessibility to stakeholders and the nature of the topic under investigation.

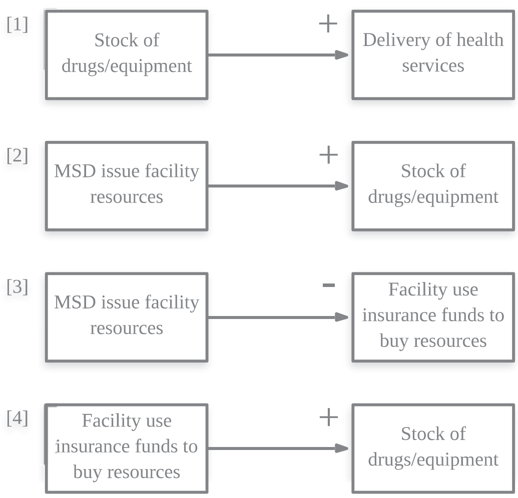
Other examples of validation using primary sources of data include comparison of CLD structure to key informant interview transcripts or the original primary data source used for CLD development (Xu and Mills, 2017; Zablith *et al.*,

Box 4. Example of applying Purposive Text Analysis to text

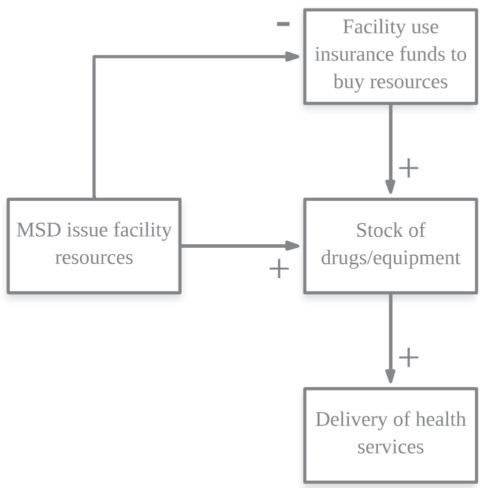
(1a) Question: Are there any strategies being implemented that aim to address these challenges (to provision of quality health services)?
(1b) Quotation: 'Yes, there is strategy done in the district, which is community health fund. We realized that the shortage of equipments and drugs was becoming a common problem which resulted in poor health service delivery [1], the community health fund was established as alternative to solve those problems. So once the government supply insufficient medicine [2] the community health fund money are used to substitute [3/4]'.
Main argument: When the Medical Stores Department (autonomous government department that procures and distributes health commodities to facilities, MSD) cannot provide drugs and equipment, facilities must draw on other sources of funding like the community health fund (community-based health insurance scheme) to buy medical commodities.
(1c) Causal structure:

[1]	*Causal variable* Stock of drugs/equipment	*Relationship* Increase	*Effect variable* Delivery of health services
[2]	*Causal variable* MSD issue facility resources	*Relationship* Increase	*Effect variable* Stock of drugs/equipment
[3]	*Causal variable* MSD issue facility resources	*Relationship* Decrease	*Effect variable* Facility use insurance funds to buy resources
[4]	*Causal variable* Facility use insurance funds to buy resources	*Relationship* Increase	*Effect variable* Stock of drugs/equipment

(1d) Causal structure diagrams



(1e) Part of a larger causal loop diagram



Notes to Box: In the example, the interviewer asked the stakeholder how health providers addressed challenges to the provision of quality health care in their facilities (1a) during a payment for performance programme. Quotations were deemed relevant and extracted if they described events or scenarios that furthered understanding of how stakeholders responded to the programme or demonstrated health system behaviour that facilitated or hindered facilities delivering quality health care (1b). Isolated cause and effect statements, with their associated quotations were extracted from transcripts and stored in an Excel file. The direction of the relationship (positive or negative, see Box 1 for details on interpretation of CLDs) was also noted; in the given example, an increase in the stock of drugs and equipment at facilities resulted in providers being able to deliver health services (1c).

At the end of this data extraction process, all cause and effect statements were drawn as simple diagrams with a polarity indicating the direction of the relationship (1d). Each of these simple diagrams were then combined to form a single CLD representative of an individual's mental model of the system (1e). Stakeholders may not use the same terminology in relaying information; as coding progresses, it becomes easier to standardise variable names assigned to cause and effect statements. Medical Stores Department (MSD).

2021) and multiple group model building sessions to validate structure (Trani *et al.*, 2016). Secondary sources of data can also be used to validate the CLD, with CLD structure compared to findings in published or grey literature (Alonge *et al.*, 2017; Ahmad *et al.*, 2019), organization reports or policy documents (Paina *et al.*, 2014; Jamal *et al.*, 2020).

It is recommended that for analysis and validation, large CLD structures are broken down into smaller segments (Sterman, 2000b). Cassidy *et al.* (2021) initially split the CLD into three smaller diagrams, related to three broad mechanisms responsible for facility achievement of targets during a payment for performance programme and presented these

Box 5. Extract of the causal loop diagram validation tool to guide interviews with stakeholders. Original tool adapted from Rwashana *et al.* (2014) and Andersen *et al.* (2012). Source: Cassidy *et al.* (2021), adapted with permission

The interviewer does not have to explicitly run through these questions while discussing the diagram, can instead probe 'Does this make sense? Are we missing anything important in this section of the diagram? Is there anything that you feel should be removed in the diagram?'. When an interviewee gives their feedback, it will generally fall into these compartments and help the modeller to go back and make modifications to the diagram:

- Does this part of the system exist to your knowledge?
- Are appropriate system variables represented? If not, what variables are missing or should be removed?
- Are appropriate in- and outflows represented? If not, what flows are missing or should be removed?
- Is the polarity of in- and outflows accurately represented? If not, what changes would you make?
- Are appropriate delays in the system represented? If not, what delays are missing or should be removed?

individual segments to stakeholders for validation. This also allowed presentation of parts of the CLD to stakeholders with knowledge of that sector (rather than presenting the entire CLD for validation). However, initial stakeholder feedback indicated that they were interested in seeing how this smaller segment fed into the wider CLD. The research team felt this was an important issue—in presenting a single segment of the CLD, knowledge of how that segment operates within the wider CLD structure is lost and stakeholders are unable to see the 'bigger picture'. In future interviews, stakeholders were still asked to comment and provide feedback on one of the three mechanisms, but the mechanism was now highlighted in the wider CLD. An extract of the validation tool used in this study is shown in Box 5, where stakeholders were verbally taken round the CLD to elicit their feedback.

Conclusion

CLDs are a valuable tool for research or decision making, enabling consideration of problem behaviour, its drivers, and potential health systems policies or interventions as part of a wider, dynamic system. CLDs can identify bottlenecks and leverage points, areas where it would be opportune to intervene to produce optimal system behaviour. They can also be used as direct input to other research tools [e.g. to develop a system dynamics model (Pruyt, 2017)] or complement other research methods [such as realist reviews (Singh *et al.*, 2021) or case studies (Jamal *et al.* 2020)]. Increased familiarity and understanding on how to use systems thinking tools, strengthened science-policy partnerships and dissemination of findings to appropriate audiences are essential to ensure their application to evaluate complex health system behaviour and use of findings (Kwamie *et al.* 2021).

Funding

This research is funded by the Medical Research Council under the Health Systems Research Initiative grant (MR/R013454/1).

Authorship statement

Conception or design of the work: Rachel Cassidy, Karl Blanchet, Josephine Borghi, Neha S. Singh
Drafting the article: Rachel Cassidy
Critical revision of the article: All authors
Final approval of the version to be submitted: All authors

Reflexivity statement

Article authors consist of four female and two male researchers, with both early career (RC) and senior researchers. Two authors (ARS and PB) are from the article study region.

Ethical approval. This study received a favourable ethical opinion from the Observational/Interventions Research Ethics Committee at The London School of Hygiene and Tropical Medicine (LSHTM Ethics Ref: 16139 – 3), the Institutional Review Board at Ifakara Health Institute (IHI/IRB/No:17 -2021) and National Institute for Medical Research (NIMR/HQ/R.8a/Vol. IX/3154) in Tanzania.

Conflict of interest statement. None declared.

References

Adam T. 2014. Advancing the application of systems thinking in health. *Health Research Policy and Systems* **12**: 50.

Adam T, de Savigny D. 2012. Systems thinking for strengthening health systems in LMICs: need for a paradigm shift. *Health Policy and Planning* **27**: iv1–3.

Agyepong IA, Kodua A, Adjei S *et al.* 2012. When "solutions of yesterday become problems of today": crisis-ridden decision making in a complex adaptive system (CAS) - the additional duty hours allowance in Ghana. Special Issue System Thinking for Health Systems Strengthening LMICs Seizing Opportunity. *Health Policy and Planning* **27**: iv20–31.

Ahmad R, Zhu NJ, Lebcir RM *et al.* 2019. How the health-seeking behaviour of pregnant women affects neonatal outcomes: findings of system dynamics modelling in Pakistan. *BMJ Global Health* **4**: e001242.

Alonge O, Lin S, Igusa T *et al.* 2017. Improving health systems performance in low- and middle-income countries: a system dynamics model of the pay-for-performance initiative in Afghanistan. *Health Policy and Planning* **32**: 1417–26.

Andersen DL, Luna-Reyes LF, Diker VG *et al.* 2012. The disconfirmatory interview as a strategy for the assessment of system dynamics models. *System Dynamics Review* **28**: 255–75.

Barasa EW, Cloete K, Gilson L. 2017. From bouncing back, to nurturing emergence: reframing the concept of resilience in health systems strengthening. *Health Policy and Planning* **32**: iii91–4.

Borghi J, Chalabi Z. 2017. Square peg in a round hole: re-thinking our approach to evaluating health system strengthening in low-income and middle-income countries. *BMJ Global Health* **2**: e000406.

Cassidy R, Singh NS, Schiratti P-R *et al.* 2019. Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models. *BMC Health Services Research* **19**: 845.

Cassidy R, Tomoaia-Cotisel A, Semwanga AR *et al.* 2021. Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach. *Social Science and Medicine* **285**: 114277.

Elliott V. 2018. Thinking about the Coding Process in Qualitative Data Analysis. *The Qualitative Report* **23**: 2850–61.

Gates EF. 2016. Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Evaluation and Program Planning* **59**: 62–73.

Glenn J, Kamara K, Umar ZA *et al.* 2020. Applied systems thinking: a viable approach to identify leverage points for accelerating progress towards ending neglected tropical diseases. *Health Research Policy and Systems* **18**: 1–15.

Hovmand P, Rouwette E, Andersen D *et al.* 2011. *Scriptapedia: A handbook of scripts for developing structured group model building sessions*. *29th International Conference of the System Dynamics Society* Washington, DC.

Isee Systems Inc. 2021. STELLA Architect v2.1.4.

Jamal Z, Alameddine M, Diaconu K *et al.* 2020. Health system resilience in the face of crisis: analysing the challenges, strategies and capacities for UNRWA in Syria. *Health Policy and Planning* **35**: 26–35.

Kanniyan B, Govindarajan PK, Haveri SP *et al.* 2021. Enactment of system thinking framework for convolution of immunization service governance. *International Journal of Research in Pharmaceutical Sciences* **12**: 849–56.

Kim D. 2000. Part II: Dynamic thinking tools. In: O'Reilly K (ed). *Systems Thinking Tools: A User's Reference Guide*. Waltham: Pegasus Communications, Inc., 18.

Kim H, Andersen DF. 2012. Building confidence in causal maps generated from purposive text data: mapping transcripts of the Federal Reserve. *System Dynamics Review* **28**: 311–28.

Kurnianingtyas D, Santosa B, Siswanto N. 2020. *A system dynamics for financial strategy model assessment in national health insurance system*. *Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering*. New York: ACM, 44–8.

Kwamie A, Ha S, Ghaffar A. 2021. Applied systems thinking: unlocking theory, evidence and practice for health policy and systems research. *Health Policy and Planning* **36**: 1715–7.

Kwamie A, van Dijk H, Agyepong IA. 2014. Advancing the application of systems thinking in health: realist evaluation of the Leadership Development Programme for district manager decision-making in Ghana. *Health Research Policy and Systems* **12**: 1–12.

Lembani M, de Pinho H, Delobelle P *et al.* 2018. Understanding key drivers of performance in the provision of maternal health services in eastern cape, South Africa: a systems analysis using group model building. *BMC Health Services Research* **18**: 912.

Li M, Yu W, Tian W *et al.* 2019. System dynamics modelling of public health services provided by China CDC to control infectious and endemic diseases in China. *Infection and Drug Resistance* **12**: 613–25.

Lipsitz LA 2012. Understanding health care as a complex ystem. *JAMA* **308**: 243.

McGill E, Er V, Penney T *et al.* 2021. Evaluation of public health interventions from a complex systems perspective: a research methods review. *Social Science & Medicine* **272**: 113697.

Namatovu HK, Semwanga AR. 2020. A systems dynamics approach to understanding the determinants of antenatal care utilization in low-and middle-income countries. *International Journal of System Dynamic Applications* **9**: 111–28.

Nigenda G, González-Robledo LM, Juárez-Ramírez C *et al.* 2015. Understanding the dynamics of the Seguro Popular de Salud policy implementation in Mexico from a complex adaptive systems perspective. *Implementation Science* **11**: 68.

Noubani A, Diaconu K, Ghandour L *et al.* 2020. A community–based system dynamics approach for understanding factors affecting mental health and health seeking behaviours in Beirut and Beqaa regions of Lebanon. *Globalization and Health* **16**: 28.

Ozawa S, Paina L, Qiu M 2016. Exploring pathways for building trust in vaccination and strengthening health system resilience. *BMC Health Services Research* **16**: 639.

Paina L, Bennett S, Ssengooba F *et al.* 2014. Advancing the application of systems thinking in health: exploring dual practice and its management in Kampala, Uganda. *Health Research Policy and Systems* **12**: 41.

Peters DH 2014. The application of systems thinking in health: why use systems thinking?. *Health Research Policy and Systems* **12**: 51.

Pruyt E. 2017. Systems dynamics: a tool for modelling and testing solutions. In: de Savigny D, Blanchet K, Adam T (eds). *Applied Systems Thinking for Health Systems Research: A Methodological Handbook*. London, England: McGraw-Hill Education, pp. 173–94.

Renmans D, Holvoet N, Criel B. 2017. Combining theory-driven evaluation and causal loop diagramming for opening the 'Black Box' of an intervention in the health sector: a case of performance-based financing in Western Uganda. *International Journal of Environmental Research and Public Health* **14**: 1–20.

Rwashana AS, Nakubulwa S, Nakakeeto-Kijjambu M *et al.* 2014. Advancing the application of systems thinking in health: understanding the dynamics of neonatal mortality in Uganda. *Health Research Policy and Systems* **12**: 36.

Rwashana AS, Williams DW, Neema S. 2009. System dynamics approach to immunization healthcare issues in developing countries: a case study of Uganda. *Health Informatics Journal* **15**: 95–107.

Sarriot E, Morrow M, Langston A *et al.* 2015. A causal loop analysis of the sustainability of integrated community case management in Rwanda. *Social Science & Medicine* **131**: 147–55.

Sharma SR, Matheson A, Lambrick D *et al.* 2020. The role of tobacco and alcohol use in the interaction of social determinants of non-communicable diseases in Nepal: a systems perspective. *BMC Public Health* **20**: 1368.

Singh NS, Kovacs RJ, Cassidy R *et al.* 2021. A realist review to assess for whom, under what conditions and how pay for performance programmes work in low- and middle-income countries. *Social Science and Medicine* **270**: 113624.

Skivington K, Matthews L, Simpson SA. *et al.* 2021. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ* 374: n2061.

Skjott Linneberg M, Korsgaard S. 2019. Coding qualitative data: a synthesis guiding the novice. *Qualitative Research Journal* **19**: 259–70.

Sterman J 2000a. Part 1: Perspective and process. In: *Business Dynamics: Systems Thinking and Modelling for a Complex World*. Illinois, U.S.A.: McGraw-Hill Companies Inc, 22.

Sterman J. 2000b. Part II: Tools for systems thinking. In: *Business Dynamics: Systems Thinking and Modelling for a Complex World*. Illinois, U.S.A.: McGraw-Hill Companies Inc, 154.

Sterman J. 2000c. Part II: Tools for systems thinking. In: *Business Dynamics: Systems Thinking and Modelling for a Complex World*. Illinois, U.S.A.: McGraw-Hill Companies Inc, 137–41.

Tomoaia-Cotisel A. 2018. The journey towards the patient-centered medical home: a grounded, dynamic theory of primary care transformation.

Tomoaia-Cotisel A, Hyunjung K, Allen S *et al.* 2017. Causal loop diagrams: a tool for visualizing emergent system behaviour. In: de Savigny D, Blanchet K, Adam T. (eds). *Applied Systems Thinking for Health Systems Research: A Methodological Handbook*. London, England: McGraw-Hill Education, pp. 97–114.

Trani JF, Ballard E, Bakhshi P *et al.* 2016. Community based system dynamic as an approach for understanding and acting on messy problems: a case study for global mental health intervention in Afghanistan. *Conflict and Health* **10**: 1–11

Valcourt N, Walters J, Javernick-Will A *et al.* 2020. Assessing the efficacy of group model building workshops in an applied setting through purposive text analysis. *System Dynamics Review* **36**: 135–57.

Varghese J, Kutty VR, Paina L *et al.* 2014. Advancing the application of systems thinking in health: understanding the growing complexity governing immunization services in Kerala, India. *Health Research Policy and Systems* **12**: 1–12.

Ventana Systems Inc. 2015. Vensim PLE Plus.

Wikibooks Contributors. 2022. "*Scriptapedia*." Wikibooks, Free Textb Proj.

Wilkerson B, Aguiar A, Gkini C *et al.* 2020. Reflections on adapting group model building scripts into online workshops. *System Dynamics Review* **36**: 358–72.

Xu J, Mills A. 2017. Challenges for gatekeeping: a qualitative systems analysis of a pilot in rural China. *International Journal for Equity in Health* **16**: 106.

Yu W, Lv Y, Hu C *et al.* 2018. Research of an emergency medical system for mass casualty incidents in Shanghai, China: a system dynamics model. *Patient Preference and Adherence* **12**: 207–22.

Zablith N, Diaconu K, Naja F *et al.* 2021. Dynamics of non-communicable disease prevention, diagnosis and control in Lebanon, a fragile setting. *Conflict and Health* **15**: 4.

Zimmerman L, Lounsbury DW, Rosen CS *et al.* 2016. Participatory system dynamics modelling: increasing stakeholder engagement and precision to improve implementation planning in systems. *Administration and Policy in Mental Health and Mental Health Services Research* **43**: 834–49.

8. System dynamics model

8.1 Introduction

Abbreviations: 2 doses of intermittent preventative treatment (IPT2); Antiretroviral therapy (ART); Antenatal care (ANC); Causal loop diagram (CLD);Community health fund (CHF); Community health worker (CHW); Council Health Management Team (CHMT); District Executive Director (DED); District Health Information Software (DHIS); Health Facility Governing Committee (HFGC); Health Management Information System (HMIS); Human Immunodeficiency Virus (HIV); Low- and middle-income country (LMIC); Maternal and child health (MCH); Medical Stores Department (MSD); Ministry of Health and Social Welfare (MoHSW); Payment for performance (P4P); Regional Health Management Team (RHMT); System dynamics model (SDM).

In Chapter 8, the results from analysis of the SDM are presented, fulfilling **Objective 4** of this thesis:

> 4. Explore how variations in the implementation, design and context of P4P could result in different outcomes to inform future design of P4P programmes.

The results of the study are presented in a paper, 'Using mathematical modelling to identify the active ingredients in payment for performance programmes', which will shortly be submitted for publication.

The chapter appendix contains supplementary material, including detailed views of model sectors, model equations and data, the study interview tool and further analysis (model calibration and sensitivity analysis results).

8.2 Research paper 4: Using mathematical modelling to identify the active ingredients in payment for performance programmes

(Cover sheet on next page)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed <u>for each</u> research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1401642	**Title**	Miss
First Name(s)	Rachel		
Surname/Family Name	Cassidy		
Thesis Title	Using systems thinking to optimise health system interventions for improved maternal and child health in low-resource settings		
Primary Supervisor	Prof. Josephine Borghi		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Social Science & Medicine
Please list the paper's authors in the intended authorship order:	Rachel Cassidy, Agnes Rwashana Semwanga, Peter Binyaruka, Karl Blanchet, Neha S. Singh, John Maiba, Josephine Borghi
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Conceptualised study with PhD supervisors, collected new data to validate model, led analysis of secondary and newly collected data, developed simulation model. Wrote first draft of the manuscript and led revisions with input from co-authors.

SECTION E

Student Signature	[redacted]
Date	05/12/22

Supervisor Signature	[redacted]
Date	05/12/22

Title: Using mathematical modelling to identify the active ingredients in payment for performance programmes

Authors: Rachel Cassidy*[a], Agnes Rwashana Semwanga[b], Peter Binyaruka[c], Karl Blanchet[d], Neha S. Singh[a], John Maiba[c], Josephine Borghi[a]

**Corresponding author, rachel.cassidy@lshtm.ac.uk, +44 (0)20 7927 2818*

Author institution address:

[a]Department of Global Health and Development, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London, WC1H 9SH, UK

[b]Information Systems Department, College of Computing and Information Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda

[c]Ifakara Health Institute, PO Box 78373, Dar es Salaam, Tanzania

[d]Geneva Centre of Humanitarian Studies, University of Geneva and the Graduate Institute, Rue Rothschild 22, 1211 Genève, Switzerland

Abstract: Payment for performance (P4P) is not a uniform intervention, with programme effect dependent on several variables, including programme design, implementation and context. The aim of this study was to develop a system dynamics model (SDM) to explore the pathways to improved outcomes and how changes in the design, implementation and context of a P4P programme affected maternal and child health (MCH) service delivery outcomes in Tanzania. The Tanzania P4P programme that is the focus of this study aimed to improve the coverage and delivery of MCH services through financial incentives for health workers and managers. Healthcare targets for primary care providers were to improve content of care (such as percentage of women who receive intermittent preventative treatment (IPT) as part of antenatal care (ANC)) and coverage of services (facility-based deliveries) for primary care providers. A previously developed causal loop diagram of the programme effects was used as a framework to inform model development, with both primary and secondary data sources (including an impact evaluation of programme, country and district-level health surveys, stakeholder feedback and evidence drawn from the literature) used to build the model.

A number of pathways to improved MCH services under P4P were identified, with increased availability of drugs underpinning the IPT during ANC outcome, which together with increased supervision, enhanced health worker motivation. This in turn increased perceived quality of care at the facility which improved facility-based deliveries, and with increased outreach, increased

awareness of services also boosted demand. Where there were delays in payment, provider capacity to achieve targets for the modelled content of care indicator was severely limited and there was reduced provider purchasing power for medicines. With severe payment delays, there was erosion of provider trust in the programme and reduced motivation for programme participation. Increasing the share of funds for facility operations relative to health worker incentives can enhance performance effects, particularly for those targeted services that rely on efficient drug administration. Context also shapes programme effect, with limited baseline provision of essential medications, lower community awareness of facility services and dispersed/distant populations limiting programme effect. The feasibility of building a SDM is demonstrated, highlighting the potential of such models to inform the design of effective health system interventions.

1. Introduction

Payment for performance (P4P) has been implemented in many low- and middle-income countries (LMICs) to improve the quality and coverage of maternal and child health (MCH) services (Das *et al.* 2016). P4P encourages achievement of pre-defined indicators through the provision of incentives to health workers and managers for performance attained (Mannion and Davies 2008; Kovacs *et al.* 2020). A recently updated Cochrane review found positive effects of P4P on certain indicators (such as child mortality, quality of child healthcare, medicine availability) and mixed effects on other indicators (including MCH vaccinations, neonatal mortality, and antenatal care (ANC) utilisation) in articles which compare P4P to a status quo control group (Diaconu *et al.* 2021). The conclusions drawn from the review intimate that P4P is not a uniform intervention, with programme effect dependent on several variables, including programme design and context within which the programme is implemented.

P4P is a complex intervention acting to influence a complex system, the health system (Lipsitz 2012; Diaconu *et al.* 2022). The properties of the intervention itself are complex; implementers have autonomy in how they respond to and tailor the intervention to their local context, and require expertise, many individuals and groups (cadres of health worker, informal care providers, managers etc.) and service delivery indicators are targeted, and data reporting and measurements are required for performance evaluation (Skivington *et al.* 2021). The health system in which the intervention is implemented is also complex, exhibiting dynamic, non-linear, emergent behaviour that changes over time in response to numerous stimuli (Paina and Peters 2012). The relationship between the intervention and health system give rise to an additional layer of complexity; the mechanisms through which the intervention aims to change health system behaviour, and the context in which the intervention is implemented will influence success or failure (Skivington *et al.* 2021). Tracing the mechanism for impact in response to an intervention is difficult using conventional methods for evaluation that assume linear cause effect relationships and do not account for complexity within the analysis (Borghi and Chalabi 2017). Furthermore, implementation of a complex intervention can result in unexpected or paradoxical behaviour with suboptimal outcomes as a result of discounting system complexity (Adam and de Savigny 2012; Paina and Peters 2012).

Systems thinking methods can be employed to explore the mechanisms through which complex interventions act to influence complex systems (such as the health system), and to better understand what works in a given context. Systems thinking is an umbrella term used to describe a range of tools that can be used for health systems research (Peters 2014), where system complexity is retained in the analysis. The choice of systems thinking tool depends on the research question (de Savigny *et al.* 2017). For example, causal loop diagrams (CLDs) can be used to identify and visualise drivers of health system behaviour and pathways to impact for interventions on key health and system outcomes (Baugh Littlejohns *et al.* 2018; Sahin *et al.* 2020; Cassidy *et al.* 2022). CLDs can be used to identify system bottlenecks, catalytic variables (those that have wide spread impact on the rest of the system and should be carefully considered in the design of interventions) and system levers (variables not currently targeted by an intervention but could be incorporated to maximise impact) (Rwashana *et al.* 2014; Cassidy *et al.* 2021). If there is interest in investigating how the behaviour of the system changes over time in response to new interventions or changes in context and quantifying the effects of such changes on health system outcomes, quantitative system dynamics modelling (SDM) is required (Pruyt 2017). SDMs can also explore the effects of potential changes in intervention design on health or service delivery outcomes, to determine how programme effects could be maximised. When developed with a user-friendly interface, SDM can be used as a tool to guide policy and support dialogue between stakeholders and researchers (Semwanga *et al.* 2016). CLDs can be used to develop SDMs, providing a blueprint of dynamic drivers for behaviour that can inform model structure (Pruyt 2017).

Use of CLDs and SDMs to explore health system behaviour is on the rise (Currie *et al.* 2018; Cassidy *et al.* 2019, 2022; Darabi and Hosseinichimeh 2020). To our knowledge, five studies have used CLD to study the effect of P4P on health and service outcomes (Meker and Barlas 2015; Alonge *et al.* 2017; Renmans *et al.* 2017; Cassidy *et al.* 2021; Singh *et al.* 2021). However, only two studies also developed a simulation model (SDM) to explore the effect of P4P on health systems, in Afghanistan (Alonge *et al.* 2017) and Turkey (Meker and Barlas 2015). In a recent systematic review on application of SDM for health systems research (Cassidy *et al.* 2019), nine articles described simulation of health system behaviour in LMIC settings. Specific to policy evaluation, and in addition to those already mentioned that model

P4P, SDM was used to explore policies that would alleviate delays in care for serious heart events in Brazil (Andrade *et al.* 2014), policies for optimisation of healthcare waste management in Turkey (Ciplak and Barton 2012) and Indonesia (Chaerul *et al.* 2008), and interventions to reduce neonatal mortality in Uganda (Semwanga *et al.* 2016). Given the resource constraints facing many LMIC, there is urgent need for further use of SDM to study health system reforms, such as P4P, in these settings.

The aim of this study was to develop a SDM to explore the mechanisms for impact within a P4P programme, and examine how changes in programme design, implementation and context affect MCH service delivery outcomes in a LMIC setting, Tanzania. This study uses a previously documented CLD of a P4P programme in Tanzania (Cassidy *et al.* 2021) to inform the development of a SDM.

2. Methods

2.1 Study setting

The P4P programme in Pwani region of Tanzania is described in detail elsewhere (Binyaruka *et al.* 2015; Borghi *et al.* 2021), but a summary is provided here. The programme was introduced by the Ministry of Health and Social Welfare in 2011, with funding from the Norwegian Ministry of Foreign Affairs. The programme aimed to improve the coverage and delivery of MCH services through financial incentives for health providers, district and regional managers based on targets achieved. For health providers, targets were aimed at improving the coverage of services (such as percentage of facility-based deliveries), content of care (such as percentage of women who receive IPT as part of ANC) and data reporting practices (Supplementary File 1). Performance was measured every 6 months. To be eligible for a bonus payment, providers needed to either improve by a specified amount in relation to previous performance or achieve an absolute amount of service coverage. For primary health care providers (health centres and dispensaries), 75% of the incentive payment was to be split between staff at the facility, with the remaining 25% to be used to improve facility operations (e.g., purchasing additional medicine where needed). District and regional level managers were responsible for supporting facilities and were also eligible to receive incentives based on the performance of facilities within their district/region (Supplementary File 1).

2.2 Payment for performance to improve health outcomes and service delivery in low- and middle-income settings

Economic and psychological theories are predominantly used to justify utilisation of P4P to improve health and service delivery outcomes, most often the principal agent-theory to enact positive system change (Paul *et al.* 2021). The agency theory stipulates that one organisation (the principal, e.g. Ministry of Health or donor group) issues work to another individual or collective (the agent, e.g. health provider), with compensation then paid to the agent for the work (Sekwat 2000; Lohmann *et al.* 2016). Through use of financial rewards or penalties to the agent, P4P programmes are designed to enhance health provider motivation and minimise inherent goal conflict, where the agent and principal's interests and priorities are not aligned.

There are several potential unintended effects of programmes that use rewards-based systems to induce positive behaviour and outcomes (Miller and Singer Babiarz 2013). Positive unintended consequences of programme implementation include improvements to other services not directly targeted by the programme, 'spill over effects' (Sherry *et al.* 2017) and improvements in overall patient satisfaction of quality of care and patient-provider interactions (Diaconu *et al.* 2021).

'Tunnel vision' is an example of a negative unintended consequence of the programme, where health providers who are required to carry out multiple tasks may shift focus and effort away from non-incentivised activities (Holmstrom and Milgrom 1991; Aryankhesal *et al.* 2015). With programme designs that have different reward amounts for each targeted service, there is a risk that health providers might focus on performing tasks that have the highest marginal return (Sherry *et al.* 2017). Where programmes stipulate the same reward for different services, this can also lead to focus of effort towards the easier to accomplish targets (Lagarde *et al.* 2013); both programme designs can lead to cherry picking of patients to improve health provider performance and boost incentive payments.

Crowding out of intrinsic motivation is also a concern, where a rewards-based system appeals to health providers extrinsic rather than intrinsic motivation to conduct service provision; externally controlled motivation (extrinsic) is thought to be more unstable and prone to changes in one's environment compared to internal (intrinsic) motivation (Deci and Ryan 2000; Lohmann *et al.* 2016). With a rewards system that requires submission of data, there is a risk that data may be misreported or distorted to conceal true performance (Kalk *et al.* 2010; Aryankhesal *et al.* 2015; Turcotte-Tremblay *et al.* 2020), although there is limited evidence within health on the incidence of gaming, with further uncertainty garnered by lack of knowledge on gaming in non-P4P health providers for comparison purposes (Van Herck *et al.* 2010). There are certain programme design features that aim to mitigate undesirable unintended consequences of P4P, for example, implementing gaming safeguarding measures such as auditing or introduction of penalties (Kovacs *et al.* 2020).

Pathways to impact specific to P4P programme in Pwani, Tanzania

The theory of change describing the expected pathways to impact for the P4P programme in Tanzania are described elsewhere (Anselmi *et al.* 2017) but summarised here and represented in the previously developed CLD (Cassidy *et al.* 2021*)* (Figure 1 and 2). During programme implementation, financial incentives to health workers were expected to induce a more motivated workforce (Figure 2b, Note 1), with investment in training expected to increase knowledge levels of health providers. Strengthened supervision of health workers by district level managers was thought to increase health worker motivation (Figure 2b, Note 2), with strengthened supervision as result of increased verification activities (Figure 2c, Note 3). Financial incentives were also to be used for increasing the availability of resources at the facility (including medicine) (Figure 2b, Note 4), with increased availability and utility of resources expected to increase health worker motivation (Figure 2b, Note 5) and reduce the costs of care for patients. Where health workers were feeling further motivated, this was expected to translate into improved patient-provider interactions (Figure 2a, Note 6) and increased mobility for outreach activities (Figure 2a, Note 7). Health providers including the Health Facility Governing Committee (HFGC) encouraged community enrolment in a voluntary insurance scheme to increase funding available to the facility (Figure 2a, Note 8); patients were more likely to engage with the scheme (Figure 2a, Note 9) and seek facility services with the improvements to quality of care (Figure 2a, Note 10) as a result of successful programme implementation.

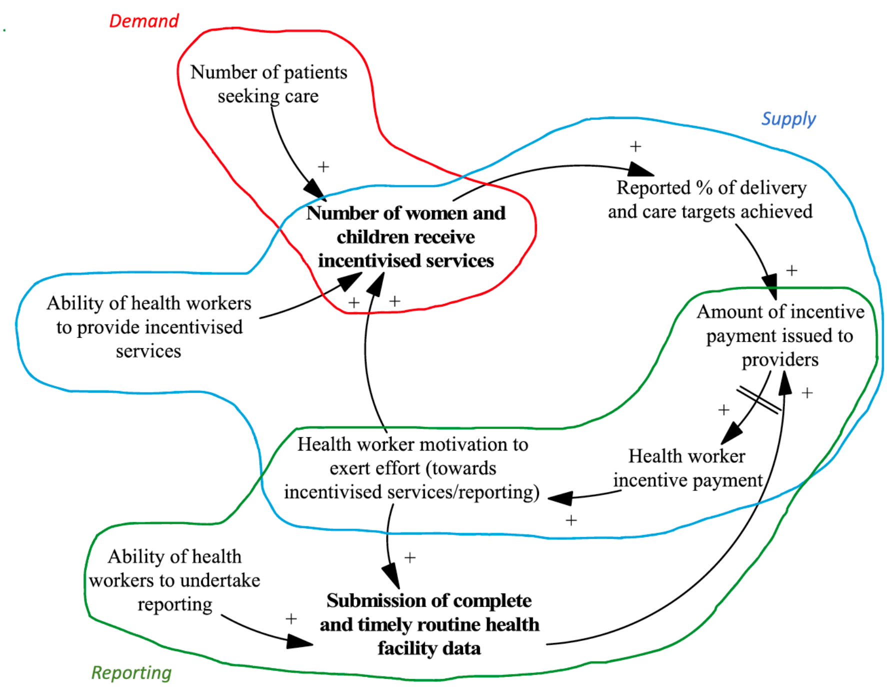
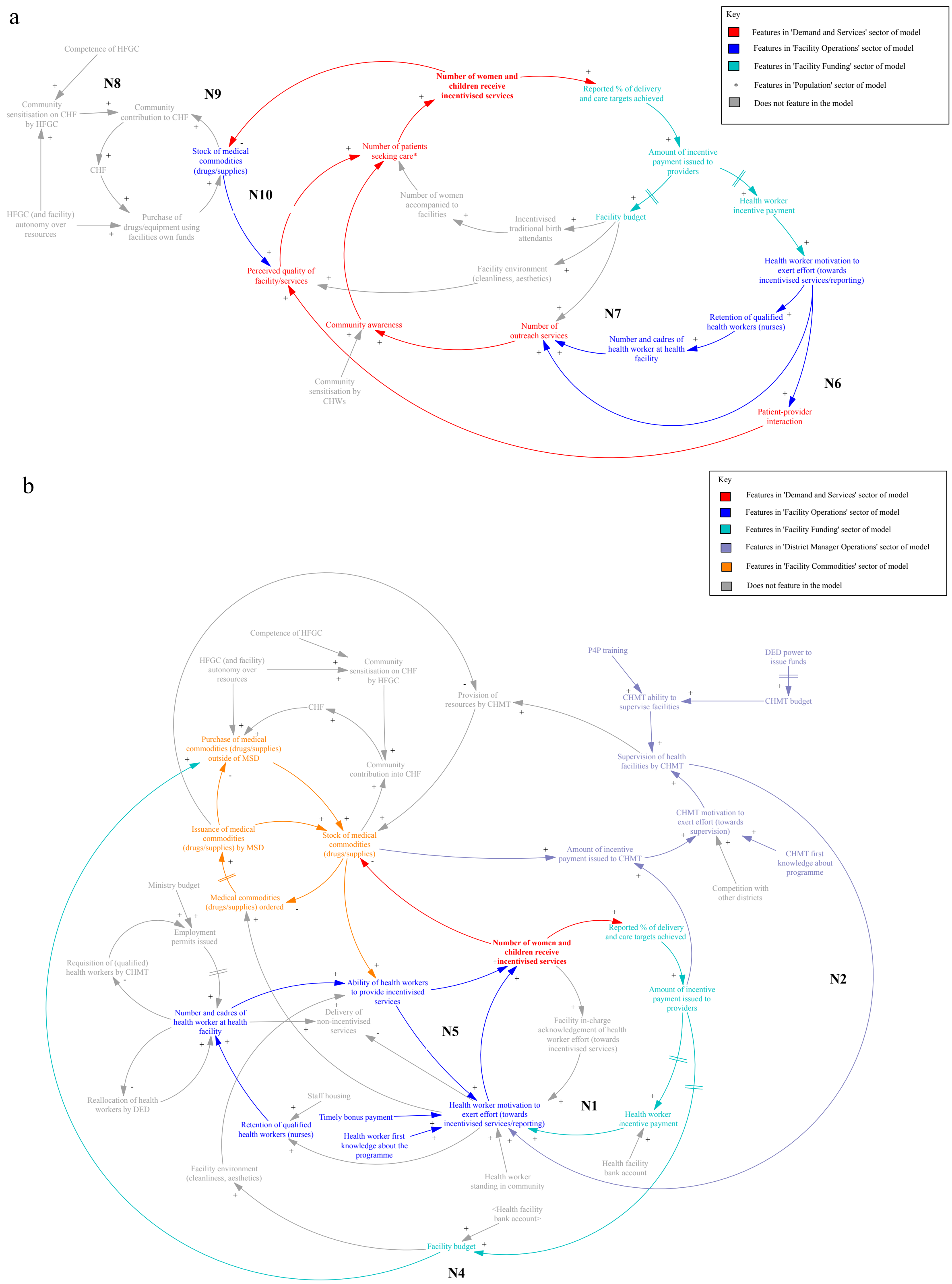


Figure 1: High level snapshot of causal loop diagram used as a blueprint to develop SDM. Notes to Figure: Three main mechanisms responsible for facility achievement of (or failure to reach) targets during P4P are shown in different colours. Changes in the supply (blue), changes to facility reporting (green), and changes in demand for services (red). Source: Cassidy *et al.* 2021.


a
Key
Features in 'Demand and Services' sector of model
Features in 'Facility Operations' sector of model
Features in 'Facility Funding' sector of model
Features in 'Population' sector of model
Does not feature in the model
Competence of HFGC
Community sensitisation on CHF by HFGC
Community contribution to CHF
CHF
HFGC (and facility) autonomy over resources
Purchase of drugs/equipment using facilities own funds
Stock of medical commodities (drugs/supplies)
Perceived quality of facility/services
Number of women and children receive incentivised services
Reported % of delivery and care targets achieved
Number of patients seeking care*
Number of women accompanied to facilities
Incentivised traditional birth attendants
Facility budget
Amount of incentive payment issued to providers
Health worker incentive payment
Health worker motivation to exert effort (towards incentivised services/reporting)
Facility environment (cleanliness, aesthetics)
Community awareness
Community sensitisation by CHWs
Number of outreach services
Number and cadres of health worker at health facility
Retention of qualified health workers (nurses)
Patient-provider interaction
N6
N7
N8
N9
N10
b
Key
Features in 'Demand and Services' sector of model
Features in 'Facility Operations' sector of model
Features in 'Facility Funding' sector of model
Features in 'District Manager Operations' sector of model
Features in 'Facility Commodities' sector of model
Does not feature in the model
Competence of HFGC
HFGC (and facility) autonomy over resources
Community sensitisation on CHF by HFGC
Provision of resources by CHMT
P4P training
DED power to issue funds
CHMT ability to supervise facilities
CHMT budget
CHF
Purchase of medical commodities (drugs/supplies) outside of MSD
Community contribution into CHF
Supervision of health facilities by CHMT
CHMT motivation to exert effort (towards supervision)
CHMT first knowledge about programme
Competition with other districts
Issuance of medical commodities (drugs/supplies) by MSD
Stock of medical commodities (drugs/supplies)
Amount of incentive payment issued to CHMT
Ministry budget
Medical commodities (drugs/supplies) ordered
Employment permits issued
Requisition of (qualified) health workers by CHMT
Number of women and children receive incentivised services
Reported % of delivery and care targets achieved
Amount of incentive payment issued to providers
Ability of health workers to provide incentivised services
Number and cadres of health worker at health facility
Delivery of non-incentivised services
Facility in-charge acknowledgement of health worker effort (towards incentivised services)
Reallocation of health workers by DED
Staff housing
Timely bonus payment
Health worker motivation to exert effort (towards incentivised services/reporting)
Retention of qualified health workers (nurses)
Health worker first knowledge about the programme
Health worker incentive payment
Health facility bank account
Health worker standing in community
Facility environment (cleanliness, aesthetics)
<Health facility bank account>
Facility budget
N1
N2
N4
N5

c

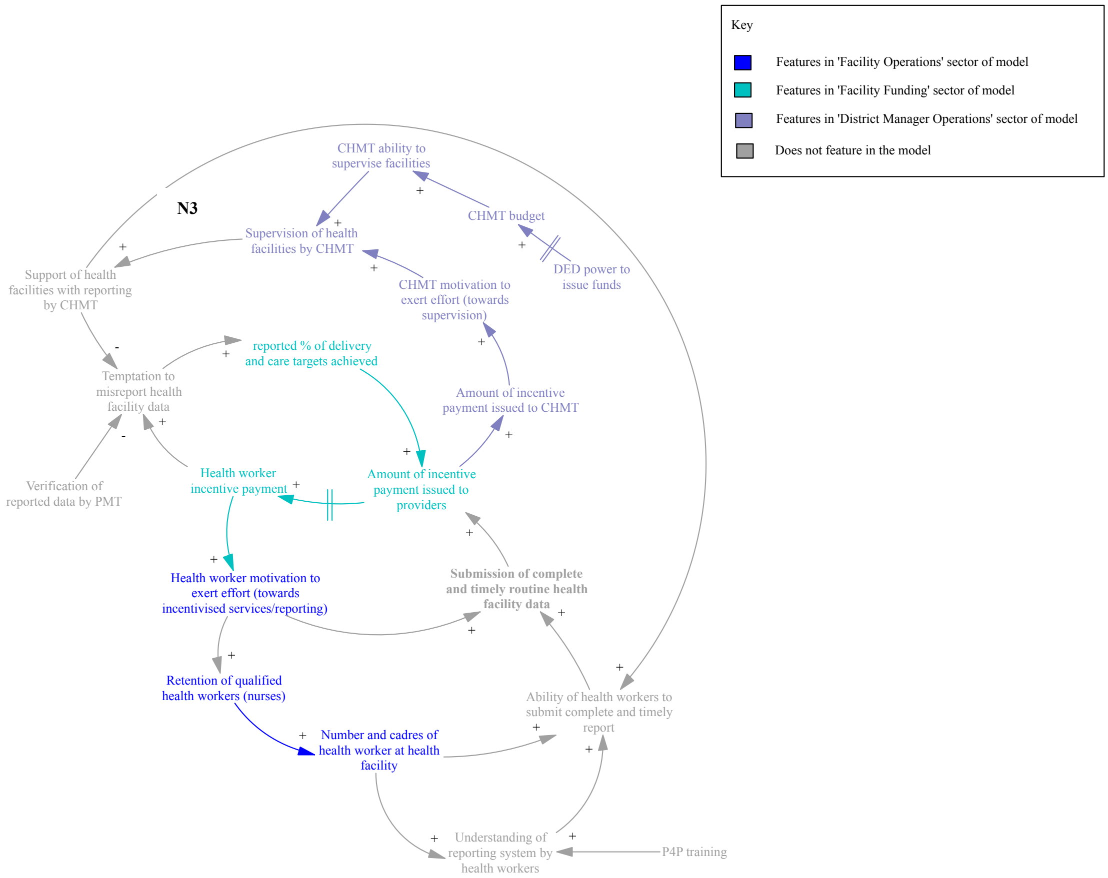
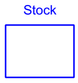
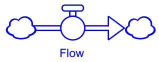


Figure 2: Structure identified in the previously developed CLD, specifically drawn from the (a) ‘demand’ (b) ‘supply’ and (c) ‘reporting’ components that feature in the simulation model. Abbreviations: Community health fund (CHF); Community health worker (CHW); Council Health Management Team (CHMT); District Executive Director (DED); Health Facility Governing Committee (HFGC); Medical Stores Department (MSD); Payment for performance (P4P).

A previous study was designed to evaluate Pwani programme effectiveness to improve content of care and coverage of services (Borghi *et al.* 2013). The study identified supporting evidence for improvements to certain services (percentage of women who have a facility-based delivery and percentage of women who received two doses of IPT during ANC) (Binyaruka *et al.* 2015), with no evidence of crowding out of intrinsic motivation (unpublished). The SDM in this study was designed to formally test pathways to programme effect and explore how the design of the programme could be adapted to optimise outcomes.

2.3 System dynamics modelling

The development of the quantitative SDM involved the development of a CLD and the adaptation of the CLD into a stock and flow diagram (Pruyt 2017). Whilst CLD notation consists of variables, arrows with attributed polarity (direction of relationship) and feedback loops (Cassidy *et al.* 2022), stock and flow diagrams consist of stocks 

, flows 

, auxiliary variables 

and constants 

(Pruyt 2017).

A simple example of a stock and flow diagram is presented in Figure 3, demonstrating replenishment and depletion of medicine at a health facility. 'Stock of medicine' represents a single stock; a container which changes value over time based on the in and out flows 'replenishment of medicine' and 'depletion of medicine', respectively. The behaviour of the inflow is dependent on the auxiliary variable 'medicine procured', a dynamic variable that changes over time in response to the constant variables 'availability of medicine from supplier' and 'medicine requested', whose values remain fixed during the simulation. The behaviour of the outflow is dependent on the constant variable 'medicine used'; in reality, 'medicine requested', 'medicine used' and 'availability of medicine from supplier' are likely to fluctuate over time but for simplicity are given constant variable status here. In this example, a request for medicine is placed every three months (300 items of medicine) but the supplier can only provide 75% of items requested. Medicine at the health facility is depleting at a steady rate of 100 items per month. The impact is felt on the stock of medicine, which is never fully replenished and often leaves the facility with stockouts.

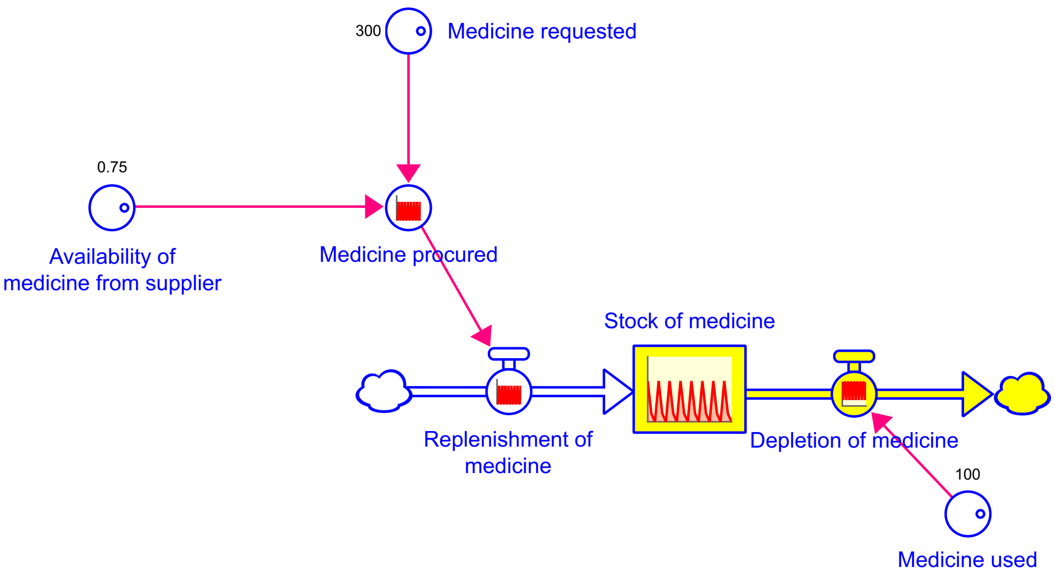


Figure 3: A simple example of a stock and flow diagram.

2.4 Model software

The SDM presented in this paper (and in the example above) were developed in STELLA Architect (version 2.1.4) (isee systems inc 2021). STELLA was chosen as the preferred modelling software due to the extensive guidance literature available for model development and functionality that allows users to develop interfaces which can be used to support model testing and discussions with stakeholders.

2.5 Development and validation of the SDM

Development and validation of the model can be broadly summarised as following four stages; (i) defining the purpose and goal of the model (ii) creation of model sectors (iii) validation of the model (iv) sensitivity analysis.

(i) Model purpose

The first step for SDM development was to (i) define the problem/health system behaviour to be investigated and (ii) define the goals of the model. A previously developed CLD, which identified pathways to impact of P4P on delivery and coverage of MCH using the Pwani programme as a case study (Cassidy *et al.* 2021), was used as a blueprint for determining model purpose, sector selection and creation (Figure 1, Figure 2). The health system behaviour explored with SDM was the performance of facilities during the P4P programme in Pwani. The goal of the SDM model was to (i) explore how facility performance responded to the P4P programme and (ii) test whether changes to implementation of the programme or its' design can result in improved performance in a 'typical' primary care facility, (iii) explore how context affects programme outcomes. In the model, the performance of a facility for two incentivised services is monitored; a content of care indicator (percentage of women who received two doses of IPT during ANC) and a coverage indicator (percentage of women who had a facility-based delivery), as these indicators showed some improvement during the P4P programme in Pwani and were the primary outcomes in the CLD. The model time step is months (the performance reporting unit), with simulation start time January 2011 and a time horizon of 54 months. The simulation period covers programme commencement (January 2011), the period of programme evaluation (January 2012 – March 2015) and a short period post-evaluation (up to July 2015), to consider both the short- and long-term effects of the programme which may fluctuate over time (Borghi *et al.* 2021).

(ii) Model sectors

Sector selection

The second step for development was the creation of model sectors that drive behaviour in different compartments of the model. The CLD was used as a framework to inform development of the simulation model. Six model sectors were generated from the structures identified in the CLD (structure taken from the CLD and used to develop the SDM are shown in different colours in Figure 3, mapped to the different sectors of the SDM). Structures identified in the 'demand' component of the CLD (Figure 2a) fed into development of the '**Demand and Services**', '**Facility Operations'**, **'Facility Funding'** and **'Population'** sectors (Figure 4). Structures identified in the 'supply' component of the CLD (Figure 2b) fed into development of the '**Demand and Services**', '**Facility Operations'**, **'Facility Funding'**, '**District Manager Operations**' and '**Facility Commodities**' sectors (Figure 4).

In this version of the model, dynamics related to facility reporting (Figure 2c) are not included. The main focus was on the facility level supply side dynamics related to facility performance as this was the primary target of P4P. As a result, the SDM does not capture a number of demand side elements, shown in 'grey' in Figure 2 including: the dynamics around payment into a community health fund (voluntary community health insurance fund that was used to support provision of services at the facility), mechanisms for employing health workers and the activities of community health workers and traditional birth attendants in service demand creation. An agent-based model is currently under development which will explore the effect of community and service demand dynamics on facility-based deliveries.

Simulation model overview

The purpose of each **sector**, *key sector outputs used as input to other sectors* and a description of how sectors pass information is given here (Figure 4), with detailed individual model sector diagrams and description of model equations given in Supplementary Files 2 and 3, respectively. This section outlines the model functioning in the absence of P4P.

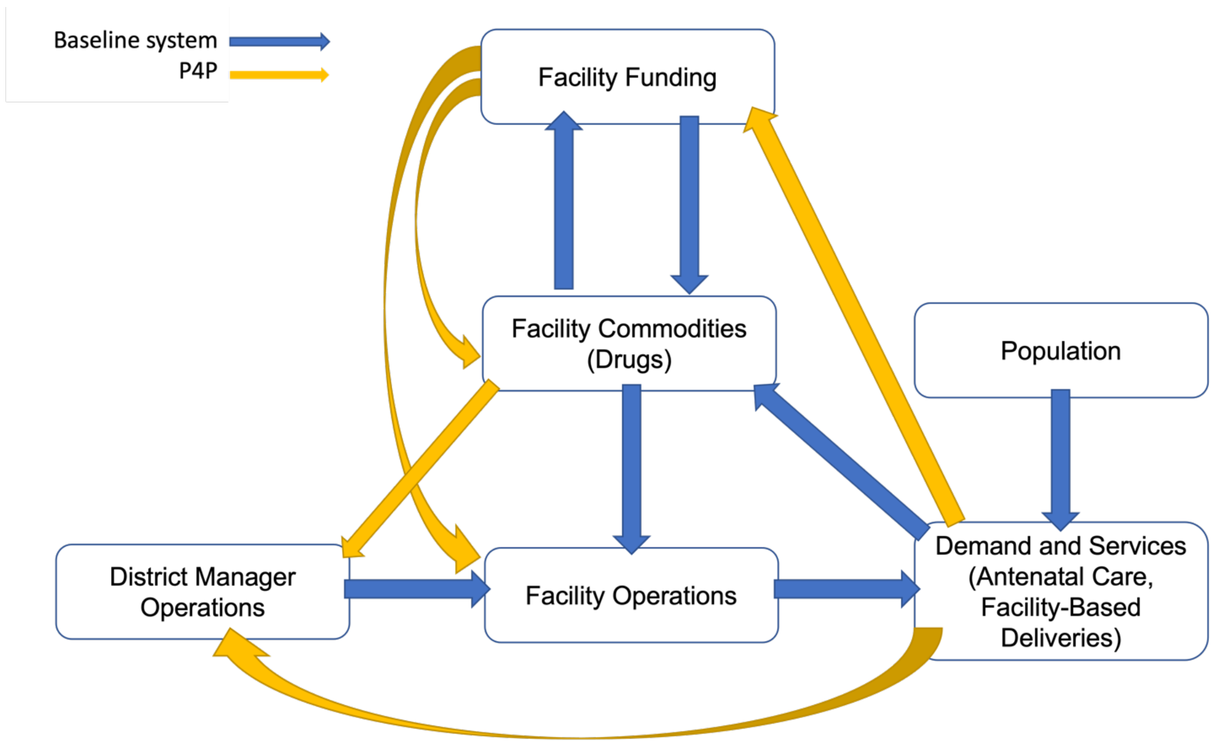


Figure 4: High level overview of simulation model. The model contains six subsectors which pass information to each other (arrows). The user can run the model with payment for performance switched 'on' (P4P, yellow) and 'off' (baseline system, blue).

The **Population** sector controls population dynamics that feed into the **Demand and Services** sector**.** It controls ageing in the population (neonates, infants, pre-schoolers, children, reproductive age adults and adults above 50) over time, which is driven by: (i) the respective age mortality rates; (ii) fertility rate. The population dynamics for this sector mirror the natural phenomenon of ageing in the Tanzania population. The general function for an ageing population was adapted from Semwanga *et al.* (2016). The sector generates the following key output and population group of interest *number of newly pregnant women,* which contributes to the flow of patients seeking care in the **Demand and Services** sector. The population sector has been structured so that the model can be later adapted to focus on other types of service provision (e.g. infant vaccination).

The **Demand and Services** sector controls the number of ANC patients that receive services and facility-based deliveries. Patients can attend up to four ANC appointments, with three possible pathways for each ANC visit (i) dropping out and not attending ANC visit, (ii) receiving treatment (up to two doses of IPT across all ANC visits, with the goal of two doses for each patient during pregnancy) or (iii) do not receive treatment. These three options reflect the possible outcomes for ANC visits by patients in facilities. Treatment receipt is dependent on (i) provider readiness to deliver care (controlled in the **Facility Operations** sector) and (ii) attendance rates for each antenatal care visit.

The percentage of facility-based deliveries is determined by (i) the number of antenatal care visits; (ii) distance to facility; (iii) awareness of maternal and child health and healthcare in the community (in part estimated from ability to perform outreach controlled in **Facility Operations** sector and fraction of women attending antenatal care); (iv) perceived quality of facility/services (estimated as an average of availability of drugs in **Facility Commodities** sector and patient-provider interaction from **Facility Operations** sector). This function input was modelled on the function for facility-based deliveries described in Semwanga *et al.* (2016), where the purpose of the model was to identify system strengthening policies to address neonatal mortality in Uganda, and adjusted during model calibration to assign reduced weight to perceived quality of facility/services to ensure a better model fit (see Supplementary File 5 for further details).

For each patient who receives a service (ANC or facility-based delivery), a single unit of a drug is 'used' with drug availability depleting in the **Facility Commodities** sector (based on expected dispensing of medication per visit). The **Demand and Services** sector generates key outputs *percentage of women who receive IPT2* and *percentage of women who seek facility-based delivery.*

The **Facility Commodities** sector controls the replenishment and depletion of malaria (IPT) and labour drugs at the facility level. The expected number of ANC and facility-based delivery patients is fed in from the **Demand and Services** sector and used to place orders for drugs on a quarterly basis to the Medical Stores Department (autonomous government department responsible for provision of medical commodities), as is the practice for health providers in Tanzania. Depending on availability of drugs at Medical Stores, facilities may need to try and address the deficit of drugs. Facilities can use funds (facility held funds, managed in the **Facility Funding** sector) where available to purchase additional drugs. Key outputs in the **Facility Commodities** sector are the *availability of IPT drugs* and *availability of labour drugs,* which deplete depending on the number of patients treated in the **Demand and Services** sector.

The **Facility Operations** sector manages facility-level dynamics including provider readiness (related to delivery of IPT). Provider readiness related to facility-based deliveries is not captured in the SDM but will be the focus of the in-development ABM, built to simulate service demand side dynamics related to facility-based deliveries. Provider readiness (related to delivery of IPT) is calculated as the minimum of availability of IPT drugs or average of (i) knowledge of health workers (IPT); (ii) number of health workers at health facility (percentage of positions filled); (iii) availability of IPT drugs fed in from the **Facility Commodities** sector; (iv) health worker motivation to exert effort towards incentivised services (as observed in the Tanzania CLD). This is so provider readiness to deliver services (related to IPT) does not exceed the stock of medicine available. Health worker motivation is calculated as an average of availability of drugs (IPT and labour) fed in from **Facility Commodities** sectors, district manager supervision (quality) fed in from **District Manager Operations** sector and number of health workers at health facility (percentage of positions filled). Key outputs in this sector are *provider readiness (related to delivery of IPT)* and *health worker motivation to exert effort towards incentivised services.*

The **Facility Funding** sector manages the funding that is held and used at the facility level and can be used to purchase additional drugs where needed. The key output from this sector is *facility held funds available*.

The **District Manager Operations** sector manages supervision visits by members of the Council Health Management Team to facilities. The district manager supervision (quality) is dependent on district level resources, management team motivation and the skill level of district managers. Supervision visits affect knowledge of health workers related to IPT and health worker motivation. The key output for this sector is *district manager supervision (quality).*

Introduction of P4P intervention

The P4P programme (as modelled) is described here, granular details of how the P4P programme and model scenarios were incorporated into the simulation model can be found in Supplementary File 3 (relevant equations/descriptions of functions highlighted). Health facilities are set targets they need to reach each cycle (6 months) to receive P4P incentive payments. Payment was to be made within three months of the conclusion of the six month performance cycle (Borghi *et al.* 2013), however, in practice payments were often delayed. In the model, the performance targets are for specific services monitored in the **Demand and Services** sector. These are (i) percentage of women who receive IPT2 and (ii) percentage of women who seek facility-based delivery. Depending on performance against these targets, providers may receive incentive payments which are deposited in the **Facility Funding** sector.

The payment is split 75:25, with the larger portion allocated for health worker incentive payments and the remaining portion to be used to improve facility operations (e.g. purchasing additional medical commodities where needed) (Binyaruka *et al.* 2015). The health worker incentive payment is fed from **Facility Funding** to the **Facility Operations** sector. Incentive payments (specifically timeliness of payments) influence health worker trust in the programme and health worker motivation to exert effort towards incentivised services. The remainder of the incentive payment, in the model, supplements facility held funds (**Facility Funding sector**) and can be used to purchase drugs (malaria and labour drugs) where needed in the **Facility Commodities** sector. A new key output from the **Facility Funding** sector is *staff incentives*.

The district management team are also eligible for incentive payments, which are processed in the **District Manager Operations** sector, with payments influencing district manager motivation to support facilities. In the simulation model, the district management targets (and determinant of incentive payment issued) are to reduce stockouts of medicine (observed in the **Facility Commodities** sector) and overall performance of health facilities (observed in **Demand and Services** sector) (Borghi *et al.* 2013).

Data

The model was populated with both primary and secondary data sources, see Supplementary File 4 for details. Examples of secondary data include population and housing census reports, country and district-level health surveys, data from the impact evaluation conducted on the Pwani P4P programme and evidence drawn from the literature. The previous evaluation conducted on the Pwani P4P programme is described elsewhere (Borghi *et al.* 2013, 2021; Binyaruka *et al.* 2015), with a summary provided here. The impact evaluation investigated the effect of the P4P programme on all targeted MCH services (including percentage of women who receive IPT2 and percentage of women who seek facility-based delivery) through a controlled before and after study design. Surveys were conducted in all six districts of Pwani region (where P4P had been implemented) and in five control districts in neighbouring regions. The evaluation consisted of a health facility survey, health worker survey, exit survey of patients and survey of women who had delivered in the last 12 months. Data collection took place at three time points: ‘baseline’ (January 2012), ‘short term’ (February 2013) and ‘long term’ (February and March 2015).

During model development, two members of the original programme evaluation team were consulted to provide insight into model dynamics related to impact of district manager supervision on health worker skill level. Model equations reflect this discussion, where effect of district manager supervision on health worker knowledge is dependent on the ‘base level’ of knowledge at the facility. Where this is lower, it will take a few supervision visits to raise the health worker knowledge (specifically related to provision of IPT during ANC).

(iii) Model validation

The third step for model development was subjecting the model to a series of verification and validation tests to build confidence in the structure, behaviour, and robustness of the model. To check for internal validity, every equation in the model was reviewed for dimensional consistency i.e. that model units were appropriate for the given variable i.e. population parameters are measured in units of ‘persons’, and that units used for outputs were appropriate based on variable input units. The model was also subjected to extreme condition testing, whereby selected model parameters were adjusted to extreme values and model output was evaluated to ensure expected results. For example, when the dropout rate for attending a first ANC visit is 0.999, only a handful of patients are expected to attend this first visit and move through the ANC part of the demand and services sector; or when the provision of medicine by the Medical Stores is severely impacted, a drastic depletion of medicine available at facilities should be observed. The model performed well when subjected to testing, producing expected behaviour under extreme conditions. Model equations and structure were also independently reviewed by a team member.

To check for external validity, selected model output projections were also compared to real data where available, with equation and parameter adjustments made where required so that model outputs were aligned with data (model calibration). The model was adequately able to replicate known trends, see Supplementary File 5 for further details on how selected outputs were calibrated to data.

To check model face validity, the resulting model was presented to nine key stakeholders involved in the implementation or evaluation of the Pwani programme during virtual interviews (conducted via Zoom) as a final validation step. A model interface was developed using Stella Architect to assist with presentation of model outputs and key assumptions, see Supplementary File 6 for the interview guide and details on model interface. The interview consisted of two segments; (i) stakeholders were shown key model output and dynamics and asked to comment on whether model behaviour was realistic and aligned with their experience of the P4P programme and (ii) stakeholders were shown model assumptions and asked to provide feedback on their validity.

The feedback received during these interviews resulted in some new additions and adjustments to existing model structure (see Supplementary File 7 for details). Stakeholders also reflected on the presentation of the model, commenting that a high-level diagram showing how the model worked would be useful to them (see Figure 4).

Stakeholders remarked on the importance of community health workers and traditional birth attendants in increasing community awareness of services and escorting women to facilities for facility-based deliveries. These dynamics are not included in this current version of the model for the reasons set out above (see sector selection).

(iv) Sensitivity analysis

The final step for model development was subjecting the model to sensitivity analyses to determine the sensitivity of key outcomes (percentage of women who receive at least two doses of IPT during ANC, percentage of women who seek facility-based delivery) to changes in model parameters. Model parameters deemed appropriate for analysis (see Supplementary File 8) were adjusted by 10%, with key outcome results recorded. The following scale was used to determine sensitivity to changes in model variables; sensitive (5% ≤ change in outcome < 15%), very sensitive (15% ≤ change in outcome < 25%) and highly sensitive (change in outcome ≥ 25%). The scale is adapted from Semwanga *et al.* (2016) and presented with smaller intervals for higher sensitivity categories, to further distinguish 'very sensitive' from 'highly sensitive' results. As the variables included in the sensitivity analysis reflect health system characteristics, this analysis also shed light on the likely effect of changes to the health system context in which P4P is implemented on key outcomes.

2.6 Simulation and scenario testing

Model scenarios were selected to contribute evidence towards the knowledge gap identified by reviews of P4P effects in LMIC settings (Das *et al.* 2016; Patel 2018; Diaconu *et al.* 2021); to further understanding on pathways to effect for P4P, acknowledging the influence of programme design, implementation and context.

The model was first used to explore how health system performance changed under P4P, to examine the effects of the programme as it was implemented (as mentioned, there were delays in issuing programme incentive payments) on pathways to effect for the programme (e.g. health worker motivation, availability of medicine) and targeted services, percentage of women who receive at least two doses of IPT during ANC and percentage of women who seek facility-based delivery. The effect of changes in programme implementation (payments made on time vs. with delays) and design (adjusting the share of funds between staff incentives and funds to strengthen facility operations) on these outcomes were then tested in the model. The effect of payment delays on procurement of additional medicines in the **Facility Commodities** sector is logical (health providers can purchase medicine when funding is available, unable to purchase when funds are unavailable), the effect of payment delays on health worker motivation and trust in the **Facility Operations** sector was determined through stakeholder consultation (see Supplementary File 7 for details). The variable 'use of incentives' in the **Facility Funding** sector was altered to adjust the share of funds between staff incentives and funds to strengthen facility operations.

Finally, the sensitivity analysis results were used to explore the effect of changes to programme and health system contextual factors (including provision of medicine from Medical Stores, amount of alternative facility held funding and staffing levels) on targeted services, percentage of women who receive at least two doses of IPT during ANC and percentage of women who seek facility-based delivery.

Results

Unpacking the mechanisms of P4P

As in the original intervention evaluation, the stark contrast between intervention and control sites at the short-term evaluation is observed, before performance starts to drop off in the intervention group at the long term evaluation for percentage of women who received two doses of IPT (Figure 5). The change in performance observed in the intervention group in the model is attributed to changes in provider readiness (related to the delivery of IPT) (Figure 6), which is a factor of knowledge of health workers in delivery of IPT (Figure 7), number of health workers at the facility (% filled) (Figure 8), availability of IPT drugs (Figure 9) and health worker motivation (Figure 10). Availability of IPT drugs and health worker motivation to deliver incentivised services experienced the most change as a result of the programme, and are driving improvements in the IPT during ANC outcome.

Availability of IPT drugs increases when a delivery is made from the Medical Stores Department (every three months) or when facility held funding is used to purchase drugs (Figure 9). Where incentive payments are received by facilities (months 13, 18, 22, 27, 34 and 42) these are used to purchase additional drugs and improve drug availability. Availability of IPT drugs is volatile, exhibiting improved behaviour where funds are available to purchase more drugs outside Medical Stores, and behaviour closer to the control group where additional funds are not available. This extreme volatility is reflected in provider readiness (Figure 6), increasing when drugs are procured from the Medical Stores every three months, depleting over a three month period as drugs are dispensed to patients, with volatility somewhat stemmed when P4P payments are used to purchase additional medicines. Availability of all drugs exhibits a similar trend to availability of IPT drugs. Health worker motivation (Figure 10) is fluctuating as a result of changes in district manager supervision (quality) (Figure 11), trust in the P4P programme (Figure 12) and availability of all drugs (Figure 13). Trust in the programme gently increases whenever payments are made but decreases when there are severe (4+ months) delays in payment, with district manager supervision also exhibiting this trend.

For facility-based deliveries, an improvement between the intervention and control sites for the short term and long term evaluations is observed (Figure 14). The change in performance observed in the intervention group is attributed to changes in community awareness of facility and services (Figure 15) and perceived quality of facility and services (Figure 16). In the model, P4P has very limited impact on provider behaviour in performing outreach activities. If outreach activities were to increase, it would drive improvement in community awareness of facility services, which would lead to an increase in facility-based deliveries. The change in percentage of women as a result of P4P is therefore driven by the perceived quality of facility and services. Perceived quality is constrained by medicine availability, with improvements seen as a result of increased drug availability and health worker motivation (taken as a proxy for patient-provider interaction in the model).

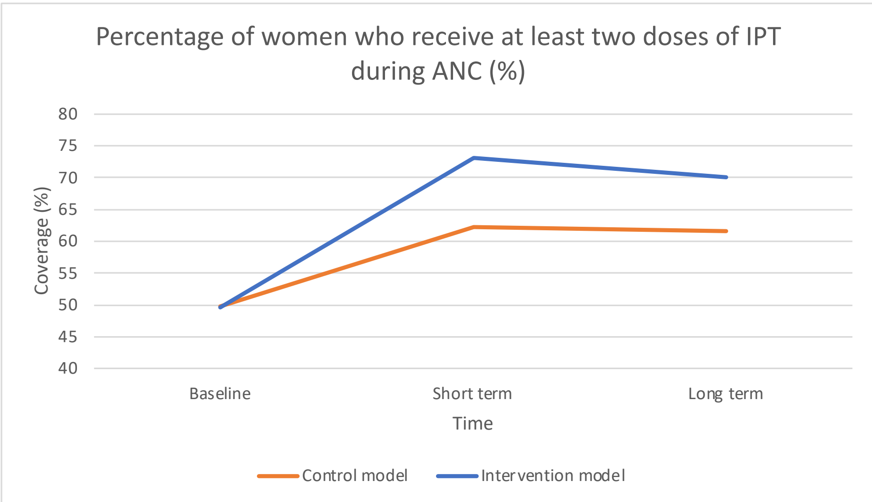


Figure 5: Model output for percentage of women who receive at least two doses of IPT during ANC (%) in the control and intervention groups.

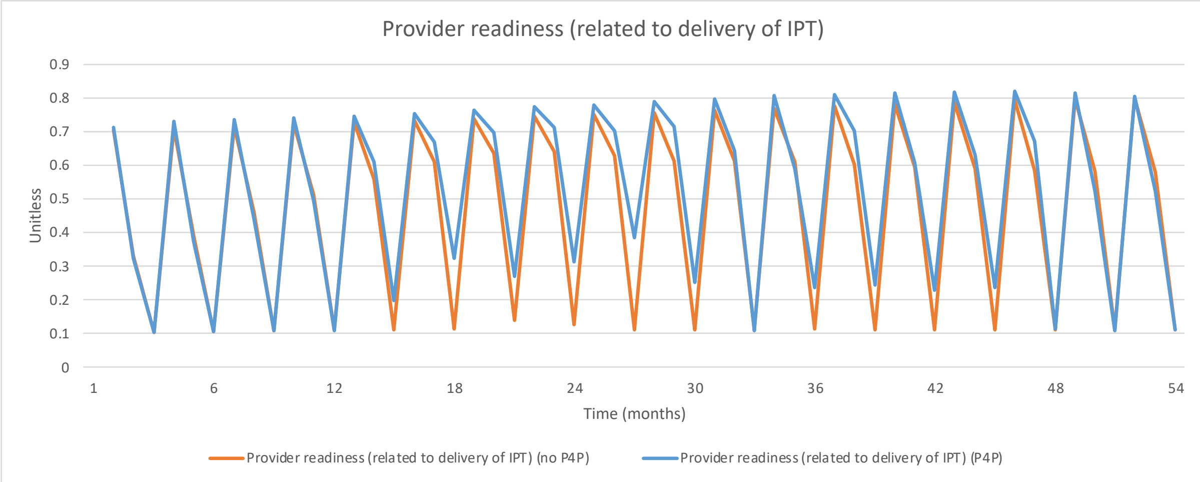


Figure 6: Model output for provider readiness (related to delivery of IPT during ANC) when P4P is turned off and on in the model.

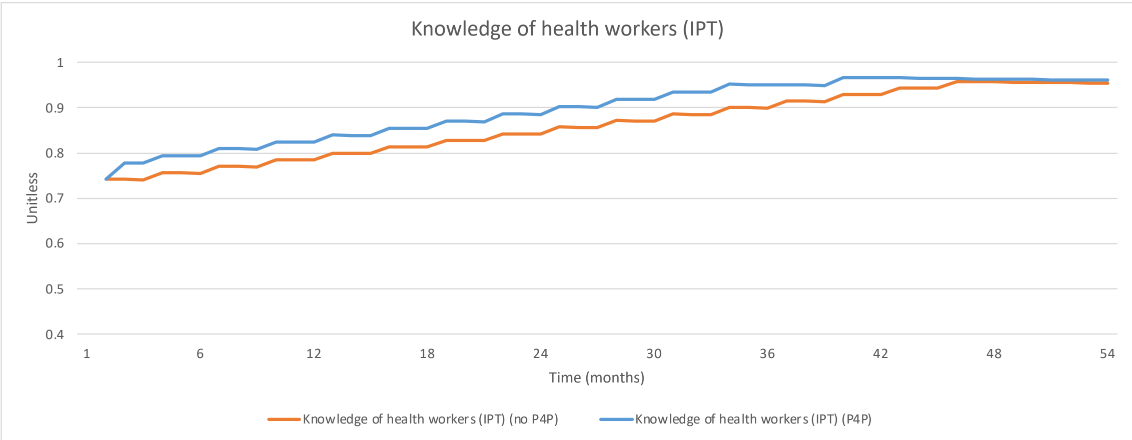


Figure 7: Model output for knowledge of health workers (IPT) when P4P is turned off and on in the model.

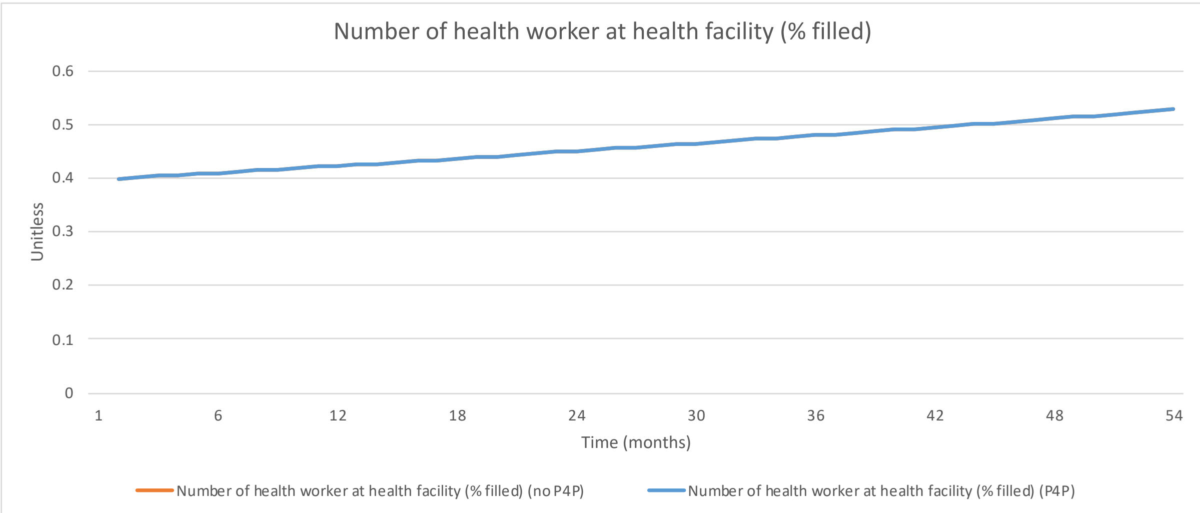


Figure 8: Model output for number of health workers at health facility (% filled) when P4P is turned off and on in the model. Note: In the model, number of health workers at health facility (% filled) is unaffected by P4P but included as input to provider readiness to deliver services (related to delivery of IPT).

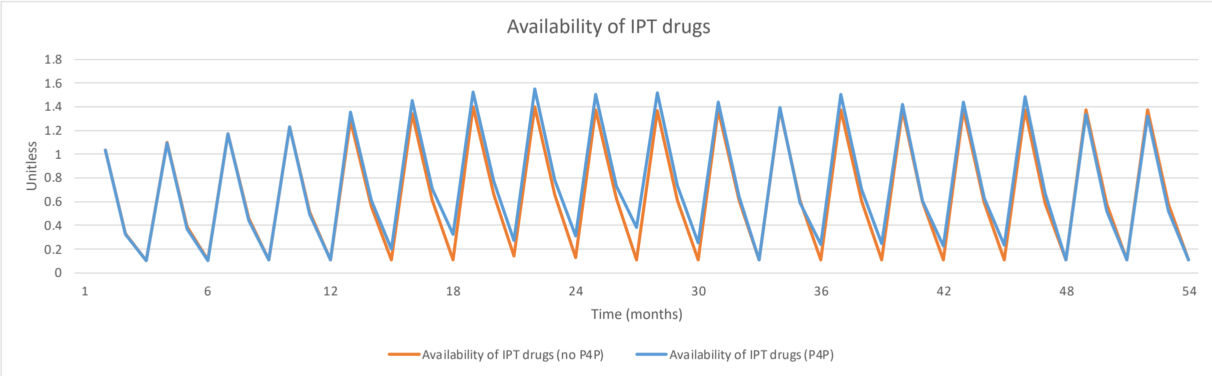


Figure 9: Model output for availability of IPT drugs when P4P is turned off and on in the model.

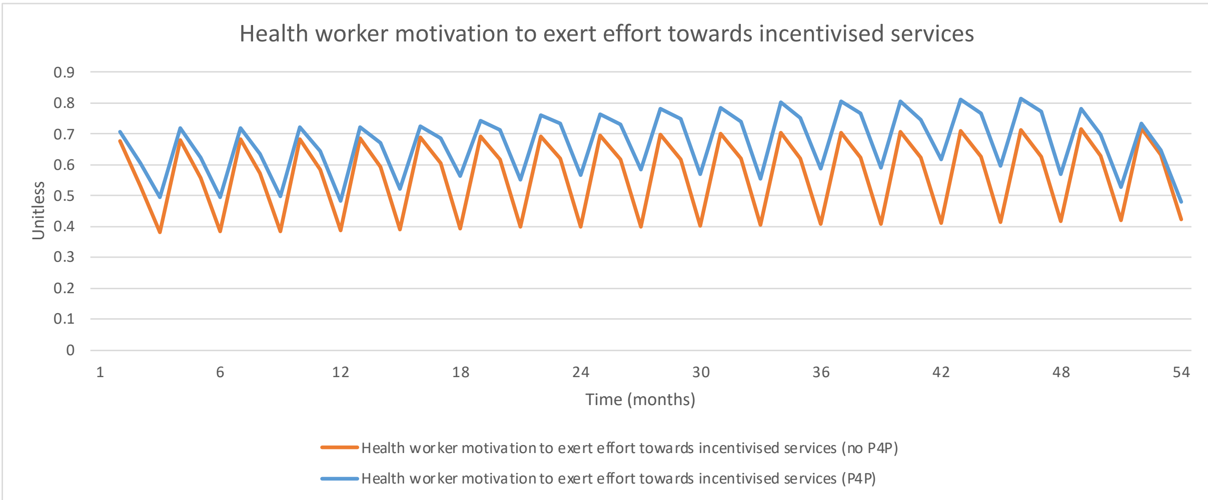


Figure 10: Model output for health worker motivation to exert effort towards incentivised services when P4P is turned off and on in the model.

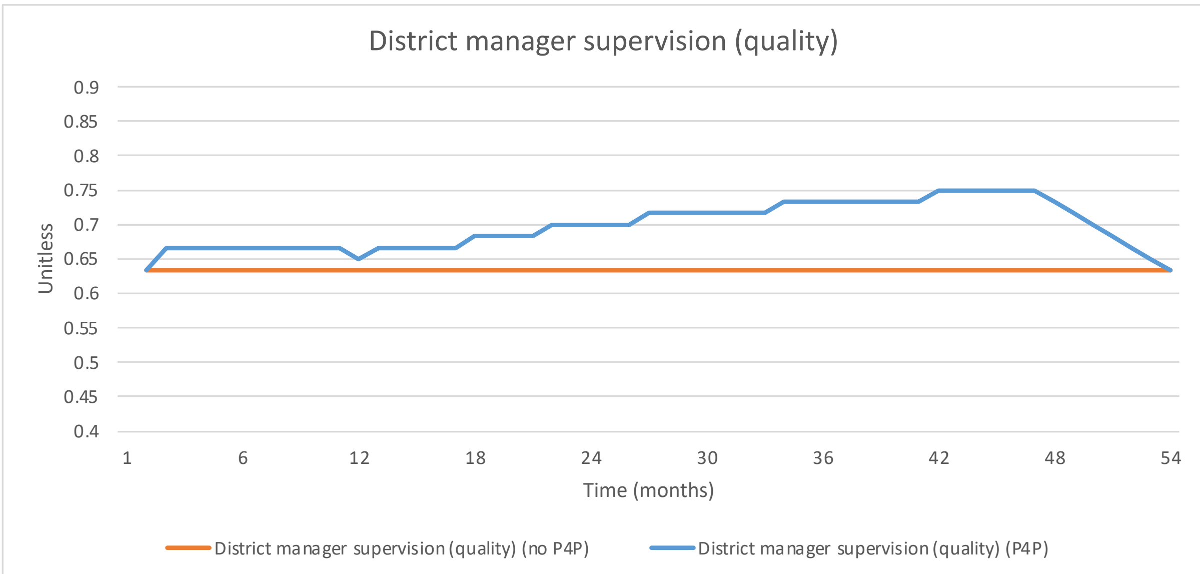


Figure 11: Model output for district manager supervision (quality) when P4P is turned off and on in the model.

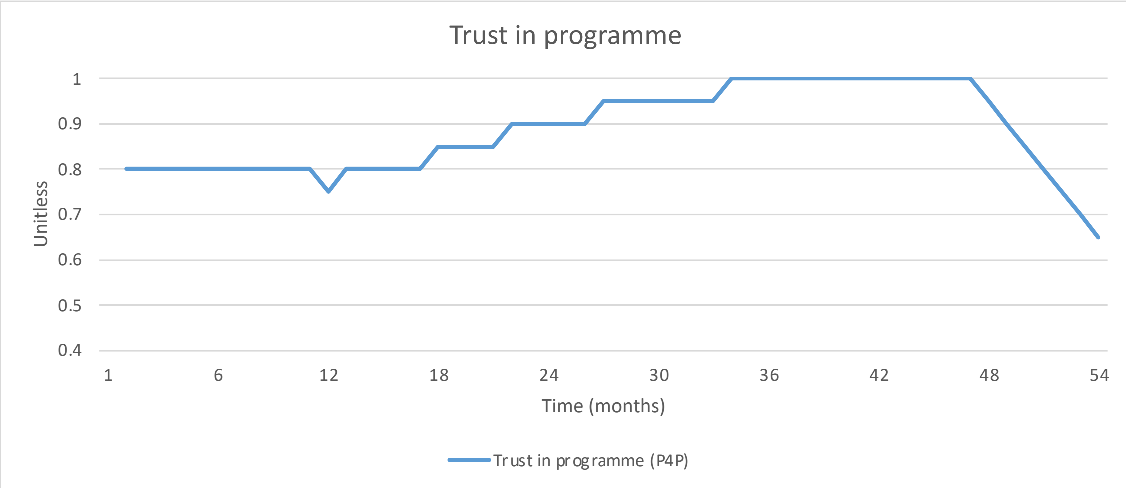


Figure 12: Model output for health worker trust in programme when P4P is turned off and on in the model. Note: This graph only shows one data set as 'Trust in programme' only exists (and impacts health worker motivation) when the P4P programme is switched 'on' in the model.

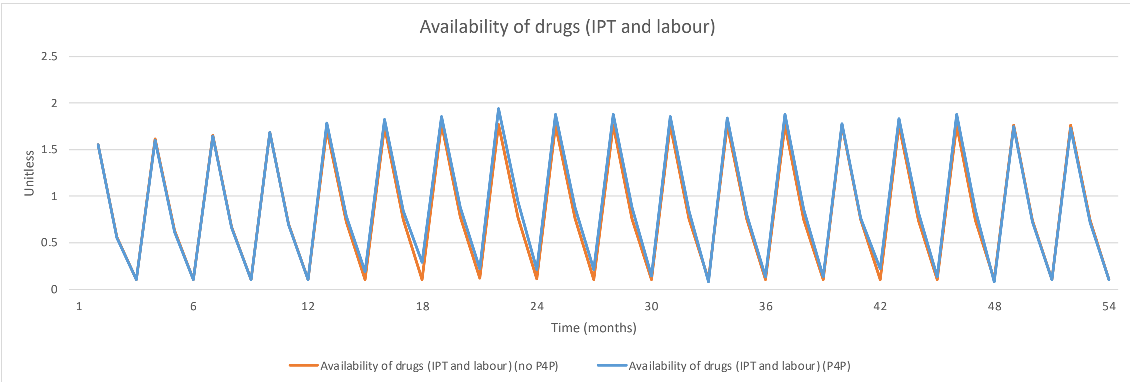


Figure 13: Model output for availability of drugs (IPT and labour) when P4P is turned off and on in the model.

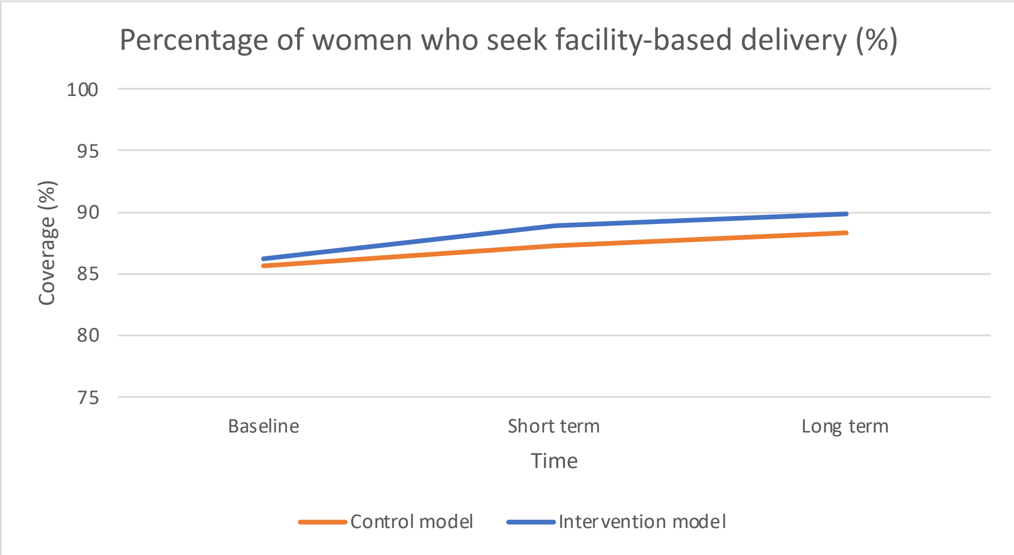


Figure 14: Model output for percentage of women who seek facility-based deliveries (%) in the control and intervention groups.

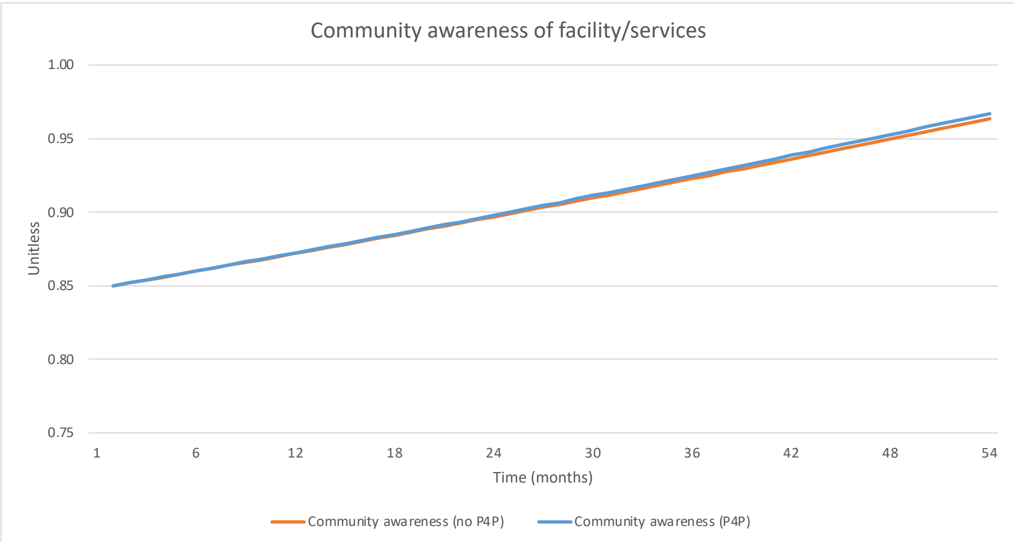


Figure 15: Model output for community awareness when P4P is turned off and on in the model.

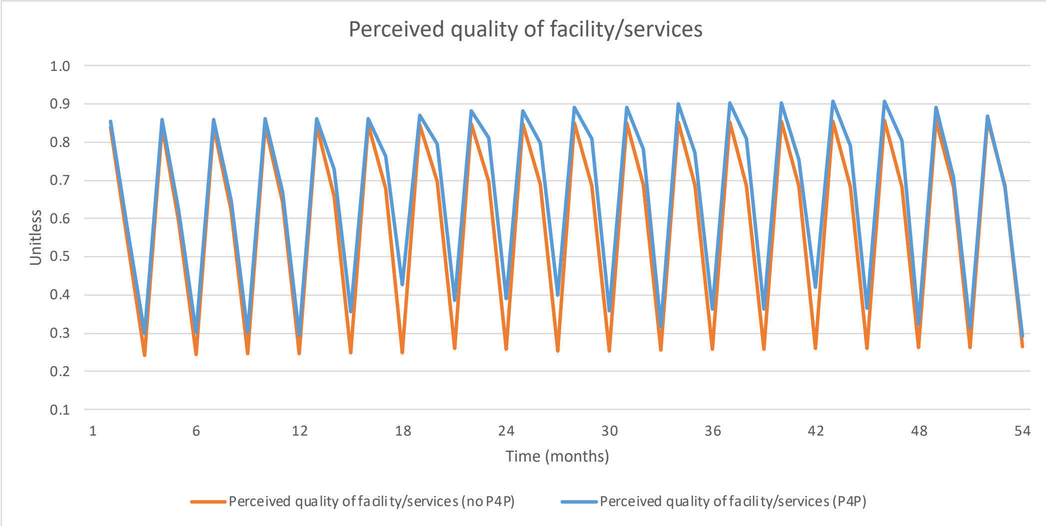


Figure 16: Model output for perceived quality of facility/services when P4P is turned off and on in the model.

Effect of changes to programme implementation and design

Model with and without payment delays

The model was run with payment delays switched ‘on’ and ‘off’ to observe the impact of delays on the delivery of incentivised services. Where delay in payments is switched ‘off’, payments are made as was originally planned, every six months (3 months after each performance cycle, i.e. month 9, 15, 21 etc.). Where delay in payments is switched ‘on’, payments are made according to the actual schedule of payments that took place (months 13, 18, 22, 27, 34 and 42). When there are no payment delays, percentage of women who receive two doses of IPT (Figure 17) consistently increase and decrease primarily as a result of changes in availability of IPT drugs (drug replenishment by the Medical Stores, drug depletion through dispensing to patients and procurement of drugs using P4P payments) and health worker motivation (district manager supervision visits, increasing trust in the programme as payments are being made on time and availability of all drugs). In the absence of delays, further improvement in availability of IPT and labour drugs (Figure 18, Figure 19) is observed as a result of P4P payments. There are also improvements (compared to the no P4P scenario) in health worker motivation (Figure 20), due to continued availability of drugs and periodic increases in health worker trust in the programme (Figure 21) and quality of supervision (Figure 22) each time a payment is made. Short term delays in payment (those less than 4 months) suspend improvement in health worker trust in programme (and therefore motivation) and quality of supervision, which then improve when payments are made. Impact on availability of drugs from short terms delays is minimal as payments are used in the next month or so to recover stock. When there are severe delays in payment (4 or more months), as seen from month 48, trust in the programme and quality of supervision decrease and struggle to recover. Prolonged reduction in funding also negatively impacts availability of drugs and ability to deliver incentivised services.

Percentage of women who seek facility-based delivery slightly improves under the no delay scenario (Figure 23). There is no noticeable improvement in community awareness of facility and services (Figure 24), with a slight improvement in perceived quality of facility/services (through consistent improvements in drug availability and health worker motivation) (Figure 25).

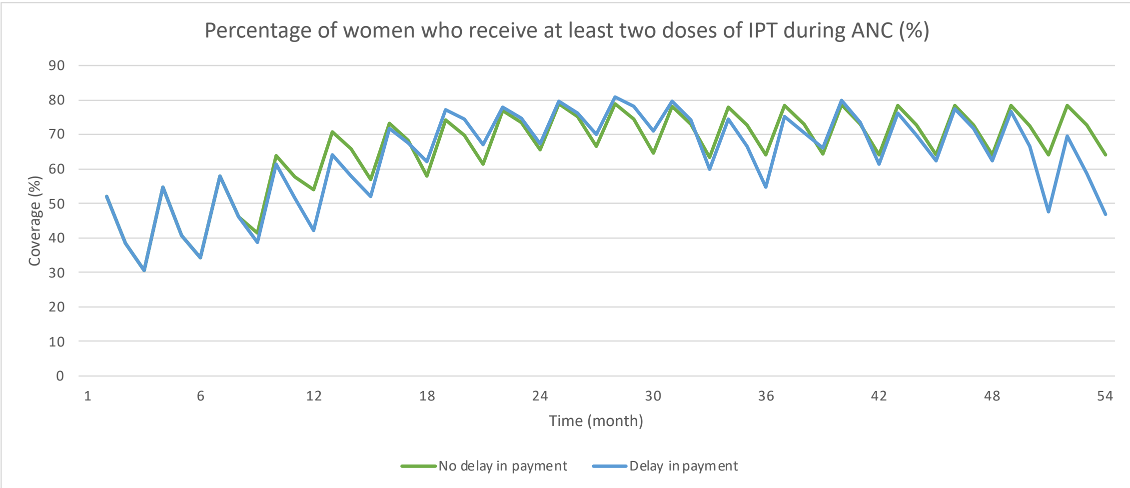


Figure 17: Model output for percentage of women who receive at least two doses of IPT during ANC (%) when (i) payments are made on time and (ii) when payments are delayed.

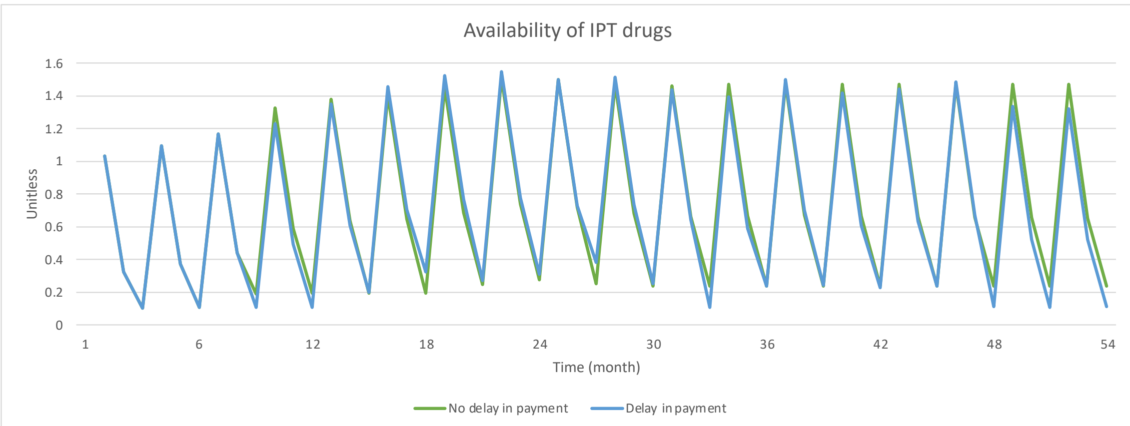


Figure 18: Model output for availability of IPT drugs when (i) payments are made on time and (ii) when payments are delayed.

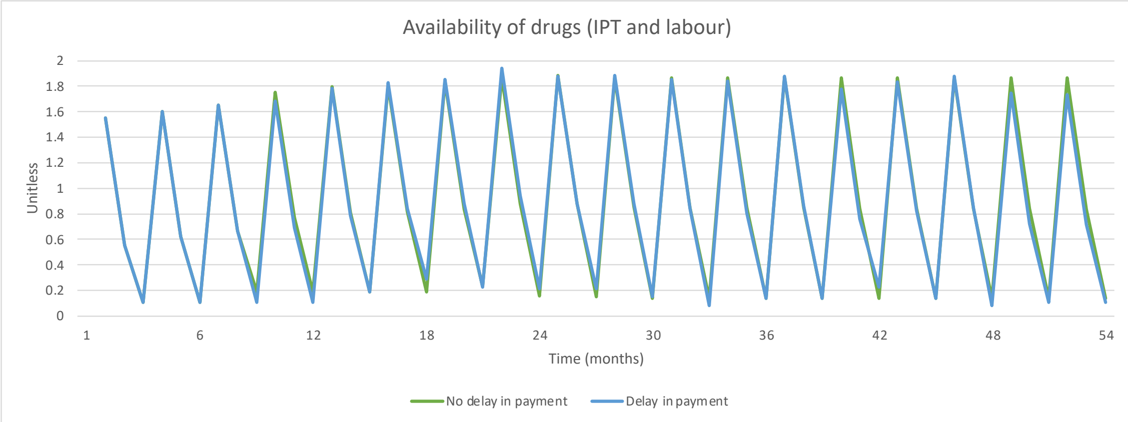


Figure 19: Model output for availability of drugs (IPT and labour) when (i) payments are made on time and (ii) when payments are delayed.

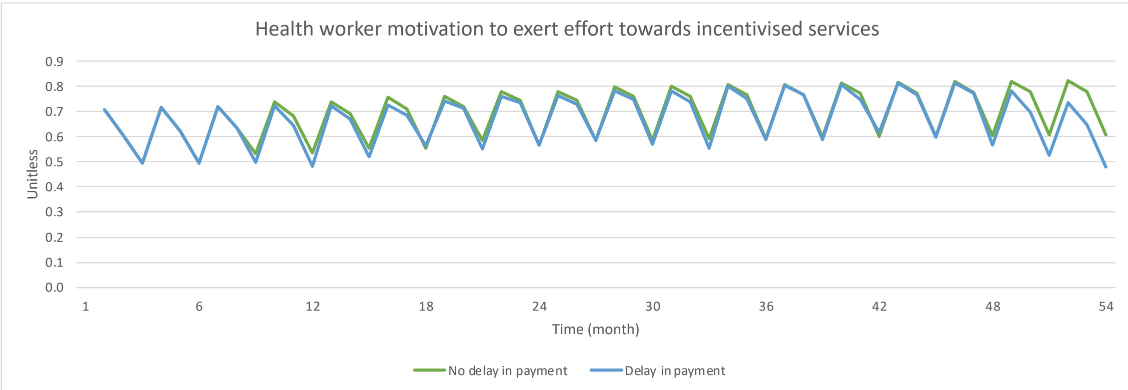


Figure 20: Model output for health worker motivation to exert effort towards incentivised services when (i) payments are made on time and (ii) when payments are delayed.

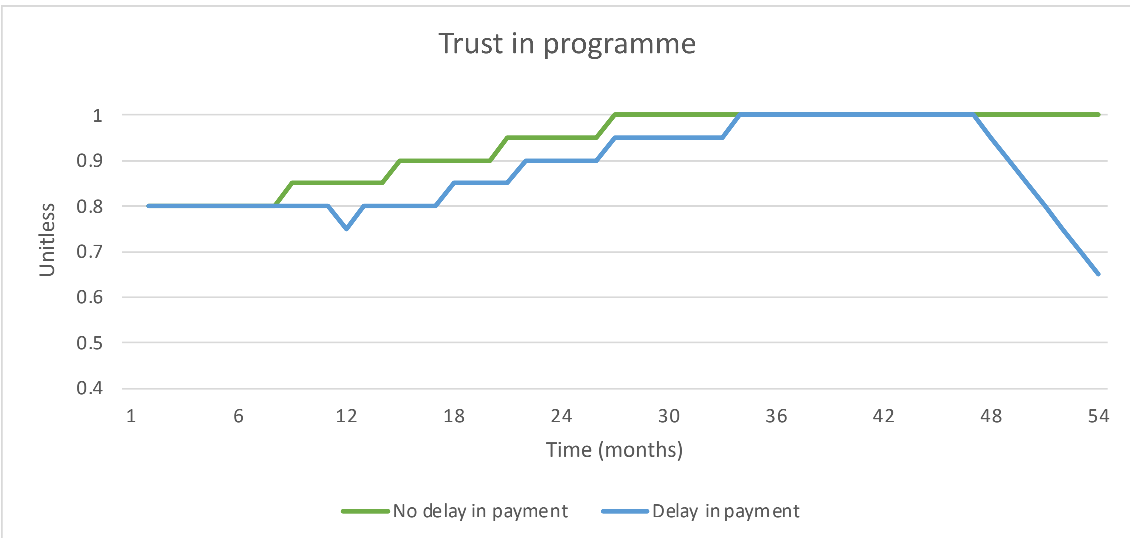


Figure 21: Model output for health worker trust in programme when (i) payments are made on time and (ii) when payments are delayed.

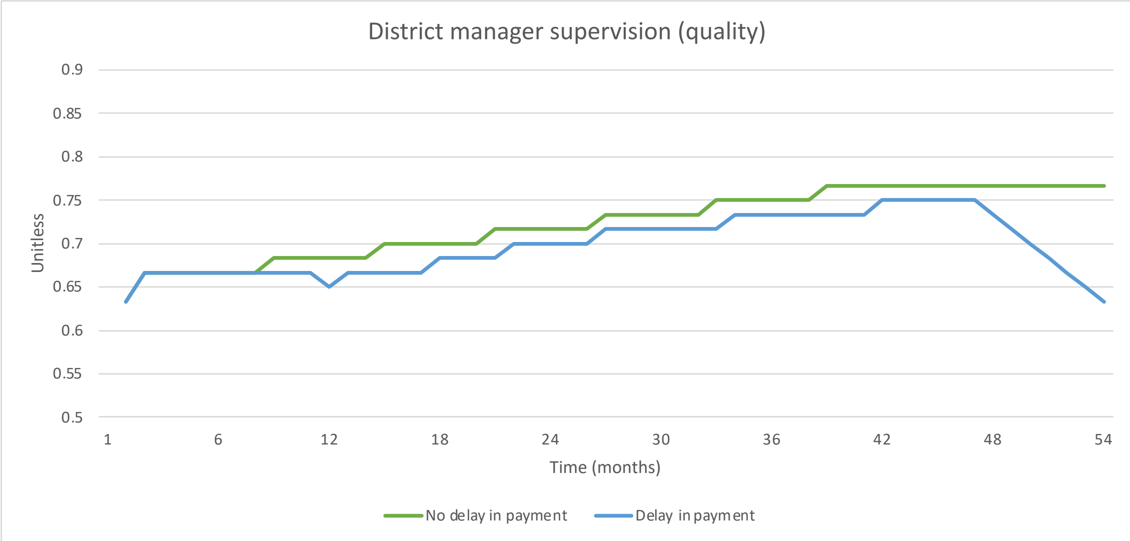


Figure 22: Model output for district manager supervision (quality) when (i) payments are made on time and (ii) when payments are delayed.

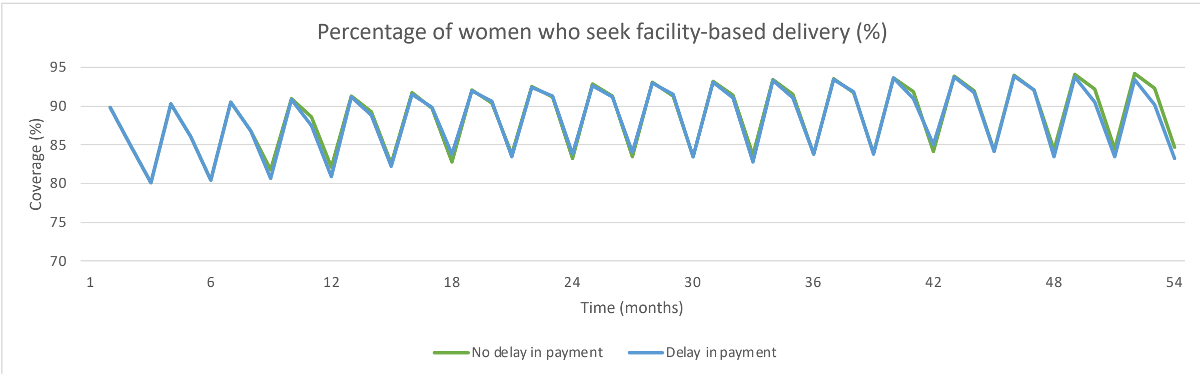


Figure 23: Model output for percentage of women who seek facility-based delivery when (i) payments are made on time and (ii) when payments are delayed.

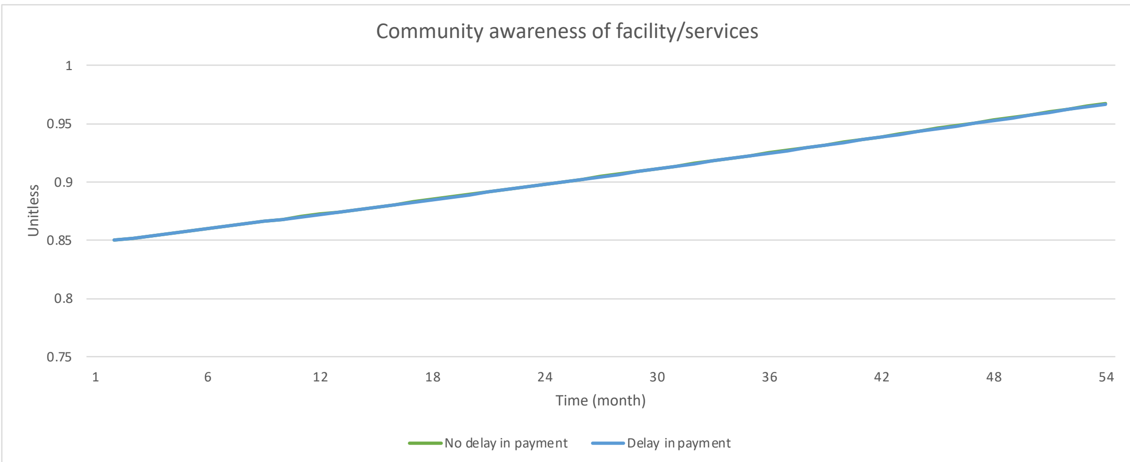


Figure 24: Model output for community awareness when (i) payments are made on time and (ii) when payments are delayed.

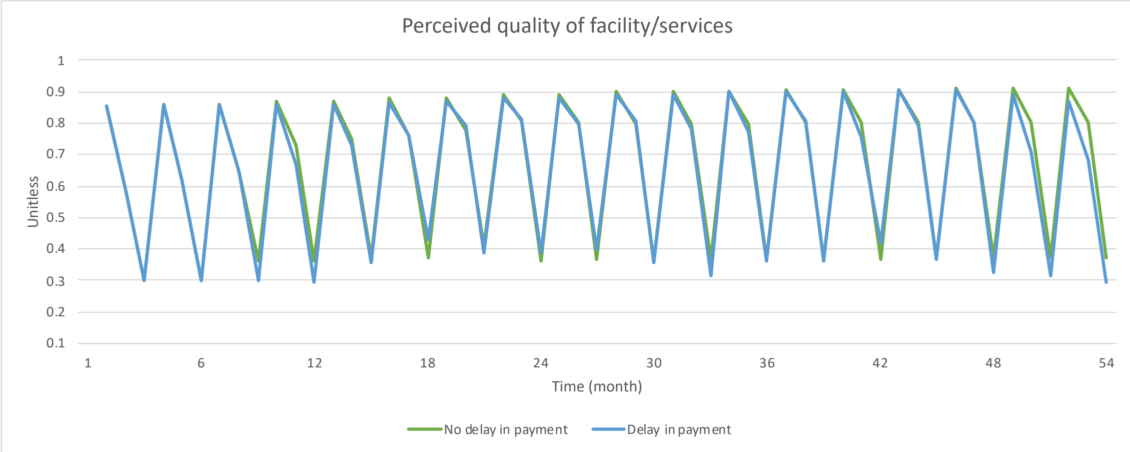


Figure 25: Model output for perceived quality of facility and services when (i) payments are made on time and (ii) when payments are delayed.

Model with changes to allocation and use of payment

When the allocation of incentive payments between staff and facility operations is adjusted under the no payment delay scenario, a direct impact is observed on both incentivised services (Figure 26, Figure 27). Figures 26-31 demonstrate the importance of the facility operations component of the incentive for achieving model outcomes – the 10:90 share of funds between staff and facility operations design resulted in the greatest improvement to incentivised services and key outcomes, with the 90:10 share performing worse than the original P4P design. With a higher level of facility operation funding, more funds are available to purchase drugs. This eases the burden of inadequate stock (Figure 28) and enables provision of care for more patients, as we see for the outcome, percentage of women who receive two doses of IPT outcome (Figure 26). As availability of drugs also affects health worker motivation, this also improves the IPT outcome through this pathway (Figure 29).

When payments are made according to the actual schedule of payments observed during programme implementation (with delays), the 10:90 share of funds design still produces the greatest improvement in incentivised services over time (Figure 32, Figure 33), but the overall improvement observed during the simulation period is worse than the simulations without payment delays (Figures 32-37). Over the time horizon of the simulation, the 10:90 share of funds design outperforms the 25:75 share of funds design for improvement in incentivised services (Figures 26, Figure 27). However, the 25:75 design intermittently supersedes the 10:90 design, which is attributed to the threshold for ordering additional drugs in the model, observed in Figure 28 at time-step 26-28 (see diagram notes for further details).

As observed in previous simulations, percentage of women who seek facility-based deliveries is less sensitive to changes in drug availability as the target is not entirely dependent on availability of labour drugs (Figure 27). The change in payment design has little effect on community awareness of facility and services (Figure 30). An improvement is observed in perceived quality of care as a result of increasing allocation of payment towards facility operations, but the overall effect on facility-based deliveries is less acute than for the IPT outcome (Figure 31).

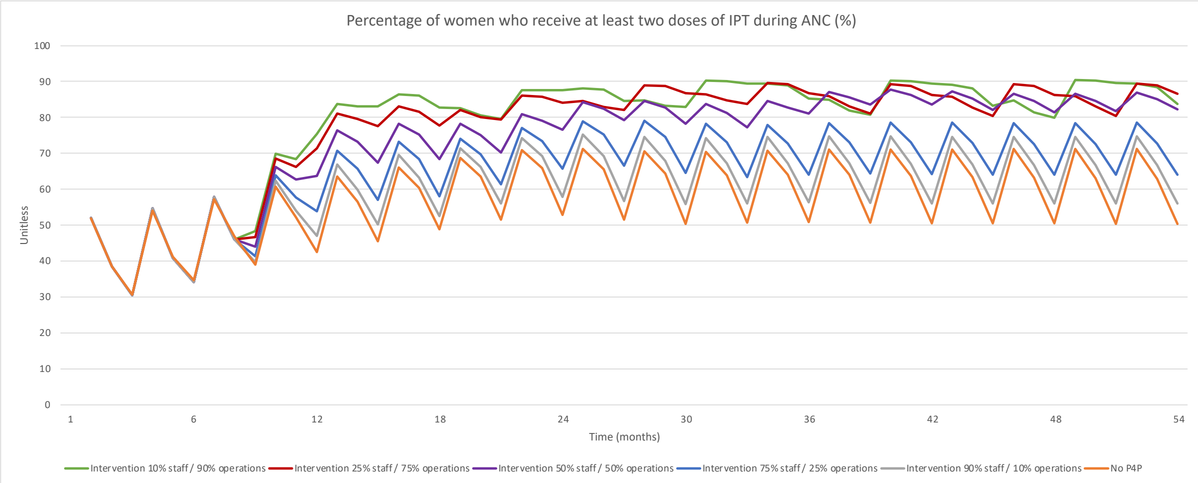


Figure 26: Model output for percentage of women who receive at least two doses of IPT during ANC (%) when (i) allocation of payments is adjusted and (ii) there are no delays in payment.

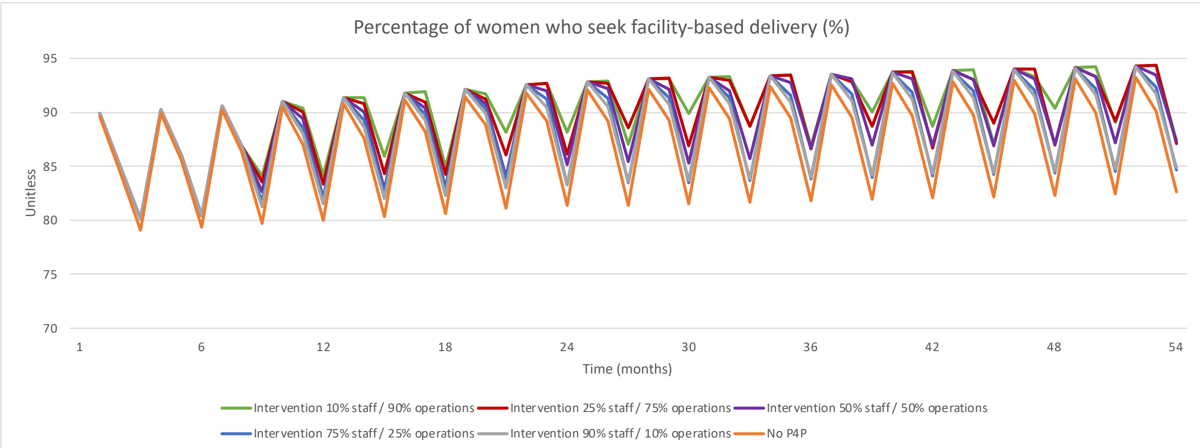


Figure 27: Model output for percentage of women who seek facility-based delivery (%) when (i) allocation of payments is adjusted and (ii) there are no delays in payment.

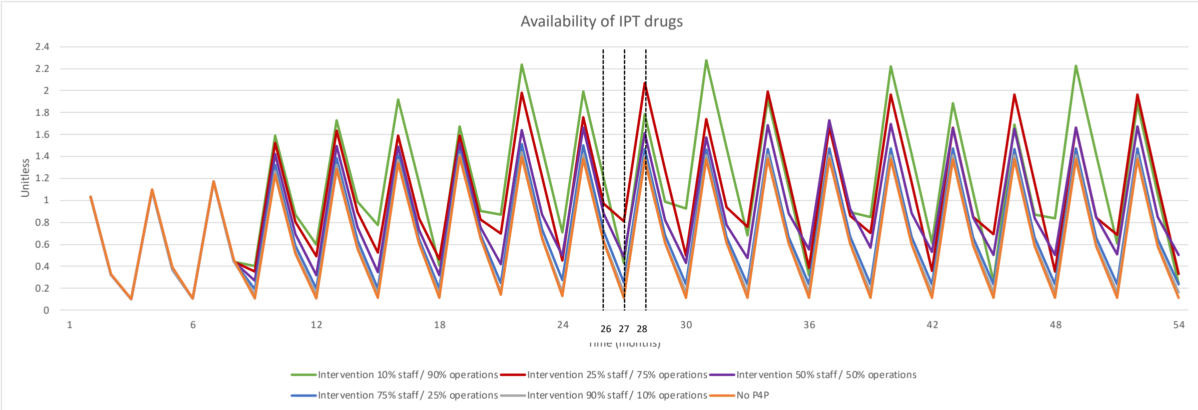


Figure 28: Model output for availability of IPT drugs when (i) allocation of payments is adjusted and (ii) there are no delays in payment.

Note: Time step 26, 27 and 28 (identified with dotted lines) illustrate where the 25:75 design performs better than the 10:90 design. In the model, when drug availability falls below '1' (when there is an inadequate supply of drugs for patients seeking care), providers can use P4P payments to purchase additional drugs. At time step 26, this criteria for purchasing additional drugs is met for the simulation with the 25:75 design but not for the simulation with the 10:90 design. At time step 27, the availability of drugs is therefore lower for the 10:90 simulation than 25:75 simulation, because providers did not use additional funding to purchase drugs at the previous time step. For both simulations at time step 27, providers use additional funding to purchase drugs and also receive a delivery from the Medical Stores Department. At time step 28, availability of drugs is still most improved under the 25:75 simulation before dropping back below the performance of the 10:90 design simulation at time step 29.

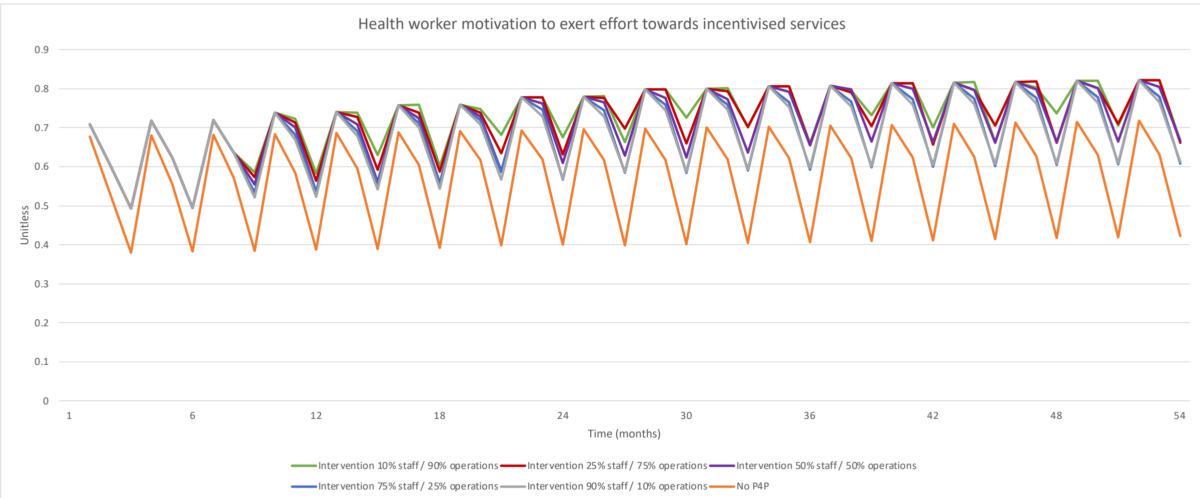


Figure 29: Model output for health worker motivation to exert effort towards incentivised services when (i) allocation of payments is adjusted and (ii) there are no delays in payment.

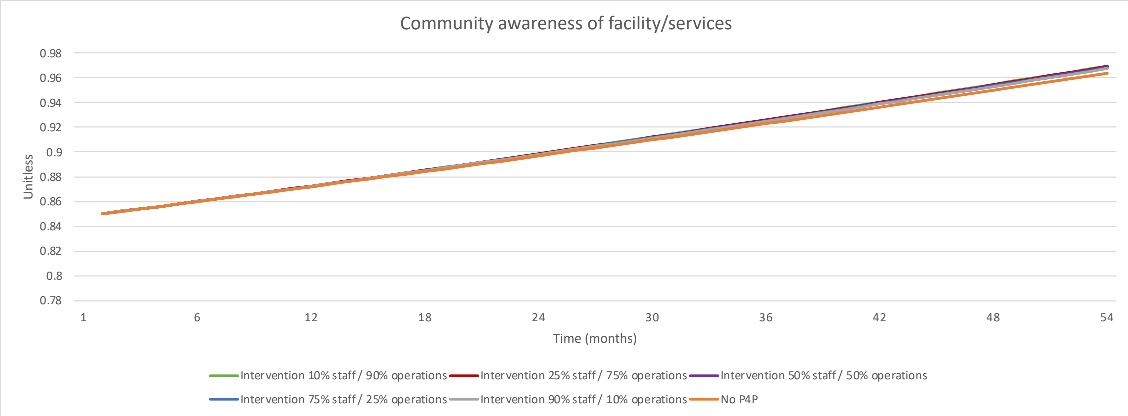


Figure 30: Model output for community awareness when (i) allocation of payments is adjusted and (ii) there are no delays in payment.

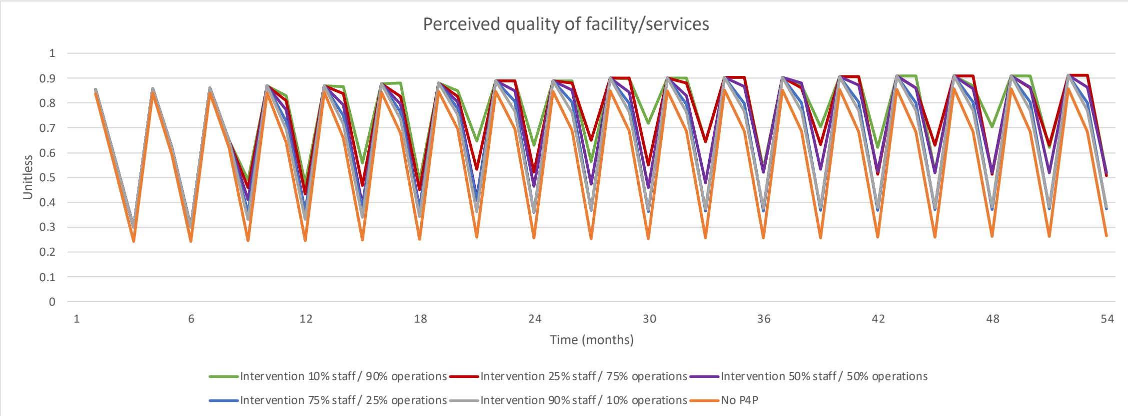


Figure 31: Model output for perceived quality of facility and services when (i) allocation of payments is adjusted and (ii) there are no delays in payment.

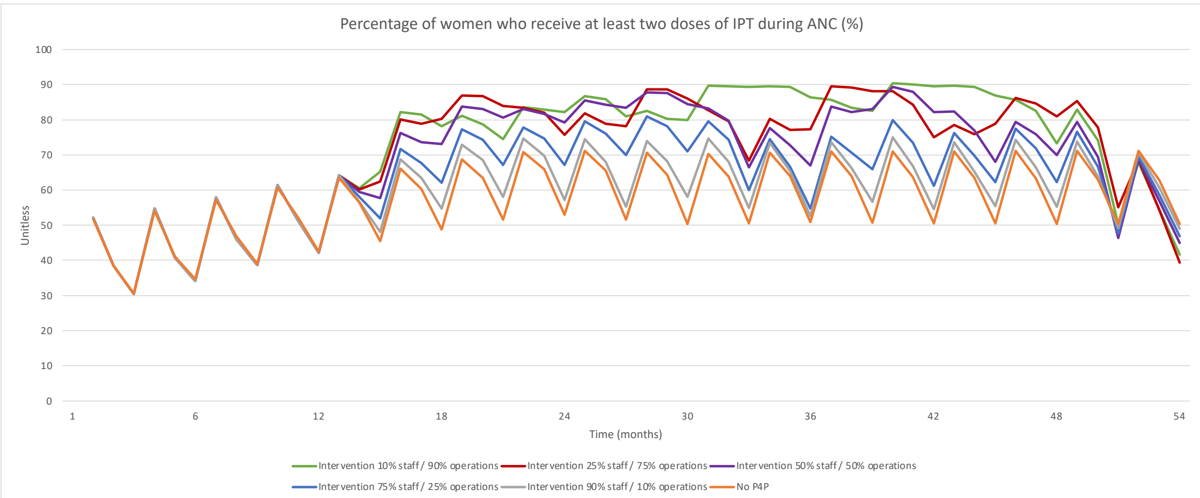


Figure 32: Model output for percentage of women who receive at least two doses of IPT during ANC (%) when (i) allocation of payments is adjusted and (ii) there are delays in payment.

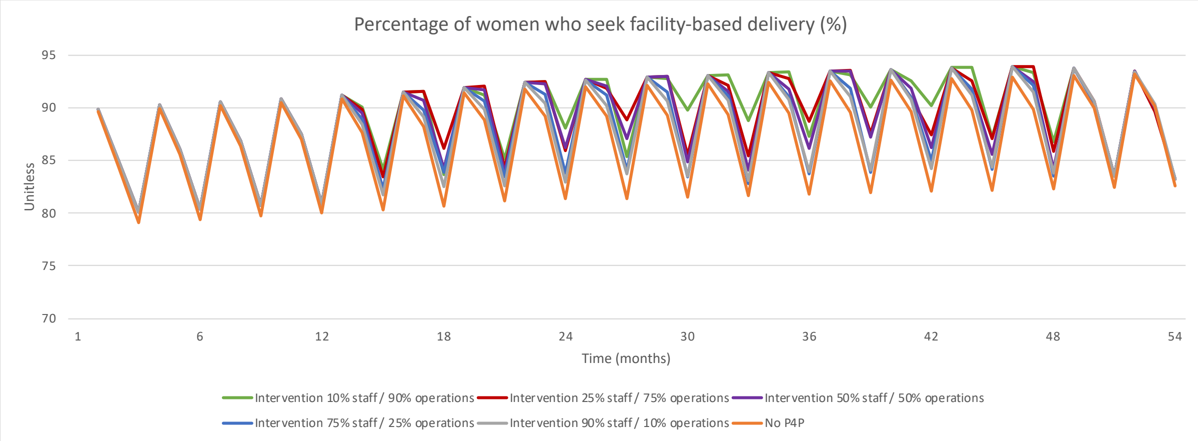


Figure 33: Model output for percentage of women who seek facility-based delivery (%) when (i) allocation of payments is adjusted and (ii) there are delays in payment.

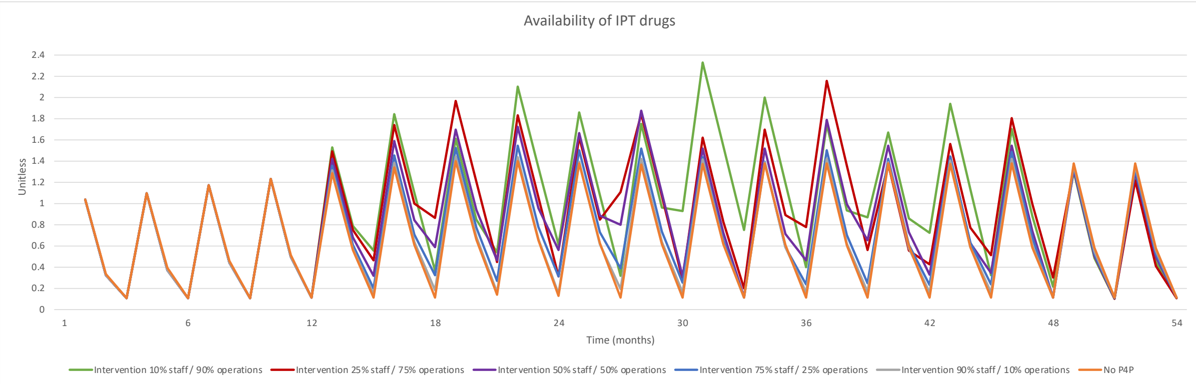


Figure 34: Model output for availability of IPT drugs when (i) allocation of payments is adjusted and (ii) there are delays in payment.

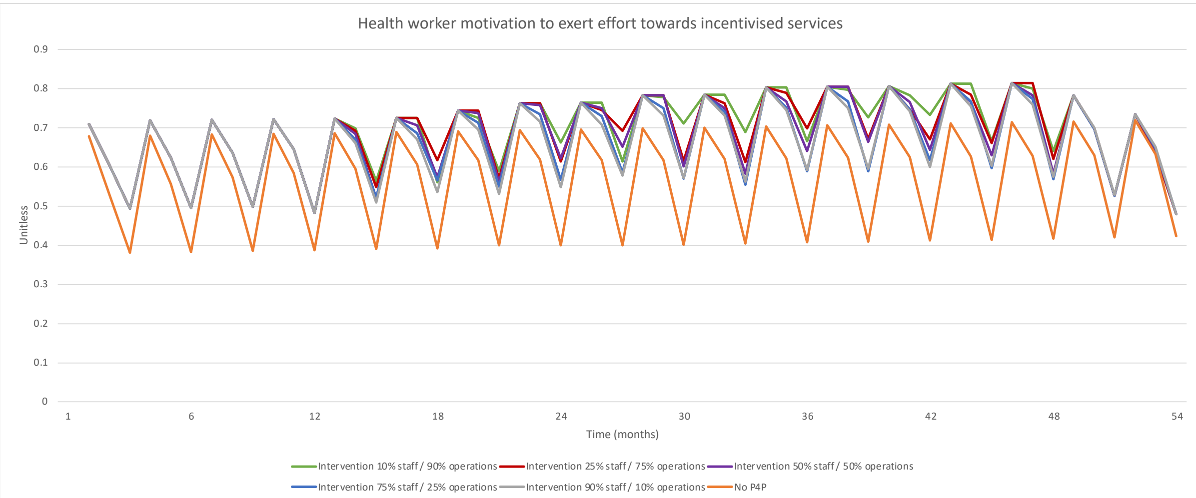


Figure 35: Model output for health worker motivation to exert effort towards incentivised services when (i) allocation of payments is adjusted and (ii) there are delays in payment.

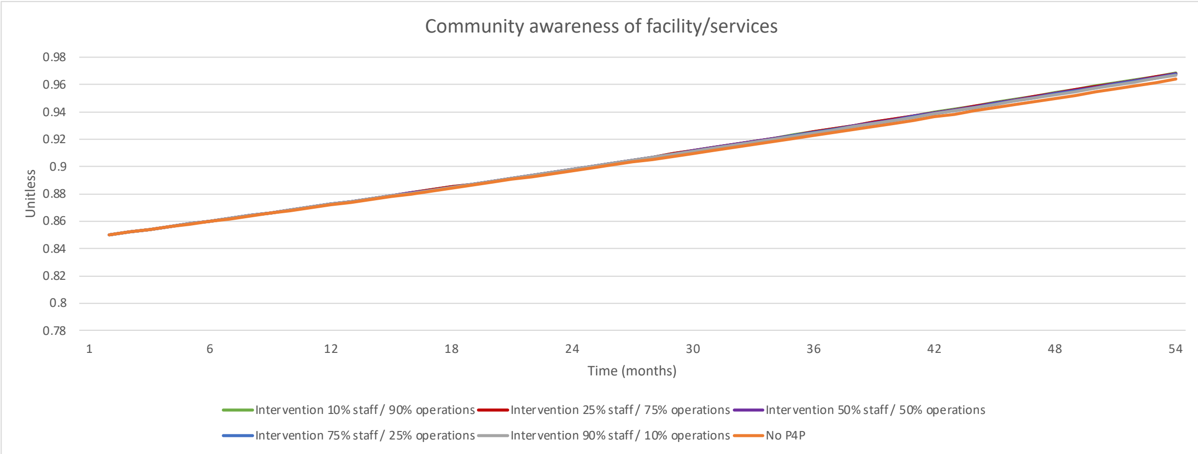


Figure 36: Model output for community awareness when (i) allocation of payments is adjusted and (ii) there are delays in payment.

Figure 37: Model output for perceived quality of facility and services when (i) allocation of payments is adjusted and (ii) there are delays in payment.

Sensitivity analysis

Further analyses were performed to examine the sensitivity of key outcomes (percentage of women who receive at least two doses of IPT during ANC, percentage of women who seek facility-based delivery) to changes in model parameters, including the health system context, (see Supplementary File 8 for details and full results) under the programme. Adjustments to community awareness, distance to facility and MSD provision of drugs prompted a ‘sensitive’ response from key outcome ‘percentage of women who receive at least two doses of IPT during ANC’ (Table 1), with low sensitivity observed across all variables for the other key outcome, percentage of women who seek facility-based delivery.

Table 1: Sensitivity analysis to examine the sensitivity of key outcomes (percentage of women who receive at least two doses of IPT during ANC, percentage of women who seek facility-based delivery) to changes in model parameters, with results indicating sensitivity to model parameters presented here.

		Percentage of women who receive at least two doses of IPT during ANC			**Percentage of women who seek facility-based delivery**		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Community_awareness(t)[(a)]	-10%	0.0	-12.8	0.2	-2.8	-3.6	-3.0
	10%	0.0	0.0	0.0	2.8	2.8	1.6
Distance_to_facility[(b)]	-10%	0.0	-12.8	0.2	-3.1	-3.9	-3.1
	10%	0.0	0.0	0.0	1.2	0.7	1.1
MSD_provision_of_IPT_ ordered[(b)]	-10%	-11.3	-8.8	-9.3	-0.3	-0.6	-0.2
	10%	14.7	7.5	7.6	0.3	0.6	0.2

Note 1: The following scale was used to determine sensitivity to changes in model variables; sensitive (5% ≤ change in outcome < 15%), very sensitive (15% ≤ change in outcome < 25%) and highly sensitive (25% ≥ change in outcome). Table cells are highlighted where outputs are categorised as sensitive (yellow), very sensitive (orange) and highly

It is clear that provision of IPT by the Medical Stores Department has a marked effect on the percentage of women who receive at least two doses of IPT during ANC, that can facilitate or prevent facility achievement of the target, producing a 'sensitive' response from this key outcome. There is some effect on percentage of women who seek facility-based delivery, through the perceived quality of facility and services pathway. For percentage of women who receive at least two doses of IPT during ANC, it is around the 'baseline' period that additional provision of drugs by the Medical Stores has the greatest effect (although a marked effect is noted throughout the simulation). This is attributed to facilities having more alternative funding (P4P incentive payments) after this period, where providers could purchase drugs, funding permitting.

Reducing the initial stock value for community awareness of services and constant variable distance to facility has a small negative effect on percentage of women who seek facility-based delivery, but surprisingly a significant short-term knock-on effect to percentage of women who receive at least two doses of IPT during ANC. Reducing the values of these parameters by a small amount resulted in 'sensitive' model behaviour and a large reduction to IPT during ANC outcome; closer inspection of model behaviour revealed that performance on the facility-based delivery target was reduced in previous months, which led to targets (in month 6 and 12) being missed. This resulted in a reduction in P4P payments and therefore in funding available to purchase medicine, leading to a reduction in percentage of women who received at least two doses of IPT during ANC. Once percentage of women who sought facility-based delivery surpassed the 85% target later in the simulation (month 18), providers just needed to maintain this average to achieve targets (Supplementary File 1, Table S1.1) and consequently P4P funding returned to previous levels.

Adjustments to all other model parameters did not elicit sensitive model behaviour, including other contextual factors, such as alternative sources of funding available to facilities and number of health workers. Altering the amount of alternative facility held funding in the model produced only minor changes in key outcomes (Supplementary File 8, Table S8.14). The amount of assumed alternative funding in the model is already limited, with modifications by 10% only adjusting the funding amount by 1%. Altering the amount prevents or facilitates providers from purchasing much needed medications, impacting facility performance on incentivised targets. Impact is greater for percentage of women who receive at least two doses of IPT during ANC, as this target is heavily dependent on availability of drugs.

Adjusting number of health workers at health facility (% of positions filled) seems, initially, to have little effect on percentage of women who receive at least two doses of IPT during ANC (Supplementary File 8, Table S8.12). Small improvement in number of health workers leads to a small initial improvement in provider readiness to deliver services, resulting in more patients treated. However, this means that facilities have a reduced level of drugs to treat patients with, balancing out the previous improvement in the IPT during ANC outcome. Adjusting number of health workers has very little effect on percentage of women who seek facility-based deliveries in the model; although this improves ability to perform outreach, the impact on community awareness of services and fraction who seek facility-based deliveries is minimal.

Discussion

This paper describes the development and application of a SDM to explore the mechanisms through which a P4P programme affected MCH service delivery outcomes in a primary care facility setting and the effect of changes in the design, implementation, and context of a P4P programme in a LMIC setting, Tanzania. The feasibility of building a SDM from a CLD is demonstrated, with the model subjected to internal, external and face validity testing.

Each time a payment is made in the model, health worker trust in the programme and motivation to deliver incentivised services gently increase, and facilities use the funding to purchase additional needed medications to support service delivery, this also improves perceived quality of care, influencing demand for institutional deliveries. Minor delays in programme payments (less than 4 months) had a minimal impact on provider performance of incentivised services. Severe delays in programme payments limited provider capacity to achieve targets for the IPT during ANC indicator, due to reduced provider purchasing power for medicines. Prolonged delays also resulted in erosion of provider trust in the programme and reduced motivation for programme participation. As the facility-based deliveries indicator was not entirely beholden to drug availability, a smaller negative effect on the outcome was observed here (through the perceived quality of care pathway, which is affected by changes in drug availability and health worker motivation). Model results show facility funding is a key driver of P4P programme success, with increased allocation of funding towards strengthening facility operations (e.g. purchasing additional drugs for service delivery) leading to greater improvements in coverage and content of care for MCH services. For the content of care indicator (2 doses of IPT during ANC), allocating a higher proportion of funding for facility operations alleviates the burden of inadequate stock and enabled provision of care for more patients, whilst also impacting health worker motivation. Allocating a higher proportion of funding for facility operations also had a positive effect on the coverage of care indicator (facility-based deliveries), through the perceived quality of care pathway, but the effect is less acute when compared to the content of care indicator.

The sensitivity analyses also identified three relevant contextual factors which have a significant effect on facility ability to achieve targets: with P4P schemes being more effective where there is adequate provision of IPT by the medical stores department at baseline, community awareness of facility services, and where facilities don't serve very dispersed/distant populations. Dependencies between the two target indicators were identified, with lower performance on facility-based deliveries resulting from lower community awareness or greater distance to facilities, leading to a reduction in performance payments which impacted provider purchasing power for ANC medication, limiting the IPT during ANC target.

In this study, programme implementation, design and context were shown to be critical determinants of provider performance during P4P. This study contributes further evidence on how, why and under what circumstances P4P does (or does not) work in LMIC settings, and how P4P design influences pathways to impact and health system outcomes, cited as critical areas for future research by a recent realist review of P4P in LMIC settings (Singh *et al.* 2021). Although the P4P programme accounted for baseline performance of facilities in setting targets, study results indicate further refinement of how funding is allocated for facilities may produce further improvements in performance; for example, those facilities who have low drug availability before programme implementation would benefit from a higher share of funds towards facility operations. Study results also indicate that the effect of certain design features is not necessarily uniform across performance targets within a given P4P scheme, while incentives for strengthening facility operations (specifically purchasing of essential medicines) was a critical pathway for improvement for the content of care indicator (2 doses of IPT during ANC) this had less impact for the coverage indicator (facility-based deliveries) which depended on demand stimulation. Model results also demonstrate that programme effects are not constant over time and can vary substantially, fluctuating in response to stimuli and events in the wider system overcoming the limitation of cross-sectional or one time evaluation assessments that struggle to identify and disentangle such dynamic system behaviour.

To our knowledge, there have been five applications of CLD and SDM methodology to explore the effect of P4P programmes in LMICs; all five articles present research with CLDs (Meker and Barlas 2015; Alonge *et al.* 2017; Renmans *et al.* 2017; Cassidy *et al.* 2021; Singh *et al.* 2021), with two articles also using the CLD to develop a SDM (Meker and Barlas 2015; Alonge *et al.* 2017). Renmans *et al.* (2017) mapped a P4P programme in Uganda, similarly identifying supervision and work environment (availability of equipment and medicine) as key mechanisms influencing health worker motivation and performance during P4P. Work environment was aggregated at a high level in the Uganda CLD, without teasing out procurement and supply chain processes, which proved to be a critical bottleneck for provider performance in the current study. Singh *et al.* (2020) used a CLD to visualise the results from a realist review of P4P in LMIC settings. The realist review and current study both highlight patient provider interactions, availability of medicine and outreach activity as pathways through which P4P programmes impact patient uptake of services; the current study contributes further evidence on attributes of provider readiness (staffing, drug availability) influencing health worker motivation, and through this pathway, provider performance during P4P. As the framework for the CLD presented in Cassidy *et al.* (2021) was used to develop the current model, there are broad similarities in their structural composition. What the current study adds, and adds to the aforementioned CLDs of P4P, is simulation of the programme over time and therefore capacity to test design and implementation changes and effect of contextual factors on health system behaviour.

The current study and Alonge *et al.* (2017) both model the effect of P4P on health worker motivation and quality of services. Although availability of drugs, an important input to facility readiness in the current study, features in Alonge *et al.* (2017) as part of an aggregated quality variable, supply chain mechanics are not present in the model. In the current study, exploration of this process proved key to identifying where bottlenecks were occurring, with reflection on how support of procurement and supply of medicines may be integrated into the design of P4P. Alonge et al. (2017) and the current study both simulate the effect of payment delays on provider performance. Whilst Alonge *et al.* (2017) assume system performance will eventually follow the same trajectory as when there are no delays in payment, in the current study, minor delays impact service delivery (procurement of additional medicines) but have minimal impact on provider motivation and trust in the programme. Alonge *et al.* (2017) do not explore the effect of major payment delays on provider behaviour and service outcomes, or impact of changes to allocation and use of payments, results which are presented in the current study. There is little overlap between the content and results from Meker and Barlas (2015) and the current study, aside from observation on the effect of P4P on providers seeking to treat more patients. The model crucially doesn't feature provider readiness to deliver services which was critical in the current SDM to understanding the effect of the programme on service delivery. The resource constraints faced by providers in lower income settings is also not accounted for in the model, making it difficult to generalise results to settings like Tanzania.

There are several limitations to this study. The model does not capture patient morbidity, mortality or health outcomes (likely to be affected by the programme), instead focussing on coverage and content of care for facility-based services as these were the primary targets measured by the programme, providing data on which to build the model. Certain community-level and care-seeking dynamics, such as the role and impact of community health workers and effect of peer networks on patient decision making, could not be captured in the current version of the model due to the level of aggregation required. The composition of heterogenous drivers for motivation was also difficult to capture in the SDM, including how individual health worker characteristics impact motivation and are affected by P4P programmes. An agent-based model that focusses on care-seeking behaviour for maternal services and health worker behaviour during P4P programmes is currently under development. Agent-based models enable simulation of individual 'agents' (patients, health workers), each with their own characteristics, that make decisions based on these attributes, the actions of other agents and events that take place in the wider system (Badham *et al.* 2018; Tracy *et al.* 2018). Agent-based models are ideally suited to capture these micro-level dynamics, such as the drivers and behaviour for individual actors, relevant for studying the impact of schemes like payment for performance. The model is currently being developed as a standalone model, with plans for a hybrid simulation that will enable analysis of both micro and macro-level health system behaviour during P4P programmes.

In the previously developed CLD (Cassidy *et al.* 2021), reporting of performance data (Figure 2) was identified as a key mechanism for provider success during the P4P programme. If data were not compiled and submitted in a timely manner, this would affect the amount of bonus payment issued to health providers, in turn impacting motivation and funding available to facilities for procurement of essential resources (e.g. medicines). There is also the potential for perverse behaviour as a result of payments linked to service provision (e.g. misreporting data) which can mask real utilisation of services and trigger unwarranted payments to health providers. The reporting mechanism was excluded in the current iteration of the model due to the primary focus on impact of the programme on supply side dynamics, but should be acknowledged as a limitation of this work as data reporting behaviour would be expected to impact key outcomes in the model (real utilisation of services, resource availability etc.).

Assumptions were made for certain model parameters and functions where it proved difficult to draw from existing data sources. Stakeholder feedback was used to shape certain assumptions to induce realistic system behaviour (such as impact of payment delays on delivery of services and trust in the programme). The model underwent various verification and validity tests (internal, external and face validity) but was not subjected to a test of generalisability, checking model robustness and ability to replicate system structure and behaviour in another setting. A test of generalisability is currently underway for the previously described CLD of the programme in Tanzania (Cassidy *et al.* 2021) to a comparable P4P programme in Zambia (Shen *et al.* 2017), with motivation to also test the generalisability of the SDM using the Zambia programme as a case study.

There is a global movement underway, with focus shifting from P4P style health system strengthening programmes towards Direct Health Facility financing (DHFF) (Kapologwe *et al.* 2019; de Walque and Kandpal 2022). In line with goals for P4P, DHFF programmes also aim to improve healthcare quality, reduce health system and service inefficiencies, and better mobilise facility and community human resources for strengthened service delivery (Mæstad *et al.* 2021); however, the design of DHFF programmes place more weight on provider autonomy and funding to improve facility operations. The results from this current study potentially support this change in programme design, with clear benefits to higher allocation of funding towards facility operations in low resource settings. Study results indicated that this funding design would have greatest improvement on content of care services such as IPT2 during ANC; coverage of services targets like facility-based deliveries would see greater improvement with focussed funding and support for outreach activities to enhance service coverage. Effectual implementation of the programme, specifically timely bonus payments, will strengthen pathways to impact for the programme to improve healthcare service delivery outcomes.

Ethics approval and consent to participate: This study received a favourable ethical opinion from the Observational/Interventions Research Ethics Committee at The London School of Hygiene and Tropical Medicine (LSHTM Ethics Ref: 16139 – 2), the Institutional Review Board at Ifakara Health Institute (IHI/IRB/No:15 -2019) and National Institute for Medical Research (NIMR/HQ/R.8a/Vol. IX/3154) in Tanzania.

Declarations of interest: None.

Consent for publication: Consent for publication on any individual data was sought from study participants (those interviewed for model validation purposes).

Availability of data and materials: Model description and equations are given in the Appendix.

Funding: This research is funded by the Medical Research Council under the Health Systems Research Initiative grant (MR/R013454/1).

Supplementary Files

Supplementary File 1 – Facility and district manager indicators and performance targets during the Pwani P4P programme in Tanzania

Supplementary File 2 – Detailed views of model sectors

Supplementary File 3 – Description of model equations

Supplementary File 4 – Description of data

Supplementary File 5 – Model calibration graphs

Supplementary File 6 – Stakeholder interview validation tool

Supplementary File 7 – Changes to model post stakeholder consultation

Supplementary File 8 – Sensitivity analysis results

References

Adam T, de Savigny D. Systems thinking for strengthening health systems in LMICs: need for a paradigm shift. *Health Policy Plan* **27**: iv1–3.

Alonge O, Lin S, Igusa T *et al.* Improving health systems performance in low- and middle-income countries: a system dynamics model of the pay-for-performance initiative in Afghanistan. *Health Policy Plan* **32**: 1417–26.

Andrade L de, Lynch C, Carvalho E *et al.* System Dynamics Modeling in the Evaluation of Delays of Care in ST-Segment Elevation Myocardial Infarction Patients within a Tiered Health System. Pizzi C (ed.). *PLoS One* **9**: e103577.

Anselmi L, Binyaruka P, Borghi J. Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania. *Implement Sci* **12**: 10.

Aryankhesal A, Sheldon TA, Mannion R *et al.* The dysfunctional consequences of a performance measurement system: the case of the Iranian national hospital grading programme. *J Health Serv Res Policy* **20**: 138–45.

Badham J, Chattoe-Brown E, Gilbert N *et al.* Developing agent-based models of complex health behaviour. *Health Place* **54**: 170–7.

Baugh Littlejohns L, Baum F, Lawless A *et al.* The value of a causal loop diagram in exploring the complex interplay of factors that influence health promotion in a multisectoral health system in Australia. *Heal Res Policy Syst* **16**: 126.

Binyaruka P, Patouillard E, Powell-Jackson T *et al.* Effect of Paying for Performance on Utilisation, Quality, and User Costs of Health Services in Tanzania: A Controlled Before and After Study. Ostermann J (ed.). *PLoS One* **10**: e0135013.

Borghi J, Binyaruka P, Mayumana I *et al.* Long-term effects of payment for performance on maternal and child health outcomes: evidence from Tanzania. *BMJ Glob Heal* **6**: e006409.

Borghi J, Chalabi Z. Square peg in a round hole: re-thinking our approach to evaluating health system strengthening in low-income and middle-income countries. *BMJ Glob Heal* **2**: e000406.

Borghi J, Mayumana I, Mashasi I *et al.* Protocol for the evaluation of a pay for performance programme in Pwani region in Tanzania: A controlled before and after study. *Implement*

Sci **8**: 80.

Cassidy R, Borghi J, Semwanga AR *et al.* How to do (or not to do)…using causal loop diagrams for health system research in low and middle-income settings. *Health Policy Plan*, DOI: 10.1093/heapol/czac064.

Cassidy R, Singh NS, Schiratti P-R *et al.* Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models. *BMC Health Serv Res* **19**: 845.

Cassidy R, Tomoaia-Cotisel A, Semwanga AR *et al.* Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach. *Soc Sci Med* **285**: 114277.

Chaerul M, Tanaka M, Shekdar A V. A system dynamics approach for hospital waste management. *Waste Manag* **28**: 442–9.

Ciplak N, Barton JR. A system dynamics approach for healthcare waste management: a case study in Istanbul Metropolitan City, Turkey. *Waste Manag Res J a Sustain Circ Econ* **30**: 576–86.

Currie DJ, Smith C, Jagals P. The application of system dynamics modelling to environmental health decision-making and policy - a scoping review. *BMC Public Health* **18**: 402.

Darabi N, Hosseinichimeh N. System dynamics modeling in health and medicine: a systematic literature review. *Syst Dyn Rev* **36**: 29–73.

Das A, Gopalan SS, Chandramohan D. Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review. *BMC Public Health* **16**: 321.

Deci EL, Ryan RM. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychol Inq* **11**: 227–68.

Diaconu K, Falconer J, Verbel A *et al.* Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev* **2021**, DOI: 10.1002/14651858.CD007899.pub3.

Diaconu K, Witter S, Binyaruka P *et al.* Appraising pay-for-performance in healthcare in low- and middle-income countries through systematic reviews: reflections from two teams. *Cochrane Database Syst Rev* **2022**, DOI: 10.1002/14651858.ED000157.

Van Herck P, De Smedt D, Annemans L *et al.* Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res* **10**: 247.

Holmstrom B, Milgrom P. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *J Law, Econ Organ* **7**: 24–52.

isee systems inc. STELLA Architect.

Kalk A, Paul FA, Grabosch E. "Paying for performance" in Rwanda: does it pay off? *Trop Med Int Heal* **15**: 182–90.

Kapologwe NA, Kalolo A, Kibusi SM *et al.* Understanding the implementation of Direct Health Facility Financing and its effect on health system performance in Tanzania: a non-controlled before and after mixed method study protocol. *Heal Res Policy Syst* **17**: 11.

Kovacs RJ, Powell-Jackson T, Kristensen SR *et al.* How are pay-for-performance schemes in healthcare designed in low- and middle-income countries? Typology and systematic literature review. *BMC Health Serv Res* **20**: 291.

Lagarde M, Wright M, Nossiter J *et al. Challenges of Payment-for-Performance in Health Care and Other Public Services – Design, Implementation and Evaluation*. Date of access: June 28, 2023. URL: https://piru.ac.uk/assets/files/Challenges%20of%20payment%20for%20performance%20in%20healthcare%20and%20other%20public%20services%20final.pdf

Lipsitz LA. Understanding Health Care as a Complex System. *JAMA* **308**: 243.

Lohmann J, Houlfort N, De Allegri M. Crowding out or no crowding out? A Self-Determination Theory approach to health worker motivation in performance-based financing. *Soc Sci Med* **169**: 1–8.

Mæstad O, Borghi J, Hassan F *et al. CMI Report: Direct Financing of Health Facilities: Experiences from Financing Reforms in Tanzania.* Date of access: December 15, 2022. URL:https://open.cmi.no/cmi-xmlui/handle/11250/2838354

Mannion R, Davies HTO. Payment for performance in health care. *BMJ* **336**: 306–8.

Meker T, Barlas Y. Dynamic Consequences of Performance-Based Payment Systems in Public Hospitals. *Syst Res Behav Sci* **32**: 459–80.

Miller G, Singer Babiarz K. *Pay-for-Performance Incentives in Low- and Middle-Income*

Country Health Programs. Date of access: June 28, 2023. URL: https://ssrn.com/abstract=2245441

Paina L, Peters DH. Understanding pathways for scaling up health services through the lens of complex adaptive systems. *Health Policy Plan* **27**: 365–73.

Patel S. Structural, institutional and organizational factors associated with successful pay for performance programmes in improving quality of maternal and child health care in low and middle income countries: a systematic literature review. *J Glob Health* **8**, DOI: 10.7189/jogh.08.021001.

Paul E, Bodson O, Ridde V. What theories underpin performance-based financing? A scoping review. *J Health Organ Manag* **35**: 344–81.

Peters DH. The application of systems thinking in health: why use systems thinking? *Heal Res Policy Syst* **12**: 51.

Pruyt E. System dynamics: a tool for modelling and testing solutions. In: de Savigny D, Blanchet K, Adam T. *Applied Systems Thinking for Health Systems Research : A Methodological Handbook*. McGraw-Hill Education, 173–94.

Renmans D, Holvoet N, Criel B. Combining Theory-Driven Evaluation and Causal Loop Diagramming for Opening the 'Black Box' of an Intervention in the Health Sector: A Case of Performance-Based Financing in Western Uganda. *Int J Environ Res Public Health* **14**: 1007.

Rwashana AS, Nakubulwa S, Nakakeeto-Kijjambu M *et al.* Advancing the application of systems thinking in health: understanding the dynamics of neonatal mortality in Uganda. *Heal Res Policy Syst* **12**: 36.

Sahin O, Salim H, Suprun E *et al.* Developing a Preliminary Causal Loop Diagram for Understanding the Wicked Complexity of the COVID-19 Pandemic. *Systems* **8**: 20.

de Savigny D, Blanchet K, Adam T. Conclusion: mapping methods to research challenges. In: de Savigny D, Blanchet K, Adam T. *Applied Systems Thinking for Health Systems Research : A Methodological Handbook*. McGraw-Hill Education, 226–7.

Sekwat A. Principal-agent theory: a framework for improving health care reform in Tennessee. *J Health Hum Serv Adm* **22**: 277–91.

Semwanga AR, Nakubulwa S, Adam T. Applying a system dynamics modelling approach to explore policy options for improving neonatal health in Uganda. *Heal Res Policy Syst* **14**: 35.

Shen GC, Nguyen HTH, Das A *et al.* Incentives to change: effects of performance-based financing on health workers in Zambia. *Hum Resour Health* **15**: 20.

Sherry TB, Bauhoff S, Mohanan M. Multitasking and Heterogeneous Treatment Effects in Pay-for-Performance in Health Care: Evidence from Rwanda. *Am J Heal Econ* **3**: 192–226.

Singh NS, Kovacs RJ, Cassidy R *et al.* A realist review to assess for whom, under what conditions and how pay for performance programmes work in low- and middle-income countries. *Soc Sci Med* **270**: 113624.

Skivington K, Matthews L, Simpson SA *et al.* A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ*: n2061.

Tracy M, Cerdá M, Keyes KM. Agent-Based Modeling in Public Health: Current Applications and Future Directions. *Annu Rev Public Health* **39**: 77–94.

Turcotte-Tremblay A-M, Gali Gali IA, Ridde V. An Exploration of the Unintended Consequences of Performance-Based Financing in 6 Primary Healthcare Facilities in Burkina Faso. *Int J Heal Policy Manag*, DOI: 10.34172/ijhpm.2020.83.

de Walque D, Kandpal E. Reviewing the evidence on health financing for effective coverage: do financial incentives work? *BMJ Glob Heal* **7**: e009932.

Anselmi L, Binyaruka P, Borghi J. 2017. Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania. *Implement Sci* **12**: 10.

Bhatnagar A, George AS. 2016. Motivating health workers up to a limit: partial effects of performance-based financing on working environments in Nigeria. *Health Policy Plan* **31**: 868–77.

Deci EL, Ryan RM. 1985. Cognitive Evaluation Theory. *Intrinsic Motivation and Self-Determination in Human Behavior*. Boston, MA: Springer US.

Deci EL, Ryan RM. 2000. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychol Inq* **11**: 227–68.

Lohmann J, Houlfort N, De Allegri M. 2016. Crowding out or no crowding out? A Self-

Determination Theory approach to health worker motivation in performance-based financing. *Soc Sci Med* **169**: 1–8.

Munger MC. 2015. Public Choice Economics. *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, 534–9.

Ogundeji YK, Sheldon TA, Maynard A. 2018. A reporting framework for describing and a typology for categorizing and analyzing the designs of health care pay for performance schemes. *BMC Health Serv Res* **18**: 686.

Paul E, Bodson O, Ridde V. 2021. What theories underpin performance-based financing? A scoping review. *J Health Organ Manag* **35**: 344–81.

Preker AS, Liu X, Velenyi EV *et al.* 2007. *Public Ends, Private Means : Strategic Purchasing of Health Services.* World Bank (ed.). Washington DC.

Sekwat A. 2000. Principal-agent theory: a framework for improving health care reform in Tennessee. *J Health Hum Serv Adm* **22**: 277–91.

8.3 Supplementary material to research paper 4

Supplementary File 1 - Facility and district manager indicators and performance targets during the Pwani P4P programme in Tanzania

Table S1.1: Facility coverage, content of care and HMIS strengthening indicators and performance targets set during the Pwani P4P programme in Tanzania.

Indicator	**Measure**	**Baseline coverage (previous cycle)**				
		0-20%	**21-40%**	**41-70%**	**71-85%**	**85%+***
Coverage indicators						
% of institutional deliveries	Percentage point increase	15%	10%	5%	5%	Maintain
% of mothers attending a facility within 7 days of delivery	Percentage point increase	15%	10%	5%	5%	Maintain
% of women using long term contraceptives	Percentage point increase	20%	15%	10%	Maintain above 71%	Maintain
% children under 1 year receiving Penta3 vaccine	Overall result	50%	65%	75%	80% +	Maintain
% children under 1 year receiving measles vaccine	Overall result	50%	65%	75%	80% +	Maintain
Content of care indicators						
% ANC clients receiving IPT2	Overall result	80%	80%	80%	80%+	Maintain above 80%
% HIV+ ANC clients on ART	Overall result	40%	60%	75%	75%+	Maintain

% of newborns receiving polio vaccine (OPV0)	Overall result	60%	75%	80%	80%+	Maintain
HMIS strengthening						
HMIS monthly reports correctly filled and submitted on time to CHMT	Overall result	100%	100%	100%	100%	100%

Notes to Table: +*85% or more. Antiretroviral therapy (ART), Antenatal care (ANC), Council Health Management Team (CHMT), Management Information System (HMIS), Human Immunodeficiency Virus (HIV), Intermittent Preventative Treatment (IPT2).

Source: Binyaruka *et al.* (2015), Ministry of Health and Social Welfare (2012) and Cassidy *et al.* (2021).

Table S1.2: Council Health Management Team and Regional Health Management Team performance

CHMT/RHMT/Both	**Indicator**	**Measure**
	Coverage indicators	
Both	% of maternal and newborn deaths that are appropriately audited on time	Overall result
	Health system strengthening	
CHMT	% of facilities reporting stock-outs of either one or more of the tracer medicines in a specified period (< 8 days)	Overall result
	HMIS strengthening	
CHMT	% of facilities included in the HMIS monthly reports exported through DHIS to RHMT in timely manner	Overall result
	Management	
RHMT	Submission to MoHSW of a Semi-Annual Regional Health Profile report, based on DHIS	Overall result
CHMT	% of facilities receiving a copy of a Quarterly District Health Profile report, based on DHIS	Overall result

	Overall	
Both	Overall performance along P4P facility-based indicators	Overall result

Notes to Table: Council Health Management Team (CHMT), District Health Information Software (DHIS), Health Management Information System (HMIS), Ministry of Health and Social Welfare (MoHSW), Regional Health Management Team (RHMT).

Source: Ministry of Health and Social Welfare (2012) and Cassidy *et al.* (2021).

Supplementary File 2 – Detailed views of model sectors

This Supplementary file provides detailed views of each model sector; **Population, Demand and Services, Facility Commodities**, **Facility Operations, Facility Funding and District Manager Operations**. Supplementary files 3 and 4 provide further information on model equations and descriptions of data used.

N.B. The **Demand and Services** and **Facility Commodities** sectors are each presented here with two sub-sectors, describing functions related to the two different health services of interest in the model (percentage of women who received two doses of IPT during ANC and percentage of women who had a facility-based delivery). This was an artifact of model development (for ease of viewing and analysis), for all intents and purposes they can be considered 'subsectors' of a single model sector.

Population sector

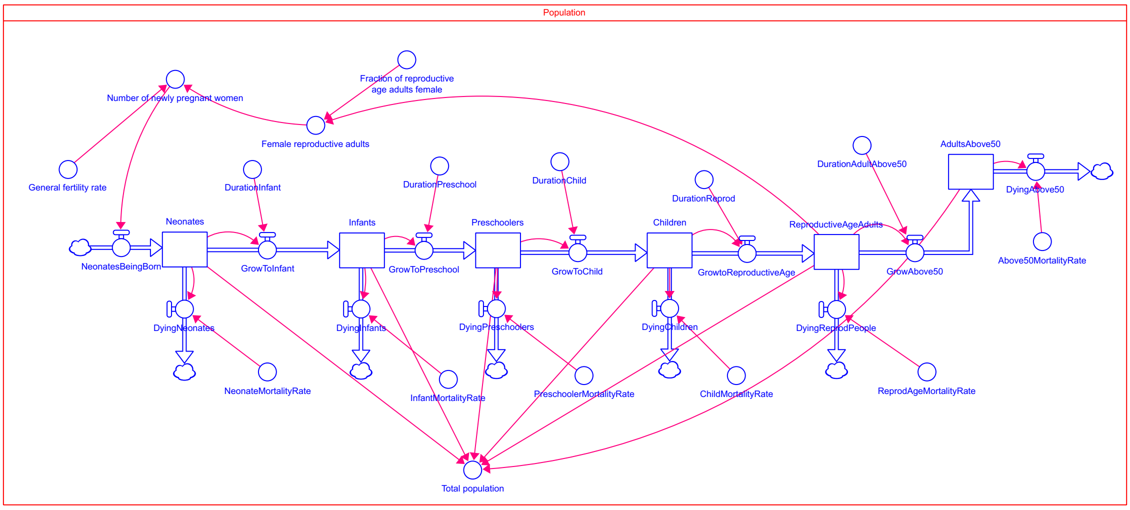


Figure S2.1: Detailed model view of Population sector

Demand and Services sector (ANC)

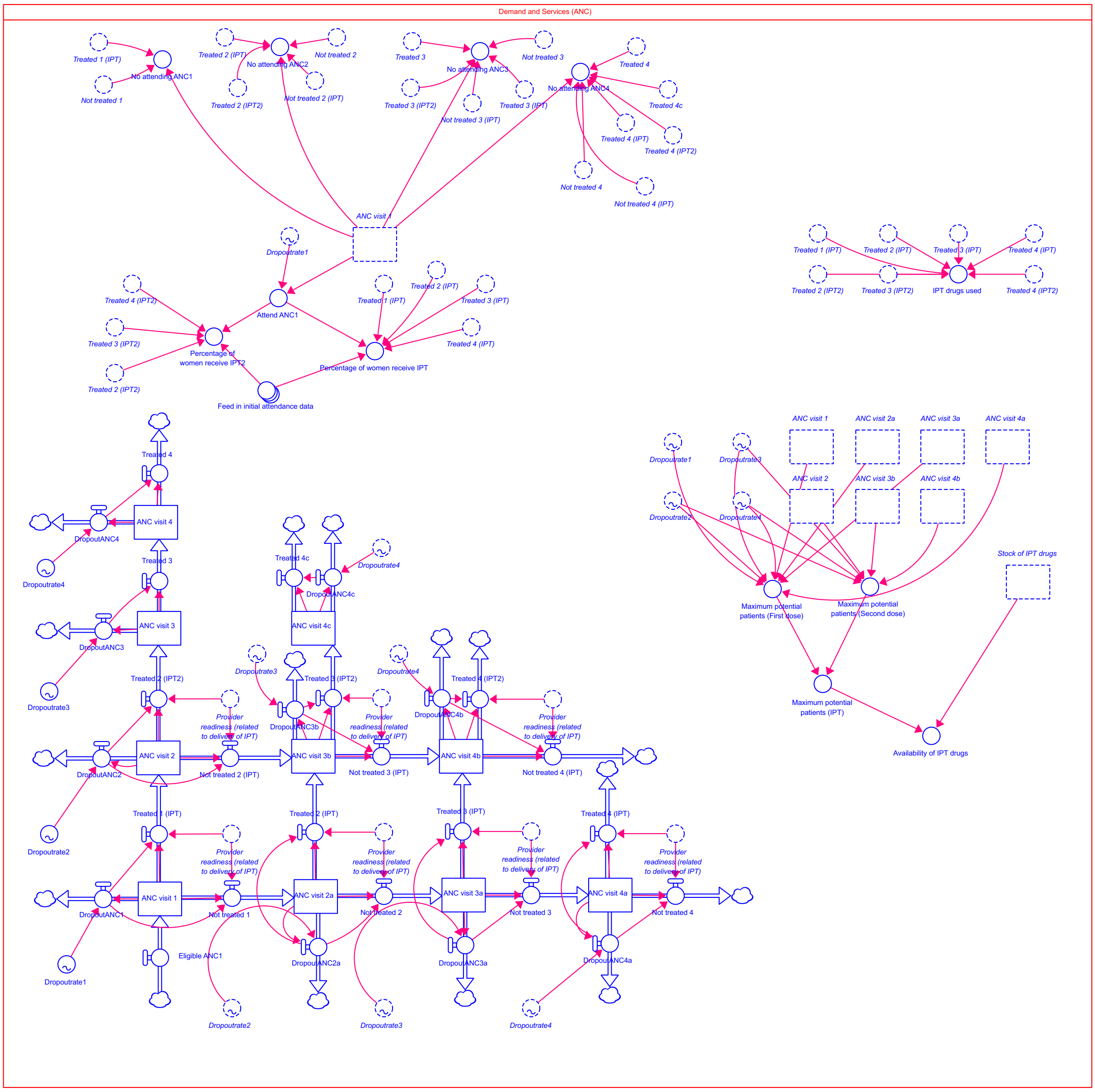


Figure S2.2: Detailed model view of Demand and Services sector

Demand and Services sector (Facility-based deliveries)

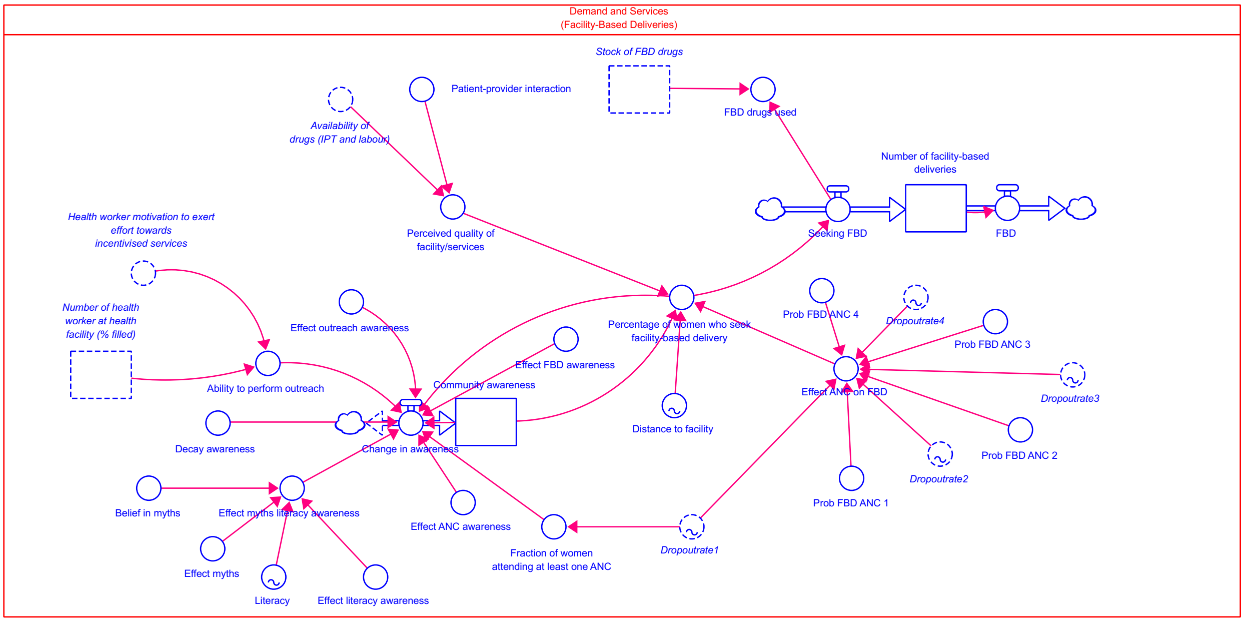


Figure S2.3: Detailed model view of Demand and Services sector

Facility Commodities sector (ANC)

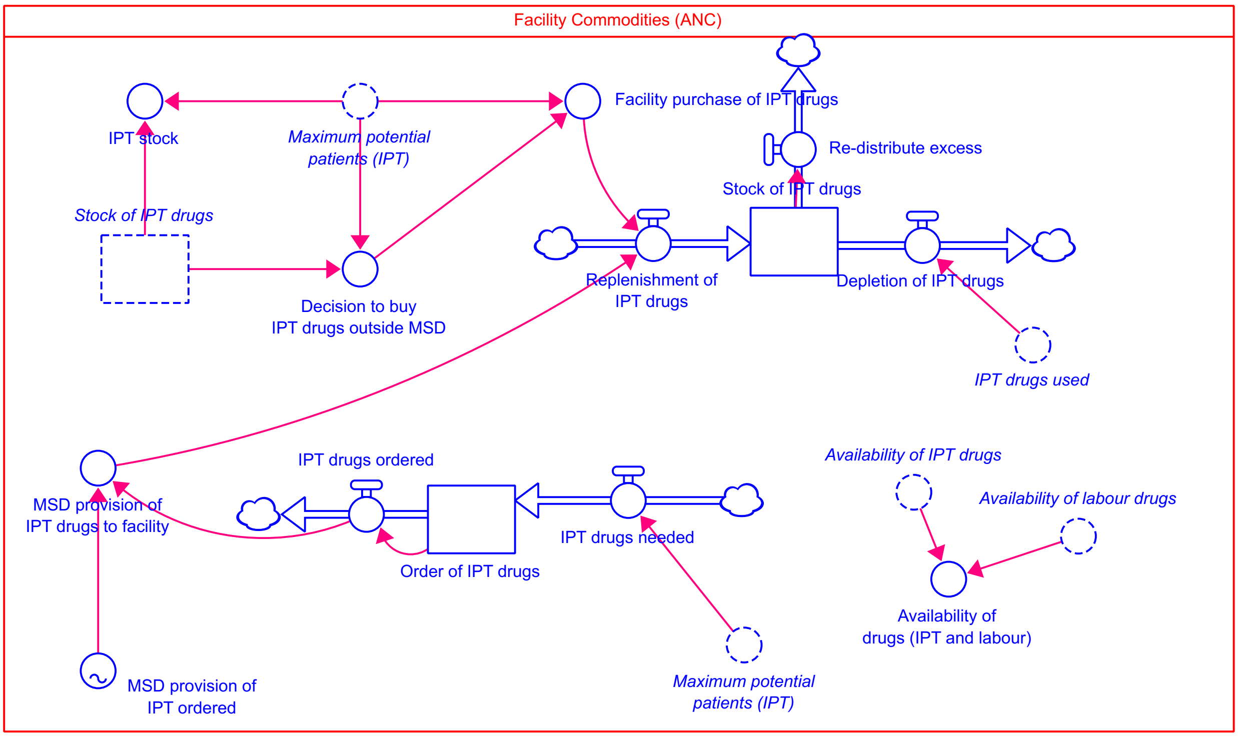


Figure S2.4: Detailed model view of Facility Commodities sector

Facility Commodities sector (Facility-based deliveries)

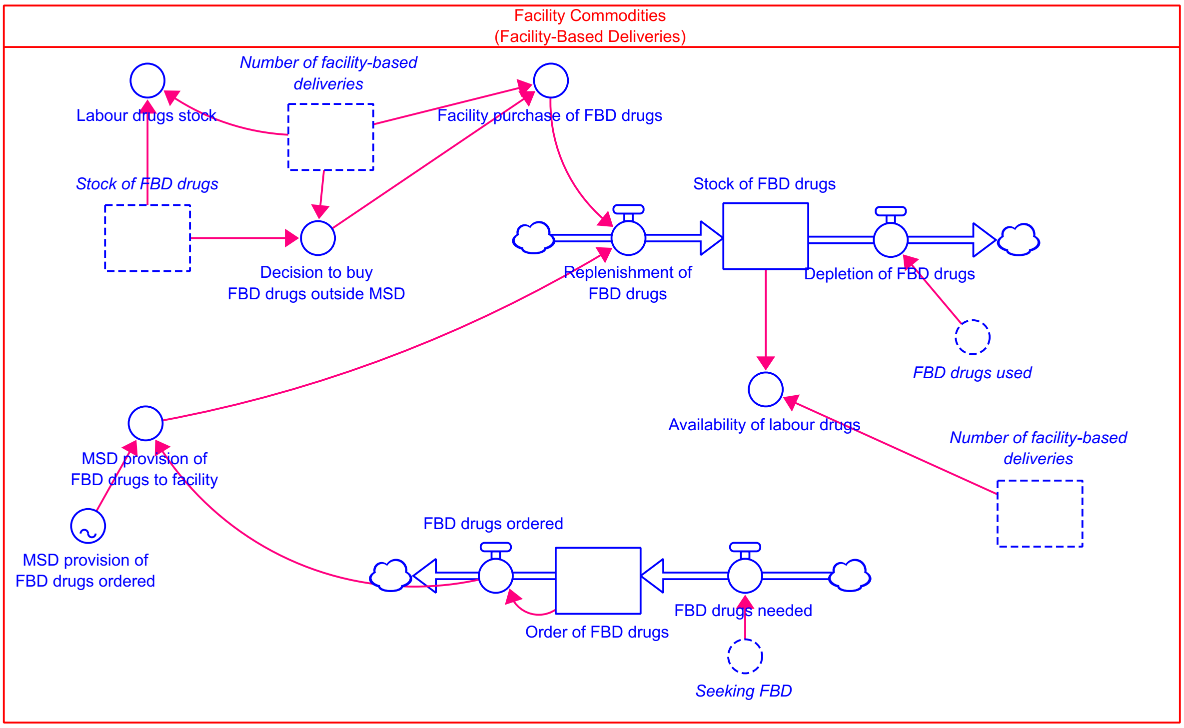


Figure S2.5: Detailed model view of Facility Commodities sector

Facility Operations sector

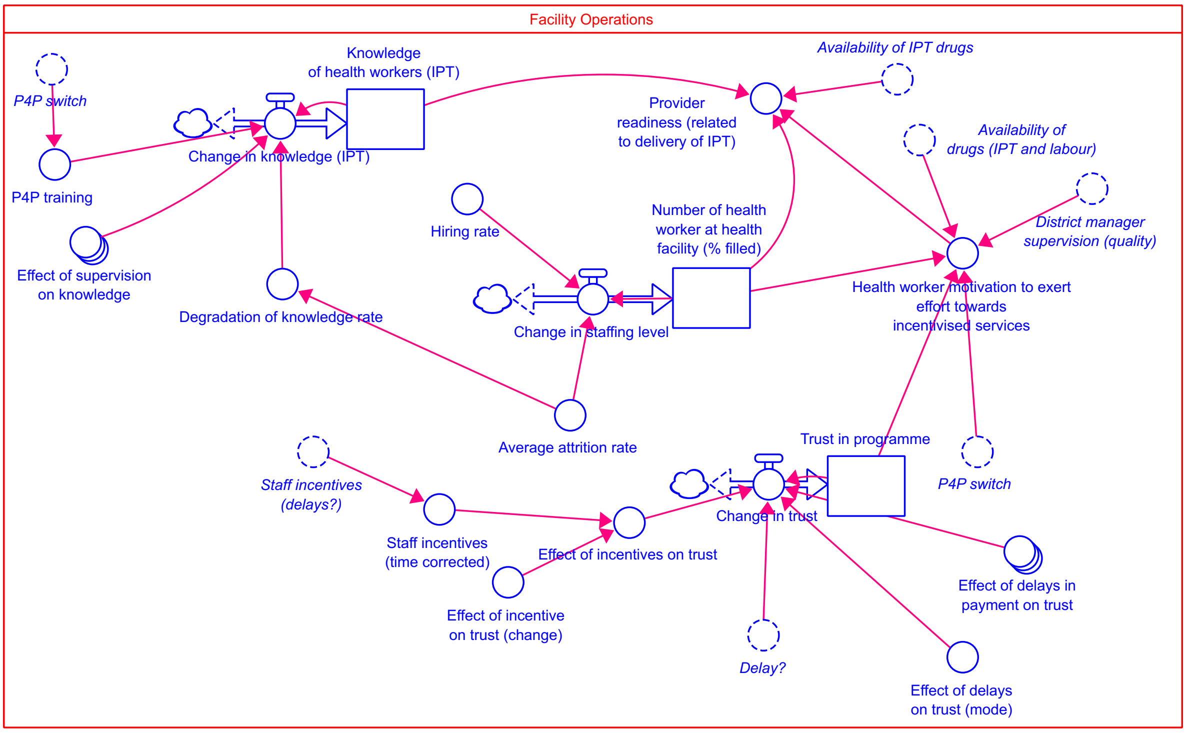


Figure S2.6: Detailed model view of Facility Operations sector

Facility Funding sector

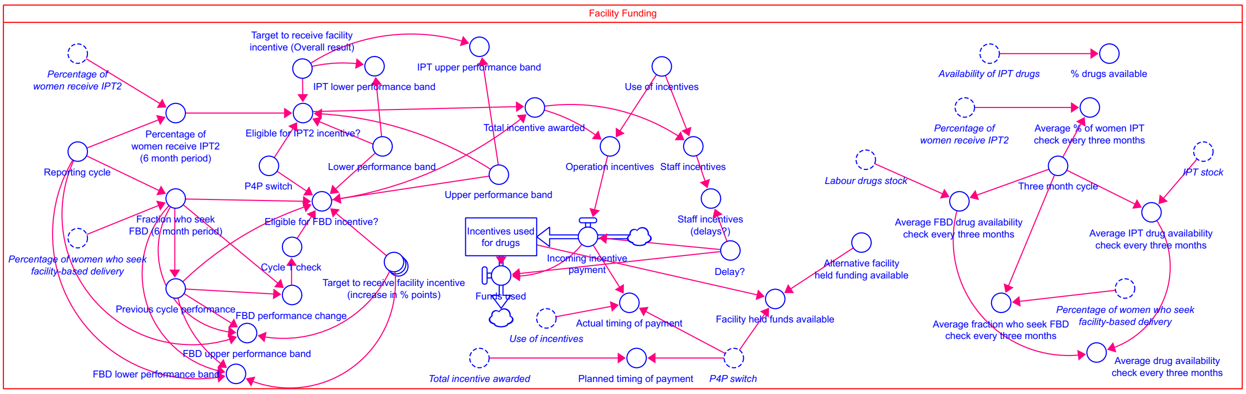


Figure S2.7: Detailed model view of Facility Funding sector

District Manager Operations sector

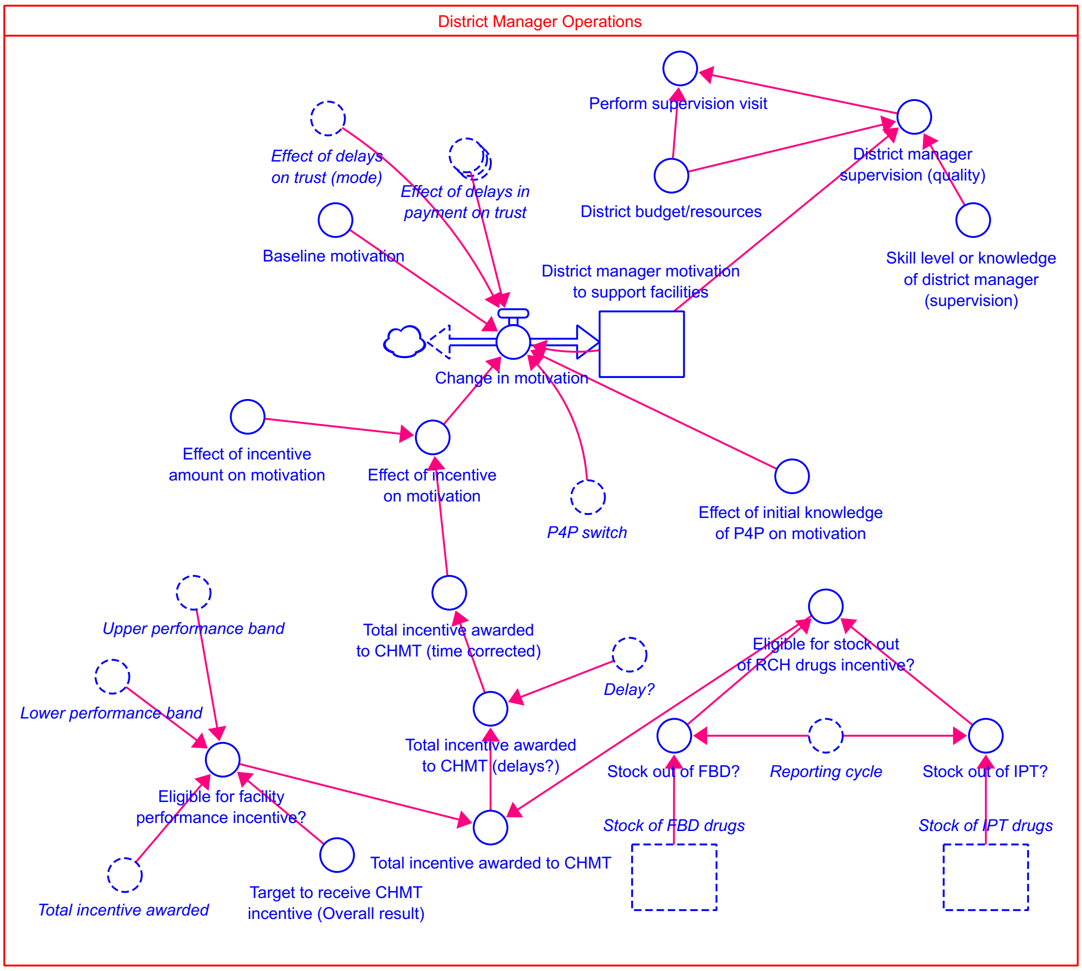


Figure S2.8: Detailed model view of District Manager Operations sector

Supplementary File 3 - Description of COSMIC Model Equations

Model summary description
Time horizon for model: 54 months
Time step for model: 1 month
Time units: Months

Model equations and documentation
Note 1: Elements highlighted in green reflect model equations and variables that describe how the P4P programme was incorporated and modelled in the simulation model.
Note 2: Elements highlighted in blue reflect where adjustments were made to simulate the described model scenario, to explore the effect of changes in programme implementation (payments made on time vs. with delays).
Note: 3: Elements highlighted in red reflect where adjustments were made to simulate the described model scenario, to explore the effect of changes in programme design (adjusting the share of funds between staff incentives and funds to strengthen facility operations).

Population sector

Stocks
Neonates(t) = Neonates(t - dt) + (NeonatesBeingBorn - GrowtoInfant - DyingNeonates) * dt
INIT Neonates = 4,740 {people}
DOCUMENT: Number of neonates (infants who are less than a month old).
INFLOWS:
NeonatesBeingBorn = Number of newly pregnant women {people/month}
OUTFLOWS:
GrowtoInfant = Neonates/DurationInfant {people/month}
DyingNeonates = Neonates*NeonateMortalityRate {people/month}

Infants(t) = Infants(t - dt) + (GrowtoInfant - DyingInfants - GrowtoPreschool) * dt
INIT Infants = 49,500 {people}
DOCUMENT: Number of infants over 1 month and up to 1 year.
INFLOWS:
GrowtoInfant = Neonates/DurationInfant {people/month}
OUTFLOWS:
DyingInfants = Infants*InfantMortalityRate {people/month}
GrowtoPreschool = Infants/DurationPreschool {people/month}

Preschoolers(t) = Preschoolers(t - dt) + (GrowtoPreschool - DyingPreschoolers - GrowtoChild) * dt
INIT Infants = 175,000 {people}
DOCUMENT: Number of preschoolers over 1 year and up to 5 years.
INFLOWS:
GrowtoPreschool = Infants/DurationPreschool {people/month}
OUTFLOWS:
DyingPreschoolers = Preschoolers*PreschoolerMortalityRate {people/month}
GrowtoChild = Preschoolers/DurationChild {people/month}

Children(t) = Children(t - dt) + (GrowtoChild - GrowtoReproductiveAge - DyingChildren) * dt
INIT Children = 284,308 {people}
DOCUMENT: Number of children aged between 5 and up to 15 years.
INFLOWS:
GrowtoChild = Preschoolers /DurationChild {people/month}
OUTFLOWS:
GrowtoReproductiveAge = Children/DurationReprod {people/month}
DyingChildren = Children*ChildMortalityRate {people/month}

ReproductiveAgeAdults(t) = ReproductiveAgeAdults(t - dt) + (GrowtoReproductiveAge - Growabove50 - DyingReprodPeople) * dt
INIT ReproductiveAgeAdults = 509,934 {people}
DOCUMENT: This is reproductive population aged between 15 and up to 50 years.

INFLOWS:
GrowtoReproductiveAge = Children/DurationReprod {people/month}
OUTFLOWS:
Growabove50 = ReproductiveAgeAdults/DurationAdultAbove50 {people/month}
DyingReprodPeople = ReproductiveAgeAdults*ReprodAgeMortalityRate {people/month}

AdultsAbove50(t) = AdultsAbove50(t - dt) + (Growabove50 - DyingAbove50) * dt
INIT AdultsAbove50 = 150,094 {people}
DOCUMENT: Number of adults above 50 years.
INFLOWS:
Growabove50 = ReproductiveAgeAdults/DurationAdultAbove50 {people/month}
OUTFLOWS:
DyingAbove50 = AdultsAbove50*Above50MortalityRate {people/month}

Auxiliary variables
NeonateMotalityRate = 0.026/12 {1/month}
DOCUMENT: The rate at which neonates die per month.

InfantMortalityRate = 0.025/12 {1/month}
DOCUMENT: The rate at which infants die per month.

PreschoolerMortalityRate = 0.032/12 {1/month}
DOCUMENT: The rate at which children of preschool age die per month.

ChildMortalityRate = 0.0024/12 {1/month}
DOCUMENT: The rate at which children die per month.

ReprodAgeMortalityRate = 0.005/12 {1/month}
DOCUMENT: The rate at which people from reproductive age group die per month.

Above50MortalityRate = 0.0575/12 {1/month}
DOCUMENT: The rate at which people aged 50 and above die per month.

GeneralFertilityRate = 0.2148/12 {1/month}
DOCUMENT: This is the birth rate per 1000 people.

DurationInfant = 1 {month}
DOCUMENT: Time it takes a neonate to become an infant.

DurationPreschool = 11 {month}
DOCUMENT: Time it takes an infant to become a preschooler.

DurationChildren = 48 {month}
DOCUMENT: Duration it takes for a preschooler to become a child.

DurationReprod = 120 {month}
DOCUMENT: Duration it takes for a child to become reproductive.

DurationAdultAbove50 = 420 {month}
DOCUMENT: Duration it takes for an adult to stop being reproductive.

FractionofReproductiveAgeAdultsFemale = 0.52 {unitless}
DOCUMENT: Fraction of reproductive age adults who are female per year.

Femalereproductiveadults = FractionofReproductiveAgeAdultsFemale * ReproductiveAgeAdults(t) {people}
DOCUMENT: Number of female reproductive age adults.

Numberofnewlypregnantwomen = GeneralFertilityRate * Femalereproductiveadults {people}
DOCUMENT: Number of newly pregnant women of reproductive age per month.

Stocks
ANC visit 1(t) = ANC visit 1(t - dt) + (Eligible ANC1 – DropoutANC1 - Treated 1 (IPT) – Not treated 1) * dt
INIT ANC visit 1 = 4750 {people}
DOCUMENT: Number of women eligible to attend first ANC visit. Either drop out, receive a dose of malaria treatment or do not receive treatment depending on provider readiness and dropout rate.
INFLOWS:
Eligible ANC1 = Number of newly pregnant women {people/month}
OUTFLOWS:
DropoutANC1 = ANC visit 1*Dropoutrate1 {people/month}
Treated 1 (IPT) = (ANC visit 1-DropoutANC1)*Provider readiness (IPT) {people/month}
Not treated 1 = (ANC visit 1-DropoutANC1)*(1- Provider readiness (IPT)) {people/month}

ANC visit 2(t) = ANC visit 2(t - dt) + (Treated 1 (IPT) – DropoutANC2 – Treated 2 (IPT2) - Not treated 2 (IPT)) * dt
INIT ANC visit 2 = 470 {people}
DOCUMENT: Number of women eligible to attend second ANC visit (having received their first dose of malaria treatment in previous visit). Either drop out, receive a second dose of malaria treatment or receive no treatment depending on provider readiness and dropout rate.
INFLOWS:
Treated 1 (IPT) = (ANC visit 1-DropoutANC1)*Provider readiness (IPT) {people/month}
OUTFLOWS:
DropoutANC2 = ANC visit 2 * Dropoutrate2 {people/month}
Treated 2 (IPT2) = (ANC visit 2-DropoutANC2)*Provider readiness (IPT) {people/month}
Not treated 2 (IPT) = (ANC visit 2-DropoutANC2)*(1- Provider readiness (IPT)) {people/month}

ANC visit 2a(t) = ANC visit 2a(t - dt) + (Not treated 1 – DropoutANC2a – Treated 2 (IPT) - Not treated 2) * dt
INIT ANC visit 2a = 4190 {people}
DOCUMENT: Number of women eligible to attend second ANC visit (who did not receive their first dose of malaria treatment in previous visit). Either drop out, receive a first dose of malaria treatment or receive no treatment depending on provider readiness and dropout rate.
INFLOWS:
Not treated 1 = (ANC visit 1-DropoutANC1)*(1- Provider readiness (IPT)) {people/month}
OUTFLOWS:
DropoutANC2a = ANC visit 2a * Dropoutrate2 {people/month}
Treated 2 (IPT) = (ANC visit 2a-DropoutANC2a)*Provider readiness (IPT) {people/month}
Not treated 2 = (ANC visit 2a-DropoutANC2a)*(1- Provider readiness (IPT)) {people/month}

ANC visit 3(t) = ANC visit 3(t - dt) + (Treated 2 (IPT2) - DropoutANC3 – Treated 3) * dt
INIT ANC visit 3 = 156 {people}
DOCUMENT: Number of women eligible to attend third ANC visit (who have had their second dose of IPT). Either drop out or become eligible to attend ANC visit 4 depending on dropout rate.
INFLOWS:
Treated 2 (IPT2) = (ANC visit 2-DropoutANC2)*Provider readiness (IPT) {people/month}
OUTFLOWS:
DropoutANC3 = ANC visit 3 * Dropoutrate3 {people/month}
Treated 3 = (ANC visit 3-DropoutANC3) {people/month}

ANC visit 3a(t) = ANC visit 3a(t - dt) + (Not treated 2 – DropoutANC3a – Treated 3 (IPT) - Not treated 3) * dt
INIT ANC visit 3a = 2700 {people}
DOCUMENT: Number of women eligible to attend third ANC visit (who did not receive their first dose of malaria treatment in previous visit). Either drop out, receive a first dose of malaria treatment or receive no treatment depending on provider readiness and dropout rate.
INFLOWS:
Not treated 2 = (ANC visit 2a-DropoutANC2a)*(1- Provider readiness (IPT)) {people/month}
OUTFLOWS:

DropoutANC3a = ANC visit 3a * Dropoutrate3 {people/month}
Treated 3 (IPT) = (ANC visit 3a-DropoutANC3a)*Provider readiness (IPT) {people/month}
Not treated 3 = (ANC visit 3a-DropoutANC3a)*(1- Provider readiness (IPT)) {people/month}

ANC visit 3b(t) = ANC visit 3b(t - dt) + (Not treated 2 (IPT) + Treated 2 (IPT) – DropoutANC3b – Treated 3 (IPT2) - Not treated 3 (IPT)) * dt
INIT ANC visit 3b = 1650 {people}
DOCUMENT: Number of women eligible to attend third ANC visit (who did receive their second dose of malaria treatment in previous visit). Either drop out, receive a second dose of malaria treatment or do not receive a second dose of treatment depending on provider readiness and dropout rate.
INFLOWS:
Not treated 2 (IPT) = (ANC visit 2-DropoutANC2)*(1- Provider readiness (IPT)) {people/month}
Treated 2 (IPT) = (ANC visit 2a-DropoutANC2a)*Provider readiness (IPT) {people/month}
OUTFLOWS:
DropoutANC3b = ANC visit 3b * Dropoutrate3 {people/month}
Treated 3 (IPT2) = (ANC visit 3b-DropoutANC3b)*Provider readiness (IPT) {people/month}
Not treated 3 (IPT) = (ANC visit 3b-DropoutANC3b)*(1- Provider readiness (IPT)) {people/month}

ANC visit 4(t) = ANC visit 4(t - dt) + (Treated 3 – DropoutANC4 – Treated 4) * dt
INIT ANC visit 4 = 1000 {people}
DOCUMENT: Number of women eligible to attend forth ANC visit (who have had their second dose of IPT). Either drop out or become eligible to attend ANC visit 5.
INFLOWS:
Treated 3 = (ANC visit 3-DropoutANC3) {people/month}
OUTFLOWS:
DropoutANC4 = ANC visit 4 * Dropoutrate4 {people/month}
Treated 4 = (ANC visit 4 – DropoutANC4) {people/month}

ANC visit 4a(t) = ANC visit 4a(t - dt) + (Not treated 3 – DropoutANC4a – Treated 4 (IPT) - Not treated 4) * dt
INIT ANC visit 4a = 740 {people}
DOCUMENT: Number of women eligible to attend fourth ANC visit (who did not receive their first dose of malaria treatment in previous visit). Either drop out, receive a first dose of malaria treatment or receive no treatment depending on provider readiness and dropout rate.
INFLOWS:
Not treated 3 = (ANC visit 3a-DropoutANC3a)*(1- Provider readiness (IPT)) {people/month}
OUTFLOWS:
DropoutANC4a = ANC visit 4a * Dropoutrate4 {people/month}
Treated 4 (IPT) = (ANC visit 4a-DropoutANC4a)*Provider readiness (IPT) {people/month}
Not treated 4 = (ANC visit 4a-DropoutANC4a)*(1- Provider readiness (IPT)) {people/month}

ANC visit 4b(t) = ANC visit 4b(t - dt) + (Not treated 3 (IPT) + Treated 3 (IPT) – DropoutANC4b – Treated 4 (IPT2) - Not treated 4 (IPT)) * dt
INIT ANC visit 4b = 2170 {people}
DOCUMENT: Number of women eligible to attend fourth ANC visit (who received their first dose of malaria treatment in previous visit). Either drop out, receive a second dose of malaria treatment or do not receive a second dose of treatment depending on provider readiness and dropout rate.
INFLOWS:
Not treated 3 (IPT) = (ANC visit 3b-DropoutANC3b)*(1- Provider readiness (IPT)) {people/month}
Treated 3 (IPT) = (ANC visit 3a-DropoutANC3a)*Provider readiness (IPT) {people/month}
OUTFLOWS:
DropoutANC4b = ANC visit 4b * Dropoutrate4
Treated 4 (IPT2) = (ANC visit 4b-DropoutANC4b)*Provider readiness (IPT) {people/month}
Not treated 4 (IPT) = (ANC visit 4b-DropoutANC4b)*(1- Provider readiness (IPT)) {people/month}

ANC visit 4c(t) = ANC visit 4c(t - dt) + (Treated 3 (IPT2) – DropoutANC4c – Treated 4c) * dt
INIT ANC visit 4c = 235 {people}
DOCUMENT: Number of women eligible to attend fourth ANC visit (who did receive their second dose of malaria treatment in previous visit). Either drop out or become eligible to attend ANC visit 5.
INFLOWS:

Treated 3 (IPT2) = (ANC visit 3b-DropoutANC3b)*Provider readiness (IPT) {people/month}
OUTFLOWS:
DropoutANC4c = ANC visit 4c * Dropoutrate4
Treated 4c = (ANC visit 4c-DropoutANC4c) {people/month}

Auxiliary variables
Dropoutrate1 = GRAPH(TIME{people/month})
(1, 0.0163), (54, 0.0146)
DOCUMENT: Fraction of pregnant women who do not attend one ANC appointment.

Dropoutrate2 = GRAPH(TIME{ people/month })
(1, 0.0237), (54, 0.026)
DOCUMENT: Fraction of pregnant women who do not attend a second ANC appointment.

Dropoutrate3 = GRAPH(TIME{ people/month })
(1, 0.0705), (54, 0.0646)
DOCUMENT: Fraction of pregnant women who do not attend a third ANC appointment.

Dropoutrate4 = GRAPH(TIME{ people/month })
(1, 0.2455), (54, 0.2876)
DOCUMENT: Fraction of pregnant women who do not attend a fourth ANC appointment.

Maximum potential patients (First dose) =
(ANC visit 1-(ANC visit 1*Dropoutrate1))+
(ANC visit 2a-(ANC visit 2a*Dropoutrate2))+
(ANC visit 3a-(ANC visit 3a*Dropoutrate3))+
(ANC visit 4a-(ANC visit 4a*Dropoutrate4)) {people}
DOCUMENT: Maximum number of patients who present for treatment (first dose IPT)

Maximum potential patients (Second dose) =
(ANC visit 2-(ANC visit 2*Dropoutrate2))+
(ANC visit 3b-(ANC visit 3b*Dropoutrate3))+
(ANC visit 4b-(ANC visit 4b*Dropoutrate4)) {people}
DOCUMENT: Maximum number of patients who present for treatment (second dose IPT)

Maximum potential patients (IPT) =
Maximum potential patients (First dose)+Maximum potential patients (Second dose) {people}
DOCUMENT: Maximum number of patients who present for treatment (first and second dose IPT)

Percentage of women receive IPT =
IF TIME <= 4
THEN
IF TIME <= 3
THEN
IF TIME <=2
THEN
IF TIME <= 1
THEN
("Treated 4 (IPT)"+
Feed in initial attendance data[Treated 3 IPT, 4]+
Feed in initial attendance data[Treated 2 IPT, 3]+
Feed in initial attendance data[Treated 1 IPT, 2])/
Feed in initial attendance data[Attend ANC1,1]

ELSE
("Treated 4 (IPT)"+
HISTORY("Treated 3 (IPT)", TIME-1)+
Feed in initial attendance data[Treated 2 IPT, 4]+
Feed in initial attendance data[Treated 1 IPT, 3])/
Feed in initial attendance data[Attend ANC1,2]

ELSE
("Treated 4 (IPT)"+
HISTORY("Treated 3 (IPT)", TIME-1)+
HISTORY("Treated 2 (IPT)", TIME-2)+
Feed in initial attendance data[Treated 1 IPT, 4])/
Feed in initial attendance data[Attend ANC1,3]

ELSE
("Treated 4 (IPT)"+
HISTORY("Treated 3 (IPT)", TIME-1)+
HISTORY("Treated 2 (IPT)", TIME-2)+
HISTORY("Treated 1 (IPT)", TIME-3))/
Feed in initial attendance data[Attend ANC1,4]

ELSE
("Treated 4 (IPT)"+
HISTORY("Treated 3 (IPT)", TIME-1)+
HISTORY("Treated 2 (IPT)", TIME-2)+
HISTORY("Treated 1 (IPT)", TIME-3))/
HISTORY(Attend ANC1, TIME-4) {unitless}
DOCUMENT: Percentage of women who receive 1st dose of IPT (of those who present for treatment). Patients have four opportunities to receive a first dose of IPT; this variable computes the percentage who receive their first dose for each cohort of ANC patients that passes through the model, once all opportunities for treatment have passed (at visit 4).

Percentage of women receive IPT2 =
IF TIME <= 4
THEN
IF TIME <= 3
THEN
IF TIME <=2
THEN
IF TIME <= 1
THEN
("Treated 4 (IPT2)"+
Feed in initial attendance data[Treated 2 IPT2,3]+
Feed in initial attendance data[Treated 3 IPT2,4])/
Feed in initial attendance data[Attend ANC1,1]

ELSE
("Treated 4 (IPT2)"+
Feed in initial attendance data[Treated 2 IPT2,4]+
HISTORY("Treated 3 (IPT2)", TIME-1))/
Feed in initial attendance data[Attend ANC1,2]

ELSE
("Treated 4 (IPT2)"+
HISTORY("Treated 2 (IPT2)", TIME-2)+
HISTORY("Treated 3 (IPT2)", TIME-1))/
Feed in initial attendance data[Attend ANC1,3]

ELSE
("Treated 4 (IPT2)"+
HISTORY("Treated 2 (IPT2)", TIME-2)+
HISTORY("Treated 3 (IPT2)", TIME-1))/
Feed in initial attendance data[Attend ANC1,4]

ELSE
("Treated 4 (IPT2)"+

HISTORY("Treated 2 (IPT2)", TIME-2)+
HISTORY("Treated 3 (IPT2)", TIME-1))/
HISTORY(Attend ANC1, TIME-4) {unitless}
DOCUMENT: Percentage of women who receive 2nd dose of IPT (of those who present for treatment). Patients who receive a first dose of IPT have up to three opportunities to receive a second dose of IPT; this variable computes the percentage who receive their second dose for each cohort of ANC patients that passes through the model, once all opportunities for treatment have passed (at visit 3).

Attend ANC1 = ANC visit 1-(ANC visit 1*Dropoutrate1) {people/month}
DOCUMENT: Number attending ANC visit 1.

Drugs used =
Treated 1 (IPT)+
Treated 2 (IPT)+
Treated 3 (IPT)+
Treated 4 (IPT)+
Treated 2 (IPT2)+
Treated 3 (IPT2)+
Treated 4 (IPT2) {people/month}
DOCUMENT: Number of drugs used in treatment for IPT.

No attending ANC1 = ("Treated 1 (IPT)"+Not treated 1)/HISTORY(ANC visit 1, TIME-1) {per month}
DOCUMENT: Number of patients attending 1 ANC visit.

No attending ANC2 = ("Treated 2 (IPT)"+"Treated 2 (IPT2)"+"Not treated 2 (IPT)"+Not treated 2)/
HISTORY(ANC visit 1, TIME-2) {per month}
DOCUMENT: Number of patients attending 2 ANC visit.

No attending ANC3 = (Treated 3+"Treated 3 (IPT2)"+"Not treated 3 (IPT)"+"Treated 3 (IPT)"+Not treated 3)/
HISTORY(ANC visit 1, TIME-3) {per month}
DOCUMENT: Number of patients attending 3 ANC visit.

No attending ANC4 =
(Not treated 4+"Treated 4 (IPT)"+"Not treated 4 (IPT)"+"Treated 4 (IPT2)"+Treated 4c+Treated 4)/
HISTORY(ANC visit 1, TIME-4) {per month}
DOCUMENT: Number of patients attending 4 ANC visit.

Stocks
Order of IPT drugs(t) = Order of IPT drugs(t – dt) + (IPT drugs needed –IPT drugs ordered) * dt
INIT Order of drugs = 14,600 {drugs}
DOCUMENT: Tracking the number of drugs used quarterly, this is then used to calculate the drugs ordered by the facility from the Medical Stores Department. Orders are made quarterly basis, stock is wiped every three months as orders are sent.
INFLOWS:
IPT drugs needed = Maximum potential patients (IPT) {drugs/month}
OUTFLOWS:
IPT drugs ordered = PULSE(Order of IPT drugs, 3, 3){drugs/month}

Stock of IPT drugs(t) = Stock of IPT drugs(t – dt) + (Replenishment of IPT drugs – Depletion of IPT drugs) * dt
INIT Stock of IPT drugs = 16,000 {drugs}
DOCUMENT: The stock of drugs available at the facility. The number of drugs reduces on a monthly basis based on services rendered and is topped up on a quarterly basis from MSD and monthly basis where facilities use own funds to purchase drugs.
INFLOWS:
Replenishment of drugs = MSD provision of IPT drugs to facility + Facility purchase of IPT drugs {drugs/month}
OUTFLOWS:
Depletion of IPT drugs = IPT drugs used{drugs/month}

Auxiliary variables
MSD provision of IPT drugs to facility = IPT drugs ordered * MSD provision of IPT ordered {drugs/month}
DOCUMENT: Number of IPT drugs ordered that are then supplied by the MSD.

MSD provision of IPT ordered = {drugs}
GRAPH(TIME{drugs/month})
(1, 0.33), (18, 0.45), (54, 0.45)
DOCUMENT: Percentage of drugs ordered that are supplied by the MSD.

Decision to buy IPT drugs outside MSD = {unitless}
IF Stock of IPT drugs/Maximum potential patients (IPT) < 1
THEN 1
ELSE 0
DOCUMENT: Facility decision on whether to buy IPT drugs outside MSD. If Stock of IPT drugs/Maximum potential patients (IPT) falls below 1 i.e. stocks fall below what is required for service delivery, facility held funding can be used to purchase drugs if available.

Facility purchase of IPT drugs = {unitless}
IF Decision to buy IPT drugs outside MSD = 1
THEN
(Facility held funds available)*Maximum potential patients (IPT)
ELSE
0
DOCUMENT: Use of facility held funding to purchase drugs. If a decision is made on using facility held funds to purchase drugs in 'Decision to buy IPT drugs outside MSD', facility held funds are used to purchase drugs.

Stocks

Number of health worker at health facility (% filled)(t)= Number of health worker at health facility (% filled)(t-dt) + (Change in staffing level) * dt
INIT Number of health worker at health facility (% filled) = 0.3991 {unitless}
DOCUMENT: Percentage of positions filled at health facilities.
BIFLOW = {per month}
IF "Number of health worker at health facility (% filled)" >= 1
THEN
-(Average attrition rate*"Number of health worker at health facility (% filled)")
ELSE (Hiring rate*"Number of health worker at health facility (% filled)")-(Average attrition rate*"Number of health worker at health facility (% filled)")

Knowledge of health workers (IPT)(t) = Knowledge of health workers (IPT)(t-dt) + (Change in knowledge (IPT)) * dt
INIT Knowledge of health workers (IPT) = 0.742 {unitless}
DOCUMENT: Percentage of time health workers are prescribing IPT during ANC visits, 0 (not prescribing at all) to 1 (prescribing to all patients).
BIFLOW = {per month}
(Perform supervision visit*(Effect of supervision on knowledge[1,ROUND(10*"Knowledge of health workers (IPT)")])+(P4P training*"Knowledge of health workers (IPT)"))
-
(Degradation of knowledge rate)

Trust in programme(t) = Trust in programme(t-dt) + (Change in trust) * dt
INIT 0.8 {unitless}
DOCUMENT: Trust in P4P changes over time in response to timing of incentive payments.
BIFLOW = {per month}
IF Trust in programme <1
THEN
Effect of incentives on trust-(Effect of delays in payment on trust["Effect of delays on trust (mode)",TIME])*Delay?
ELSE
MIN(0, (Effect of incentives on trust-(Effect of delays in payment on trust["Effect of delays on trust (mode)",TIME])*Delay?))

Auxiliary variables

Effect of supervision on knowledge = TABLE[Supervision on knowledge, Current knowledge] {unitless}
DOCUMENT = Table function, see table at end of section. Effect of supervision on knowledge, depending on current knowledge of provider.

Effect of incentives on trust = {unitless}
IF "Staff incentives (time corrected)">0
THEN Effect of incentive on trust (change)
ELSE 0
DOCUMENT = How payment of incentives changes health worker trust in P4P over time.

Effect of incentive on trust (change)= 0.05
DOCUMENT = Change in trust (either negative or positive) based on incentive payment.

Effect of delays in payment on trust = TABLE[Delays on trust mode, Delay P4P payment] {unitless}
DOCUMENT: Table function, see table at end of section. Effect of delays in payment on trust of health workers in P4P programme.

Effect of delays on trust (mode) = 3
DOCUMENT: There are three possible effects of payment delays on trust (1,2,3).

Hiring rate = 0.12/12 {unitless}
DOCUMENT: The rate at which health worker positions are filled.

Average attrition rate = 0.057/12 {unitless}
DOCUMENT = The rate at which health workers are leaving their positions.

Degradation of knowledge rate = (0.057/12)/10 {unitless}
DOCUMENT = Decrease in health worker skill over time. Related to attrition rate. Average attrition rate/10

P4P training = {unitless}
IF TIME = 1
THEN 0.05*P4P switch
ELSE 0
DOCUMENT = Initial increase in health worker knowledge attributed to training at start of programme. Only viable at time 1.

Provider readiness (IPT) = {unitless}
MIN(Availability of IPT drugs, MEAN("Number of health worker at health facility (% filled)", Health worker motivation to exert effort towards incentivised services, Availability of IPT drugs, "Knowledge of health workers (IPT)"))
DOCUMENT = Provider readiness to deliver services to patients, range 0 (low provider readiness) to 1 (max provider readiness, can treat all patients who present for care.

Health worker motivation = {unitless}
IF P4P switch = 1
THEN MEAN("Availability of drugs (IPT and labour)", "Number of health worker at health facility (% filled)", "District manager supervision (quality)", Trust in programme)
ELSE
MEAN("Availability of drugs (IPT and labour)", "Number of health worker at health facility (% filled)", "District manager supervision (quality)")
DOCUMENT = Health worker motivation to deliver incentivized services, range 0 (low motivation) to 1 (high motivation).

Staff incentives (time corrected) = {unitless}
DELAY("Staff incentives (delays?)", 1)
DOCUMENT = There is a lag between incentive payment issued and payment received (payment is issued month 6, received by month 9 etc.).

Auxiliary variables - table functions

Effect of supervision on knowledge = TABLE[Supervision on knowledge, Current knowledge] {unitless}
DOCUMENT = Table function, see table at end of section. Effect of supervision on knowledge, depending on current knowledge of provider.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0.1	0.1	0.1	0.1	0.2	0.2	0.025	0.025	0.025	0

Effect of delays in payment on trust = TABLE[Delays on trust mode, Delay P4P payment] {unitless}
DOCUMENT: Table function, see table at end of section. Effect of delays in payment on trust of health workers in P4P programme.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0.05	0.05	0.05	0.05	0	0	0.05	0.05	0.05	0	0	0	0.05
3	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0

21	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0	0	0.05	0.05	0.05
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

41	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0	0	0	0	0	0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Stocks
District manager motivation to support facilities = District manager motivation to support facilities(t-dt) + Change in motivation * dt
INIT: 0.6 Scale of 0 to 1, 0 (low motivation), 1 (high motivation). {unitless}
DOCUMENT: District manager motivation to supervise facilities.
BIFLOW: {per month}
IF District manager motivation to support facilities > Baseline motivation
THEN

IF District manager motivation to support facilities < 1
THEN
(Effect of initial knowledge of P4P on motivation+Effect of incentive on motivation-Effect of delays in payment on trust["Effect of delays on trust (mode)",TIME])*P4P switch
ELSE
MIN(0, (Effect of initial knowledge of P4P on motivation+Effect of incentive on motivation-Effect of delays in payment on trust["Effect of delays on trust (mode)",TIME])*P4P switch)

ELSE
MAX(0, (Effect of initial knowledge of P4P on motivation+Effect of incentive on motivation-Effect of delays in payment on trust["Effect of delays on trust (mode)",TIME])*P4P switch)

Auxiliary variables
District manager supervision (quality) = {unitless}
MEAN(District manager motivation to support facilities, "Skill level or knowledge of district manager (supervision)", "District budget/resources")
DOCUMENT: Quality of district manager supervision, dependent on motivation, knowledge and budget. Scale of 0 to 1, 0 (low quality), 1 (high quality).

Skill level or knowledge of district manager (supervision) = 0.8 {unitless}
DOCUMENT: Skill level or knowledge of managers regarding supervision. Scale of 0 to 1, 0 (low skill), 1 (high skill).

District budget/resources = 0.5 {unitless}
DOCUMENT: Percentage of allocated supervision budget that is available. Scale of 0 to 1, 0 (no funds), 1 (all allocated funds available).

Perform supervision visit = {unitless}
IF District budget/resources>= 0.5
THEN PULSE(District manager supervision (quality), 4, 4)
ELSE 0
DOCUMENT: This auxiliary variable passes the quality of supervision variable to the change in health worker knowledge stock every 4 months (if there are enough resources available for visits).

Baseline motivation = 0.6 {unitless}
DOCUMENT: Minimum level for motivation.

Effect of incentive on motivation = {unitless}
IF "Total incentive awarded to CHMT (time corrected)">0
THEN
Effect of incentive amount on motivation
ELSE
0
DOCUMENT: Effect of incentive payment on motivation to supervise.

Effect of incentive amount on motivation = 0.05 {unitless}
DOCUMENT: Increase (or decrease) in motivation induced by incentive payments.

Effect of initial knowledge of P4P on motivation = {unitless}
IF TIME = 1
THEN 0.1
ELSE 0
DOCUMENT: Effect of initial knowledge of P4P on motivation

Total incentive awarded to CHMT = {unitless}
DELAY((Eligible for facility performance incentive?+Eligible for stock out of RCH drugs incentive?)/2, 1)
DOCUMENT: Total incentive amount awarded to CHMT.

Total incentive awarded to CHMT (time corrected) = {unitless}
DELAY("Total incentive awarded to CHMT (delays?)", 1)
DOCUMENT: There is a lag between incentive payment issued and payment received (payment is issued month 6, received by month 9 etc.).

Total incentive awarded to CHMT (delays?) = {unitless}
IF TIME>6 AND Delay? = 1
THEN
IF TIME>6 AND TIME<13
THEN DELAY(Total incentive awarded to CHMT, 4)
ELSE
IF TIME>12 AND TIME<19
THEN DELAY(Total incentive awarded to CHMT, 3)
ELSE
IF TIME>18 AND TIME<25
THEN DELAY(Total incentive awarded to CHMT, 1)
ELSE
IF TIME>24 AND TIME<31
THEN Total incentive awarded to CHMT
ELSE
IF TIME>30 AND TIME<37
THEN DELAY(Total incentive awarded to CHMT, 1)
ELSE
IF TIME>36 AND TIME<43
THEN DELAY(Total incentive awarded to CHMT, 3)
ELSE
IF TIME>42
THEN 0
ELSE 999
ELSE Total incentive awarded to CHMT
DOCUMENT: Implements a delay in payment if the delay switch is turned 'on'.

Eligible for facility performance incentive? = {unitless}
IF (Total incentive awarded/"Target to receive CHMT incentive (Overall result)") >= Lower performance band
THEN
(IF (Total incentive awarded/"Target to receive CHMT incentive (Overall result)") >= Upper performance band
THEN 1
ELSE 0.5)
ELSE 0
DOCUMENT: Checking whether CHMT is eligible to receive incentive payment for this target.

Target to receive CHMT incentive (Overall result) = 0.8 {unitless}
DOCUMENT: Facility performance needed to induce CHMT payment for this target.

Eligible for stock out of RCH drugs incentive? = {unitless}
(Stock out of IPT?*Stock out of FBD?)
DOCUMENT: Checking if CHMT are eligible to receive incentive payment for this target.

```
Stock out of FBD? =          {unitless}
IF Reporting cycle = 1
THEN
IF
Stock of FBD drugs = 0 OR
HISTORY(Stock of FBD drugs, TIME-1) = 0 OR
HISTORY(Stock of FBD drugs, TIME-2) = 0 OR
HISTORY(Stock of FBD drugs, TIME-3) = 0 OR
HISTORY(Stock of FBD drugs, TIME-4) = 0 OR
HISTORY(Stock of FBD drugs, TIME-5) = 0
THEN 0
ELSE 1
ELSE 0
DOCUMENT: Checking if facilities have had a stock out in the last 6 months.
```

```
Stock out of IPT? =          {unitless}
IF Reporting cycle = 1
THEN
IF
Stock of IPT drugs = 0 OR
HISTORY(Stock of IPT drugs, TIME-1) = 0 OR
HISTORY(Stock of IPT drugs, TIME-2) = 0 OR
HISTORY(Stock of IPT drugs, TIME-3) = 0 OR
HISTORY(Stock of IPT drugs, TIME-4) = 0 OR
HISTORY(Stock of IPT drugs, TIME-5) = 0
THEN 0
ELSE 1
ELSE 0
DOCUMENT: Checking if facilities have had a stock out in the last 6 months.
```

Stocks

Community awareness(t) = Community awareness(t-dt) + (Change in awareness) * dt
INIT Community awareness = 0.85 {unitless}
DOCUMENT = Represents awareness of maternal and child health and healthcare in the community. Awareness is improved through outreach, attendance at ANC and facility-based deliveries in community and depleted by belief in myths and decay of awareness (in absence of activities that promote awareness).
BIFLOW = {per month}
IF Community awareness < 1
THEN (
((Fraction of women attending at least one ANC*Effect ANC awareness)
+(Ability to perform outreach*Effect outreach awareness)
+(Percentage of women who seek facility-based delivery*Effect FBD awareness))
-(Effect myths literacy awareness*Decay awareness))*Community awareness
ELSE
-(Effect myths literacy awareness*Decay awareness)*Community awareness

Number of facility-based deliveries(t) = Number of facility-based deliveries(t-dt) + (Seeking FBD – FBD) * dt
INIT Number of facility-based deliveries = 3670 {people}
DOCUMENT = Number of facility based deliveries.
INFLOWS:
Seeking FBD = Number of Newly Pregnant Women*Fraction who seek FBD {people/month}
OUTFLOWS:
FBD = Number of facility-based deliveries {people/month}

Auxiliary variables

Percentage of women who seek facility-based delivery = {unitless}
0.18*"Perceived quality of facility/services"+
0.273*Community awareness+
0.273*Effect ANC on FBD+
0.273*Distance to facility
DOCUMENT = Fraction of women who seek facility-based delivery.

Patient-provider interaction = Health worker motivation to exert effort towards incentivised services {unitless}
DOCUMENT: Interaction between health workers and patients (tied to health worker motivation). Scale of 0 (low quality interaction), 1 (high quality interaction).

Perceived quality of facility/services =
MEAN("Patient-provider interaction", "Availability of drugs (IPT and labour)") {unitless}
DOCUMENT: Patient perceived quality of services, 0 (low quality) to 1 (high quality).

Distance to facility =
GRAPH(TIME{unitless })
(1, 0.948), (18, 0.983), (54, 0.957)
DOCUMENT: Percentage of women that do not perceive distance to being a barrier for facility-based delivery.

Effect ANC on FBD = {unitless}
((Dropoutrate2*Prob FBD ANC 1)+
(Dropoutrate3*Prob FBD ANC 2)+
(Dropoutrate4*Prob FBD ANC 3)+
((1-Dropoutrate1-Dropoutrate2-Dropoutrate3-Dropoutrate4)*Prob FBD ANC 4))*1.2
DOCUMENT: Effect of number of ANC visits on facility-based deliveries.

Prob FBD ANC 1= EXP(0.47)/1+EXP(0.47) =61.54 {Unitless}
DOCUMENT = Probability of facility-based delivery for those who attend one ANC visit.

Prob FBD ANC 2= EXP(0.74)/1+EXP(0.74) = 67.70 {unitless}
DOCUMENT = Probability of facility-based delivery for those who attend two ANC visit.

Prob FBD ANC 3= EXP(1.13)/1+EXP(1.13) = 75.58 {unitless}
DOCUMENT = Probability of facility-based delivery for those who attend three ANC visit.

Prob FBD ANC 4= EXP(1.52)/1+EXP(1.52) = 82.05{unitless}
DOCUMENT = Probability of facility-based delivery for those who attend four + ANC visits.

FBD drugs used = Number of facility-based deliveries {drugs}
DOCUMENT: Number of facility-based deliveries and facility-based delivery drugs used per month.

Fraction of women attending at least one ANC = 1 – Dropoutrate1 {unitless}
DOCUMENT: The fraction of women who attend at least 1 ANC appointment.

Effect ANC awareness = 0.001{unitless}
DOCUMENT: Effect of ANC on awareness of maternal and child health and healthcare in the community.

Literacy = TIME [(2011, 0.730), (2012, 0.781), (2013, 0.780), (2014, 0.780), (2015, 0.779)] {unitless}
DOCUMENT: Literacy rates for Tanzania adults (15+).

Effect literacy awareness = 0.001 {unitless}
DOCUMENT: Effect of literacy on awareness of maternal and child health and healthcare.

Belief in myths = 0.01{unitless}
DOCUMENT: Percentage of the population that believes in myths.

Effect myths = 0.001{unitless}
DOCUMENT: Effect of myths on awareness of maternal and child health and healthcare in the community.

Effect myths literacy awareness = (Belief in myths*Effect myths)*(Literacy*Effect literacy awareness) {unitless}
DOCUMENT: Effect of myths and literacy rate on awareness of maternal and child health and healthcare in the community.

Decay awareness = 0.015{unitless}
DOCUMENT: Decay of awareness in community over time.

Ability to perform outreach = {unitless}
MEAN(Number of health worker at health facility % filled), Health worker motivation to exert effort towards incentivised services)
DOCUMENT = Ability of health workers to perform outreach activities, dependent on number of health workers at facility and motivation.

Effect outreach awareness = 0.001{unitless}
DOCUMENT = Effect of outreach on awareness of maternal and child health and healthcare in the community.

Effect FBD awareness = 0.001{unitless}
DOCUMENT = Effect of FBD on awareness of maternal and child health and healthcare in the community.

Stocks
Order of FBD drugs(t) = Order of FBD drugs(t – dt) + (FBD drugs needed –FBD drugs ordered) * dt
INIT Order of FBD drugs = 3,720 {drugs}
DOCUMENT: Tracking the number of drugs used quarterly, this is then used to calculate the drugs ordered by the facility from the Medical Stores Department. Orders are made on quarterly basis, stock is wiped every three months as orders are sent.
INFLOWS:
FBD drugs needed = Seeking FBD {drugs/month}
OUTFLOWS:
FBD drugs ordered = PULSE(Order of FBD drugs, 3, 3){drugs/month}

Stock of FBD drugs(t) = Stock of FBD drugs(t – dt) + (Replenishment of FBD drugs – Depletion of FBD drugs) * dt
INIT Stock of FBD drugs = 7,600 {drugs}
DOCUMENT: The stock of drugs available at the facility. The number of drugs reduces on a monthly basis based on services rendered and is topped up on a quarterly basis from MSD and monthly basis where facilities use own funds to purchase drugs.
INFLOWS:
Replenishment of FBD drugs = MSD provision of FBD drugs to facility + Facility purchase of FBD drugs {drugs/month}
OUTFLOWS:
Depletion of FBD drugs = FBD drugs used{drugs/month}

Auxiliary variables
MSD provision of FBD drugs to facility = FBD drugs ordered * MSD provision of FBD drugs ordered {drugs}
DOCUMENT: Number of FBD drugs ordered that are then supplied by the MSD.

MSD provision of FBD drugs ordered =
GRAPH(TIME{drugs/month})
(1, 0.63), (18, 0.63), (54, 0.63)
DOCUMENT: Percentage of drugs ordered that are supplied by the MSD.

Decision to buy FBD drugs outside MSD =
IF Stock of FBD drugs/Number of facility-based deliveries < 1
THEN 1
ELSE 0
DOCUMENT: Facility decision on whether to buy FBD drugs outside MSD. If Stock of FBD drugs/Number of facility-based deliveries falls below 1 i.e. stocks fall below what is required for service delivery, facility held funding can be used to purchase drugs if available.

Facility purchase of FBD drugs =
IF Decision to buy FBD drugs outside MSD = 1
THEN
(Facility held funds available)* Number of facility-based deliveries
ELSE
0
DOCUMENT: Use of facility held funding to purchase drugs. If a decision is made on using facility held funds to purchase drugs in ‘Decision to buy FBD drugs outside MSD’, facility held funds are used to purchase drugs.

Facility Funding sector

Stocks
Incentives used for drugs(t) = Incentives used for drugs(t - dt) + (Incoming incentive payment – Funds used) * dt
INIT Incentives used for drugs = 0 {unitless}
DOCUMENT: Incentives available to purchase drugs where needed.
INFLOW:
Incoming incentive payment = IF TIME>6 AND Delay? = 1 {per month}
THEN
 IF TIME>6 AND TIME<13
 THEN DELAY(Operation incentives*0.6, 4)
 ELSE
  IF TIME>12 AND TIME<19
  THEN DELAY(Operation incentives*0.6, 3)
  ELSE
   IF TIME>18 AND TIME<25
   THEN DELAY(Operation incentives*0.6, 1)
   ELSE
    IF TIME>24 AND TIME<31
    THEN Operation incentives*0.6
    ELSE
     IF TIME>30 AND TIME<37
     THEN DELAY(Operation incentives*0.6, 1)
     ELSE
     IF TIME>36 AND TIME<43
     THEN DELAY(Operation incentives*0.6, 3)
     ELSE
     IF TIME>42
     THEN 0
     ELSE 999
ELSE Operation incentives*0.6
OUTFLOW:
Funds used = IF TIME>5 AND Delay? = 1 {per month}
THEN
 IF TIME = 17 OR TIME= 22 OR TIME = 26 OR TIME =
31 OR TIME = 38 OR TIME = 46
THEN HISTORY(Incoming incentive payment, TIME-6)
 ELSE 0
ELSE
PULSE(Incentives used for drugs, 13, 6)

Auxiliary variables
Actual timing of payment =
((DELAY(Incoming incentive payment, 2)*2)/(1-Use of incentives))*P4P switch {unitless}
DOCUMENT: This number is used to represent the actual timing of incentive payments when there are delays, which is represented graphically in the model interface.

Alternative facility held funding available = 0.1 {unitless}
DOCUMENT: Facilities can buy 10% of the monthly drugs they need with alterative facility held funding.

Cycle 1 check =
IF TIME = 6
THEN
  IF FBD performance change >= 0.85
  THEN FBD performance change
  ELSE 0
ELSE
FBD performance change

DOCUMENT: There is no historical performance preceding first cycle of facility-based deliveries; if cycle 1 performance is 0.85 or better, they achieve the target. If it is less, we assume they did not achieve the target for cycle 1.

Delay? = 1 {unitless}
DOCUMENT: A switch that turns payment delays on and off.

Eligible for FBD incentive? =
IF "Fraction who seek FBD (6 month period)" >= 0.85
THEN
1*P4P switch
ELSE
IF
(Cycle 1 check/(("Target to receive facility incentive (increase in % points)"[1,ROUND((Previous cycle performance*100)+1)]))) >= Lower performance band
THEN
(IF
(Cycle 1 check/"Target to receive facility incentive (increase in % points)"[1,ROUND((Previous cycle performance*100)+1)]) >= Upper performance band
THEN 1*P4P switch
ELSE 0.5*P4P switch)
ELSE 0
DOCUMENT: Calculating how much incentive payment should be paid for FBD target.

Eligible for IPT2 incentive? = {unitless}
IF ("Percentage of women receive IPT2 (6 month period)"/"Target to receive facility incentive (Overall result)") >= Lower performance band
THEN
(IF ("Percentage of women receive IPT2 (6 month period)"/"Target to receive facility incentive (Overall result)") >= Upper performance band
THEN 1*P4P switch
ELSE 0.5*P4P switch)
ELSE 0
DOCUMENT: Calculating how much incentive payment should be paid for IPT target.

Facility held funds available = (Incentives used for drugs*P4P switch)+Alternative facility held funding available
DOCUMENT: Facility held funds available to purchase drugs.

FBD performance change = {unitless}
IF Previous cycle performance >= 0.85
THEN
IF "Percentage of women who seek facility-based delivery (6 month period)" >= 0.85
THEN 1
ELSE 0
ELSE
IF ("Percentage of women who seek facility-based delivery (6 month period)"-Previous cycle performance)<=0
THEN 0
ELSE ("Percentage of women who seek facility-based delivery (6 month period)"-Previous cycle performance)
DOCUMENT: The change in performance between cycles for the FBD target.

FBD lower performance band = {unitless}
IF TIME = 6
THEN NAN
ELSE
IF Reporting cycle = 1
THEN
IF "Percentage of women who seek facility-based delivery (6 month period)" >= 0.85

THEN 0.85
ELSE
IF "Percentage of women who seek facility-based delivery (6 month period)">= 0.85 AND Previous cycle performance >= 0.85
THEN 0.85
ELSE
Previous cycle performance+
("Target to receive facility incentive (increase in % points)"[1,ROUND((Previous cycle performance*100)+1)]*0.75)
ELSE PREVIOUS(SELF, NAN)
DOCUMENT: This number is used to represent the lower performance band for this target on the performance graph presented on the model interface.

FBD upper performance band = {unitless}
IF TIME = 6
THEN NAN
ELSE
IF Reporting cycle = 1
THEN
IF "Percentage of women who seek facility-based delivery (6 month period)" >= 0.85
THEN 0.85
ELSE
IF "Percentage of women who seek facility-based delivery (6 month period)" >= 0.85 AND Previous cycle performance >= 0.85
THEN 0.85
ELSE
Previous cycle performance+
"Target to receive facility incentive (increase in % points)"[1,ROUND((Previous cycle performance*100)+1)]
ELSE PREVIOUS(SELF, NAN)
DOCUMENT: This number is used to represent the upper performance band for this target on the performance graph presented on the model interface.

"Fraction who seek FBD (6 month period)" = {unitless}
IF Reporting cycle = 1
THEN
MEAN(
Percentage of women who seek facility-based delivery,
HISTORY(Percentage of women who seek facility-based delivery, TIME-1),
HISTORY(Percentage of women who seek facility-based delivery, TIME-2),
HISTORY(Percentage of women who seek facility-based delivery, TIME-3),
HISTORY(Percentage of women who seek facility-based delivery, TIME-4),
HISTORY(Percentage of women who seek facility-based delivery, TIME-5)
)
ELSE 0
DOCUMENT: Calculating performance for FBD.

IPT lower performance band = {unitless}
"Target to receive facility incentive (Overall result)"*Lower performance band
DOCUMENT: This number is used to represent the lower performance band for this target on the performance graph presented on the model interface.

IPT upper performance band = {unitless}
"Target to receive facility incentive (Overall result)"*Upper performance band
DOCUMENT: This number is used to represent the upper performance band for this target on the performance graph presented on the model interface.

Lower performance band = 0.75 {unitless}
DOCUMENT: Facilities need to achieve at least 75% of the target to receive 50% of the payment for that target.

Operation incentives = DELAY(Total incentive awarded*(1-Use of incentives), 1)
DOCUMENT: Percentage of incentive used for facility operations (drugs).

P4P switch = 0 {unitless}
DOCUMENT: The switch to turn P4P programme on (1) or off (0).

"Percentage of women receive IPT2 (6 month period)" = {unitless}
IF Reporting cycle = 1
THEN
MEAN(
Percentage of women receive IPT2,
HISTORY(Percentage of women receive IPT2, TIME-1),
HISTORY(Percentage of women receive IPT2, TIME-2),
HISTORY(Percentage of women receive IPT2, TIME-3),
HISTORY(Percentage of women receive IPT2, TIME-4),
HISTORY(Percentage of women receive IPT2, TIME-5)
)
ELSE 0
DOCUMENT: Performance of facility on the IPT2 target.

Planned timing of payment = {unitless}
DELAY(Total incentive awarded, 3)*P4P switch
DOCUMENT: This number is used to represent the planned timing of incentive payments, which is represented graphically in the model interface.

Previous cycle performance = {unitless}
IF TIME = 6
THEN 0
ELSE
HISTORY("Percentage of women who seek facility-based delivery (6 month period)", TIME-6)
DOCUMENT: Previous cycles FBD performance.

Reporting cycle = PULSE(1, 6, 6) {unitless}
DOCUMENT: Facilities report performance data every six months in order to be eligible for incentive payments.

Staff incentives = DELAY(Total incentive awarded*(Use of incentives), 1)
{unitless}
DOCUMENT: Incentives are paid three months after incentive amount is agreed.

Staff incentives (delays?) = {unitless}
IF TIME>6 AND Delay? = 1
THEN
 IF TIME>6 AND TIME<13
 THEN DELAY(Staff incentives, 4)
 ELSE
  IF TIME>12 AND TIME<19
  THEN DELAY(Staff incentives, 3)
  ELSE
   IF TIME>18 AND TIME<25
   THEN DELAY(Staff incentives, 1)
   ELSE
    IF TIME>24 AND TIME<31
    THEN Staff incentives
    ELSE
     IF TIME>30 AND TIME<37
     THEN DELAY(Staff incentives, 1)
     ELSE
     IF TIME>36 AND TIME<43

THEN DELAY(Staff incentives, 3)
ELSE
IF TIME>42
THEN 0
ELSE 999
ELSE Staff incentives
DOCUMENT: If ‘delay’ is switched on, delays in payment are made.

"Target to receive facility incentive (increase in % points)" = TABLE[Increase in % required, Previous cycle] {unitless}
DOCUMENT: Table function, see table at end of section. Outputs the change in performance needed to achieve incentive payment.

"Target to receive facility incentive (Overall result)" = 0.8 {unitless}
DOCUMENT: Target for IPT2 delivery to achieve incentive payment.

Total incentive awarded = (Eligible for IPT2 incentive?*0.5)+(Eligible for FBD incentive?*0.5) {unitless}
DOCUMENT: Total incentive awarded based on performance of IPT2 and FBD.

Upper performance band = 1 {unitless}
DOCUMENT: Facilities need to achieve 100% of the target to receive 100% of the payment for that target.

Use of incentives = 0.75 {unitless}
DOCUMENT: 75% of the incentive payment goes to health workers, 25% for facility operations.

Auxiliary variables - table functions

"Target to receive facility incentive (increase in % points)" = TABLE[Increase in % required, Previous cycle] {unitless}

DOCUMENT: Table function. Outputs the change in performance needed to achieve incentive payment.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.1

21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.05	0.05

42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62
0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101
0.05	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001	0.000 001

Supplementary File 4 - Description of COSMIC Parameters Document

This Supplementary file provides a description of model parameters and data used for the following model sectors; **Population, Demand and Services, Facility Commodities**, **Facility Operations, Facility Funding and District Manager Operations**. Supplementary files 2 and 3 provide further information on model sectors and model equations.

N.B. The **Demand and Services** and **Facility Commodities** sectors are each presented here with two sub-sectors, describing data related to the model equations for two different health services of interest in the model (percentage of women who received two doses of IPT during ANC and percentage of women who had a facility-based delivery). This was an artifact of model development (for ease of viewing and analysis), for all intents and purposes they can be considered ‘subsectors’ of a single model sector.

Population sector

Table S4.1: Population sector variable descriptions and data sources

	Variable Name	Description	Initial Value	Units	Reference
1	Neonates	Number of neonates (infants who are less than a month old) in Pwani region. Estimated at a 1/12 of the population of infants. (32830/12) = 2736 $0 \leq x < 1/12$	2,736 Changed to 4740 during calibration	People	National Bureau of Statistics (2013).
2	Infants	Number of infants over 1 month and up to 1 year in Pwani region.. The number of infants below 1 year were 32830. (32830-neonates(2736)=30094) $1/12 \leq x < 1$	30,094 Changed to 49500 during calibration	People	National Bureau of Statistics (2013).
3	Preschoolers	Number of children aged $1 \leq x < 5$ in Pwani region.	154,332 Changed to 175000 during calibration	People	National Bureau of Statistics (2013).
4	Children	Number of children aged $5 \leq x < 15$ in Pwani region.	284,308	People	National Bureau of Statistics (2013).
5	ReproductiveAge Adults	Number of adults of reproductive age $15 \leq x < 50$ in Pwani region.	509,934	People	National Bureau of Statistics (2013).
6	AdultsAbove50	Number of adults above 50 years. $x \geq 50$ in Pwani region.	150,094	People	National Bureau of Statistics (2013).
	DurationInfant	The duration it takes a neonate to become an infant. 1 month	1	Months	--
	DurationPrescho ol	The duration it takes an infant to become a preschooler. 11 months	11	Months	--
7	DurationChild	The duration it takes for a preschooler to become a child. 4 years	48	Months	--

		4*12 = 48			
8	DurationReprod	Duration it takes for a child to become reproductive. 10 years 10*12 = 120	120	Months	--
9	DurationAdultAbove50	The duration it takes for an adult to stop being reproductive. 35 years 35*12 = 420	420	Months	--
10	FractionofReproductiveAgeAdultsFemale	The fraction of reproductive age adults who are female.	0.52	Unitless	National Bureau of Statistics (2013).
11	NeonateMortalityRate	The rate at which neonates (less than one month) die per year. For Tanzania - five years preceding the survey (2006-2010) neonatal mortality is 26 deaths per 1000 live births.	0.026/12	1/month	National Bureau of Statistics (2011).
12	InfantMortalityRate	The rate at which infants (1 month to 1 year) die per year. For Tanzania - five years preceding the survey (2006-2010) infant mortality (1 month to 1 year) is 25 deaths per 1000 live births.	0.025/12	1/month	National Bureau of Statistics (2011).
	PreschoolMortRate	The rate at which preschoolers die per year. For Tanzania - five years preceding the survey (2006-2010) preschool mortality (1 year to 5 years) is 32 deaths per 1000 live births.	0.032/12	1/month	National Bureau of Statistics (2011).
13	ChildMortRate	The rate at which children die per year. Number of children dying between age 5 and 14 for Tanzania mainland is 2.4 per 1000 persons.	0.0024/12	1/month	National Bureau of Statistics (2015).
14	ReprodAgeMortality	The rate at which people from reproductive age group die per year.	0.005/12	1/month	National Bureau of

		Age-specific mortality rates for women and men age 15-49 based on the survivorship of sisters and brothers of survey respondents, for the ten-year period preceding the survey, Tanzania 2010. Women (age standardized rate) 5.1 deaths per 1000 years of exposure) Men (age standardized rate) 5 deaths per 1000 years of exposure)			Statistics (2011).
15	Above50MortalityRate	The rate at which people over the age of 50 die per year. Tanzania mainland mortality 60+ is 57.5 per 1000 persons.	0.0575/12	1/month	National Bureau of Statistics (2015).
16	GeneralFertilityRate	Number of live births per 1000 women of reproductive age. Tanzania (2010 DHS) general fertility rate three years preceding the survey, number of live births per 1000 women of reproductive age (15-49), 189 per 1000 women. NOTES DURING CALIBRATION: Changing 0.189/12 to 0.2148/12	0.2148/12	1/month	National Bureau of Statistics (2011).
17	Projected population growth rates for Tanzania mainland	Using the raw data, numbers for each age category reported, percentage change per year: 2013 3.2 2014 3.2 2015 3.2 2016 3.2 2017 3.2 2018 3.1 2019 3.1 2020 3.1 2021 3.1	Used raw data percentage change to calibrate growth in population	1/year	National Bureau of Statistics (2018).

		2022 3.1 2023 3.1 2024 3.1 2025 3.0			

Demand and Services (Antenatal Care) sector

Table S4.2: Demand and Services sector variable descriptions and data sources

	Variable Name	Description	Initial Value	Units	Reference
1	ANC visit 1	Number of women eligible to attend first ANC visit.	Calibrated based on incoming flow. 4,750	People	--
2	ANC visit 2	Number of women eligible to attend second ANC visit (who have had their first dose of IPT)	Calibrated based on incoming flow. 470	People	--
3	ANC visit 2a	Number of women eligible to attend second ANC visit (who have not had first dose of IPT)	Calibrated based on incoming flow. 4190	People	--
4	ANC visit 3	Number of women eligible to attend third ANC visit (who have had their second dose of IPT)	Calibrated based on incoming flow. 156	People	--
5	ANC visit 3a	Number of women eligible to attend third ANC visit (who have not had first dose of IPT)	Calibrated based on incoming flow. 2700	People	--
6	ANC visit 3b	Number of women eligible to attend third ANC visit (who have had their first dose of IPT).	Calibrated based on incoming flow. 1650	People	--
7	ANC visit 4	Number of women eligible to attend fourth ANC visit (who have had their second dose of IPT)	Calibrated based on incoming flow. 1000	People	--
8	ANC visit 4a	Number of women eligible to attend fourth ANC visit	Calibrated based on incoming flow.	People	--

		(who have not had first dose of IPT)	740		
9	ANC visit 4b	Number of women eligible to attend fourth ANC visit (who have had their first dose of IPT).	Calibrated based on incoming flow. 2170	People	--
10	ANC visit 4c	Number of women eligible to attend fourth ANC visit (who have had their second dose of IPT).	Calibrated based on incoming flow. 235	People	--
11	Dropoutrate1	The fraction of pregnant women who do not attend 1 ANC visit.	TIME [(2011, 0.0163), (2015, 0.0146)]	People/ Month	Pwani IHI/LSHTM impact evaluation, (2011, 2015)
12	Dropoutrate2	Fraction of pregnant women who do not attend a second ANC appointment.	TIME [(2011, 0.0237), (2015, 0.026)]	People/ Month	Pwani IHI/LSHTM impact evaluation, (2011, 2015)
13	Dropoutrate3	Fraction of pregnant women who do not attend a third ANC appointment.	TIME [(2011, 0.0705), (2015, 0.0646)]	People/ Month	Pwani IHI/LSHTM impact evaluation, (2011, 2015)
14	Dropoutrate4	Fraction of pregnant women who do not attend a fourth ANC appointment.	TIME [(2011, 0.2455), (2015, 0.2876)]	People/ Month	Pwani IHI/LSHTM impact evaluation, (2011, 2015)

Demand and Services variables (Facility-based delivery care)

Table S4.3: Demand and Services sector variable descriptions and data sources

	Variable Name	Description	Initial Value	Units	Reference
1	Community awareness	Knowledge and awareness of maternal and child health and healthcare (including facility-based deliveries).	0.85 Assumed.	Unitless	--
2	Number of facility-based deliveries	Number of facility-based deliveries.	Calibrated based on incoming flow. 3670	People	--
3	Distance to facility	Percentage of women that do not perceive distance to being a barrier for facility-based delivery.	TIME [(2011, 0.948), (2013, 0.983), (2015, 0.957)]	Unitless	Pwani IHI/LSHTM impact evaluation, (2011, 2013, 2015)
4	Belief in myths	Percentage of the female population that believes in myths. Model assumes 1% of the female population believes in myths.	0.01	Unitless	Semwanga *et al.* (2016)
5	Literacy	Percent of adult population (15+) that are literate.	TIME [(2011, 0.730), (2012, 0.781), (2013, 0.780), (2014, 0.780), (2015, 0.779)] Assumed 2011, 2013 and 2014 based on adjacent years data (2010, 2012, 2015).	Unitless	UNESCO Institute for Statistics
6	Decay awareness	Decay in awareness in community over time.	Assumed 0.015	Unitless	--
7	Prob FBD ANC 1	Probability of facility-based delivery for those	EXP(0.47)/ 1+EXP(0.47)=	Unitless	Ensor *et al.* (2014)

		who attend one ANC visit.	61.54		
8	Prob FBD ANC 2	Probability of facility-based delivery for those who attend two ANC visit.	EXP(0.74)/ 1+EXP(0.74)= 67.70	Unitless	Ensor *et al.* (2014)
9	Prob FBD ANC 3	Probability of facility-based delivery for those who attend three ANC visit.	EXP(1.13)/ 1+EXP(1.13)= 75.58	Unitless	Ensor *et al.* (2014)
10	Prob FBD ANC 4+	Probability of facility-based delivery for those who attend four + ANC visits.	EXP(1.52)/ 1+EXP(1.52)= 82.05	Unitless	Ensor *et al.* (2014)

Facility Commodities variables, specifically for antenatal care

Table S4.4: Facility Commodities sector variable descriptions and data sources

	Variable Name	Description	Initial Value	Units	Reference
1	Order of drugs	Number of drugs used quarterly (every four months).	Calibrated based on incoming flow. 14,600	Drugs	--
2	Stock of drugs	Stock of drugs available at the facility.	Calibrated based on incoming flow. 16,000	Drugs	--
3	MSD provision of IPT drugs ordered (%)	Percentage of drugs ordered that are then supplied by the MSD.	Original assumption: 0.5 Calibrated to percentage of women who receive IPT2. GRAPH(TIME{drugs/month}) (1, 0.33), (18, 0.45), (54, 0.45)	Drugs/ Month	--

Facility Commodities variables, specifically for facility-based delivery care

Table S4.5: Facility Commodities sector variable descriptions and data sources

	Variable Name	Description	Initial Value	Units	Reference
1	Order of FBD drugs	Number of drugs used quarterly (every four months).	Calibrated based on incoming flow. 3,720	Drugs	--
2	Stock of FBD drugs	Stock of drugs available at the facility.	Calibrated based on incoming flow. 7,600	Drugs	--
3	MSD provision of drugs ordered (%)	Percentage of drugs ordered that are then supplied by the MSD.	Original assumption: 0.5 Calibrated to fraction of women who seek facility-based delivery. GRAPH(TIME{drugs/month}) (1, 0.63), (18, 0.63), (54, 0.63)	Drugs/ Month	--

Facility Operations variables

Table S4.6: Facility Operations sector variable descriptions and data sources

	Variable Name	Description	Initial Value	Units	Reference
1	Number of health worker at health facility (% filled)	Percentage of positions filled at health facilities. (0 no positions filled, 1 all positions filled)	Used the following data to calibrate hiring rate: 2011 – 39.91 2012 – 42.37 2013 – 48.25 2014 – 47.00 Assumed 2011, based on adjacent years 2009 (35.00) and 2012 (42.37) data.	Unitless	Ministry of Health and Social Welfare (2013) Ministry of Health and Social Welfare (2014a)
2	Hiring rate	The rate at which health worker positions are filled from hiring	2011 – 0.11 2012 – 0.09 2013 – 0.11 With calibration, 0.12/12	Unitless	Ministry of Health and Social Welfare (2014b)
4	Average attrition rate	The rate at which health workers are leaving their positions.	0.057 per year 0.057/12	Unitless	Kurowski *et al.* (2007)
5	Skill level/knowledge of health workers (IPT2)	The skill level of health workers at facilities (percentage, 0 low skill, 1 high skill).	0.742	Unitless	Pwani IHI/LSHTM impact evaluation, (2011).
6	Effect of supervision on knowledge	Effect district manager supervision has on health worker	Determined through qualitative	Unitless	--

		knowledge to prescribe IPT during ANC.	interview with stakeholder. There will be some improvement after even one visit. Depends on what knowledge level is like at facility already. Need reminders to keep up knowledge. For those with high knowledge/skill, didn't need several visits for improvement, single visit enough. Table function (see page 11 of model equations document)		
6	Degradation of skill rate	Decrease in health worker skill over time. Related to attrition rate. Average attrition rate/10	Assumed. Attrition/10 (0.057/12)/10	Unitless	--
7	P4P training	Initial increase in health worker skill level/knowledge attributed to training at start of programme	Assumed. 0.05	Unitless	--
8	Effect of incentive on trust (change)	Change in trust (either negative or positive)	Assumed. 0.05	Unitless	--

		based on payment of incentives.			
9	Effect of delays in payment on trust	Effect of delays in payment on trust, incremental increase or decrease.	Determined through qualitative interview with stakeholders. Only severe delays in payment negatively impact trust (4+ months delay). Table function (see page 11 of model equations document)	Unitless	--

Facility Funding variables

Table S4.7: Facility Funding sector variable descriptions and data sources

	Variable Name	Description	Initial Value	Units	Reference
1	P4P switch	Switch for turning P4P on (1) and off (0) in the model.	0	Unitless	--
2	Use of incentives	75% of the incentive payment goes to health workers, 25% for facility operations.	0.75	Unitless	Binyaruka *et al.* (2015)
3	Target to receive facility incentive (Overall result)	Target for IPT2 delivery to achieve incentive payment.	0.8	Unitless	Binyaruka *et al.* (2015)
4	Upper performance band	Facilities need to achieve 100% of the target to receive 100% of the payment for that target.	1	Unitless	Binyaruka *et al.* (2015)
5	Lower performance band	Facilities need to achieve at least 75% of the target to receive 50% of the payment for that target.	0.75	Unitless	Binyaruka *et al.* (2015)
6	Target to receive facility incentive (increase in % points)	Table function. Outputs the change in performance needed to achieve incentive payment.	Table function (see page 23 of model equations document)	Unitless	Binyaruka *et al.* (2015)

District Manager Operations variables

Table S4.8: District Manager Operations sector variable descriptions and data sources

	Variable Name	Description	Initial Value	Units	Reference
1	Skill level or knowledge of district manager (supervision)	The skill or knowledge of district managers related to supervision activities (percentage, 0 low skill, 1 high skill).	0.8 assumed.	Unitless	--
2	District budget/resources	The percentage of the budget (allocated for supervision activities) that district managers can effectively use for supervision (percentage, 0 none of the budget, 1 all of the budget).	0.5 assumed.	Unitless	--
3	District manager motivation	Motivation of district level managers to perform supervision activities (percentage, 0 low motivation, 1 high motivation).	0.6 assumed.	Unitless	--
4	Effect of incentive amount on motivation	Increase (or decrease) in motivation induced by incentive payments.	0.05 assumed.	Unitless	--
5	Target to receive CHMT incentive (Overall result)	Facility performance needed to induce CHMT payment for this target.	0.8	Unitless	Ministry of Health and Social Welfare (2012)
6	Baseline motivation	Minimum level for motivation.	0.6 assumed.	Unitless	--

Supplementary File 5 – COSMIC calibration graphs

Population sector

The percentage population change was calculated for each year in the model (Figure S5.1) and compared to projected population growth rates for Tanzania (Supplementary File 4, Table S4.1). The initial values for each age category and fertility rate for adults of reproductive age were adjusted until a good fit for population change was achieved (original and calibrated initial values and rates are shown in Supplementary File 4, Table S4.1). The model output and data were also compared for population growth within each age category over time (Figures S5.2 – S5.7). The model output for neonate and infant population growth is a close fit (apart from the first year of data, 2012/13). The data for older age groups reads as quite volatile; without understanding the cause for the dramatic change year on year, it is difficult to replicate in the model. The model instead follows a general smoothed trend for these age categories.

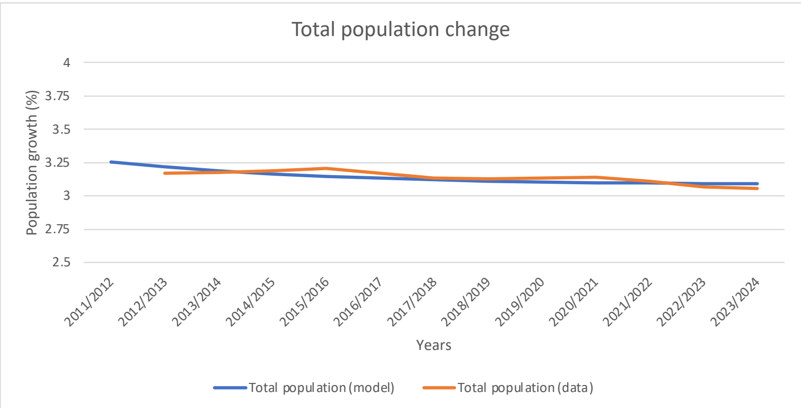


Figure S5.1: Model calibration results for total population change each year.
Source: National Bureau of Statistics (2013, 2018).

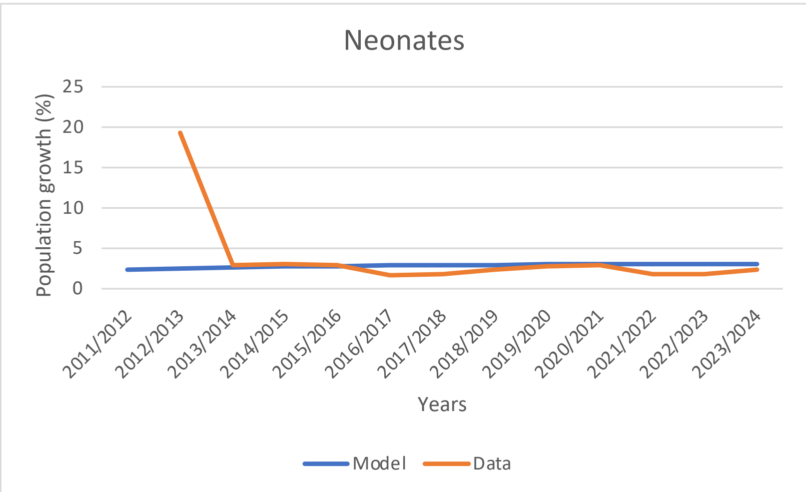


Figure S5.2: Model calibration results for number of neonates each year.
Source: National Bureau of Statistics (2013, 2018).

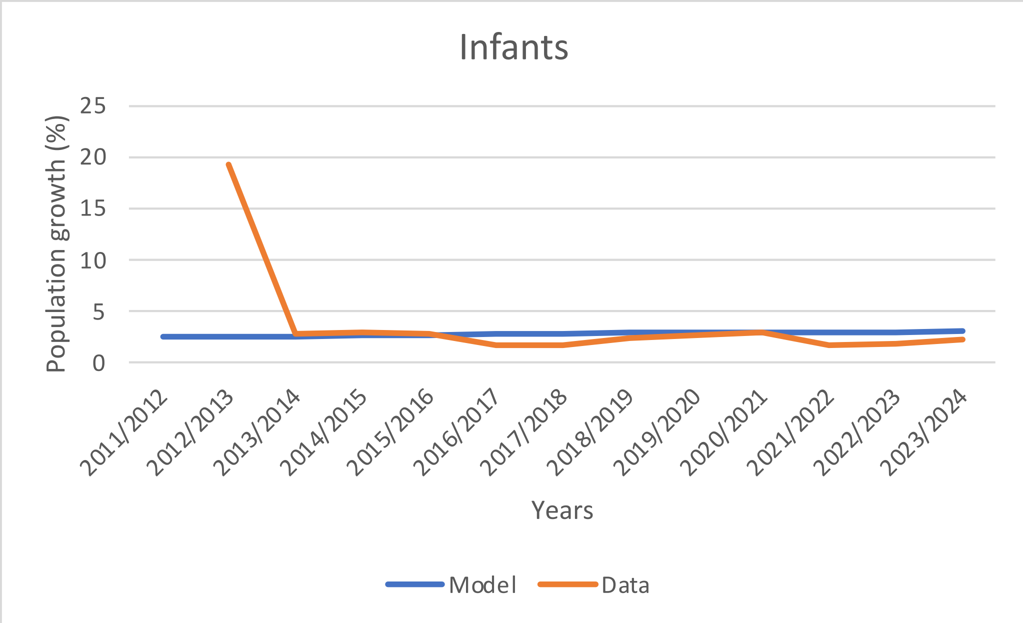


Figure S5.3: Model calibration results for number of infants each year. Source: National Bureau of Statistics (2013, 2018).

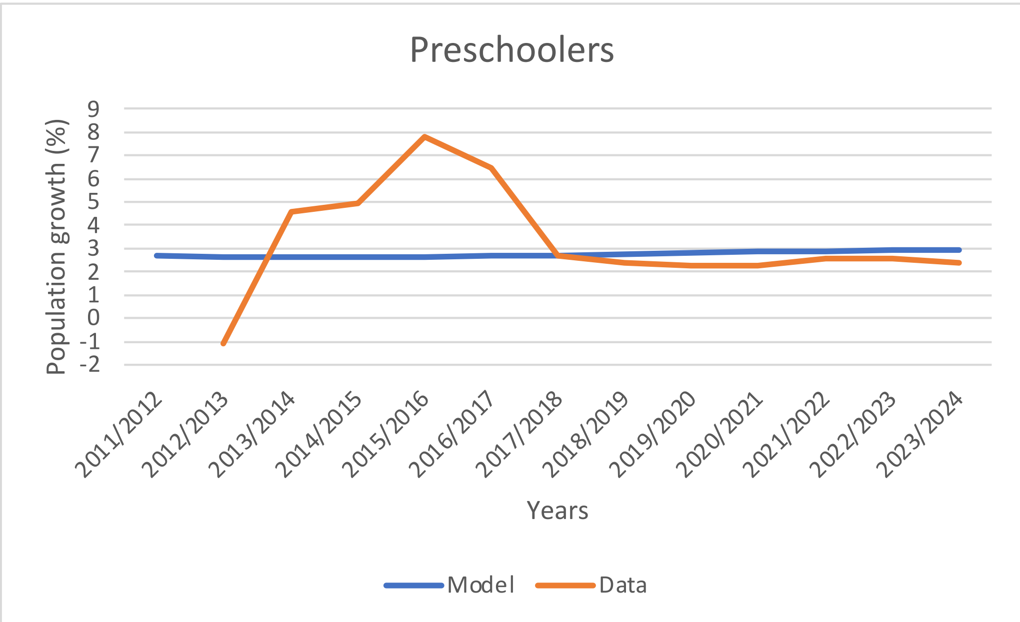


Figure S5.4: Model calibration results for number of preschoolers each year. Source: National Bureau of Statistics (2013, 2018).

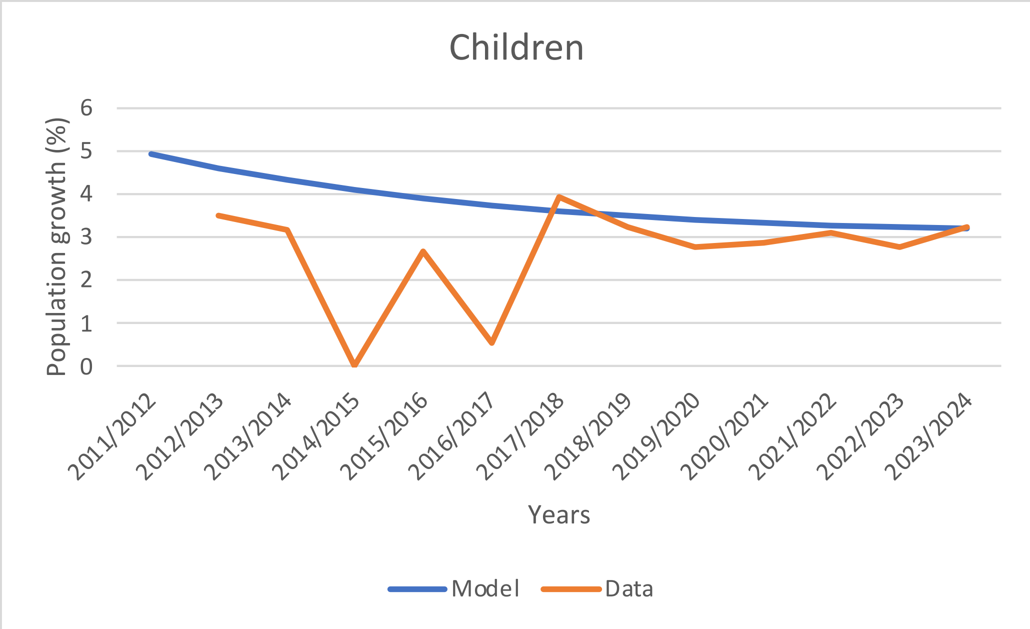


Figure S5.5: Model calibration results for number of children each year. Source: National Bureau of Statistics (2013, 2018).

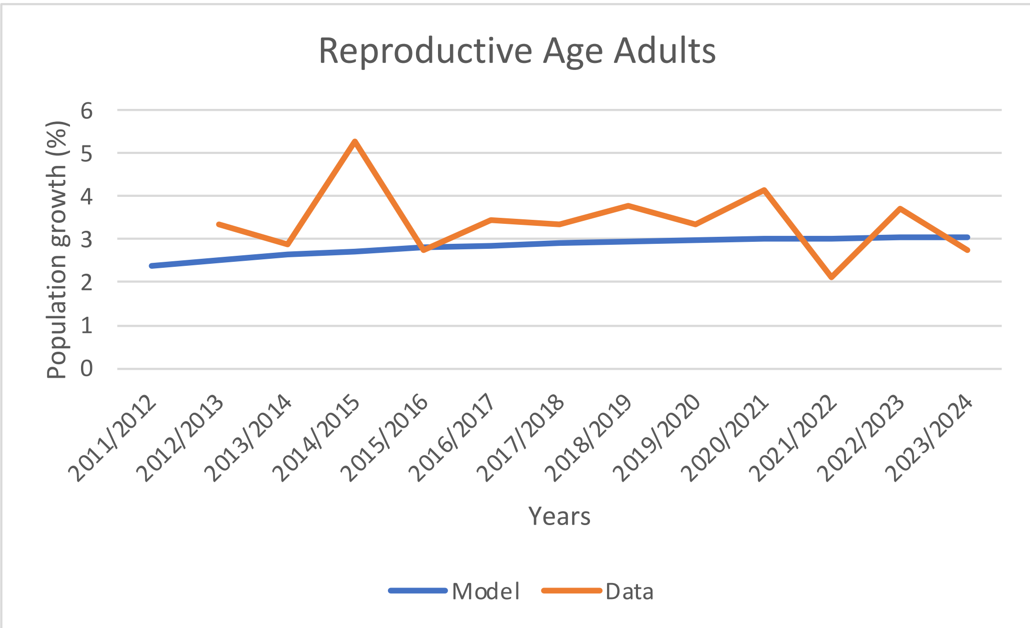


Figure S5.6: Model calibration results for number of reproductive age adults each year. Source: National Bureau of Statistics (2013, 2018).

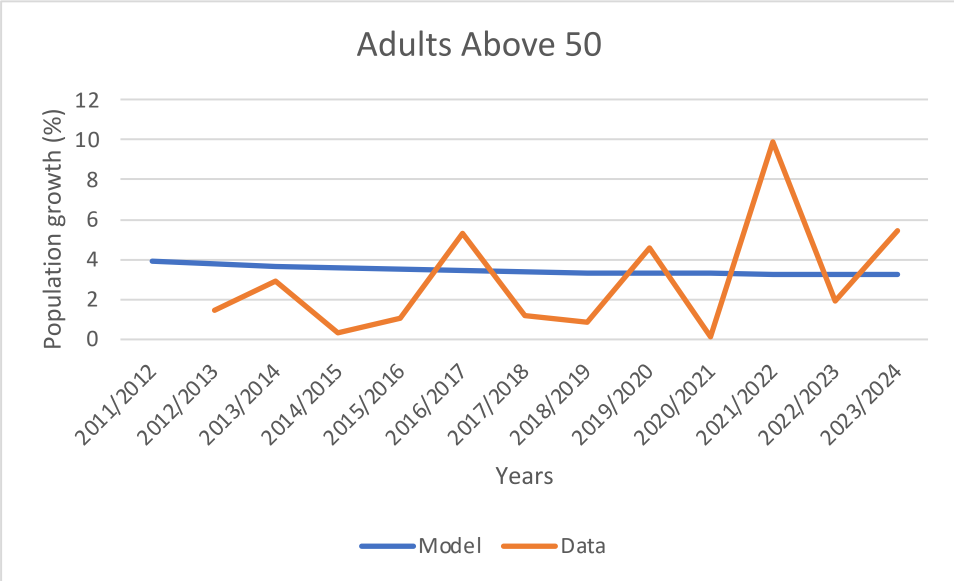


Figure S5.7: Model calibration results for number of adults above 50 each year. Source: National Bureau of Statistics (2013, 2018).

Facility operations

The percentage of positions filled at health facilities was output for each year in the model (Figure S5.8) and compared to data (Supplementary File 4, Table S4.6). The hiring rate for health worker positions was adjusted until a good fit for percentage of positions filled at health facilities was achieved (original and calibrated rates are shown in Supplementary File 4, Table S4.6). The model output is a good fit for the data aside from year 2013, where it is slightly underestimated.

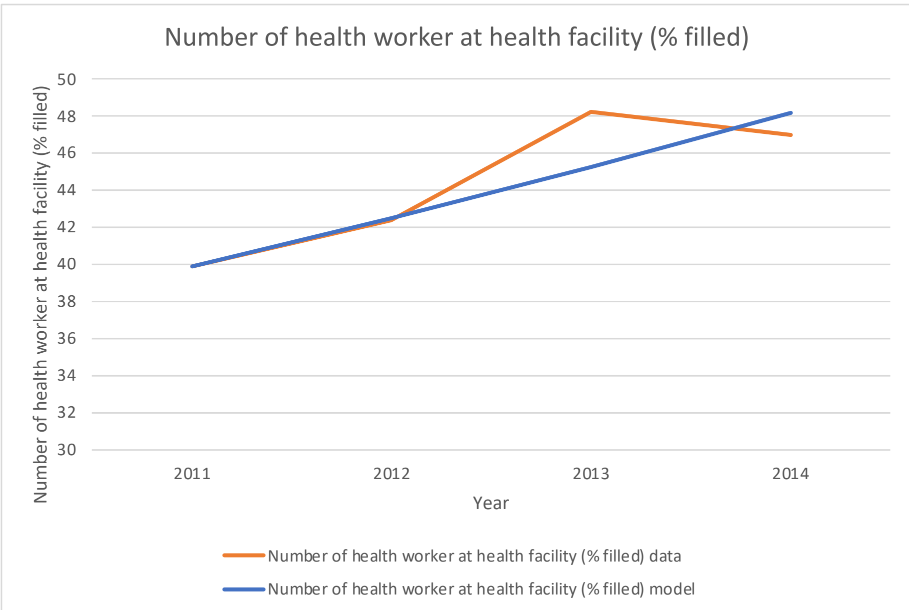


Figure S5.8: Model calibration results for number of health workers at health facility (% filled) each year. Source: Ministry of Health and Social Welfare (2013, 2014a, 2014b).

Facility funding

The performance of targeted services (percentage of women who received two doses of IPT and percentage of women who had a facility-based delivery) when P4P is switched 'on' and 'off' in the model was output for 'baseline', 'short term' and long term' time points and compared to data (Table S5.1) (Borghi *et al.* 2021). The improvement and decline between intervention and control sites was noted (Table S5.2) and used to calculate model calibration 'aim' for when P4P is switched 'off' in the model (Table S5.3), as initialisation of the model is the same regardless of whether P4P is then switched 'on' or 'off' in the model. The constant and uncertain (unable to extrapolate from data) variables 'MSD provision of IPT ordered' and 'MSD provision of FBD drugs ordered' were adjusted until a relatively good fit for both outputs were achieved (original and calibrated rates are shown in Supplementary File 4, Table S4.4 and Table S4.5).

The equation to calculate percentage of women who seek facility-based deliveries was further adjusted (from an average of inputs) to ensure a better fit for the intervention case, namely assigning greater weight to (i) the number of antenatal care visits; (ii) distance to facility; (iii) awareness of maternal and child health and healthcare in the community (in part estimated from ability to perform outreach controlled in **Facility Operations** sector and fraction of women attending antenatal care), over perceived quality of facility/services (estimated as an average of availability of drugs in **Facility Commodities** sector and patient-provider interaction from **Facility Operations** sector). See Supplementary 3 for equation details.

Table S5.1: Performance at intervention and control sites for targeted services (at least two doses of IPT during ANC (%), institutional delivery rate (%)) at baseline, short term and long term evaluation. Source Borghi et al. 2021.

	Intervention			**Control**		
	Baseline	Short term	Long term	Baseline	Short term	Long term
Targeted services						
At least two doses of IPT during ANC (%)	49.5	72.9	67.5	56.7	69.2	68.4
Institutional delivery rate (%)	84.7	89.2	92.2	86.8	83.1	89.4

Table S5.2: The difference in performance between baseline and short term, and short term and long term evaluations in the intervention and control sites.

	Intervention		Control	
	Baseline/Short term	Short term/Long term	Baseline/Short term	Short term/Long term
Targeted services				
At least two doses of IPT during ANC (%)	23.4	-5.4	12.5	-0.8
Institutional delivery rate (%)	4.5	3	-3.7	6.3

Table S5.3: The goal for model calibration for both intervention and control model runs.

	Intervention			Control		
	Baseline (aim)	Short term (aim)	Long term (aim)	Baseline (aim)	Short term (aim)	Long term (aim)
Targeted services						
At least two doses of IPT during ANC (%)	49.5	72.9	67.5	49.5	62.0	61.2
Institutional delivery rate (%)	84.7	89.2	92.2	84.7	81	87.3

The model output is a good fit for the percentage of women who received two doses of IPT when P4P is switched ‘on’ and ‘off’ in the model (Figures S5.9 and S5.10). The model overestimates the percentage of women who had a facility-based delivery when P4P is turned ‘off’ at the short term time point (Figure S5.11). In the real control evaluation site, facility-based deliveries actually decreased by 3.7% between baseline and the short-term evaluation (Borghi *et al.* 2021). Stakeholders were asked during validation interviews what might have caused this decrease in the control group, without reaching a definitive conclusion on the cause of the decrease which was in contrast to the national trend in facility-based deliveries during this period. Instead, a modest increase in the control group is assumed for this period. The model slightly underestimates the percentage of women who had a facility-based delivery when P4P is turned ‘on’ at the long-term time point (Figure S5.12).

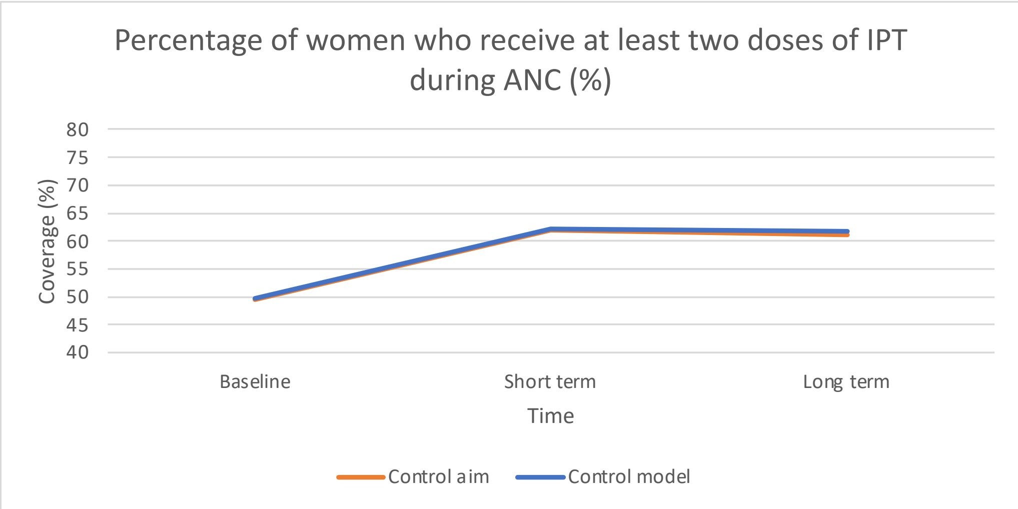


Figure S5.9: Model calibration results for percentage of women who receive at least two doses of IPT during ANC (%) in the control group.

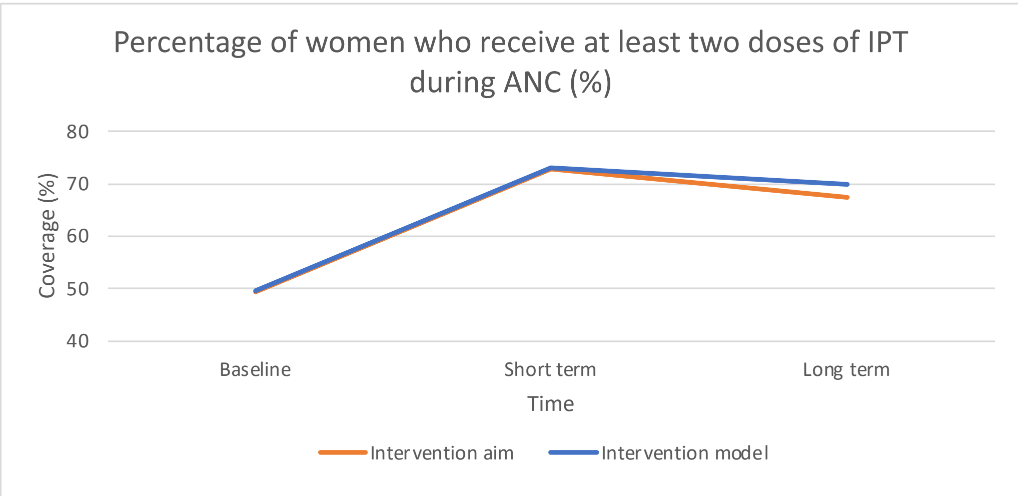


Figure S5.10: Model calibration results for percentage of women who receive at least two doses of IPT during ANC (%) in the intervention group.

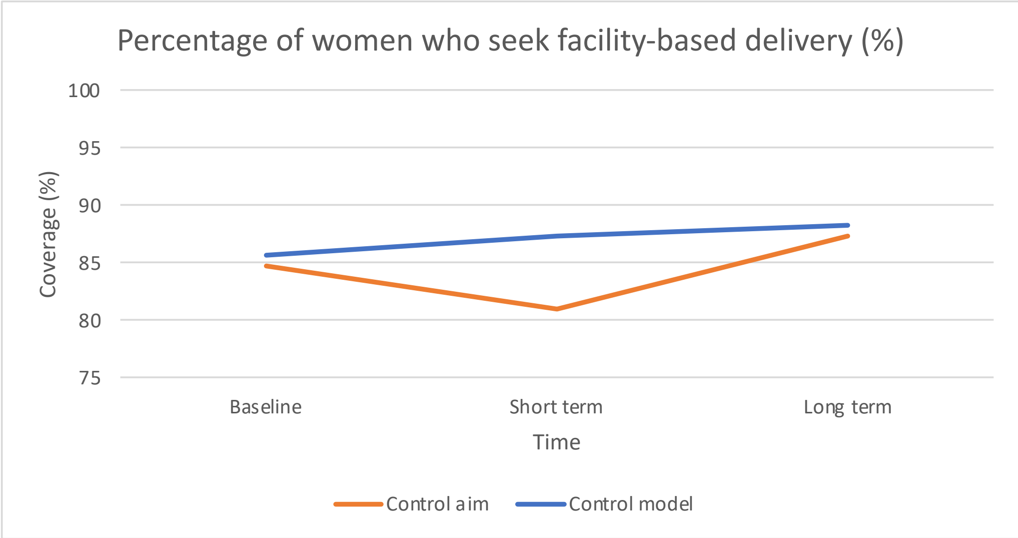


Figure S5.11: Model calibration results for percentage of women who seek a facility-based delivery (%) in the control group.

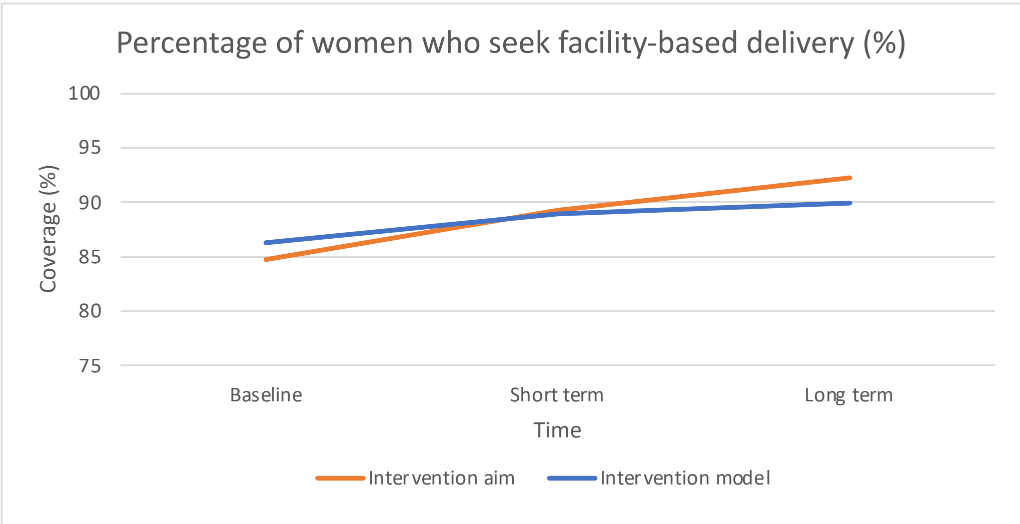


Figure S5.12: Model calibration results for percentage of women who seek a facility-based delivery (%) in the intervention group.

Supplementary File 6 – Stakeholder interview validation tool

Interviewer: This series of interviews have been organised by researchers from Ifakara Health Institute (IHI) and the London School of Hygiene and Tropical Medicine (LSHTM). We are hoping to conduct interviews with experts, such as yourself, to validate a simulation model we have created of the Tanzania maternal and child health (MCH) system response to payment for performance (P4P).

Interviewer: In an earlier interview (June/November/December 2020), we presented the system map we had developed using data collected during the P4P programme in Pwani that ran from 2011-2013. Using this system map, data collected during the programme, evidence from the literature and your feedback, we have now developed a computer model. In the model, we focus on the impact of P4P on the two programme targets where improvements were documented: provision of two doses of intermittent preventive treatment and facility-based deliveries.

Interviewer: To ensure our model is accurately representing the behaviour of the Tanzanian health system and its response to the Pwani P4P scheme, we now require model output to be validated by experts. We also intend to further develop the model to examine health system response to more recent health financing programmes:

(i) The national Results-Based Finance programme (RBF, 2016-2020)

(ii) Direct Health Facility Financing programme (DHFF, 2019-Present)

Interviewer: During this interview we will present and describe the key results from our simulation model related to the health system response to RBF (impact on health worker motivation, availability of medical commodities etc.). Using your knowledge of health system operation, your role will be to evaluate our results and determine their credibility. This crucial step in a series of model validation stages will provide confidence in the model, which will be used to develop policy recommendations for the implementation of results-based finance programmes in Tanzania and other settings.

Interviewer: Just before we begin, I have received a hard copy of your consent form but I would just like to seek your verbal consent that you are happy to continue with the interview and you are happy for me to take written notes and an audio-recording of this session. This is only for our records and shared only with our research team. You can change your mind or stop the interview at any time.

OR

Interviewer: Just before we begin, I am going to take verbal consent for the interview, I can then collect a hard copy of the consent form after the interview. I will now read through the informed consent form with you.

You understand that your participation is voluntary.
You understand that you can withdraw from the interview at any time.
You understand that your feedback during the interview will be recorded via written notes and audio-recorded by a member of the research team (using either encrypted web-based software or a handheld recording device).
You understand that your feedback collected in this interview will be kept confidential and be used for research purpose only. Also, you understand that your responses will be only shared with/by authorised individuals in the research team from LSHTM and IHI and any information included in the report, academic presentation or in published work will not be identified as the respondent.
You confirm that you are happy for the research team to contact you in the future to ask follow up questions.
You agree to take part in the above-named study.

Interviewer: *Interviewer then proceeds with presentation and discussion of key model results and output periodically stopping to check interviewee understanding and to ask if any modifications should be made to the behaviour of the model to reflect their experience of the Tanzania programme.*

Interviewer: At the moment, we are focused on Pwani P4P programme but also hope to model DHFF and RBF. We would like to ask you:

(i) if there other outputs you would like to see from the model

(ii) if there are things you would like to see varied in the model (to see impact on key outcomes).

(iii) do you think this model and types of simulation could be useful for decision making around the design of P4P programmes?

(iv) would you be happy for us to contact you with follow up questions?

Interviewer: Thank you for your time today, this has been incredibly useful and it is very much appreciated by the COSMIC team.

Model interface slides

Figure S6.1: Slide 1 – Introduction slide.

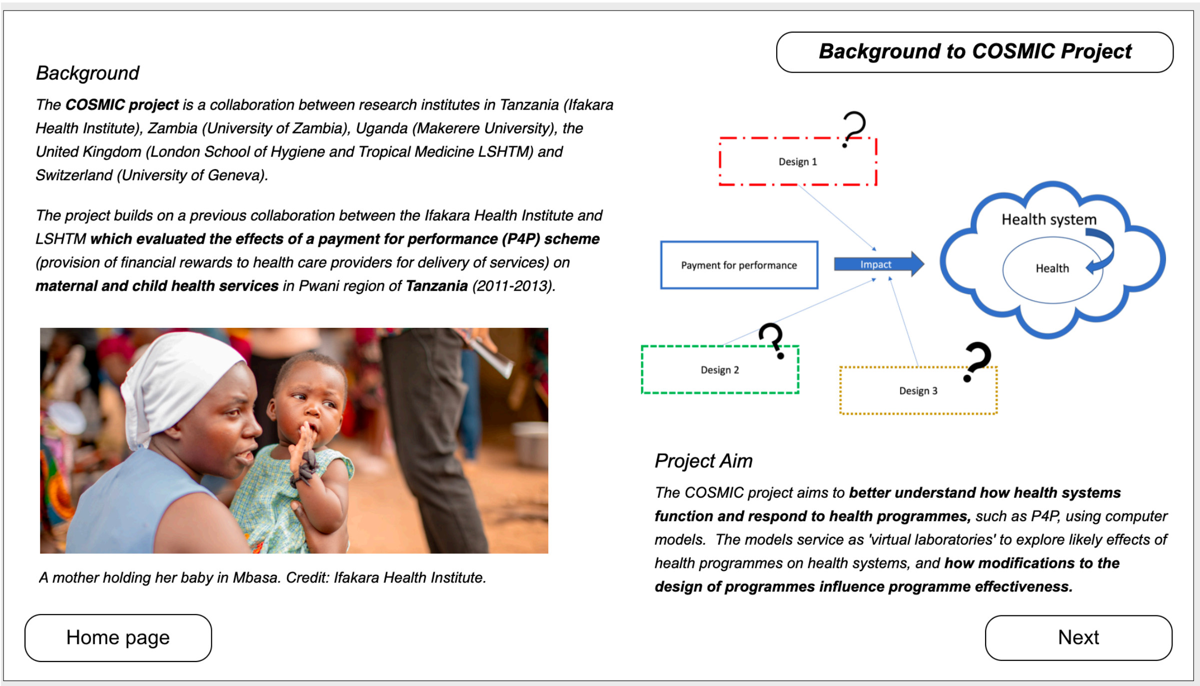


Figure S6.2: Slide 2 – Background (1) slide.

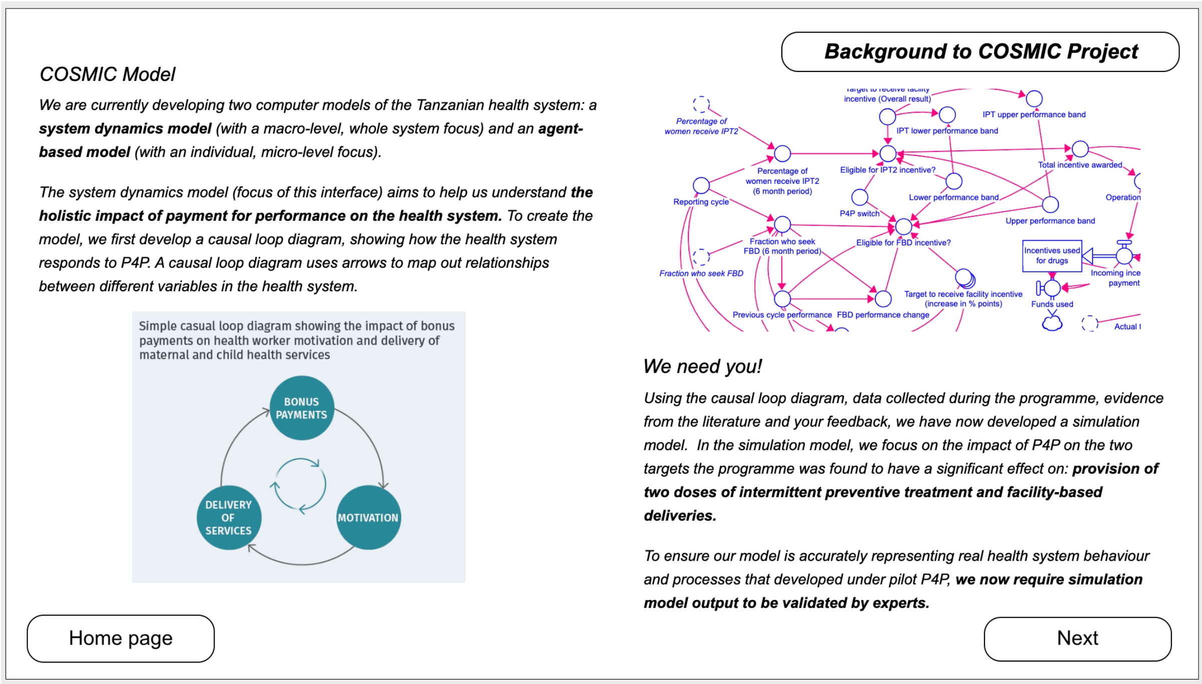


Figure S6.3: Slide 3 – Background (2) slide.

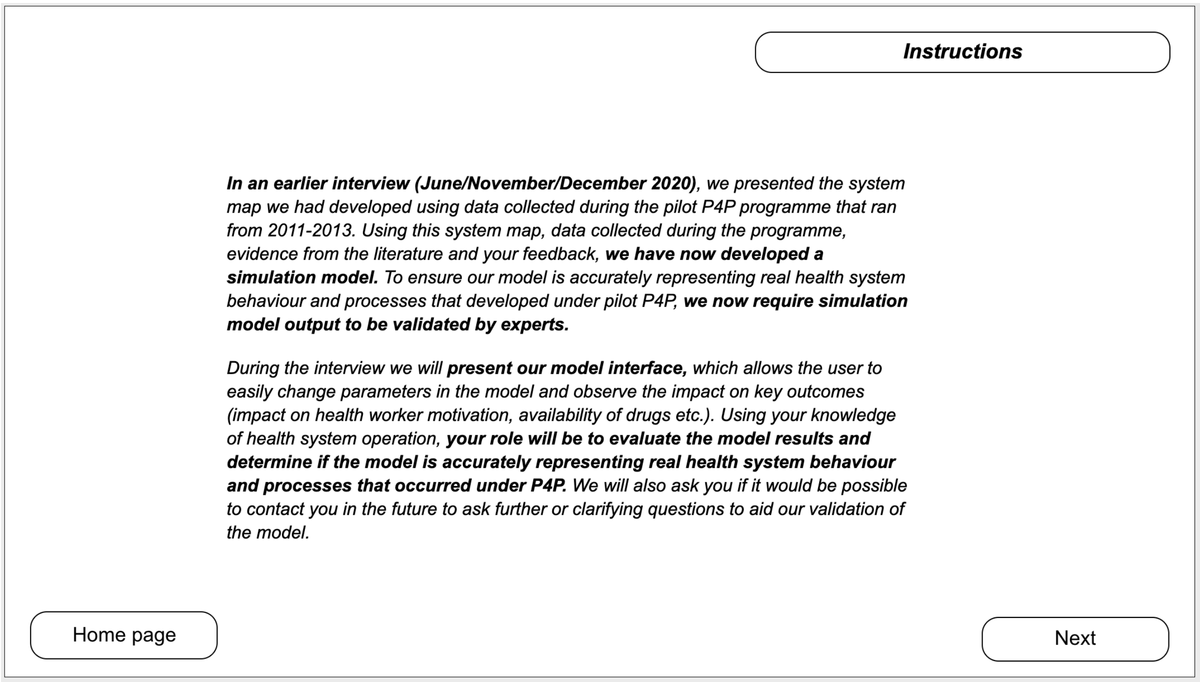


Figure S6.4: Slide 4 – Instruction slide.

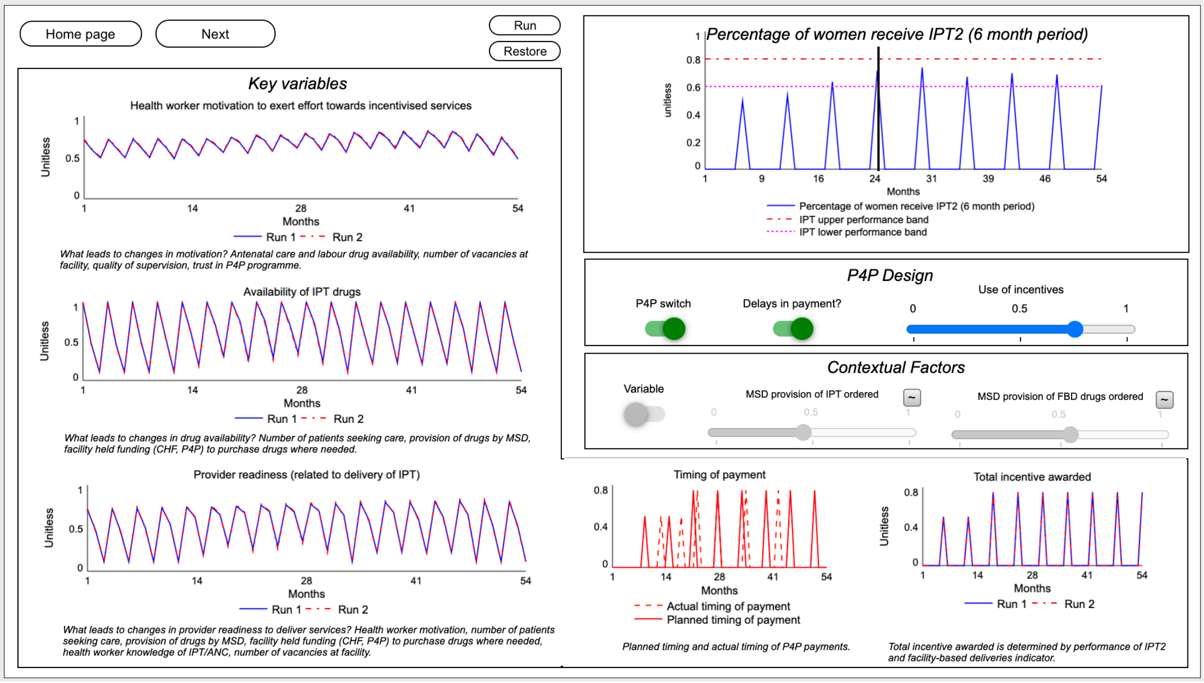


Figure S6.5: Slide 5 – Interactive slide to discuss dynamics around percentage of women who receive at least two doses of IPT during ANC.

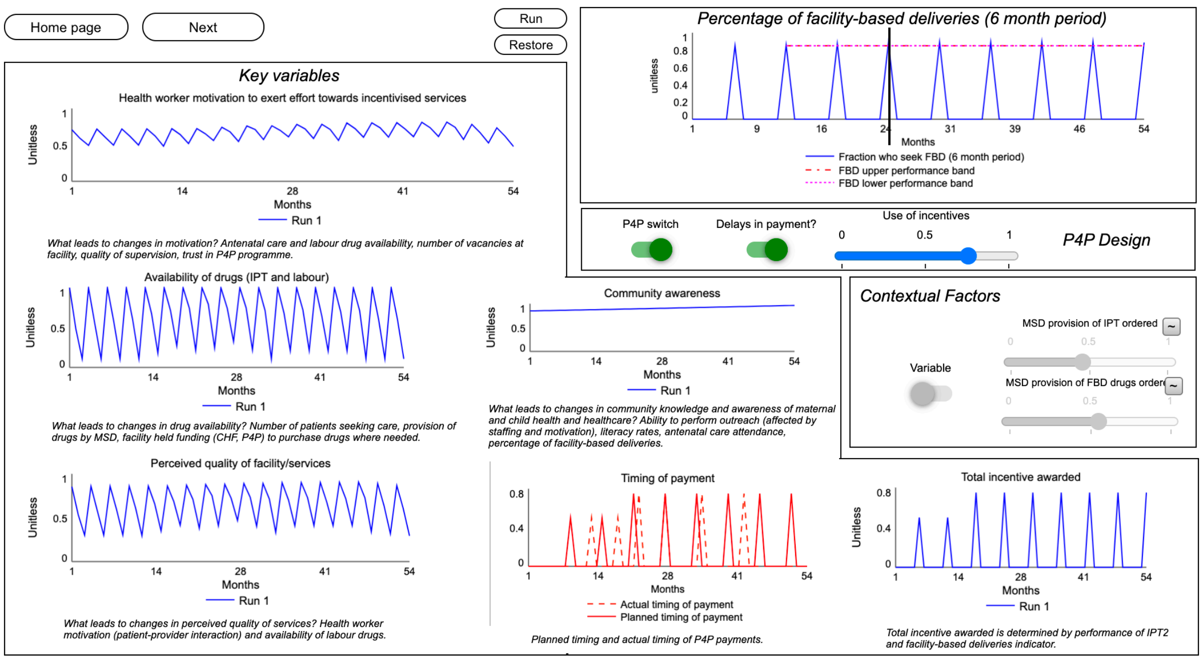


Figure S6.6: Slide 6 – Interactive slide to discuss dynamics around percentage of women who seek facility-based deliveries.

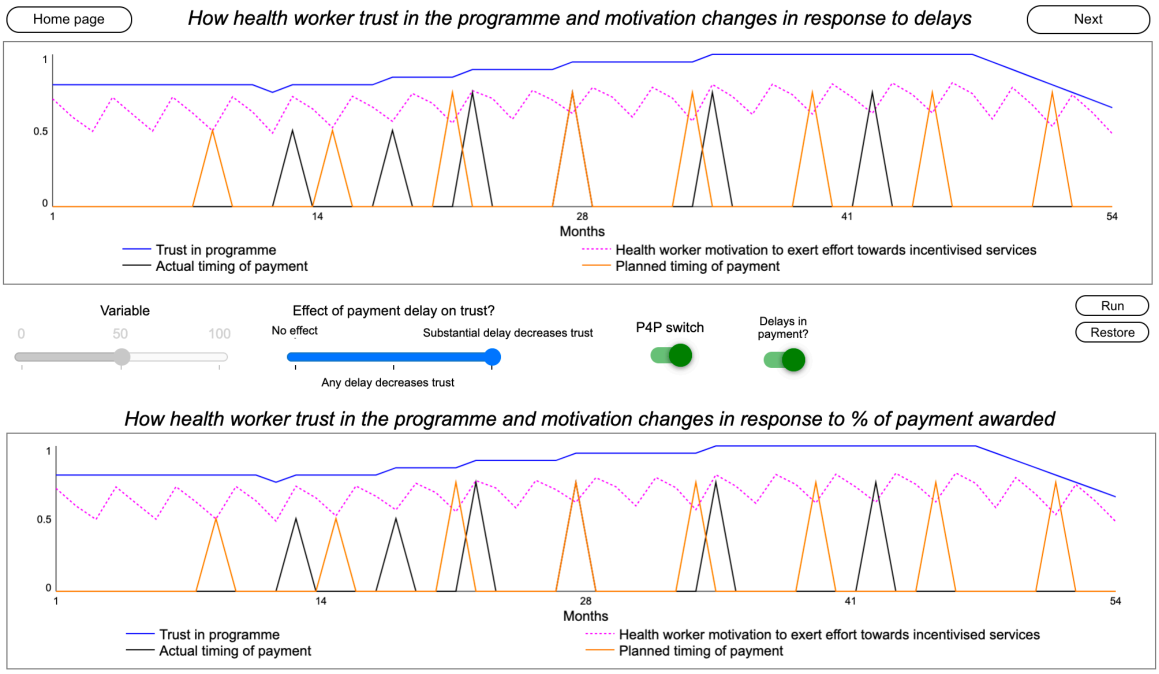


Figure S6.7: Slide 7 – Interactive slide to discuss how health worker trust and motivation responds to changes in timing of payments and amount of payment.

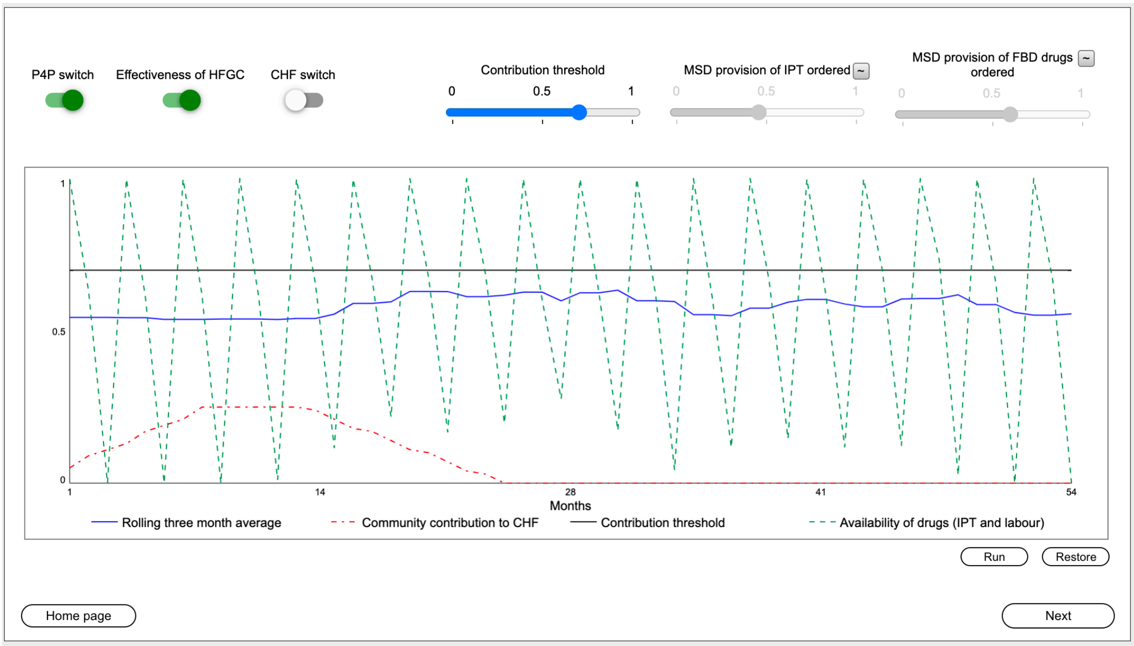


Figure S6.8: Slide 8 – Interactive slide to discuss how drug availability, perceived threshold of acceptable drug availability and effectiveness of Health Facility Governing Committee (HFGC) affect community payment into the Community Health Fund (CHF).

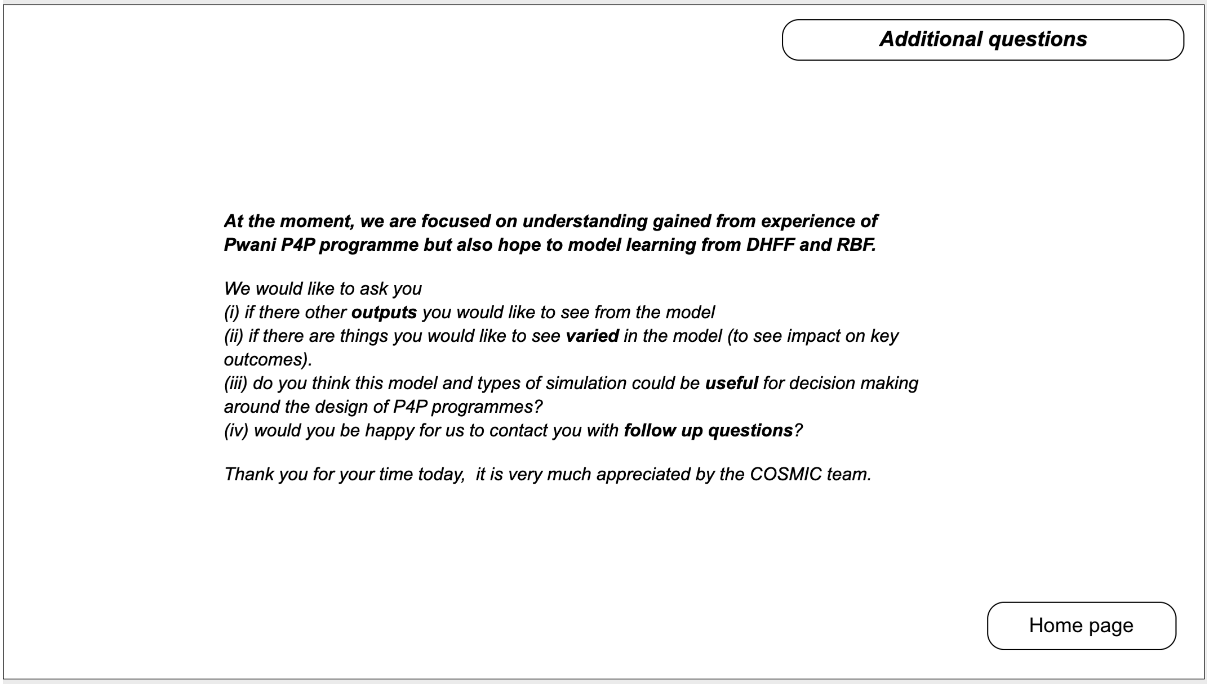


Figure S6.9: Slide 9 – Additional questions and conclusion slide.

Supplementary File 7 – Changes to model post stakeholder consultation

The feedback received during stakeholder consultations resulted in some new additions and adjustments to existing model structure, described here.

Key changes consisted of: (i) inclusion of an 'alternative facility held funding' variable (ii) adjustment of the effect of incentives on trust in programme and health worker motivation, and (iii) adjustment of effect of payment delays on trust in programme and health worker motivation. For change (i), stakeholders suggested inclusion of other types of facility held funding (other than P4P), as this would affect the purchasing power of facilities for buying additional drugs during the programme. In the current version of the model, this variable 'alternative facility held funding available' is static, but in future iterations of the model this will be dynamic. For change (ii), stakeholders commented that health workers would want to improve their performance, so a lower incentive payment (reflective of performance) would not be demotivating but would spur health workers on to try and improve their performance.

For change (iii), stakeholders were presented with three scenarios related to effect of payment delays in the model: i) payment delays do not affect trust and motivation, ii) any delay affects trust and motivation and iii) only severe delays (4+ months) affect trust and motivation. The consensus from stakeholders was that the third scenario, severe delays affect trust and motivation, was most likely, with communication of delays and expected payment dates sustaining trust and motivation up to a point. If payments are not made after a certain period (assumed 4+ months in the model), trust and motivation decrease until a payment is made. This relationship between payment delays, trust and motivation is retained in the model when payment delays are enacted.

Supplementary File 8 – Sensitivity analysis

The model was subjected to sensitivity analysis to determine the sensitivity of key outcomes (percentage of women who receive at least two doses of IPT during ANC, percentage of women who seek facility-based delivery) to changes in model parameters. Model parameters deemed appropriate for analysis were adjusted by 10%, with key outcome results recorded. Initial stock values, constant variable values (including table and graphical function values) were adjusted and simulated. Equation based parameters (flows and auxiliary variables) and constant variables where the value was not appropriate for adjustment (such as the 'on' and 'off' switch for turning the intervention on and off in the model, represented by '0' and '1' in the model) were not subjected to sensitivity analysis.

The following scale was used to determine sensitivity to changes in model variables; sensitive (5% ≤ change in outcome < 15%), very sensitive (15% ≤ change in outcome < 25%) and highly sensitive (25% ≥ change in outcome). Table cells are highlighted where outputs are categorised as sensitive (yellow), very sensitive (orange) and highly sensitive (red). The scale is adapted from Semwanga *et al.* (2016) and presented with smaller intervals for higher sensitivity categories, to further distinguish 'very sensitive' from 'highly sensitive' results. This analysis shed light on the likely effect of changes to programme design as well as contextual factors on key outcomes.

Population sector

Table S8.1: Sensitivity analyses for population sector parameters, change in key outcomes.

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Above50MortalityRate	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
AdultsAbove50(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
ChildMortalityRate	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Children(t)	-10%	0.497	0.731	0.700	0.862	0.889	0.899
	10%	0.495	0.730	0.699	0.862	0.889	0.899
Fraction_of_reproductive_age_adults_female	-10%	0.519	0.733	0.700	0.863	0.889	0.899
	10%	0.477	0.729	0.700	0.862	0.889	0.899
General_fertility_rate	-10%	0.519	0.733	0.700	0.863	0.889	0.899
	10%	0.477	0.729	0.700	0.862	0.889	0.899
InfantMortalityRate	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Infants(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
NeonateMortalityRate	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Neonates(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899

PreschoolerMortalityRate	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Preschoolers(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
ReprodAgeMortalityRate	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
ReproductiveAgeAdults(t)	-10%	0.518	0.732	0.699	0.862	0.889	0.899
	10%	0.478	0.730	0.700	0.862	0.889	0.899

Table S8.2: Sensitivity analyses for population sector parameters, change in key outcomes (% change).

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Above50MortalityRate	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
AdultsAbove50(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
ChildMortalityRate	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Children(t)	-10%	0.1	0.1	0.1	0.0	0.0	0.0
	10%	-0.1	-0.1	-0.1	0.0	0.0	0.0
Fraction_of_reproductive_age_adults_female	-10%	4.6	0.3	0.0	0.0	0.0	0.0
	10%	-3.8	-0.2	0.0	0.0	0.0	0.0
General_fertility_rate	-10%	4.6	0.3	0.0	0.0	0.0	0.0
	10%	-3.8	-0.2	0.0	0.0	0.0	0.0

InfantMortalityRate	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Infants(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
NeonateMortalityRate	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Neonates(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
PreschoolerMortalityRate	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Preschoolers(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
ReprodAgeMortalityRate	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
ReproductiveAgeAdults(t)	-10%	4.4	0.2	-0.1	0.0	0.0	0.0
	10%	-3.6	-0.1	0.1	0.0	0.0	0.0

Demand and Services sector

Table S8.3: Sensitivity analyses for demand and services (antenatal care) sector parameters, change in key outcomes.

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
ANC_visit_1(t)	-10%	0.487	0.731	0.700	0.862	0.889	0.899
	10%	0.505	0.731	0.700	0.863	0.889	0.899
ANC_visit_2(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
ANC_visit_2a(t)	-10%	0.489	0.730	0.700	0.862	0.889	0.899
	10%	0.503	0.731	0.700	0.863	0.889	0.899
ANC_visit_3(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
ANC_visit_3a(t)	-10%	0.493	0.731	0.700	0.862	0.889	0.899
	10%	0.499	0.731	0.700	0.862	0.889	0.899
ANC_visit_3b(t)	-10%	0.495	0.731	0.700	0.862	0.889	0.899
	10%	0.497	0.731	0.700	0.862	0.889	0.899
ANC_visit_4(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
ANC_visit_4a(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
ANC_visit_4b(t)	-10%	0.495	0.731	0.700	0.862	0.889	0.899
	10%	0.497	0.731	0.700	0.862	0.889	0.899
ANC_visit_4c(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899

	10%	0.496	0.731	0.700	0.862	0.889	0.899
Dropoutrate1	-10%	0.496	0.731	0.700	0.863	0.890	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.898
Dropoutrate2	-10%	0.497	0.733	0.702	0.863	0.889	0.899
	10%	0.495	0.729	0.698	0.862	0.889	0.899
Dropoutrate3	-10%	0.499	0.734	0.703	0.863	0.890	0.899
	10%	0.493	0.728	0.697	0.862	0.889	0.898
Dropoutrate4	-10%	0.501	0.735	0.705	0.863	0.890	0.899
	10%	0.491	0.726	0.695	0.862	0.889	0.898

Table S8.4: Sensitivity analyses for demand and services (antenatal care) sector parameters, change in key outcomes (% change).

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
ANC_visit_1(t)	-10%	-1.7	0.0	0.0	0.0	0.0	0.0
	10%	1.7	0.0	0.0	0.0	0.0	0.0
ANC_visit_2(t)	-10%	-0.1	0.0	0.0	0.0	0.0	0.0
	10%	0.1	0.0	0.0	0.0	0.0	0.0
ANC_visit_2a(t)	-10%	-1.4	0.0	0.0	0.0	0.0	0.0
	10%	1.4	0.0	0.0	0.0	0.0	0.0
ANC_visit_3(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
ANC_visit_3a(t)	-10%	-0.5	0.0	0.0	0.0	0.0	0.0
	10%	0.5	0.0	0.0	0.0	0.0	0.0
ANC_visit_3b(t)	-10%	-0.2	0.0	0.0	0.0	0.0	0.0

	10%	0.2	0.0	0.0	0.0	0.0	0.0
ANC_visit_4(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
ANC_visit_4a(t)	-10%	-0.1	0.0	0.0	0.0	0.0	0.0
	10%	0.1	0.0	0.0	0.0	0.0	0.0
ANC_visit_4b(t)	-10%	-0.2	0.0	0.0	0.0	0.0	0.0
	10%	0.2	0.0	0.0	0.0	0.0	0.0
ANC_visit_4c(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Dropoutrate1	-10%	0.0	0.0	0.0	0.1	0.0	0.0
	10%	0.0	0.0	0.0	-0.1	0.0	0.0
Dropoutrate2	-10%	0.2	0.3	0.3	0.0	0.0	0.0
	10%	-0.2	-0.3	-0.3	0.0	0.0	0.0
Dropoutrate3	-10%	0.5	0.4	0.4	0.0	0.0	0.0
	10%	-0.5	-0.4	-0.4	0.0	0.0	0.0
Dropoutrate4	-10%	1.0	0.6	0.7	0.1	0.1	0.1
	10%	-1.0	-0.6	-0.7	-0.1	-0.1	-0.1

Table S8.5: Sensitivity analyses for demand and services (facility-based delivery care) sector parameters, change in key outcomes.

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Belief_in_myths	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Community_awareness(t)	-10%	0.496	0.637	0.702	0.839	0.857	0.871
	10%	0.496	0.731	0.700	0.886	0.914	0.914
Decay_awareness	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Distance_to_facility	-10%	0.496	0.637	0.702	0.836	0.854	0.871
	10%	0.496	0.731	0.700	0.873	0.895	0.909
Effect_ANC_awareness	-10%	0.496	0.731	0.700	0.862	0.889	0.898
	10%	0.496	0.731	0.700	0.863	0.890	0.900
Effect_FBD_awareness	-10%	0.496	0.731	0.700	0.862	0.889	0.898
	10%	0.496	0.731	0.700	0.863	0.890	0.900
Effect_literacy_awareness	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Effect_myths	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Effect_outreach_awareness	-10%	0.496	0.731	0.700	0.862	0.889	0.898
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Literacy	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
"Number_of_facility-based_deliveries"(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899

	10%	0.496	0.731	0.700	0.862	0.889	0.899
Prob_FBD_ANC_1	-10%	0.496	0.731	0.700	0.862	0.889	0.898
	10%	0.496	0.731	0.700	0.863	0.890	0.899
Prob_FBD_ANC_2	-10%	0.496	0.731	0.700	0.861	0.888	0.897
	10%	0.496	0.731	0.700	0.864	0.891	0.900
Prob_FBD_ANC_3	-10%	0.496	0.698	0.699	0.856	0.881	0.892
	10%	0.496	0.731	0.700	0.869	0.896	0.906
Prob_FBD_ANC_4	-10%	0.496	0.698	0.699	0.845	0.870	0.882
	10%	0.496	0.731	0.700	0.880	0.906	0.915

Table S8.6: Sensitivity analyses for (facility-based delivery care) sector parameters, change in key outcomes (% change).

		Percentage of women who receive at least two doses of IPT during ANC			**Percentage of women who seek facility-based delivery**		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Belief_in_myths	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Community_awareness(t)	-10%	0.0	-12.8	0.2	-2.8	-3.6	-3.0
	10%	0.0	0.0	0.0	2.8	2.8	1.6
Decay_awareness	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Distance_to_facility	-10%	0.0	-12.8	0.2	-3.1	-3.9	-3.1
	10%	0.0	0.0	0.0	1.2	0.7	1.1
Effect_ANC_awareness	-10%	0.0	0.0	0.0	0.0	-0.1	-0.1
	10%	0.0	0.0	0.0	0.0	0.1	0.1
Effect_FBD_awareness	-10%	0.0	0.0	0.0	0.0	0.0	-0.1

	10%	0.0	0.0	0.0	0.0	0.0	0.1
Effect_literacy_awareness	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Effect_myths	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Effect_outreach_awareness	-10%	0.0	0.0	0.0	0.0	0.0	-0.1
	10%	0.0	0.0	0.0	0.0	0.0	0.1
Literacy	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
"Number_of_facility-based_deliveries"(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Prob_FBD_ANC_1	-10%	0.0	0.0	0.0	-0.1	-0.1	-0.1
	10%	0.0	0.0	0.0	0.1	0.1	0.1
Prob_FBD_ANC_2	-10%	0.0	0.0	0.0	-0.2	-0.2	-0.2
	10%	0.0	0.0	0.0	0.2	0.2	0.2
Prob_FBD_ANC_3	-10%	0.0	-4.5	-0.1	-0.7	-0.9	-0.8
	10%	0.0	0.0	0.0	0.7	0.7	0.8
Prob_FBD_ANC_4	-10%	0.0	-4.5	-0.1	-2.0	-2.1	-1.9
	10%	0.0	0.0	0.0	2.0	1.9	1.9

Facility Commodities sector

Table S8.7: Sensitivity analyses for facility commodities (antenatal care) sector parameters, change in key outcomes.

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
MSD_provision_of_IPT_ordered	-10%	0.440	0.666	0.635	0.860	0.884	0.897
	10%	0.569	0.786	0.753	0.865	0.895	0.901
Order_of_IPT_drugs(t)	-10%	0.490	0.730	0.700	0.862	0.889	0.899
	10%	0.502	0.731	0.700	0.862	0.889	0.899
Stock_of_IPT_drugs(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.489	0.731	0.700	0.862	0.889	0.899

Table S8.8: Sensitivity analyses for facility commodities (antenatal care) sector parameters, change in key outcomes (% change).

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
MSD_provision_of_IPT_ordered	-10%	-11.3	-8.8	-9.3	-0.3	-0.6	-0.2
	10%	14.7	7.5	7.6	0.3	0.6	0.2
Order_of_IPT_drugs(t)	-10%	-1.2	-0.1	0.0	0.0	0.0	0.0
	10%	1.2	0.1	0.0	0.0	0.0	0.0
Stock_of_IPT_drugs(t)	-10%	-0.1	0.0	0.0	0.0	0.0	0.0
	10%	-1.3	0.0	0.0	0.0	0.0	0.0

Table S8.9: Sensitivity analyses for facility commodities (facility-based delivery care) parameters, change in key outcomes.

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
MSD_provision_of_FBD_drugs_ordered	-10%	0.496	0.731	0.700	0.859	0.888	0.898
	10%	0.496	0.731	0.700	0.865	0.891	0.902
Order_of_FBD_drugs(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Stock_of_FBD_drugs(t)	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899

Table S8.10: Sensitivity analyses for facility commodities (facility-based delivery care) parameters, change in key outcomes (% change).

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
MSD_provision_of_FBD_drugs_ordered	-10%	0.0	0.0	0.0	-0.4	-0.2	-0.1
	10%	0.0	0.0	0.0	0.2	0.2	0.4
Order_of_FBD_drugs(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Stock_of_FBD_drugs(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0

Facility Operations sector

Table S8.11: Sensitivity analyses for facility operations sector parameters, change in key outcomes.

		Percentage of women who receive at least two doses of IPT during ANC			**Percentage of women who seek facility-based delivery**		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Average_attrition_rate	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.730	0.700	0.862	0.889	0.899
Effect_of_delays_in_payment_on_trust [3]	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.900
"Effect_of_incentive_on_trust_(change)"	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Effect_of_supervision_on_knowledge[1, 1]	-10%	0.496	0.730	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
Hiring_rate	-10%	0.496	0.730	0.700	0.862	0.889	0.898
	10%	0.496	0.731	0.700	0.862	0.889	0.899
"Knowledge_of_health_workers_(IPT)"(t)	-10%	0.497	0.729	0.701	0.863	0.891	0.899
	10%	0.494	0.733	0.700	0.862	0.888	0.899
"Number_of_health_worker_at_health_facility_(%_filled)"(t)	-10%	0.497	0.729	0.700	0.862	0.889	0.898
	10%	0.495	0.733	0.699	0.863	0.889	0.900
Trust_in_programme(t)	-10%	0.496	0.730	0.700	0.861	0.888	0.898
	10%	0.496	0.731	0.700	0.864	0.891	0.900

Table S8.12: Sensitivity analyses for facility operations sector parameters, change in key outcomes (% change).

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Average_attrition_rate	-10%	0.0	0.0	0.1	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
Effect_of_delays_in_payment_on_trust [3]	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.1
"Effect_of_incentive_on_trust_(change)"	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.1
Effect_of_supervision_on_knowledge[1, 1]	-10%	0.0	-0.1	0.0	0.0	0.0	0.0
	10%	0.0	0.1	0.0	0.0	0.0	0.0
Hiring_rate	-10%	0.0	-0.1	0.0	0.0	0.0	0.0
	10%	0.0	0.1	0.0	0.0	0.0	0.0
"Knowledge_of_health_workers_(IPT)"(t)	-10%	0.3	-0.3	0.2	0.0	0.2	0.0
	10%	-0.3	0.4	0.0	0.0	-0.2	0.0
"Number_of_health_worker_at_health_facility_(%_filled)"(t)	-10%	0.2	-0.2	0.1	-0.1	0.0	-0.1
	10%	-0.2	0.3	-0.1	0.1	0.0	0.1
Trust_in_programme(t)	-10%	0.1	-0.1	0.1	-0.2	-0.2	-0.1
	10%	-0.1	0.1	0.0	0.2	0.2	0.1

Facility Funding sector

Table S8.13: Sensitivity analyses for facility funding sector parameters, change in key outcomes.

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Alternative_facility_held_funding_available	-10%	0.483	0.723	0.691	0.862	0.888	0.898
	10%	0.509	0.739	0.708	0.863	0.890	0.899
"Target_to_receive_facility_incentive_(increase_in_%_points)"	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
"Target_to_receive_facility_incentive_(Overall_result)"	-10%	0.496	0.731	0.693	0.862	0.889	0.898
	10%	0.496	0.727	0.700	0.862	0.888	0.899
Use_of_incentives	-10%	0.496	0.759	0.723	0.862	0.891	0.900
	10%	0.496	0.701	0.676	0.862	0.886	0.900

Table S8.14: Sensitivity analyses for facility funding sector parameters, change in key outcomes (% change).

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Alternative_facility_held_funding_available	-10%	-2.6	-1.1	-1.2	-0.1	-0.1	-0.1
	10%	2.6	1.1	1.2	0.1	0.1	0.1
"Target_to_receive_facility_incentive_(increase_in_%_points)"	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
"Target_to_receive_facility_incentive_(Overall_result)"	-10%	0.0	0.0	-1.0	0.0	0.0	0.0
	10%	0.0	-0.5	0.0	0.0	-0.1	0.0
Use_of_incentives	-10%	0.0	3.8	3.4	0.0	0.2	0.2
	10%	0.0	-4.0	-3.5	0.0	-0.4	0.1

District Manager Operations sector

Table S8.15: Sensitivity analyses for district manager operations sector parameters, change in key outcomes.

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Baseline_motivation	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
"District_budget/resources"	-10%	0.496	0.726	0.699	0.862	0.891	0.900
	10%	0.496	0.731	0.700	0.863	0.889	0.899
District_manager_motivation_to_support_facilities(t)	-10%	0.496	0.730	0.700	0.862	0.889	0.898
	10%	0.496	0.731	0.700	0.863	0.890	0.899
Effect_of_incentive_amount_on_motivation	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899
"Skill_level_or_knowledge_of_district_manager_(supervision)"	-10%	0.496	0.730	0.700	0.862	0.889	0.898
	10%	0.496	0.731	0.700	0.863	0.890	0.899
"Target_to_receive_CHMT_incentive_(Overall_result)"	-10%	0.496	0.731	0.700	0.862	0.889	0.899
	10%	0.496	0.731	0.700	0.862	0.889	0.899

Table S8.16: Sensitivity analyses for district manager operations sector parameters, change in key outcomes (% change).

		Percentage of women who receive at least two doses of IPT during ANC			Percentage of women who seek facility-based delivery		
Parameter values adjusted	**Adjustment**	*Baseline*	*Endline*	*Longterm*	*Baseline*	*Endline*	*Longterm*
Baseline_motivation	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0
"District_budget/resources"	-10%	0.0	-0.6	-0.1	0.0	0.2	0.1
	10%	0.0	0.0	0.0	0.0	0.0	0.1
District_manager_motivation_to_support_facilities(t)	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.1
Effect_of_incentive_amount_on_motivation	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.1	0.0	0.0	0.0
"Skill_level_or_knowledge_of_district_manager_(supervision)"	-10%	0.0	-0.1	0.0	-0.1	0.0	-0.1
	10%	0.0	0.1	0.0	0.1	0.0	0.1
"Target_to_receive_CHMT_incentive_(Overall_result)"	-10%	0.0	0.0	0.0	0.0	0.0	0.0
	10%	0.0	0.0	0.0	0.0	0.0	0.0

References

Binyaruka P, Patouillard E, Powell-Jackson T *et al.* 2015. Effect of Paying for Performance on Utilisation, Quality, and User Costs of Health Services in Tanzania: A Controlled Before and After Study. Ostermann J (ed.). *PLoS One* **10**: e0135013.

Borghi J, Binyaruka P, Mayumana I *et al.* 2021. Long-term effects of payment for performance on maternal and child health outcomes: evidence from Tanzania. *BMJ Glob Heal* **6**: e006409.

Cassidy R, Tomoaia-Cotisel A, Semwanga AR *et al.* 2021. Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach. *Soc Sci Med* **285**: 114277.

Ensor T, Quigley P, Green C *et al.* 2014. Knowledgeable antenatal care as a pathway to skilled delivery: modelling the interactions between use of services and knowledge in Zambia. *Health Policy Plan* **29**: 580–8.

Kurowski C, Wyss K, Abdulla S *et al.* 2007. Scaling up priority health interventions in Tanzania: the human resources challenge. *Health Policy Plan* **22**: 113–27.

Ministry of Health and Social Welfare. 2012. *The Pwani Region Pay-for-Performance (P4P) Pilot Design Document*. Dar es Salaam: MoHSW, United Republic of Tanzania

Ministry of Health and Social Welfare. 2013. *Human Resource for Health Country Profile 2012/2013*. p 34. Date of access: December 29, 2022. URL: https://www.jica.go.jp/project/tanzania/006/materials/ku57pq00001x6jyl-att/country_profile_2013.pdf

Ministry of Health and Social Welfare. 2014a. *Staffing Levels for Ministry of Health and Social Welfare Departments, Health Service Facilities, Health Training Institutions and Agencies 2014-2019.* p xi. Date of access: December 29, 2022. URL: https://www.jica.go.jp/project/tanzania/006/materials/ku57pq00001x6jyl-att/REVIEW_STAFFING_LEVEL_2014-01.pdf

Ministry of Health and Social Welfare. 2014b. *Human Resource for Health and Social Welfare Country Profile 2013/2014*. p 35. Date of access: December 29, 2022. URL: https://www.jica.go.jp/project/tanzania/006/materials/ku57pq00001x6jyl-att/COUNTRY_PROFILE.pdf

National Bureau of Statistics. 2011. *Tanzania Demographic and Health Survey*. p 55, 118, 263.

Date of access: December 29, 2022. URL: https://dhsprogram.com/publications/publication-fr243-dhs-final-reports.cfm

National Bureau of Statistics. 2013. *Population and Housing Census: Population Distribution by Age and Sex*. p 157. Date of access: December 29, 2022. URL: https://www.nbs.go.tz/index.php/en/census-surveys/population-and-housing-census/163-phc-2012-population-distribution-by-age-and-sex-report

National Bureau of Statistics. 2015. *Mortality and Health Report*. p 24-26.

National Bureau of Statistics. 2018. *National Population Projections*. p 67-76. Date of access: December 29, 2022. URL: https://www.nbs.go.tz/nbs/takwimu/census2012/Projection-Report-20132035.pdf

Semwanga AR, Nakubulwa S, Adam T. 2016. Applying a system dynamics modelling approach to explore policy options for improving neonatal health in Uganda. *Heal Res Policy Syst* **14**: 35.

UNESCO Institute for Statistics. United Republic of Tanzania: Education and Literacy. Date of access: July 20, 2021. URL: http://uis.unesco.org/en/country/tz

DISCUSSION AND CONCLUSION

9 Discussion

9.1 Introduction

Abbreviations: 2 doses of intermittent preventative treatment (IPT2); Agent-based modelling (ABM); Antenatal care (ANC); Causal loop diagram (CLD); Community health fund (CHF); Direct Health Facility financing (DHFF); Discrete-event simulation (DES); Group Model Building (GMB); Low- and middle-income country (LMIC); Maternal and child health (MCH); Payment for performance (P4P); Results Based Financing (RBF); System dynamics model (SDM).

In this final chapter, the thesis findings are summarised, accompanied by strengths and limitations of the work, closing with research and policy implications and recommendations. The goal of this thesis was to use a systems thinking approach to better understand pathways to effect for P4P programmes, leading to recommendations for more effective implementation in LMICs. Each chapter that comprises of a published (or ready for publication) paper contributes towards the thesis goal and objectives:

Paper 1: A systematic review identifying application of systems thinking (SDM) for health systems research, with reflections relevant for future research conducted in LMIC settings (**Objective 1**).

Paper 2: Application of systems thinking (CLD) to explore pathways to impact for P4P programmes, identification of bottlenecks to successful implementation and system leverage points that should be targeted by P4P in LMIC settings (**Objectives 2 and 3**)

Paper 3: A guidance piece for application of systems thinking (CLD) for health systems research in LMICs, drawing on candidates own experience of application and evidence from the literature (**Objectives 1 and 5**).

Paper 4: Application of systems thinking (SDM) to explore how variation in the design, implementation and context of P4P influence programme effectiveness, with recommendations for future implementation (**Objective 4**).

9.2 Synthesis of key findings

9.2.1 Synthesis of key findings

The Pwani P4P programme was implemented with expectation to improve the delivery and coverage of targeted MCH services in participating districts. Health workers and district managers were eligible for incentive payments when certain performance targets were met, with the allocation of funds for health providers split between incentives for staff and funding to be used to strengthen facility operations. The anticipated pathways to improving MCH services under P4P were (i) an increase in health worker motivation to deliver services (ii) an increase in district manager motivation to support facilities and (iii) additional funding to be used to strengthen facility operations (i.e. purchasing medicines for ANC and facility-based delivery services). These pathways to effect were identified in the CLD developed in this thesis, but the success of the programme was found to be dependent on many other interconnected pathways and processes in the health system. Three broad mechanisms were found to be responsible for provider success (or failure) under the programme; (i) changes in the supply of services, (ii) changes to the demand for services and (iii) changes to facility reporting of performance. All three mechanisms needed to operate effectively under P4P for the programme to be a success and positively impact MCH services.

Inspection of what facilitated (or hindered) operation of these mechanisms yielded certain catalytic variables (adjustment to the variable causes widespread change to the broader system, and therefore deserves serious consideration in the design of a P4P, or indeed any health system strengthening, programme) and system levers (system elements that should be incorporated into the design of P4P to enhance the effects of the programme).

The extent to which the P4P programme effectively targeted these system elements varied. Facility readiness, staffing and supervision of facilities were identified as catalytic variables in the CLD. Facility readiness, particularly availability of medicines, is crucial for effective MCH service delivery, not only as a direct requirement for supply of services but also due to its impact on health worker motivation. Readiness also impacts patient perceived quality of care (and through this pathway, patient uptake of care), as well as community decision-making on payment to the community-based health insurance scheme, used to financially support facility operations. The P4P programme targeted facility readiness through provision of additional funding to purchase needed medications for incentivised services. The routine supply chain for provision of medicines (from the Medical Stores Department) was regularly disrupted during the time of programme implementation and issuing of incentive payments to health providers were often delayed. As observed in results from the SDM, frequent supply chain disruption and repeated failure to issue incentive payments on time severely limit health provider performance under P4P (reduced availability of drugs and with severe delays, erosion of health worker trust in programme and motivation), weakening programme pathway to impact.

Simulation model results further indicated that refinement of how funding is allocated for facilities may produce improvements in performance; for example, those facilities who have low drug availability before programme implementation would benefit from a higher share of funds towards facility operations. Study results also indicate that the effect of certain design features is not necessarily uniform across performance targets within a given P4P scheme, while incentives for strengthening facility operations (specifically purchasing of essential medicines) was a critical pathway for improvement for the content of care indicator (2 doses of IPT during ANC) this had less impact for the coverage indicator (facility-based deliveries).

Staffing influences the behaviour of all three mechanisms that underpin provider achievement during P4P: ability to provide MCH services (supply of services); ability to perform outreach activities (influencing demand for services); and timely completion and submission of programme reports (reporting mechanism). Improvement in staffing numbers and composition was not directly supported under P4P, with district managers and health facilities having minimal power to alleviate shortages. District managers did have powers to reallocate staff to try and address imbalances in staff composition but could not hire new staff; the design of the programme therefore weakly targeted this catalytic variable.

District manager supervision of facilities also impacted supply of services and facility reporting mechanisms, with support generating a more motivated workforce and improvement in ability to undertake the required facility performance data reporting. The design of the programme effectively targeted this variable, issuing incentive payments to encourage supportive supervision (and in tying payments to facility performance and drug availability, encouraging quality supportive supervision), as a key element underpinning programme effects.

Community health workers and the health facility governing committee were identified as system levers in the CLD. Community health workers were not an incentivised group under the P4P programme, but their activities were instrumental in demand creation for healthcare services and facility performance during the programme. They are trusted members of the community, in a position to reach and spread awareness of health education, facilitating communication on health services between providers and the wider community. Likewise, the activities of health facility governing committees were integral to provider success during the programme. Providers used an additional source of facility funding, community health fund, to support facility operations during P4P, such as procurement of additional medicines when needed. The sensitisation activities of the governing committee in encouraging community members to opt in to payments and their role as signatories on use of payments were crucial pathways to provider achievement of targets during the programme. The simulation model also revealed the importance of community awareness and perceived quality of services for facility-based deliveries.

9.2.2 Accumulated learning over duration of PhD

This thesis comprises of four research papers, with the content of papers contributing towards the overall goal and objectives of the PhD (see section 9.1). The papers illustrate work that was conducted by the candidate in sequential order; the learnings and conclusions presented in the papers are therefore based on the work conducted up until development of the paper. For example, Paper 2 (Chapter 6) describes the creation and results of a CLD to explore pathways to impact for the Tanzania P4P programme. The paper does not reflect the learnings that were accumulated over the duration of the PhD about P4P or systems thinking, such as the enhanced understanding derived from Paper 4 (Chapter 8) on how variation in design, implementation, and context influence programme effectiveness.

For example, in Chapter 6, ability to purchase additional medicines when required using incentive payments and timely receipt of payments were identified in the CLD as crucial mechanisms to ensuring programme success and MCH service delivery. It was acknowledged in Chapter 6 that a certain share of the P4P incentive payment must go towards drug procurement in settings where availability of drugs is limited. In Chapter 8, the impact of adjusting the share of incentive payment allocation between incentives paid directly to health workers and payment used for procurement of medicine was explored, with a preferential allocation of payments identified to induce sustained improvement in MCH service delivery (specifically for the content of care indicator).

The length of payment delays was also identified in Chapter 8 as relevant to the consequent impact on facility performance during the programme. Feedback received during stakeholder consultation at this later stage revealed only severe delays in payment were likely to damage health worker trust and motivation, with communication of delays and payment dates able to sustain trust and motivation to a point, with deterioration then likely with lengthy delays. Minor delays also had minimal impact on provider performance of incentivised services, with prolonged delays reducing provider purchasing power for medicines.

A further example of accumulated learning is differential programme pathways to impact for different targeted health services. In Chapter 6, programme pathways to impact for service delivery were centred on the number of women and children who receive incentivised services (encompassing all incentivised services in a single variable). In Chapter 8, study results indicated that the effect of certain design features for P4P programmes were not necessarily uniform across targeted services; incentives for strengthening facility level operations (procurement of medicines) were critical for improving the content of care indicator (IPT2 during ANC) but had less impact for the coverage of services indicator (facility-based deliveries) which was more dependent on demand simulation.

The Discussion chapter of this thesis (particularly the implications and recommendations section, 9.5) reflect the accumulated learning on P4P and systems thinking gained from the thesis.

On reflection of the accumulated learnings and journey for myself as a PhD student, one of the greatest challenges I faced was becoming familiar and comfortable using a new methodology (system dynamics). I had previous experience of individual-based disease modelling but did not have hands on experience of macro-level health system modelling for policy evaluation. Building this confidence required the support of my team, extensive reading of literature related to modelling for health systems and policy evaluation, attendance at model development courses and continuous experimentation with system dynamics software. An unexpected challenge that arose during completion of this PhD was the COVID-19 pandemic. For myself and many others, the travel restrictions led to uncertainty on whether in person data collection would be possible. After months of uncertainty, the data collection tools described in this thesis were adapted and piloted to prepare for data collection through a virtual platform (Zoom). Usually, this type of data collection for model development and validation is conducted in person and I was unsure whether use of a virtual platform would yield the same type of rich, detailed information gathered through in person interviews. Although the move to a virtual platform proved an effective method for data collection, I would have preferred to conduct these consultations in person, to meet with stakeholders in country, at their organisations (as originally planned).

If I was to start the PhD journey again, I would consider from the outset development of a hybrid model, containing both SDM and agent-based model (ABM) features to model the effects of P4P on health system functioning. A standalone ABM is currently under development, to primarily model individual, heterogenous agent behaviour (patients, health workers and managers) in response to the Tanzania P4P programme. There is also intention to develop a hybrid SDM-ABM model, capitalising on the strengths of both micro- and macro-simulation models for policy evaluation whilst minimising their individual limitations. Reflecting on the capabilities of the standalone models and potential of hybrid modelling, retrospectively I would have planned to develop a hybrid simulation model from the outset of the PhD.

9.3 Contributions to knowledge

9.3.1 Payment for performance in LMICs

In employing a systems thinking approach to programme evaluation research, this thesis contributes evidence on how, why and under what circumstances P4P does (or does not) work in LMIC settings, and how P4P design influences pathways to impact and health system outcomes, cited as critical areas for future research by a recent realist review of P4P in LMIC settings (Singh et al. 2021). In addition to publications, findings from this thesis have been presented at workshops and conferences (Global Symposium on Health Systems Research, International Conference of the System Dynamics Society, International Health Economics Association), see Appendix 5 for full details.

To the candidate's knowledge, four other studies have employed CLD and/or SDM methodology to explore pathways to impact for P4P programmes in LMIC settings, modelling programmes in Turkey (Meker and Barlas 2015), Afghanistan (Alonge *et al.* 2017), Uganda (Renmans *et al.* 2017) and a realist review which used a CLD to visualise evidence on impact of P4P in LMICs (Singh *et al.* 2021).

This thesis is the first study to explore supply chain mechanisms and it's influence on facility readiness for service delivery, highlighting where bottlenecks were occurring and how support of procurement and supply of medicines may be integrated into the design of P4P. This thesis also contributes new evidence on drug availability as a condition for community engagement with the community health insurance fund and use of such additional funding to support facility operations. A version of facility readiness and reflection on its importance in the pathway to impact for P4P programmes features in publications by Alonge *et al.* (2017), Renmans *et al.* (2017) and Singh *et al.* (2021). Although availability of drugs, an important input to facility readiness, features in Singh *et al.* (2021) and as part of aggregated quality and environment variables in Alonge *et al.* (2017) and Renmans *et al.* (2017), supply chain mechanics are not represented in the models.

The role of community health workers in demand creation, and potential for their activities to be incentivised in addition to the facility governing committee is a unique contribution of this thesis. A version of a health facility governing committee feature in both Singh *et al.* (2021) and Renmans *et al.* (2017) CLDs; comparable to the current study, Renmans *et al.* (2017) refers to their key role in community sensitisation acitivities, with Singh *et al.* (2021) recommending incentivisation of the group to enhance governance functioning. This thesis also contributes evidence on staffing as a potential bottleneck for service delivery and performance under P4P, and the positive effect of managerial supervision on health provider motivation. These results concur with findings from Singh *et al.* (2021) and Renmans *et al.* (2017), although in Renmans *et al.* (2017) this was not a direct pathway to impact for the programme as managers were not incentivised.

The effect of payment delays on performance is modelled in this thesis and in the study by Alonge *et al.* (2017), but the effect on outcomes differ. Alonge *et al.* (2017) assume that with minor delays, system performance will eventually follow the same trajectory as when there are no delays in payment, but over a longer time period. In the current study, minor delays have impact on procurement of additional medicines which facilitate provider achievement of targets but has minimal impact on provider motivation and trust in the programme. Alonge *et al.* (2017) do not explore the effect of major payment delays on provider behaviour and service outcomes, or impact of changes to allocation and use of payments, results which are presented in this study.

In this thesis, a model interface was developed to support stakeholder validation of the model; in building a model 'front-end', stakeholders were able to see up close how the model responded to changes in programme implementation and design, commenting in real time on model scenarios and providing input on model assumptions. Meker and Barlas (2015), Alonge *et al.* (2017) and Singh *et al.* (2021) also employ stakeholder consultation to validate findings from their diagrams/models but did not develop a model interface to aid this process. In the wider SDM health systems literature described in Chapter 2, a single example of using a model interface to support stakeholder consultations was identified (Semwanga *et al.* 2016).

Data originally collected as part of a programme evaluation study was repurposed and contributed to development of the CLD and SDM in this thesis. There are few other examples in the wider health systems literature where previous programme evaluation data has been repurposed for CLD or SDM development (Varghese *et al.* 2014; Nigenda *et al.* 2015; Sarriot *et al.* 2015; Schuh *et al.* 2017). This route for model development is of particular importance for research conducted in LMIC settings; difficulty in accessing data or stakeholders (e.g. in humanitarian settings) is not necessarily a barrier to using a systems thinking approach to programme evaluation.

There is little overlap between the content and results from P4P SDM study Meker and Barlas (2015) and the current study, aside from observation on the effect of P4P on providers seeking to treat more patients. The model crucially doesn't feature provider readiness to deliver services which was critical in the current SDM to understanding the effect of the programme on service delivery. The resource constraints faced by providers in lower income settings is also not accounted for in the model, making it difficult to generalise results to settings like Tanzania.

9.3.2 Contribution and added value of using a system dynamics approach in comparison to other methodologies that have been used for P4P programme evaluation research in LMICs

The contribution and added value of using a system dynamics approach for P4P programme evaluation, specifically for quantitative policy impact assessment, is clear. In a recent Cochrane review (Diaconu *et al.* 2021), 59 articles were identified that used controlled before-after study designs, non-randomised and randomised control trials, and interrupted time series to contribute evidence on effect of P4P on healthcare service delivery and health outcomes in LMIC settings. The review identified differential effects of the programme for different indicators, with certainty of evidence generally low; for example, HIV testing and delivery of family planning were found to be potentially impacted by the programme, whilst inconsistent evidence was reported for other MCH service and health outcome indicators, including mother and child immunisations, total number of ANC visits and facility-based deliveries (programme vs. standard care). A key recommendation from the evidence review was further research focussed on identifying and unpacking contextually sensitive pathways to impact in addition to establishing P4P effectiveness, using dynamic approaches to policy evaluation.

System dynamics approaches, and the policy effectiveness and impact methodologies described above, can be used together to provide a more complete picture of how, why and under what circumstances does P4P work to improve health and service delivery outcomes. An example of how this can work in practice (and the added value brought by dynamic approaches) is Binyaruka *et al.* (2015), Anselmi *et al.* (2017) and work detailed in this PhD thesis.

A controlled before and after study design was implemented to assess whether a P4P programme could improve the content and coverage of MCH services in Tanzania (Borghi *et al.* 2013). Difference-in-difference regression analysis was used to explore programme impact on utilisation and quality of MCH services, with positive effects identified on two targeted outcomes (facility-based deliveries and provision of IPT2 during ANC) (Binyaruka *et al.* 2015). The impact evaluation study design was intended to contribute evidence on whether there may be positive, negative or unintended effects of the programme on MCH services; without the additional understanding on actual (in addition to intended) programme pathways to effect, it is unclear why the programme produces certain outcomes and therefore what recommendations can be made for its design and implementation. In a follow-on study, Anselmi *et al.* (2017) used casual mediation analysis (linear structural equation modelling) to identify steps on the causal pathway to programme effect on facility-based deliveries and provision of IPT2 during ANC, including identification of mediators of programme effect. Although the study contributed evidence on significant mediators to programme effect, the method used was unable to shed light on the order of the causal chain between programme implementation and outcomes, or the interactions between different causal chains identified in silo. Building on these previous impact evaluations of P4P in Tanzania, a system dynamics approach was used in this thesis (Chapter 6 and 8) to contribute evidence on pathways to programme effect, identify factors that facilitated or hindered effective programme implementation and generate evidence for health system strengthening programme design and implementation.

A system dynamics approach to P4P programme evaluation can also enhance or expand on the more traditional, qualitative approaches to policy evaluation, such as quantitative impact evaluations, theory-based or realist evaluations.

Previous studies that have included a qualitative component to their design have aimed to generate knowledge on programme effect, context and design in relation to various outcomes, such as accountability mechanisms (Falisse *et al.* 2012; Mayumana *et al.* 2017), access to care equity (Turcotte-Tremblay *et al.* 2018), service delivery (Fox *et al.* 2014; Matsuoka *et al.* 2014; Ir *et al.* 2015; Ogundeji *et al.* 2016; Feldacker *et al.* 2017), motivation (Kalk *et al.* 2010; Witter *et al.* 2011; Aryankhesal *et al.* 2013; Fox *et al.* 2014; Bhatnagar and George 2016; Ogundeji *et al.* 2016; Feldacker *et al.* 2017; Millar *et al.* 2017; Lohmann *et al.* 2018) and perception (Witter *et al.* 2011; Songstad *et al.* 2012; Miller *et al.* 2014; Paul *et al.* 2014; Bertone *et al.* 2016; Chimhutu *et al.* 2016; Wilhelm *et al.* 2016) of those who were targeted and/or participated in the programme, programme implementation (Ssengooba *et al.* 2012; Bertone and Witter 2015; Bertone *et al.* 2016; Ogundeji *et al.* 2016; McMahon *et al.* 2018) and the interest and role of donors (Chimhutu *et al.* 2015).

These studies predominantly used an analytical framework or thematic analysis as their chosen methodology for qualitative data exploration and programme evaluation. For example, Bertone *et al.* (2013) developed and tested an analytical framework (drawing on existing theories related to New Institutional Economics) that can be used to further understand and provide a narrative on the pathways to impact and challenges faced during implementation of complex programmes such as P4P. Frameworks specific to P4P, such as the PBF monitoring and evaluation framework (Witter *et al.* 2013; Paul *et al.* 2017), and other relevant frameworks such as the community participation in health framework (Rifkin *et al.* 1988; Falisse *et al.* 2012), internal and external accountability framework (Cleary *et al.* 2013; Mayumana *et al.* 2017) and political economy framework (Harris 2013; Bertone and Witter 2015) have been used to analyse data and describe P4P operation and impact on health system functioning.

A system dynamics approach (specifically use of CLDs) can enhance the learnings and results obtained from qualitative programme evaluation. In Chapter 7 of this thesis, examples of different approaches for CLD development were presented, including studies that have used thematic analysis or a framework to evaluate and extract information that is then used to create a CLD (Kwamie *et al.* 2014; Renmans *et al.* 2017; Xu and Mills 2017; Lembani *et al.* 2018). A CLD approach can enhance the learnings drawn from more traditional, linear theories of change, reflecting system feedback and non-linear dynamics related to health service provision, utilisation of care and pathways to policy impact.

An example of the value added by using a CLD approach to qualitative programme evaluation is Olafsdottir *et al.* (2014) and work detailed in this PhD thesis (Chapter 6). Olafsdottir *et al.* (2014) presented findings from a process evaluation of the Tanzania P4P programme, which aimed to improve the content and coverage of MCH services (Borghi *et al.* 2013). A thematic analysis approach was used to determine how stakeholders perceived the environment in which P4P was introduced and what influence context may have had on programme implementation. In using a CLD approach to policy evaluation in this thesis (Chapter 6), we can capture not only the one-way, linear effects of the Tanzania P4P programme on various health system elements (including the influence of context on programme implementation), but the interaction effects between these different pathways and the inherent feedback loops these pathways form. These interactions and feedback reveal a highly interconnected system, where well intended policies cause spill-over effects and emergent behaviour we need to acknowledge to avoid suboptimal programme design and implementation recommendations.

System dynamics methodology, specifically CLDs, can also be used to enhance the results drawn from realist reviews for policy evaluation (Singh *et al.* 2021). The results of the review were presented using a CLD, to visualise identified programme pathways to impact and various contextual and design features of the programme. Singh *et al.* (2021) argue that the results of the review could not be adequately represented by a linear structure, instead choosing an approach that can visualise complex system and programme dynamics. The results generated from this thesis have been compared to the existing evidence base described in Singh *et al.* (2021) (see section 9.3.1). With comparison to how CLDs were used in Singh *et al.* (2021) and this thesis (Chapter 6), the realist review presented findings from a broader evidence base, drawing on the results from 117 studies and stakeholder consultation to develop their CLD of programme pathways to impact for P4P. The updated programme theory CLD in the review has perhaps wider generalisability scope than the CLD developed in this thesis due to the wide array of source material as input, illustrating many possible pathways to impact for the programme. A separate workstream of the COSMIC Project (that the PhD candidate is supporting) is concerned with testing the generalisability of the Tanzania CLD (Chapter 6) to represent the experiences of stakeholders who participated in a P4P programme in a comparable setting, Zambia (Shen *et al.* 2017).

The CLD presented in this thesis (Chapter 6) was developed through close review of original process evaluation data for P4P in Tanzania and stakeholder consultation, using purposive text analysis to extract information from the process evaluation data for input to CLD development. Analysis of the CLD led to programme design recommendations related to targeting supply chain mechanisms (including bottlenecks) for improved facility readiness for service delivery and community health workers (potential to incentivise activities to support demand creation), and considerations for programme implementation (effect of payment delays), mechanisms that do not feature in the realist review CLD. The realist review has incorporated information reported in articles that did not employ a system dynamics approach to identify pathways to programme impact, which therefore may not have fully revealed all detailed connections and pathways to impact for the programme.

9.3.3 Guidance for application of systems thinking for health systems research in LMICs

A key contribution of this thesis is guidance on two applications of systems thinking for health systems research and programme evaluation in an LMIC setting (Chapters 6 and 8). With the expectation that the papers will be read by those interested in these methods but unsure of how to apply them, non-technical language is used where possible, with explanation accompanying modelling specific terminology. System maps and model visualisations have been purposefully designed for review by a non-technical audience, with full model documentation (model sectors, equations and data) provided in Appendices, as recommended in guidelines for publishing simulation-based research (Rahmandad and Sterman 2012).

This thesis (specifically Chapter 7) also presents a crucial resource for researchers and practitioners new to CLDs who are seeking guidance on whether CLDs are an appropriate approach for exploring a research question and if so, how to design a study utilising this method. To the candidate's knowledge, this is the first article to provide guidance (drawn from research experience and the literature) on CLD study design and application for health systems research. Littlejohns *et al.* (2021) describe approaches to CLD development for public health research based on a scoping review. The article presents examples from papers that focus on CLD application to health, disease and physiological research questions rather than health system processes, there is no overlap in the papers selected for discussion in each paper. The current study provides further details on a wider selection of potential CLD data sources and methods for development, with a goal that readers are informed on an array of potential study design approaches.

The CLD guidance article is purposely written in non-technical language, which is crucial for the aim of encouraging uptake of this method for health systems research. Based on the PhD candidate's own experience of learning a new research tool, the article also contains pointers on when it may be preferable to choose one data source or development method over another. For example, guidance is provided on when one might opt for key stakeholder interviews over group model building (GMB) (such as sensitivity of research topic), and comparison between the likely CLD output from both methods (Valcourt *et al.* 2020). Of particular relevance to research in conflict-affected and humanitarian settings (or other cases where primary data collection proves difficult), is discussion on the use of secondary data to develop and validate CLDs. In practice, researchers often use more than one source of data for CLD creation, triangulating sources. If a crises-affected setting was initially untenable for data collection but travel was later restored, an initial CLD could be developed using secondary sources before validation with in-country stakeholders.

In addition to discussion of possible routes for CLD creation, caution is also heeded for various study design approaches. For example, on use of secondary data; although it is potentially less resource intensive to obtain, if it was not collected with anticipation for use in CLD development, care should be taken to review the data source and confirm it is providing causal information on drivers for the behaviour of interest (and therefore suitable as CLD input). The inclusion of CLD validation in study design is also strongly advised. For example, where an ex-post development approach has been employed, researchers may have introduced unconscious bias during analysis or misunderstood data. Validation of the CLD, through stakeholder dialogue or other, will minimise this error. Based on the candidates own experience of CLD validation, readers are encouraged to consider how the diagram is presented to stakeholders and consider pilot testing tools; although it is advised (Sterman 2000) to break the CLD into segments for presentation to stakeholders, this was identified in pilot testing to be unsuitable for stakeholders who wanted to see the 'bigger picture' and dynamics represented in the wider CLD. An alternative solution was found which satisfied requests to see the entire CLD (highlighting specific areas of the CLD to focus their attention for the interview).

9.4 Limitations

There were several limitations to this thesis. The secondary data used to develop the CLD was not collected with the purpose of CLD development which may have resulted in the omission of relevant causal links. This risk was mitigated by stakeholder validation interviews which suggested refinements to the CLD. Our approach resulted in a very cost-effective approach to CLD creation, contributing methodological evidence on potential for data re-use for CLD development in resource or data constrained settings (as outlined in 9.3). The CLD was developed by a single researcher which may have introduced unconscious bias into the development process, although the CLD validation process helped to mitigate this risk. To try and address this limitation in the SDM, model equations and structure were independently reviewed by a team member. Mathematical graph theory approaches can be used to compare CLDs (Markóczy 1995; Schaffernicht and Groesser 2011) but due to the large number of diagrams that needed to be compared and combined, qualitative reasoning was thought to be more practical for this particular case. Assumptions were made in the SDM for parameters and equations where it was not possible to draw on existing data sources. Stakeholder dialogue was used to shape certain parts of the model to recreate realistic system behaviour (such as impact of delays in payment on provider trust and motivation).

Another limitation of the work was the lack of patient data used to develop the CLD, which may have revealed further dynamics related to uptake of facility services during P4P. The later developed simulation model (Chapter 8) was parameterised using a variety of data sources (including patient data). The CLD also does not allow visualisation of how dynamics and pathways to impact change over time under the programme, key for identifying where (and crucially when) system bottlenecks occur. However, this limitation was overcome in the SDM which allowed us to explore how the health system changes over time in response to P4P and propose recommendations for improved implementation of the programme.

The use of secondary data and positioning of stakeholders who were consulted in this thesis should be considered when interpreting thesis results and conclusions drawn. Health workers and managers who were interviewed as part of the original programme evaluation (data that was then repurposed and used as input to the CLD in Chapter 6) were asked about their experience of the programme during implementation. As they were incentivised parties who benefitted from the programme, it is possible that they would be reluctant to provide information on what would be perceived to be negative or unsavoury practices, such as misreporting data or diversion of effort or resources away from non-incentivised activities. The design of the programme was expected to deter such practices in theory, including verification of reported data by district, regional and national managers, with suspicion of misreporting resulting in an investigation and potential suspension of stakeholder incentive payments. Binyaruka *et al.* (2015) found there was not a significant change overall in the use of non-targeted services (by proxy, total outpatients visits for both under and over five years categories), but a signification reduction was found in the use of these services at dispensary level facilities.

Focus on the effect of the programme on non-incentivised services is critical to ensure programme design does not negatively impact these services. Full investigation of this programme pathway to negative impact and discussion on negation of effect was outside the scope of thesis research but should be acknowledged as high priority for future design of such health system strengthening programmes, such as possible penalty or reward for maintaining service levels. There is limited evidence within health on the incidence of gaming, with further uncertainty garnered by lack of knowledge on gaming in non-P4P health providers for comparison purposes (Van Herck *et al.* 2010). The SDM could be adapted for such exploratory analysis, with further focus on the potential pathways to unintended negative impact (effects of gaming on service delivery, potential reduction in provision or quality of non-incentivised services etc). Inclusion of a data reporting module and further service delivery modules would be required, with additional stakeholder/data consultation needed to establish how P4P impacts these services, and how the design of P4P might be altered to mitigate negative outcomes on targeted and non-targeted services.

The format for CLD development (extraction of data from individual transcripts via purposive text analysis, development of individual level CLDs and CLD combination) allowed representation of the experiences of health workers, managers and the Health Facility Governing Committees (HFGCs) in the CLD. As these stakeholder groups were not asked to comment on the experiences of individuals from other groups, critique of each other's contribution was not possible, which may have resulted in a different CLD. Stakeholders involved in the evaluation and implementation of the programme were consulted as part of the validation process for CLD development; health workers who participated in the programme were not consulted (due to the length of time between programme implementation and this study, difficulty in now identifying and consulting those who had participated) which may have resulted in missing content from the CLD.

The SDM does not capture health outcomes, patient morbidity or mortality (which are likely to be impacted by the programme), instead simulating the coverage and content of care indicators primarily targeted by the model. There are many factors that drive change in health outcomes, outside of the influence of the P4P programme; focus was centred on understanding how the programme worked and pathways to impact for health system performance. Community and demand-seeking dynamics, such as the role of community health workers in demand creation and effect of peer-to-peer interactions on patient decision-making, were not included in the model; the primary focus of the model was to capture facility level supply side dynamics related to facility performance, as this was the primary target of P4P. In a separate workstream of the COSMIC project, an ABM is currently being developed which will focus primarily on capturing the effect of P4P on care-seeking dynamics for MCH services. The model is initially being developed as a standalone model, with intention to develop a hybrid SDM-ABM simulation that can simulate both micro- and macro-level health system behaviour under P4P.

Excluding facility reporting and related mechanisms identified in the CLD from the SDM is a limitation of this study. Analysis of the CLD revealed provider ability to complete forms and submit performance reports was a key mechanism for achievement of targets during the P4P programme; if they did not submit (or submitted incomplete reports) they would not be eligible for a payment award that cycle. In this iteration of the model, the focus was modelling service coverage and content of care indicators as this was the primary goal of P4P; including feedback and activities around reporting will be retained for a future iteration of the model, to ensure a complete picture of factors that facilitate or drive provider achievement during the programme.

The SDM was used to explore pathways to effect for P4P on two incentivised services that showed some improvement during programme implementation (IPT2 during ANC and facility-based deliveries). The model could have been used to look at indicators that did not show any improvement under P4P, to see how programme design could be adjusted to better support or target services. Focus was instead placed on modelling the services that did show improvement to try and better understand the mechanisms for this positive change and if any further improvement could be gained by adjustment to programme design and implementation. A very limited set of scenarios were simulated in the model; currently it is not possible to simulate and test the effect of other financial mechanisms that are implemented to try and improve MCH services, or indeed other system strengthening programmes on service delivery. However, the model has been purposefully developed so that the P4P component of the model can be extracted and replaced with other system strengthening initiatives, and so that extension of the model to include other MCH services (post-natal care etc.) is possible.

Interviews to validate the CLD and SDM were intended to take place in person, in Tanzania, in the second and third year of PhD studies (2020/2021). The COVID-19 pandemic prohibited travel and data collection for an extended, uncertain period from March 2020. Traditionally primary data collection for CLD and SDM development is conducted in person (Cassidy *et al.* 2019; Cassidy *et al.* 2022) but the pandemic forced researchers to get innovative with stakeholder engagement (Zimmermann *et al.* 2021). Interviews were moved to a virtual platform for the study which proved to not only be an effective method for data collection but also enabled flexibility in the location and timing of interviews; stakeholders had limited time for discussions and would need to cancel and reschedule interviews at very short notice. To mitigate participant recall bias of the programme, interviewees were encouraged to voice when they were uncertain on recollection of events. Stakeholders often provided reflections and examples of why they agreed with certain diagram structure or model behaviour (or why they did not agree, based on their experience) which lent confidence to their commentary on the programme.

The generalisability of the CLD and SDM to encapsulate the effect of P4P on other types of healthcare provider or to other study settings is a limitation. Although the P4P programme was also implemented in secondary care facilities (hospitals), data on these facilities were excluded due to the differences in programme design for hospitals (differences in incentivised services and allocation of incentive payments) and the much larger number of primary care providers targeted by the programme. Testing generalisability of these models to other country settings is an important (and neglected) area of methodological research. Research articles often allude to this in the discussion section, indicating the CLD or SDM is expected to have limited generalisability attributed to the unique challenges faced in the modelled setting (Glenn *et al.* 2020), that findings may be comparable to other country or health system settings (Broekhuizen *et al.* 2020; Zablith *et al.* 2021; Brailsford *et al.* 2004) or the model can be reconfigured for a different setting (Al-Khatib *et al.* 2016). In practice, researchers rarely perform this exercise of external validation. It is critical that further research is undertaken to explore the external validity of CLDs and SDMs to different settings, and to identify methods for model adaptation to new settings to reduce the need for completely developing a CLD or SDM from the ground up. As part of the work undertaken by the wider COSMIC project, the Tanzania CLD and SDM are currently undergoing testing to see to what extent the models represent the experiences of stakeholders who participated in a P4P programme in a comparable setting, Zambia (Shen *et al.* 2017).

9.5 Implications and recommendations

9.5.1 Implications and recommendations for policy

In this thesis, drug and other medical supply availability was found to be imperative for continued delivery of incentivised services (such as malaria treatment provision). Inclusion of drug availability as a condition for district manager incentive payments was critical for provider performance during the programme. In settings or facilities where drug availability is an issue prior to programme implementation, it is essential that (i) a portion of payments can be used to support facility operations, such as purchasing medication and (ii) other stakeholder who can support access to needed supplies are incentivised to support this process. This latter targeted group could also include the medical stores autonomous government department, as has been observed in the most recent, scaled-up design of P4P in Tanzania where medical stores hubs were incentivised to reach certain targets, such as improving inventory accuracy and on time delivery rate (MoHSW (Ministry of Health and Social Welfare)). Delays, as a result of late order submission by facilities or medicine stockouts, could have potentially been mitigated if the medical stores had been directly supported during the programme. Strengthening processes for procurement and provision of medicine and medical commodities may also be a target of other system strengthening programmes outside of the P4P design, including updating logistic systems and number of transport vehicles, as was part of a series of strategic reforms implemented by the Global Health Fund in collaboration with the Tanzania government (Githendu *et al.* 2020).

Adequate staffing levels, essential for provider performance, was not directly supported under P4P, with district managers and health facilities having minimal power to alleviate shortages. District managers did have powers to reallocate staff to try and address imbalances in staff composition but could not hire new staff; reallocation rates/staff composition could be an incentivised target for district managers but without powers to hire new staff, this is arguably an unsuitable target for P4P programmes. Greater autonomy in use of funding to address local system constraints would help to alleviate system bottlenecks, including human resource shortages. Ensuring key stakeholders are targeted under P4P design was also found to be imperative to sustained supply and demand for services, and therefore success during programme implementation. The activities of community health workers and health facility governing committees were crucial for demand creation, and specifically governing committees, critical for procurement of needed medications. Study findings indicate that the current design of P4P work well in facilities where (i) there is sufficient medicine availability and (ii) adequate composition of health professionals, but will be less effective in settings where these resources are limited.

There is a global movement underway, with focus shifting from P4P style health system strengthening programmes towards Direct Health Facility financing (DHFF) (Kapologwe *et al.* 2019; de Walque and Kandpal 2022). In line with goals for P4P, DHFF programmes also aim to improve healthcare quality, reduce health system and service inefficiencies, and better mobilise facility and community human resources for strengthened service delivery (Mæstad *et al.* 2021); however, the design of DHFF programmes place more weight on provider autonomy and funding to improve facility operations.

The accumulated learning from this PhD thesis (particularly knowledge gained during development of Chapter 8) resulted in a more refined perspective on the value of P4P programmes to improve healthcare service delivery for targeted services. Thesis results potentially support a change in programme design, with clear benefits to higher allocation of funding towards facility operations in low resource settings, leaning further towards a DHFF-like funding support design. Study results indicated that this funding design would have greatest improvement on content of care services such as IPT2 during ANC; coverage of services targets like facility-based deliveries would see greater improvement with focussed funding and support for outreach activities to enhance service coverage.

Effectual implementation of health system strengthening programmes like DHFF, specifically limiting payment delays, will strengthen pathways to impact for healthcare service delivery outcomes. Frequency of bonus payments should account for the needs of health providers; as observed in this thesis, where there is already consistently low availability of medicines, funding should be released over shorter cycles to effectively support service delivery. Supportive supervision visits by district managers should continue to be incentivised; removing this component from the funding model may negatively impact health provider motivation to deliver incentivised services, particularly where programme design places less emphasis on provider performance as a condition to receive funding (e.g. DHFF). For coverage of service targets (such as facility-based deliveries), providers need to engage further with communities in addition to improving the quality of facility services; without the performance tracking element of P4P style programmes, motivation to perform such activities may be negatively impacted.

The simulation model has been carefully designed so that addition of new service sectors and targets (e.g. postnatal care, childhood vaccination) or simulation scenarios (adjusting timing of funding disbursement from quarterly to yearly basis) can be readily incorporated for use in evaluation and programme implementation research. A website is currently under development (as part of the wider COSMIC project) which will host the model with full documentation to encourage interaction with the model and reuse in future health systems research.

9.5.2 Implications and recommendations for research agenda

The majority of publications identified in the systematic review (Chapter 2) were found to model high-income health system settings, with only 9 papers presenting application of SDM to model health system behaviour in LMICs; all hybrid modelling papers were from high-income settings. This can perhaps be attributed to lack of capacity of these modelling methods and potentially perceived scarcity of suitable data; however, quantitative and qualitative data collated in previous programme and healthcare evaluations can be repurposed and used to develop these models. Strengthening capacity for using these modelling methods and guidance on model conceptualisation and development using secondary data in low-income settings should be a priority. The need to apply modelling for health system research in low-income settings is paramount; even where it is perceived that data is unavailable for model development, sensitivity analysis can help inform key data needs, accounting for resource and healthcare service delivery constraints. This research is crucial for strengthened understanding on healthcare functioning in such settings; models provide a platform for strategizing and testing the impact of possible policies, prior to resource intensive implementation, informing optimal programme design and identification of suboptimal or unintended consequences. With increased guidance (imperative that it is written/designed for a non-technical audience), visibility and examples of application, the community of those who see the benefit of systems thinking and feel empowered to apply it, either for research or in practice, will grow.

There is also hope for further research innovation through development and use of hybrid modelling for health systems research and programme evaluation. Hybrid models (such as SDM-ABM) present a unique opportunity to combine the strengths of different simulation approaches and counter their respective individual limitations. For example, SDM-ABM models have the potential to capture both the intricate level decision-making and interactions between system agents (patients, health workers etc.) as well as the high-level, aggregate processes and structures that exist in the wider health system (funding allocation and supply chain dynamics etc.) that influence agent decision-making and overall system performance. The review presented in Chapter 2 identified only a single application of SDM-ABM modelling in this field (Djanatliev *et al.* 2012). The 'holy grail' of hybrid modelling, where system elements are simulated by two or more models without clear division or distinction of processes remains an elusive target in not only the healthcare literature but the wider modelling literature (Brailsford *et al.* 2010); a review of hybrid modelling application for operational research identified only four articles of this nature (Brailsford *et al.* 2019). As part of the work undertaken in the wider COSMIC project, our intention is to develop a hybrid SDM-ABM simulation that can simulate both micro- and macro-level health system behaviour under P4P, to generate recommendations for programme design and implementation using a model that encapsulates both facility level supply side dynamics and care seeking dynamics for MCH services. Guidance on model application and development will be produced to encourage use of hybrid modelling for programme evaluation within health systems research and beyond.

9.6 Conclusion

There remains a lack of consensus on whether P4P is an effective initiative for health system strengthening, with a critical knowledge gap on how, why and under what circumstances P4P does (or does not) work in LMIC settings. In this thesis, the candidate employed systems thinking methods (CLD and SDM) to better understand the pathways to impact for the programme and explore how changes in the design, implementation and context of the programme might result in different health system outcomes. Findings from this thesis support the shift away from P4P style health system strengthening programmes towards DHFF payment modality, placing further emphasis on provider autonomy and funding to improve facility operations. Timeliness and frequency of bonus payments, supportive supervision, and accounting for the differences in pathways to impact for individual health services should be high priority for implementation and design of future programmes. Further innovation in our approach to health systems research, namely use of hybrid modelling, will generate recommendations for programme design that encapsulate both supply side and care seeking dynamics in the evaluation, to better target and support delivery and utilisation of healthcare services in LMIC settings.

References

Al-Khatib IA, Eleyan D, Garfield J. 2016. A system dynamics approach for hospital waste management in a city in a developing country: the case of Nablus, Palestine. *Environ Monit Assess* **188**: 503.

Alonge O, Lin S, Igusa T *et al.* 2017. Improving health systems performance in low- and middle-income countries: a system dynamics model of the pay-for-performance initiative in Afghanistan. *Health Policy Plan* **32**: 1417–26.

Anselmi L, Binyaruka P, Borghi J. 2017. Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania. *Implement Sci* **12**: 10.

Aryankhesal A, Sheldon TA, Mannion R. 2013. Role of pay-for-performance in a hospital performance measurement system: a multiple case study in Iran. *Health Policy Plan* **28**: 206–14.

Atkinson J-A, Wells R, Page A *et al.* 2015. Applications of system dynamics modelling to support health policy. *Public Heal Res Pract* **25**, DOI: 10.17061/phrp2531531.

Baugh Littlejohns L, Baum F, Lawless A *et al.* 2018. The value of a causal loop diagram in exploring the complex interplay of factors that influence health promotion in a multisectoral health system in Australia. *Heal Res Policy Syst* **16**: 126.

Bertone MP, Lagarde M, Witter S. 2016. Performance-based financing in the context of the complex remuneration of health workers: findings from a mixed-method study in rural Sierra Leone. *BMC Health Serv Res* **16**: 286.

Bertone MP, Meessen B. 2013. Studying the link between institutions and health system performance: a framework and an illustration with the analysis of two performance-based financing schemes in Burundi. *Health Policy Plan* **28**: 847–57.

Bertone MP, Witter S. 2015. An exploration of the political economy dynamics shaping health worker incentives in three districts in Sierra Leone. *Soc Sci Med* **141**: 56–63.

Bhatnagar A, George AS. 2016. Motivating health workers up to a limit: partial effects of performance-based financing on working environments in Nigeria. *Health Policy Plan* **31**: 868–77.

Binyaruka P, Anselmi L. 2020. Understanding efficiency and the effect of pay-for-performance across health facilities in Tanzania. *BMJ Glob Heal* **5**: e002326.

Binyaruka P, Borghi J. 2017. Improving quality of care through payment for performance: examining effects on the availability and stock-out of essential medical commodities in Tanzania. *Trop Med Int Heal* **22**: 92–102.

Binyaruka P, Maestad O, Somville V. 2021. *CMI Report: Results-Based Financing in the Health Sector in Tanzania: Process Evaluation.* Date of access: December 15, 2022. URL: https://www.cmi.no/publications/8035-results-based-financing-in-the-health-sector-in-tanzania-2

Binyaruka P, Patouillard E, Powell-Jackson T *et al.* 2015. Effect of Paying for Performance on Utilisation, Quality, and User Costs of Health Services in Tanzania: A Controlled Before and After Study. Ostermann J (ed.). *PLoS One* **10**: e0135013.

Binyaruka P, Robberstad B, Torsvik G *et al.* 2018. Who benefits from increased service utilisation? Examining the distributional effects of payment for performance in Tanzania. *Int J Equity Health* **17**: 14.

Brailsford SC, Desai SM, Viana J. 2010. Towards the holy grail: Combining system dynamics and discrete-event simulation in healthcare. *Proceedings of the 2010 Winter Simulation Conference*. IEEE, 2293–303.

Brailsford SC, Eldabi T, Kunc M *et al.* 2019. Hybrid simulation modelling in operational research: A state-of-the-art review. *Eur J Oper Res* **278**: 721–37.

Brailsford SC, Lattimer VA, Tarnaras P *et al.* 2004. Emergency and on-demand health care: modelling a large complex system. *J Oper Res Soc* **55**: 34–42.

Broekhuizen H, Lansu M, Gajewski J *et al.* 2020. Using Group Model Building to Capture the Complex Dynamics of Scaling Up District-Level Surgery in Arusha Region, Tanzania. *Int J Heal Policy Manag*, DOI: 10.34172/ijhpm.2020.249.

Cassidy R, Borghi J, Semwanga AR *et al.* 2022. How to do (or not to do)…using causal loop diagrams for health system research in low and middle-income settings. *Health Policy Plan*, DOI: 10.1093/heapol/czac064.

Cassidy R, Singh NS, Schiratti P-R *et al.* 2019. Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models. *BMC Health Serv Res* **19**: 845.

Cassidy R, Tomoaia-Cotisel A, Semwanga AR *et al.* 2021. Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach. *Soc Sci Med* **285**: 114277.

Chahal K, Eldabi T. 2008. Applicability of hybrid simulation to different modes of governance in UK healthcare. *2008 Winter Simulation Conference*. IEEE, 1469–77.

Chang AY, Ogbuoji O, Atun R *et al.* 2017. Dynamic modeling approaches to characterize the functioning of health systems: A systematic review of the literature. *Soc Sci Med* **194**: 160–7.

Chimhutu V, Songstad NG, Tjomsland M *et al.* 2016. The inescapable question of fairness in Pay-for-performance bonus distribution: a qualitative study of health workers' experiences in Tanzania. *Global Health* **12**: 77.

Chimhutu V, Tjomsland M, Songstad NG *et al.* 2015. Introducing payment for performance in the health sector of Tanzania- the policy process. *Global Health* **11**: 38.

Cleary SM, Molyneux S, Gilson L. 2013. Resources, attitudes and culture: an understanding of the factors that influence the functioning of accountability mechanisms in primary health care settings. *BMC Health Serv Res* **13**: 320.

Diaconu K, Falconer J, Verbel A *et al.* 2021. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev* **2021**, DOI: 10.1002/14651858.CD007899.pub3.

Djanatliev A, German R, Kolominsky-Rabas PL *et al.* 2012. Hybrid simulation with loosely coupled system dynamics and agent-based models for Prospective Health Technology Assessments. *Proc Title Proc 2012 Winter Simul Conf*: 1–12.

Falisse J-B, Meessen B, Ndayishimiye J *et al.* 2012. Community participation and voice mechanisms under performance-based financing schemes in Burundi. *Trop Med Int Heal* **17**: 674–82.

Feldacker C, Bochner AF, Herman-Roloff A *et al.* 2017. Is it all about the money? A qualitative exploration of the effects of performance-based financial incentives on Zimbabwe's voluntary male medical circumcision program. van Wouwe JP (ed.). *PLoS One* **12**: e0174047.

Fox S, Witter S, Wylde E *et al.* 2014. Paying health workers for performance in a fragmented, fragile state: reflections from Katanga Province, Democratic Republic of Congo. *Health Policy Plan* **29**: 96–105.

Githendu P, Morrison L, Silaa R *et al.* 2020. Transformation of the Tanzania medical store department through global fund support: an impact assessment study. *BMJ Open* **10**: e040276.

Glenn J, Kamara K, Umar ZA *et al.* 2020. Applied systems thinking: a viable approach to identify leverage points for accelerating progress towards ending neglected tropical diseases. *Heal Res Policy Syst* **18**: 56.

Günal MM, Pidd M. 2010. Discrete event simulation for performance modelling in health care: a review of the literature. *J Simul* **4**: 42–51.

Harris D. 2013. *Applied Political Economy Analysis: A Problem-Driven Framework*. London

Van Herck P, De Smedt D, Annemans L *et al.* 2010. Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res* **10**: 247.

Ifakara Health Institute. 2015. *Evolution of Results Based Financing Policy and Programmes in Tanzania: 2006 to 2015.*

Ir P, Korachais C, Chheng K *et al.* 2015. Boosting facility deliveries with results-based financing: a mixed-methods evaluation of the government midwifery incentive scheme in Cambodia. *BMC Pregnancy Childbirth* **15**: 170.

isee systems. 2022. Stella online help, tutorials, and printed documentation.

Kalk A, Paul FA, Grabosch E. 2010. â€˜Paying for performanceâ€™ in Rwanda: does it pay off? *Trop Med Int Heal* **15**: 182–90.

Kapologwe NA, Kalolo A, Kibusi SM *et al.* 2019. Understanding the implementation of Direct Health Facility Financing and its effect on health system performance in Tanzania: a non-controlled before and after mixed method study protocol. *Heal Res Policy Syst* **17**: 11.

Kwamie A, Dijk H van, Agyepong IA. 2014. Advancing the application of systems thinking in health: realist evaluation of the Leadership Development Programme for district manager decision-making in Ghana. *Heal Res Policy Syst* **12**: 29.

Lembani M, de Pinho H, Delobelle P *et al.* 2018. Understanding key drivers of performance in the provision of maternal health services in eastern cape, South Africa: a systems analysis using group model building. *BMC Health Serv Res* **18**: 912.

Lohmann J, Wilhelm D, Kambala C *et al.* 2018. ‘The money can be a motivator, to me a little, but mostly PBF just helps me to do better in my job.’ An exploration of the motivational mechanisms of performance-based financing for health workers in Malawi. *Health Policy Plan* **33**: 183–91.

Long KM, Meadows GN. 2018. Simulation modelling in mental health: A systematic review. *J Simul* **12**: 76–85.

Mæstad O, Borghi J, Hassan F et al. 2021. *CMI Report: Direct Financing of Health Facilities: Experiences from Financing Reforms in Tanzania.* Date of access: December 15, 2022. URL:https://open.cmi.no/cmi-xmlui/handle/11250/2838354

Markóczy L. 1995. A method for eliciting and comparing causal maps. *J Manage* **21**: 305–33.

Matsuoka S, Obara H, Nagai M *et al.* 2014. Performance-based financing with GAVI health system strengthening funding in rural Cambodia: a brief assessment of the impact. *Health Policy Plan* **29**: 456–65.

Mayumana I, Borghi J, Anselmi L *et al.* 2017. Effects of Payment for Performance on accountability mechanisms: Evidence from Pwani, Tanzania. *Soc Sci Med* **179**: 61–73.

McMahon SA, Muula AS, De Allegri M. 2018. "I wanted a skeleton … they brought a prince": A qualitative investigation of factors mediating the implementation of a Performance Based Incentive program in Malawi. *SSM - Popul Heal* **5**: 64–72.

Meker T, Barlas Y. 2015. Dynamic Consequences of Performance-Based Payment Systems in Public Hospitals. *Syst Res Behav Sci* **32**: 459–80.

Mielczarek B, Zabawa J. 2016. Modeling Healthcare Demand Using a Hybrid Simulation Approach. *Proceedings of the 2016 Winter Simulation Conference*. IEEE Press, 1535–1546.

Millar R, Chen Y, Wang M *et al.* 2017. It's all about the money? A qualitative study of healthcare worker motivation in urban China. *Int J Equity Health* **16**: 120.

Miller JS, Musominali S, Baganizi M *et al.* 2014. A process evaluation of performance-based incentives for village health workers in Kisoro district, Uganda. *Hum Resour Health* **12**: 19.

Mohiuddin S, Busby J, Savović J *et al.* 2017. Patient flow within UK emergency departments: a systematic review of the use of computer simulation modelling methods. *BMJ Open* **7**: e015007.

MoHSW (Ministry of Health and Social Welfare). *Result Based Financing, Operational Manual.*

Nigenda G, González-Robledo LM, Juárez-Ramírez C *et al.* 2015. Understanding the dynamics of the Seguro Popular de Salud policy implementation in Mexico from a complex adaptive systems perspective. *Implement Sci* **11**: 68.

Ogundeji YK, Jackson C, Sheldon T *et al.* 2016. Pay for performance in Nigeria: the influence of context and implementation on results. *Health Policy Plan* **31**: 955–63.

Olafsdottir AE, Mayumana I, Mashasi I *et al.* 2014. Pay for performance: an analysis of the context of implementation in a pilot project in Tanzania. *BMC Health Serv Res* **14**: 392.

Paul E, Lamine Dramé M, Kashala J-P *et al.* 2017. Performance-Based Financing to Strengthen the Health System in Benin: Challenging the Mainstream Approach. *Int J Heal Policy Manag* **7**: 35–47.

Paul E, Sossouhounto N, Eclou DS. 2014. Local stakeholders' perceptions about the introduction of performance-based financing in Benin: a case study in two health districts. *Int J Heal Policy Manag* **3**: 207–14.

Rahmandad H, Sterman JD. 2012. Reporting guidelines for simulation-based research in social sciences. *Syst Dyn Rev* **28**: 396–411.

Renmans D, Holvoet N, Criel B. 2017. Combining Theory-Driven Evaluation and Causal Loop Diagramming for Opening the 'Black Box' of an Intervention in the Health Sector: A Case of Performance-Based Financing in Western Uganda. *Int J Environ Res Public Health* **14**: 1007.

Rifkin SB, Muller F, Bichmann W. 1988. Primary health care: on measuring participation. *Soc Sci Med* **26**: 931–40.

Rusoja E, Haynie D, Sievers J *et al.* 2018. Thinking about complexity in health: A systematic review of the key systems thinking and complexity ideas in health. *J Eval Clin Pract* **24**: 600–6.

Salleh S, Thokala P, Brennan A *et al.* 2017. Discrete Event Simulation-Based Resource Modelling in Health Technology Assessment. *Pharmacoeconomics* **35**: 989–1006.

Sarriot E, Morrow M, Langston A *et al.* 2015. A causal loop analysis of the sustainability of integrated community case management in Rwanda. *Soc Sci Med* **131**: 147–55.

Schaffernicht M, Groesser SN. 2011. A comprehensive method for comparing mental models of dynamic systems. *Eur J Oper Res* **210**: 57–67.

Schuh HB, Merritt MW, Igusa T *et al.* 2017. Examining the structure and behavior of

Afghanistan's routine childhood immunization system using system dynamics modeling. *Int J Heal Gov* **22**: 212–27.

Semwanga AR, Nakubulwa S, Adam T. 2016. Applying a system dynamics modelling approach to explore policy options for improving neonatal health in Uganda. *Heal Res Policy Syst* **14**: 35.

Shen GC, Nguyen HTH, Das A *et al.* 2017. Incentives to change: effects of performance-based financing on health workers in Zambia. *Hum Resour Health* **15**: 20.

Singh NS, Kovacs RJ, Cassidy R *et al.* 2021. A realist review to assess for whom, under what conditions and how pay for performance programmes work in low- and middle-income countries. *Soc Sci Med* **270**: 113624.

Songstad N, Lindkvist I, Moland K *et al.* 2012. Assessing performance enhancing tools: experiences with the open performance review and appraisal system (OPRAS) and expectations towards payment for performance (P4P) in the public health sector in Tanzania. *Global Health* **8**: 33.

Ssengooba F, McPake B, Palmer N. 2012. Why performance-based contracting failed in Uganda – An "open-box" evaluation of a complex health system intervention. *Soc Sci Med* **75**: 377–83.

Sterman J. 2000. Part II: Tools for systems thinking. *Business Dynamics: Systems Thinking and Modelling for a Complex World*. Illinois, U.S.A.: McGraw-Hill Companies Inc, 154.

Turcotte-Tremblay A-M, De Allegri M, Gali-Gali IA *et al.* 2018. The unintended consequences of combining equity measures with performance-based financing in Burkina Faso. *Int J Equity Health* **17**: 109.

Valcourt N, Walters J, Javernick-Will A *et al.* 2020. Assessing the efficacy of group model building workshops in an applied setting through purposive text analysis. *Syst Dyn Rev* **36**: 135–57.

Varghese J, Kutty VR, Paina L *et al.* 2014. Advancing the application of systems thinking in health: understanding the growing complexity governing immunization services in Kerala, India. *Heal Res Policy Syst* **12**.

Viana J, Simonsen TB, Dahl FA *et al.* 2018. A HYBRID DISCRETE EVENT AGENT BASED OVERDUE PREGNANCY OUTPATIENT CLINIC SIMULATION MODEL. *2018 Winter Simulation Conference (WSC)*. IEEE, 1488–99.

de Walque D, Kandpal E. Reviewing the evidence on health financing for effective coverage: do financial incentives work? *BMJ Glob Heal* **7**: e009932.

Wilhelm DJ, Brenner S, Muula AS *et al.* 2016. A qualitative study assessing the acceptability and adoption of implementing a results based financing intervention to improve maternal and neonatal health in Malawi. *BMC Health Serv Res* **16**: 398.

Witter S, Toonen J, Meessen B *et al.* 2013. Performance-based financing as a health system reform: mapping the key dimensions for monitoring and evaluation. *BMC Health Serv Res* **13**: 367.

Witter S, Zulfiqur T, Javeed S *et al.* 2011. Paying health workers for performance in Battagram district, Pakistan. *Hum Resour Health* **9**: 23.

Xu J, Mills A. 2017. Challenges for gatekeeping: a qualitative systems analysis of a pilot in rural China. *Int J Equity Health* **16**: 106.

Zablith N, Diaconu K, Naja F *et al.* 2021. Dynamics of non-communicable disease prevention, diagnosis and control in Lebanon, a fragile setting. *Confl Health* **15**: 4.

Zhang X. 2018. Application of discrete event simulation in health care: a systematic review. *BMC Health Serv Res* **18**: 687.

Zimmermann N, Pluchinotta I, Salvia G *et al.* 2021. Moving online: reflections from conducting system dynamics workshops in virtual settings. *Syst Dyn Rev* **37**: 59–71.

Zulkepli J, Eldabi T, Mustafee N. 2012. Hybrid simulation for modelling large systems: An example of integrated care model. *Proceedings Title: Proceedings of the 2012 Winter Simulation Conference (WSC)*. IEEE, 1–12.

APPENDICES TO THESIS

Appendix 1: Ethical Clearances

Appendix 1 provides the ethical clearance for this project from three research bodies:

- 1a: Observational/Interventions Research Ethics Committee, London School of Hygiene & Tropical Medicine
- 1b: Ifakara Health Institute Review Board (IHI-IRB)
- 1c: Tanzania National Institute for Medical Research (NIMR)

1a: Observational/Interventions Research Ethics Committee, London School of Hygiene & Tropical Medicine

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

Miss Rachel Cassidy
LSHTM

7 February 2019

Dear Rachel,

Study Title: Novel methods for optimising health systems payment for performance interventions to improve maternal and child health in low-resource settings.

LSHTM Ethics Ref: 16139

Thank you for 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 Type	File Name	Date	Version
Investigator CV	CV - Karl Blanchet	13/11/2018	1
Investigator CV	CV - Jo Borghi	13/11/2018	1
Investigator CV	CV - Neha Singh	13/11/2018	1
Investigator CV	CV - Rachel Cassidy	13/11/2018	1
Investigator CV	CV - Peter Binyaruka	13/11/2018	1
Information Sheet	Stakeholder Information Sheet for the CLD Validation Workshop 13th November 2018	13/11/2018	1
Information Sheet	Stakeholder Information Sheet for the HDD Validation Workshop 13th November 2018	13/11/2018	1
Information Sheet	Stakeholder Consent Form 13th November 2018	13/11/2018	1
Advertisements	Workshop Letter (Tanzania) 13th November 2018	13/11/2018	1
Advertisements	Workshop Letter (UK) 13th November 2018	13/11/2018	1
Protocol / Proposal	CLD and HDD Interview Questions 13th November 2018	13/11/2018	1
Protocol / Proposal	COSMIC Study Protocol 26th November 2018	26/11/2018	1
Investigator CV	CV - Agnes Rwashana Semwanga	26/11/2018	1
Protocol / Proposal	COSMIC Study Protocol	14/01/2019	2
Protocol / Proposal	Agenda and Topic Guide for Workshop (UK)	14/01/2019	2
Protocol /	Agenda and Topic Guide for Workshop (TZ)	14/01/2019	2

Proposal			
Information Sheet	Participant Information Sheet (UK)	14/01/2019	2
Information Sheet	Participant Information Sheet (TZ)	14/01/2019	2
Information Sheet	Consent form (UK)	14/01/2019	2
Information Sheet	Consent form (TZ)	14/01/2019	2
Advertisements	Workshop Letter (UK)	14/01/2019	2
Advertisements	Workshop Letter (TZ)	14/01/2019	2
Covering Letter	Ethics Committee Cover Letter	14/01/2019	1
Covering Letter	Ethics committee clarification cover letter	01/02/2019	2
Advertisements	Workshop Letter (UK)	01/02/2019	3
Advertisements	Workshop Letter (TZ)	01/02/2019	3

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the Committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.

At the end of the study, the CI or delegate must notify the committee using an End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: http://leo.lshtm.ac.uk

Additional information is available at: www.lshtm.ac.uk/ethics

Yours sincerely,

[redacted]

Professor John DH Porter
Chair

ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Miss Rachel Cassidy
LSHTM

14 November 2019

Dear Rachel,

Study Title: Novel methods for optimising health systems payment for performance interventions to improve maternal and child health in low-resource settings.

LSHTM Ethics Ref: 16139 - 1

Thank you for your application for the above amendment to the existing ethically approved study and submitting revised documentation. The amendment application has been considered by the Observational Committee.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above amendment to 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 for the amendment having been received, where relevant.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Local Approval	COSMIC -IRB certificate	30/09/2019	1
Local Approval	COSMIC_NIMR_permit	30/09/2019	1
Other	Consent form (TZ)	15/10/2019	1
Other	COSMIC Study Protocol	15/10/2019	1
Other	Participant Information Sheet (TZ)	15/10/2019	1
Other	Workshop Letter (TZ)	15/10/2019	1
Other	Interview and Topic Guide for Workshop (TZ)	15/10/2019	1

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the Committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.

At the end of the study, the CI or delegate must notify the committee using an End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: http://leo.lshtm.ac.uk

Additional information is available at: www.lshtm.ac.uk/ethics

Yours sincerely,

[redacted]

Professor Jimmy Whitworth
Chair

ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

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

Research Ethics Committee

Miss Rachel Cassidy

LSHTM

1 May 2020

Dear Rachel,

Study Title: Novel methods for optimising health systems payment for performance interventions to improve maternal and child health in low-resource settings.

LSHTM Ethics ref: 16139 - 2

Thank you for submitting your amendment for the above research project.

Your amendment has been assessed by the Research Governance & Integrity Office and has been approved as a non-substantial change. The amendment does not require further ethical approval from the observational ethics committee.

List of documents reviewed:

Document Type	File Name	Date	Version
Other	Consent form	30/04/2020	4
Other	COSMIC Study Protocol	30/04/2020	4
Other	Interview and topic guide	30/04/2020	4
Other	Interview Letter (Other)	30/04/2020	4
Other	Interview Letter (TZ)	30/04/2020	4
Other	Participant Information Sheet	30/04/2020	4
Local Approval	COSMIC -IRB certificate	30/04/2020	1
Local Approval	COSMIC_NIMR_permit	30/04/2020	1

Any subsequent changes to the application must be submitted to the Committee via an Amendment form on the ethics online applications website: http://leo.lshtm.ac.uk .

Best of luck with your project.

Yours sincerely,

[redacted]

Rebecca Carter

Research Governance Coordinator

Ethics@lshtm.ac.uk
http://www.lshtm.ac.uk/ethics/

Improving health worldwide

1b: Ifakara Health Institute Review Board (IHI-IRB)

IFAKARA HEALTH INSTITUTE
research | training | services

INSTITUTIONAL REVIEW BOARD
P O BOX 78373 DAR ES SALAAM, TANZANIA
Tel +255 (0) 22 2774714, Fax: + 255 (0) 22 2771714 Email: irb@ihi.or.tz

13th May, 2019

National Institute for Medical Research
P O Box 9653
Dar Es Salaam
Email; headquarters@nimr.or.tz

Dr. Peter Binyaruka
Ifakara Health Institute,
P O Box 78373,
Dar es Salaam.

IHI/IRB/No: 15 - 2019

INSTITUTIONAL CLEARANCE CERTIFICATE FOR CONDUCTING HEALTH RESEARCH

On 29 March 2019, the Ifakara Health Institute Review Board (IHI-IRB) reviewed the study titled: ***"Novel methods of optimizing health systems payment for performance interventions to improve maternal and child health in low resource settings (COSMIC project)"*** Submitted by Principal Investigator: Dr. Peter Binyaruka.

The study has been approved for implementation after IRB consensus. This certificate thus indicates that; the above-mentioned study has been granted an Institutional Ethics Clearance to be conducted in (Pwani and Mwanza region) Tanzania.

The following documents were reviewed and approved:

1. Study Protocol
2. Informed Consent Forms English and Kiswahili versions
3. Data collection tools in English and Swahili
4. Budget and budget justification
5. Data Transfer Agreement Form
6. Investigators' CVs

The Principal Investigator of the study must ensure that, the following conditions are fulfilled during or after the implementation of the study:

1. PI should submit a six month progress report and the final report at the end of the project
2. Any amendment, which will be done after the approval of the protocol, must be communicated as soon as possible to the IRB for another approval
3. All research must stop after the project expiration date, unless there is prior information and justification to the IRB
4. There should be plans to give feedback to the community on the findings
5. Any publication needs to pass through the IRB
6. The approval is valid until 12th May 2020.

The IRB reserves the right to undertake field inspections to check on the protocol compliance

[redacted]
Dr. Ahmed M. Abdallah

[redacted]
IRB Secretary
Dr. Mwifadhi Mrisho

Dar es Salaam	Ifakara	Bagamoyo	Rufiji	Mtwara	Kigoma
PO Box 78373	PO Box 53	PO Box 74	PO Box 40 Ikwiriri	PO Box 1048	PO Box 1077
Tel: 022 2774756	Tel: 0232 625164	Tel: 0232 440065	Tel: 0787 364521	Tel: 0232 333487	Tel: 0282 803655
Fax: 022 2771714	Fax: 0232 625312	Fax: 0232 440064	Fax: 0232 010001		

www.ihi.or.tz

Plot 463,
Kiko Avenue, Mikocheni | P.o. Box 78,373
Dar es Salaam. Tanzania | Phone: +255222774756
Email: irb@ihi.or.tz

INSTITUTIONAL REVIEW BOARD

ISO 9001:2015 certified

IFAKARA HEALTH INSTITUTE
research | training | services

August 31, 2020

National Institute for Medical Research
P O Box 9653
Dar Es Salaam
Email; headquarters@nimr.or.tz

Dr. Khalfan Ngowo
Ifakara Health Institute,
P O Box 53,
Ifakara

IHI/IRB/EXT/No: 22 - 2020

EXTENSION APPROVAL

On 28th August, 2020, the Ifakara Health Institute Review Board (IHI-IRB) renewed Annual Extension application to a study titled: ***"Novel methods of optimizing health system payment for performance interventions to improve maternal and child health in low resources settings (COSMIC PROJECT)".*** Submitted by the Principal Investigator: Peter Binyaruka. The Annual Extension extends from 13th May 2020 to 12th May 2021. The above-named study had a previous approval number IHI/IRB/No: 15 – 2019**.**

The IRB reserves the right to undertake field inspections to check on the protocol compliance.

Sincerely,

[redacted]

Dr. Mwifadhi Mrisho

IHI - IRB Secretary

Plot 463,
Kiko Avenue, Mikocheni | P.o. Box 78,373
Dar es Salaam, Tanzania | Phone: +255222774756
Email: irb@ihi.or.tz

ISO 9001:2015 certified

IFAKARA HEALTH INSTITUTE
research | training | services

INSTITUTIONAL REVIEW BOARD

May 24, 2021

National Institute for Medical Research
P O Box 9653
Dar Es Salaam
Email; headquarters@nimr.or.tz

Dr Peter Binyaruka
Ifakara Health Institute
P O Box 78373
Dar es Salaam

IHI/IRB/EXT/No: 17 - 2021

EXTENSION APPROVAL

On 21st May, 2021, the Ifakara Health Institute Review Board (IHI-IRB) renewed Annual Extension application to a study titled: ***"Novel methods of optimizing health systems payment for performance interventions to improve maternal and child health in low resource settings (COSMIC project)."*** Submitted by Principal Investigator, Dr Peter Binyaruka. The Annual Extension extends from 14th May 2021 to 12th May 2022. The above-named study had a previous approval number IHI/IRB/No: 15 – 2019 of 13th May 2019.

The IRB reserves the right to undertake field inspections to check on the protocol compliance.

Sincerely,

[redacted]

Mwifadhi Mrisho, PhD

IHI - IRB Secretary

1c: Tanzania National Institute for Medical Research (NIMR)

THE UNITED REPUBLIC
OF TANZANIA

National Institute for Medical Research
3 Barack Obama Drive
P.O. Box 9653
11101 Dar es Salaam
Tel: 255 22 2121400
Fax: 255 22 2121360
E-mail: nimrethics@gmail.com

Ministry of Health, Community
Development, Gender, Elderly & Children
University of Dodoma, College of
Business Studies and Law
Building No. 11
P.O. Box 743
40478 Dodoma

NIMR/HQ/R.8a/Vol. IX/3154

08th July, 2019

Josephine Borghi
London School of Hygiene and Tropical Medicine
C/o Dr. Peter Binyaruka
Ifakara Health Centre
P.O. Box 78373
Dar es Salaam

RE: ETHICAL CLEARANCE CERTIFICATE FOR CONDUCTING MEDICAL RESEARCH IN TANZANIA

This is to certify that the research entitled: Novel methods for optimizing health systems payment for performance interventions to improve maternal and child health in low-resource settings (Borghi J. et al), whose local investigator is Dr. Peter Binyaruka of Ifakara Health Institute has been granted ethical clearance to be conducted in Tanzania.

The Principal Investigator of the study must ensure that the following conditions are fulfilled:

1. Progress report is submitted to the Ministry of Health, Community Development, Gender, Elderly & Children and the National Institute for Medical Research, Regional and District Medical Officers after every six months.
2. Permission to publish the results is obtained from National Institute for Medical Research.
3. Copies of final publications are made available to the Ministry of Health, Community Development, Gender, Elderly & Children and the National Institute for Medical Research.
4. Any researcher, who contravenes or fails to comply with these conditions, shall be guilty of an offence and shall be liable on conviction to a fine as per NIMR Act No. 23 of 1979, PART III Section 10(2).
5. Sites: Mwanza and Pwani regions.

Approval is valid for one year: 08th July 2019 to 07th July 2020.

Name: Prof. Yunus Daud Mgaya

[redacted]

Signature
CHAIRPERSON
MEDICAL RESEARCH
COORDINATING COMMITTEE

Name: Prof. Muhammad Bakari Kambi

[redacted]

Signature
CHIEF MEDICAL OFFICER
MINISTRY OF HEALTH, COMMUNITY
DEVELOPMENT, GENDER, ELDERLY &
CHILDREN

CC: Director, Health Services -TAMISEMI, Dodoma
RMO of Mwanza and Pwani regions
DMO/DED of respective districts

THE UNITED REPUBLIC
OF TANZANIA

National Institute for Medical Research
3 Barack Obama Drive
P.O. Box 9653
11101 Dar es Salaam
Tel: 255 22 2121400
Fax: 255 22 2121360
Email: nimrethics@gmail.com

Permanent Secretary (Health)
Ministry of Health, Community
Development, Gender, Elderly & Children
Government City Mtumba, Health Road
P.O. Box 743
40478 Dodoma

NIMR/HQ/R.8c/Vol. I /1600

08th October, 2020

Josephine Borghi
London School of Hygiene and Tropical Medicine
C/o Dr. Peter Binyaruka
Ifakara Health Institute
P O Box 78373
Dar es Salaam

RE: APPROVAL FOR EXTENSION OF ETHICAL CLEARANCE

This letter is to confirm that your application for extension on the already approved proposal: Novel methods of optimizing health systems payment for performance interventions to improve maternal and child health in low-resource settings (Borghi J. et al) whose local investigator is Dr. Peter Binyaruka of Ifakara Health Institute, has been approved.

The extension approval is based on the progress report dated 24th August, 2020 on the project, Ref. NIMR/HQ/R.8a/Vol. IX/3154, dated 08th July, 2019. Extension approval is valid until 07th July, 2021.

The Principal Investigator must ensure that other conditions of approval remain as per ethical clearance letter. The PI should ensure that progress and final reports are submitted in a timely manner.

Name: Prof. Yunus Daud Mgaya

[redacted]

Signature
CHAIRPERSON
MEDICAL RESEARCH
COORDINATING COMMITTEE

Name: Prof. Abel Nkono Makubi

[redacted]

CHIEF MEDICAL OFFICER
MINISTRY OF HEALTH, COMMUNITY DEVELOPMENT, GENDER, ELDERLY & CHILDREN

Appendix 2: Study tools, participant information sheets and consent forms

Appendix 2 provides the study tools, participant information sheets and consent forms for each modelling method:

- 2a: Stakeholder CLD validation interview tool
- 2b: Information sheet for CLD validation interviews
- 2c: Consent form for CLD validation interviews
- 2d: Flyer for stakeholder engagement
- 2d: Stakeholder simulation model validation interview tool
- 2e: Information sheet for simulation model validation interviews
- 2f: Consent form for simulation model validation interviews

2a: Stakeholder CLD validation interview tool

Interviewer: This series of interviews have been organised by researchers from The London School of Hygiene and Tropical Medicine and Ifakara Health Institute. We are hoping to conduct interviews with experts, such as yourself, to validate a map we have created of the Tanzania maternal and child health (MCH) system response to payment for performance (P4P). We developed the map using interview data that was collected during the pilot P4P programme in Tanzania (2011-2013); interviews were conducted with health workers, facility in-charges and district level managers on how the programme had been received by providers and managers, and what factors had facilitated or hindered effective implementation of the programme. We are currently focussing on the primary care facilities that offered MCH services and took part in the pilot programme (excluding up-graded health centres).

Interviewer: To ensure our system map accurately represents the real health system behaviour and processes that developed under the pilot we now require this map to be validated by experts with knowledge of the pilot programme.

Interviewer: During this interview, I will show you system maps that are representative of how we believe the health system functioned following the introduction of the pilot P4P programme. Using your knowledge, experience and feedback of health system operation we will then refine the structure of our maps to ensure they reflect the pilot P4P programme.

Interviewer: In the next phase of our project we are going to be looking at the differences between the pilot and other health system strengthening programmes that have taken place in the country, including the up-scaled Results-Based Financing programme (RBF, 2016-2019) and Direct Health Facility Financing programme (DHFF, 2019-Present). If you have time at the end of the interview, I would be very interested to hear your opinion on the core (intended and observed) differences in health system transformation and outcomes between the three programmes.

Interviewer: Just before we begin, I have received a copy of your consent form but I would just like to seek your verbal consent that you are happy to continue with the interview and you are happy for me to take written notes and an audio-recording of this session. This is only for our records and shared only with our research team. You can change your mind or stop the interview at any time.

Interviewer: *If no* That is okay I will take written notes instead.

Once participant has given consent, open the Vensim diagram that shows the system map and ask the interviewee if they can see the map on their screen

Interviewer: We have this large system map of the Tanzania health system response to P4P but to make the most use of the time we have today, I am going to focus the interview on one area of the map. The map has been split into three segments corresponding to the (i) demand, (ii) supply and (iii) reporting-side mechanisms underpinning achievement of targets during P4P, with targets represented in the diagram in bold labelled 'Number of women and children receive incentivised services' and 'submission of routine health facility data by providers'.

Interviewer: In today's interview we are going to focus on the part of the map that describes *refer to (i), (ii) or (iii)*. I have highlighted the portion of the system map that corresponds to *refer to (i), (ii) or (iii)* so that we can still see how this part of the map connects to other elements of the map (just to show it doesn't operate in isolation). I will describe what we are seeing in the map then periodically stop to check, to your knowledge, that this process occurred during the pilot P4P programme. Your feedback will help us validate our diagram and make any necessary refinements.

Interviewer: Just a few comments on what we are seeing here. We have variables and arrows connecting each of the variables. This indicates some kind of causal relationship exists between pairs of variables. You will also notice that the arrows have polarity attached to them, plus and minus signs. These indicate the direction of causality. For example, as 'Amount of inventive payment issued to providers' increases, so does 'Health worker salary top up' (i.e. health workers receive bonus payments for improved performance during P4P).

*Stop here and check if the interviewee understands what you have described – does this make sense? *

Interviewer: You will also notice there are two small dashes across the arrow; this indicates a delay in effect. Taking the same example, although facilities who improve their performance during P4P should receive a bonus payment, there were often delays between the incentive payment being issued and health workers receiving this money (particularly at the beginning of the programme). This delay in effect is represented by those two dashes across the arrow.

*Stop here and check if the interviewee understands what you have described – does this make sense? *

Interviewer then proceeds with taking the interviewee round the rest of this map segment, periodically stopping to check interviewee understanding and to ask if any modifications should be made to the map to reflect their experience of the programme

The interviewer does not have to explicitly run through these questions while discussing the map, can instead probe 'Does this make sense? Are we missing anything important in this section of the map? Is there anything that you feel should be removed in the map?'. When an interviewee gives their feedback on the map, it will generally fall into these compartments and help the modeller to go back and make modifications to the map:

- Does this part of the system exist to your knowledge?
- Are appropriate system variables represented? If not, what variables are missing or should be removed?

- Are appropriate in and out flows represented? If not, what flows are missing or should be removed?
- Is the polarity of in and out flows accurately represented? If not, what changes would you make?
- Are appropriate delays in the system represented? If not, what delays are missing or should be removed?

When interviewer has finished with validating the system map

Interviewer: We may have already touched on this during our discussion of the map but I would also be interested to hear your view on what you think could have been changed in the implementation of the pilot programme to help facilities achieve targets (and improve the delivery and coverage of MCH services?).

Interviewer: As I said earlier, in the next phase of our project we are going to be looking at the differences between the pilot and other health system strengthening programmes that have taken place in the country, including the up-scaled Results-Based Financing programme (RBF, 2016-2019) and Direct Health Facility Financing programme (DHFF, 2019-Present). If you have time now, I would be very interested to hear your opinion on:

- What are the key similarities and differences between P4P pilot program and/or 1) RBF program and (2) DHFF program?

Interviewer: Thank you very much for your time today, this has been incredibly useful. If you feel comfortable doing so, is there anyone you would recommend for us to interview next?

2b: Information sheet for CLD validation interviews

Novel methods for optimising health systems payment for performance interventions to improve maternal and child health in low-resource settings.

Participants Information Sheet

Introduction

We would like to interview you as part of our study. Participating in this interview is entirely up to you. Before you decide, you need to understand why we need your expertise and what it would involve from you. We have made this information sheet available prior to the interview in order for you have time to digest the information presented in this document and allow time for you to ask any questions. Please ask the research team (contact details given at the end of this document) if anything you read is not clear or you would like more information.

What is the purpose of the study?

The Ifakara Health Institute (IHI) and London School of Hygiene and Tropical Medicine (LSHTM) are conducting research to improve women's and children's health in low-resource settings. We are hoping to maximise the impact of health systems payment for performance (P4P) interventions through the development and use of diagrams or maps of the Tanzanian health system and the relationships between elements within the system, such as the behaviour of health workers and patients, and how they respond to the P4P programme. The system maps were developed using data collected during the pilot P4P programme that ran from 2011-2013. To ensure our models are accurately representing the real health system behaviour and processes that developed under the pilot and to assess how the pilot differs from other programmes that have recently taken place in the country to strengthen MCH services, we now require these maps to be validated by experts with knowledge of one or more of the following:

(i) The pilot P4P programme (2011-2013)
(ii) The national Results-Based Financing programme (RBF, 2016-2020)
(iii) Direct Health Facility Financing programme (DHFF, 2019-Present)

Why have I been asked to take part?

You have been invited to take part in this interview because of your expertise and experience of health system operation and/or (i) the pilot P4P programme, (2) the national RBF programme, (3) DHFF programme. We will be interviewing approximately 20 individuals, which includes representatives from the Ministry of Health (MOHCDGEC), President's Office Regional Administration and Local Government (PORALG), National Health Insurance Fund (NHIF), Medical Stores Department (MSD), Regional Health Management Team (RHMT), Development partners (e.g., World Bank, USAID, and WHO). Based on the feedback we receive from participants during interviews, we will refine the structure of our maps to ensure they are accurately representing the impact of these programmes on the Tanzania health system.

Do I have to take part?

No, your participation in this interview is voluntary. We have made this information sheet available prior to the interview in order for you have time to digest the information presented in this document and allow time for you to ask any questions. If you agree to take part, we will ask you to sign a consent form. If you change

info@ihi.or.tz | www.ihi.or.tz

Ifakara Branch
Off Mlabani Passage
P.o. Box 53 Ifakara
Phone: +255232931572

Dar es Salaam Office
Plot 463, Kiko Avenue, Mikocheni
P.o. Box 78,373 Dar es Salaam
Phone: +255222774756

Bagamoyo Branch
Inside District Hospital
P.o. Box 74 Bagamoyo
Phone: +255232625164

your mind about wanting to be interviewed, you can let us know using the contact details at the bottom of the form. You can also stop the interview at any time.

What will I have to do?

During this interview we will show you system maps that are representative of how we believe the health system functions following the introduction of P4P initiatives using data collected during the Tanzania pilot P4P programme (2011-2013). We will examine each part of the system map in segments with clear explanations given for how each part functions and contributes to the wider system. We will then ask you a series of questions related to the structure of each system map segment in turn. Using your knowledge, experience and feedback of health system operation we will then refine the structure of our maps to ensure they are able to reflect each of the programmes outlined above. We will also ask you if it would be possible to contact you in the future to ask further or clarifying questions to aid our validation of the model.

What data is collected in this study?

Our researchers will make written notes and may also take audio-recordings of feedback to ensure we have a comprehensive record of your feedback. If you would rather not have your feedback audio-recorded we will take written notes. Please indicate on this consent form if you would prefer not to have your feedback audio-recorded.

How data is collected in this study?

Due to COVID-19, there are now restrictions in place on international travel and required to observe a social distance (which restricts face to face interviews). We will therefore conduct interviews over Zoom (or alternative web call service) with measures in place to ensure the security of the web call. Our researchers will make written notes during the interview and with your permission, use a handheld audio recording device to record the interview.

What will happen to the data collected in this study?

All feedback collected from this interview will be anonymised. We will assign participants a code which will be used to refer to a participant's feedback in analysis or published written work so they cannot be identified. The participants code will confer what their respondent group is (e.g., donor, government, health care provider) but they will not be identifiable. Only the research team will have access to the secure password-protected folder containing the original feedback (either written or audio-recordings) obtained in interviews, protecting the anonymity of participants. The feedback collected in these interviews will be made available in published work, academic presentations and study reports but you will not be identified as having participated and feedback will be anonymised.

What will happen to the results of this study?

The information we get from the study will aid our knowledge and understanding of the health system impact of resource- and initiative-based programmes in Tanzania, benefiting patients, health workers and those implementing such initiatives. The study results will be published in a health systems journal and as part of academic presentations and study reports so that other researchers can learn from our findings. Your contribution to this study will be anonymised using codes only our researchers have access too.

What are the possible risks and disadvantages?

info@ihi.or.tz | www.ihi.or.tz

Ifakara Branch
Off Mlabani Passage
P.o. Box 53 Ifakara
Phone: +255232931572

Dar es Salaam Office
Plot 463, Kiko Avenue, Mikocheni
P.o. Box 78,373 Dar es Salaam
Phone: +255222774756

Bagamoyo Branch
Inside District Hospital
P.o. Box 74 Bagamoyo
Phone: +255232625164

Participants will not be exposed to any physical risks and the study is unlikely to pose any risk greater than the risk encountered in daily life. The study team will also be particularly cautious about protecting respondents' identity, and, where relevant, that of their organisations, when reporting findings. We will assign participants a code which will be used to refer to a participant's feedback in study reports, academic presentations or published written work so they cannot be identified. Only the research team will have access to the secure password-protected folder containing the original feedback (written or audio-recorded) obtained in interviews, protecting the anonymity of participants.

Who is funding this study?

Our study is funded by the Health Systems Research Initiative, jointly funded by Department of International Development (DFID), the Economic and Social Research Council (ESRC), the MRC and the Wellcome Trust.

Who has checked this study?

All research involving human participants is looked at by an independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by Ifakara Health Institute institutional review board, the National Institute for Medical Research, and the London School of Hygiene and Tropical Medicine Research Ethics Committee (ref: 16139) has also approved this study.

Further information and contact details

If you would like any further information, please contact the principal investigators:

- *Dr Peter Binyaruka (Ifakara Health Institute)*
  pbinyaruka@ihi.or.tz or +255655 363361
- *Dr Josephine Borghi (The London School of Hygiene and Tropical Medicine Research)*
  Josephine.Borghi@lshtm.ac.uk or +4420 7927 2090

What if I have a question/ concern?

If you have a question or concern about any aspect of this study, please feel free to contact the local principal investigator on this study: Dr Peter John Binyaruka (+255655363361, pbinyaruka@ihi.or.tz) from IHI. This research has been approved by board of research ethics of IHI-IRB and NIMR. For ethical issues please contact Mr. Fakih Bakari (fbakari@ihi.or.tz) representative of the IHI IRB, P.O. Box 78373 Dar es salaam, (+255 23 2625164/ +255 22 2774714); and Ms. Sia Malekia (smalekia@nimr.or.tz), representative of NIMR, P. O. Box 9653, 11101 Dar es Salaam, Tanzania, Tel: +255-22-2121400, Fax: +255-22-2121360. You can contact also the London School of Hygiene and Tropical Medicine ethics committee by telephone at +4420 7927 2221 or by email at ethics@lshtm.ac.uk.

2c: Consent form for CLD validation interviews

Novel methods for optimising health systems payment for performance interventions to improve maternal and child health in low-resource settings.

CONSENT FORM

Study Purpose:
The Ifakara Health Institute (IHI) and London School of Hygiene and Tropical Medicine (LSHTM) are conducting research to improve maternal and child health (MCH) in low-resource settings. We are hoping to maximise the impact of health systems payment for performance (P4P) interventions through the development and use of diagrams or maps of the Tanzanian health system and the relationships between elements within the system, such as the behaviour of health workers and patients, and how they respond to the P4P programme. The system maps were developed using data collected during the pilot P4P programme that ran from 2011-2013. To ensure our models are accurately representing the real health system behaviour and processes that developed under the pilot and to assess how the pilot differs from other programmes that have recently taken place in the country to strengthen MCH services, we now require these maps to be validated by experts with knowledge of one or more of the following:

(i) The pilot P4P programme (2011-2013)
(ii) The national Results-Based Finance programme (RBF, 2016-2020)
(iii) Direct Health Facility Financing programme (DHFF, 2019-Present)

Participation in study:
You have been invited to take part in this interview because of your expertise and experience of health system operation and/or (i) the pilot P4P programme, (2) the national RBF programme, (3) DHFF programme. We will be interviewing approximately 20 individuals, which includes representatives from the Ministry of Health (MOHCDGEC), President's Office Regional Administration and Local Government (PORALG), National Health Insurance Fund (NHIF), Medical Stores Department (MSD), Regional Health Management Team (RHMT), Development partners (e.g., World Bank, USAID, and WHO). Based on the feedback we receive from participants during interviews, we will refine the structure of our maps to ensure they are accurately representing the impact of these programmes on the Tanzania health system.

Your participation in this interview is entirely voluntary, if you change your mind about wanting to attend the interview you can let us know using the contact details at the bottom of the form.

During the interview we will show you system maps that are representative of how we believe the health system functions following the introduction of payment for performance initiatives using data collected during the Tanzania pilot P4P programme (2011-2013). We will examine each part of the system map in segments with clear explanations given for how each part functions and contributes to the wider system. We will then ask you a series of questions related to the structure of each system map segment in turn. Using your knowledge, experience and feedback of health system operation we will then refine the structure of our maps to ensure they are able to reflect each of the programmes outlined above.

info@ihi.or.tz | www.ihi.or.tz

Ifakara Branch
Off Mlabani Passage
P.o. Box 53 Ifakara
Phone: +255232931572

Dar es Salaam Office
Plot 463, Kiko Avenue, Mikocheni
P.o. Box 78,373 Dar es Salaam
Phone: +255222774756

Bagamoyo Branch
Inside District Hospital
P.o. Box 74 Bagamoyo
Phone: +255232625164

Due to COVID-19, there are now restrictions in place on international travel and required to observe a social distance (which restricts face to face interviews). We will therefore conduct interviews over Zoom (or an alternative web call service) with measures in place to ensure the security of the web call. Our researchers will make written notes during the interview and with your permission, use a handheld audio recording device to record the interview. Please indicate on this consent form if you would prefer not to have your feedback audio-recorded. We will also ask you if it would be possible to contact you in the future to ask further or clarifying questions to aid our validation of the model.

Study benefits and risks:
The information we get from the study will aid our knowledge and understanding of the health system impact of resource- and incentive-based programmes in Tanzania, benefiting patients, health workers and those implementing such initiatives. The study results will be published in a health systems journal and as part of academic presentations and study reports so that other researchers can learn from our findings.

Participants will not be exposed to any physical risks and the study is unlikely to pose any risk greater than the risk encountered in daily life. Your contribution to this study will be anonymised using codes only our researchers have access too. The participants code will confer what their respondent group is (e.g., donor, government, health care provider) but they will not be identifiable. We will use these codes to refer to participant feedback in study reports, academic presentations or published written work so they cannot be identified. Only the research team will have access to the secure password-protected folder containing the original feedback (written or audio-recorded) obtained in interviews, protecting the anonymity of participants.

Rights:
If you have a question/concern about any aspect of this study, please contact the Principal investigators:

- Dr Peter Binyaruka (Ifakara Health Institute, IHI)
  pbinyaruka@ihi.or.tz or +255655 363361
- Dr Josephine Borghi (The London School of Hygiene and Tropical Medicine, LSHTM)
  Josephine.Borghi@lshtm.ac.uk or +4420 7927 2090

In case you want information please feel free to contact the local principal investigator on this study: Dr Peter John Binyaruka (+255655363361, pbinyaruka@ihi.or.tz) from IHI. This research has been approved by board of research ethics of IHI-IRB and NIMR. For ethical issues please contact Mr. Fakih Bakari (fbakari@ihi.or.tz), representative of the IHI IRB, P.O. Box 78373 Dar es salaam, (+255 23 2625164/ +255 22 2774714); and Ms. Sia Malekia (smalekia@nimr.or.tz), representative of NIMR, P. O. Box 9653, 11101 Dar es Salaam, Tanzania, Tel: +255-22-2121400, Fax: +255-22-2121360. You can contact also the London School of Hygiene and Tropical Medicine ethics committee by telephone at +4420 7927 2221 or by email at ethics@lshtm.ac.uk.

Consent:

Please read the following statements and initial the boxes to provide your consent.

Statement	Please initial
I understand that my participation is voluntary.	
I understand that I can withdraw from the interview at any time.	
I understand that my feedback during the interview will be recorded via written notes and audio-recorded by a member of the research team (using either encrypted web-based software or a handheld recording device).	
I understand that my feedback collected in this interview will be kept confidential and be used for research purpose only. Also, I understand that my responses will be only shared with/by authorised individuals in the research team from LSHTM and IHI and any information included in the report, academic presentation or in published work I will not be identified as the respondent.	
I confirm that I am happy for the research team to contact me in the future to ask follow up questions.	
I agree to take part in the above-named study.	

Printed name of participant	Signature of participant	Date

Printed name of person obtaining consent	Signature of person obtaining consent	Date

info@ihi.or.tz | www.ihi.or.tz

Ifakara Branch
Off Mlabani Passage
P.o. Box 53 Ifakara
Phone: +255232931572

Dar es Salaam Office
Plot 463, Kiko Avenue, Mikocheni
P.o. Box 78,373 Dar es Salaam
Phone: +255222774756

Bagamoyo Branch
Inside District Hospital
P.o. Box 74 Bagamoyo
Phone: +255232625164

2d: Flyer for stakeholder engagement

The COSMIC Project

Using computer modelling to optimise the design of health system programmes

Background

The COSMIC project is a collaboration between research institutes in Tanzania (Ifakara Health Institute), Zambia (University of Zambia), Uganda (Makerere University), the United Kingdom (London School of Hygiene and Tropical Medicine LSHTM) and Switzerland (University of Geneva).

The project builds on a previous collaboration between the Ifakara Health Institute and LSHTM which evaluated the effects of a payment for performance (P4P) scheme (provision of financial rewards to health care providers for delivery of services) on maternal and child health services in Pwani region of Tanzania (2011-2013).[1,2]

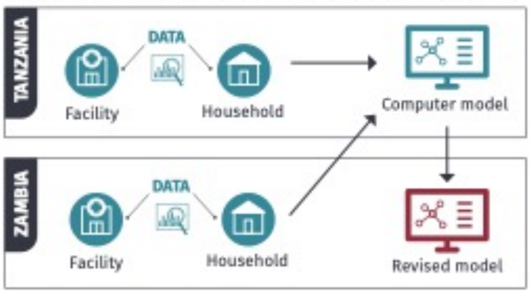
Despite a growing number of P4P schemes globally, we still understand little about how these programmes affect health systems. The schemes often vary in their design across settings, and we need to better understand how to optimise the design of these programmes for maximum impact.

Aim

The COSMIC project aims to better understand how health systems function and respond to health programmes, such as P4P, using computer models. The models serve as 'virtual laboratories' to explore likely effects of health programmes on health systems, and how modifications to the design of programmes influence programme effectiveness.

Models are being developed using data collected from facilities and households as part of an evaluation of the P4P programme in the Pwani region of Tanzania (2011-2013). After the Tanzania model has been developed, we will use data from another P4P programme (Zambia) to test the use of the model in other settings (Figure 1).

FIGURE 1: PROCESS OF DEVELOPING THE COMPUTER MODEL



We plan to develop user-friendly interfaces for our models so that they can be used by policymakers to optimise the design of payment for performance programmes and also for health system management.

Introducing the computer models

We are currently developing two computer models of the Tanzanian health system: a system dynamics model (with a macro-level, whole system focus) and an agent-based model (with an individual, micro-level focus).

TOP DOWN SYSTEM DYNAMICS MODEL

The system dynamics model aims to help us understand the holistic impact of payment for performance on the health system. To create the model, we first develop a causal loop diagram, showing how the health system responds to P4P. A causal loop diagram uses arrows to map out relationships between different variables in the health system.

Figure 2 shows a simple example, where the promise of bonus payments motivates health workers to deliver more incentivised services to woman and children, resulting in further bonus payments and the cycle continues in a loop.

The completed causal loop diagram for the P4P programme in Tanzania will be used as a blueprint to develop a quantitative simulation model, to understand how these health system variables behave and change over time in response to payment for performance.

FIGURE 2: CASUAL LOOP DIAGRAM

Simple casual loop diagram showing the impact of bonus payments on health worker motivation and delivery of maternal and child health services

1. Borghi J, Mayumana I, Mashasi I, et al. Protocol for the evaluation of a pay for performance programme in Pwani region in Tanzania: A controlled before and after study. Implement Sci 2013;8:80.

2. Binyaruka P, Patouillard E, Powell-Jackson T, et al. Effect of Paying for Performance on Utilisation, Quality, and User Costs of Health Services in Tanzania: A Controlled Before and After Study. PLoS One 2015;10:1–16.

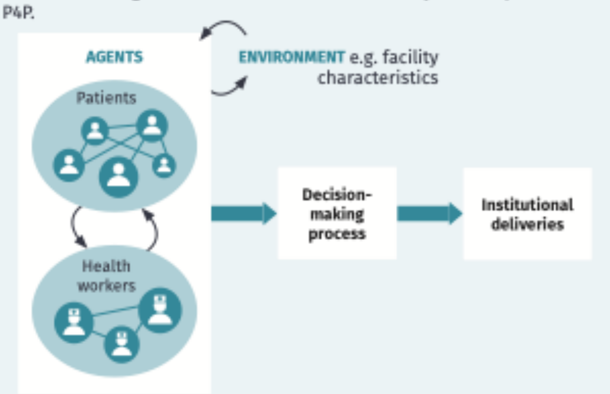
BOTTOM UP AGENT-BASED MODEL

The agent-based model explores how the decisions and behaviours of individuals (or 'agents') affect the health system and its response to P4P.

In our model, we have two groups of agents – patients and health workers (Figure 3) with specific characteristics and behaviours (based on the Pwani data). We have developed a conceptual map to consider how the behaviour of health workers (such as kindness towards patients), facility characteristics (such as availability of drugs) and patient characteristics (such as socio-economic status) affect demand for and supply of services.

FIGURE 3: AGENT-BASED MODEL

Outline of health system processes mapped and modelled using agent-based modelling to better understand the health system response to P4P.



What we need from stakeholders and policy makers

The success of this research depends on the accuracy of the causal loop diagram and conceptual map as they act as blueprints for the simulation models.

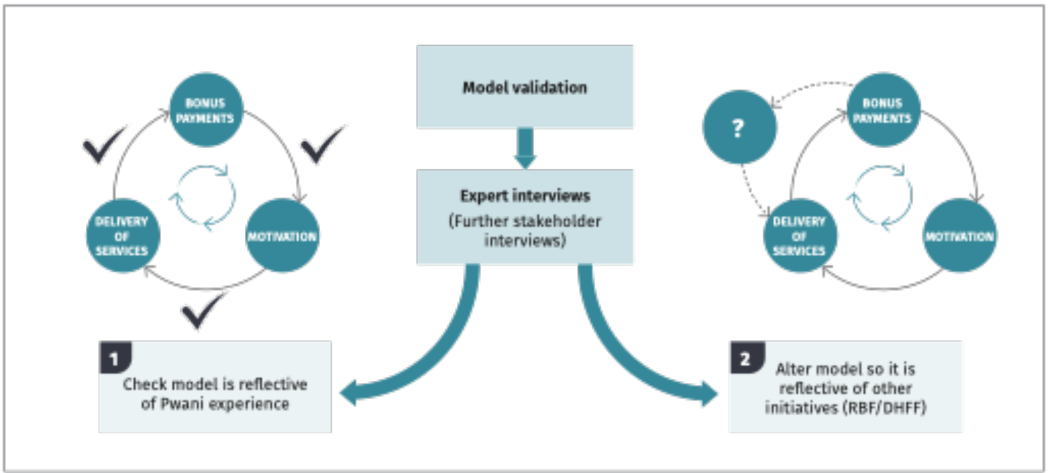
So far, the causal loop diagram we have developed has been validated by several experts that were involved in the evaluation of the pilot Pwani programme.

For the next stage of research (outlined in Figure 4), we need:

1. **The causal loop diagram and conceptual map to be validated by more stakeholders** to ensure the models fully reflect the experiences of those who were involved in implementing or managing the Pwani P4P programme.
2. **To hear from experts who were involved in implementing or managing more recent maternal and child health quality improvement programmes**, such as Results-Based Financing (RBF) and Direct Health Facility Financing (DHFF), so that our models can be used to study these programmes as well.

We very much hope you will be able to support this research by participating in the upcoming interviews to validate our models.

FIGURE 4: NEXT STAGES OF THE RESEARCH



More information: For further information and to indicate your interest in participating in this study please contact Professor Jo Borghi (Josephine.Borghi@lshtm.ac.uk) at the London School of Hygiene and Tropical Medicine or Dr Peter Binyaruka (pbinyaruka@ihi.or.tz) at Ifakara Health Institute.

Funding: The work described in this paper was funded by the Health Systems Research Initiative (HSRI).
MRC Grant Reference Number: MR/R013454/1.

2e: Stakeholder simulation model validation interview tool

Interviewer: This series of interviews have been organised by researchers from Ifakara Health Institute (IHI) and the London School of Hygiene and Tropical Medicine (LSHTM). We are hoping to conduct interviews with experts, such as yourself, to validate a simulation model we have created of the Tanzania maternal and child health (MCH) system response to payment for performance (P4P).

Interviewer: In an earlier interview (June/November/December 2020), we presented the system map we had developed using data collected during the P4P programme in Pwani that ran from 2011-2013. Using this system map, data collected during the programme, evidence from the literature and your feedback, we have now developed a computer model. In the model, we focus on the impact of P4P on the two programme targets where improvements were documented: provision of two doses of intermittent preventive treatment and facility-based deliveries.

Interviewer: To ensure our model is accurately representing the behaviour of the Tanzanian health system and its response to the Pwani P4P scheme, we now require model output to be validated by experts. We also intend to further develop the model to examine health system response to more recent health financing programmes:

(i) The national Results-Based Finance programme (RBF, 2016-2020)

(ii) Direct Health Facility Financing programme (DHFF, 2019-Present)

Interviewer: During this interview we will present and describe the key results from our simulation model related to the health system response to RBF (impact on health worker motivation, availability of medical commodities etc.). Using your knowledge of health system operation, your role will be to evaluate our results and determine their credibility. This crucial step in a series of model validation stages will provide confidence in the model, which will be used to develop policy recommendations for the implementation of results-based finance programmes in Tanzania and other settings.

Interviewer: Just before we begin, I have received a hard copy of your consent form but I would just like to seek your verbal consent that you are happy to continue with the interview and you are happy for me to take written notes and an audio-recording of this session. This is only for our records and shared only with our research team. You can change your mind or stop the interview at any time.

OR

Interviewer: Just before we begin, I am going to take verbal consent for the interview, I can then collect a hard copy of the consent form after the interview. I will now read through the informed consent form with you.

- You understand that your participation is voluntary.
- You understand that you can withdraw from the interview at any time.
- You understand that your feedback during the interview will be recorded via written notes and audio-recorded by a member of the research team (using either encrypted web-based software or a handheld recording device).
- You understand that your feedback collected in this interview will be kept confidential and be used for research purpose only. Also, you understand that your responses will be only shared with/by authorised individuals in the research team from LSHTM and IHI and any information included in the report, academic presentation or in published work will not be identified as the respondent.
- You confirm that you are happy for the research team to contact you in the future to ask follow up questions.
- You agree to take part in the above-named study.

Interviewer: *Interviewer then proceeds with presentation and discussion of key model results and output periodically stopping to check interviewee understanding and to ask if any modifications should be made to the behaviour of the model to reflect their experience of the Tanzania programme.*

Interviewer: At the moment, we are focused on Pwani P4P programme but also hope to model DHFF and RBF. We would like to ask you:

(i) if there other outputs you would like to see from the model

(ii) if there are things you would like to see varied in the model (to see impact on key outcomes).

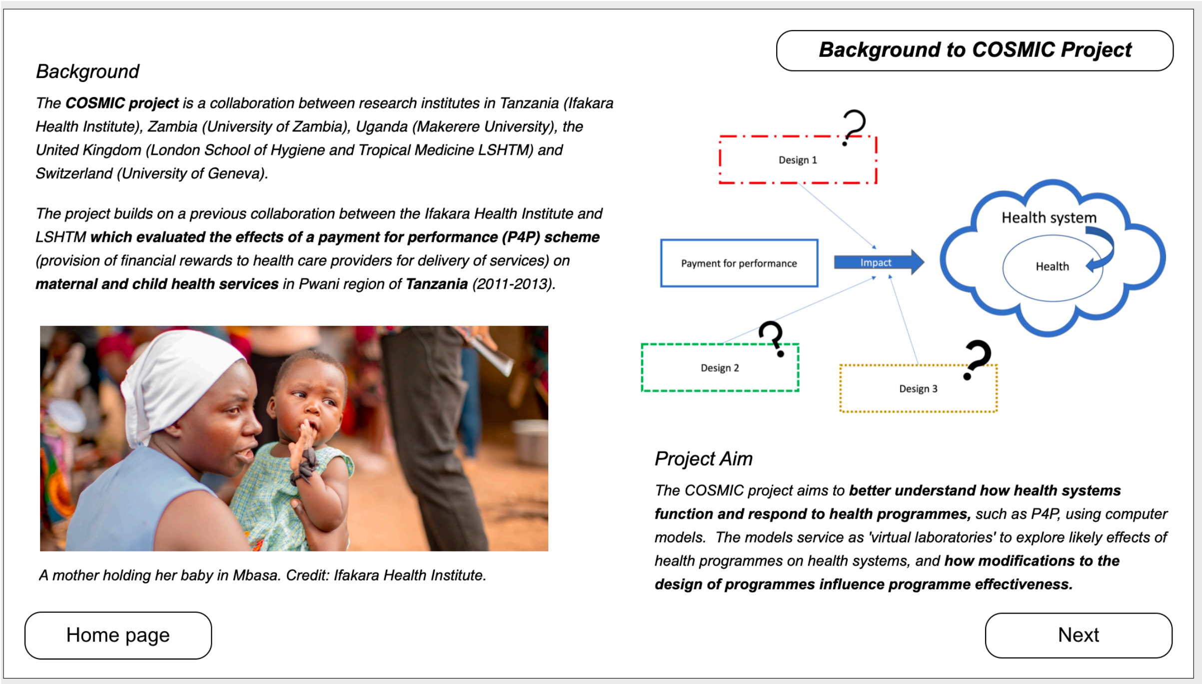
(iii) do you think this model and types of simulation could be useful for decision making around the design of P4P programmes?

(iv) would you be happy for us to contact you with follow up questions?

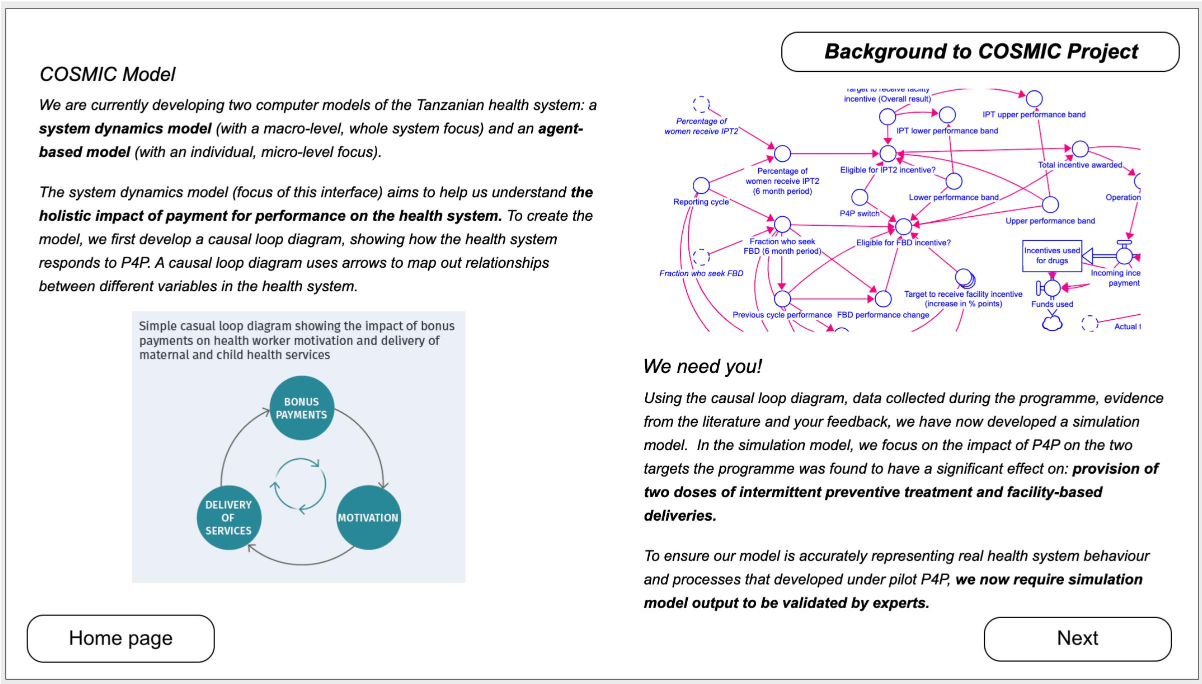
Interviewer: Thank you for your time today, this has been incredibly useful and it is very much appreciated by the COSMIC team.

Model interface slides

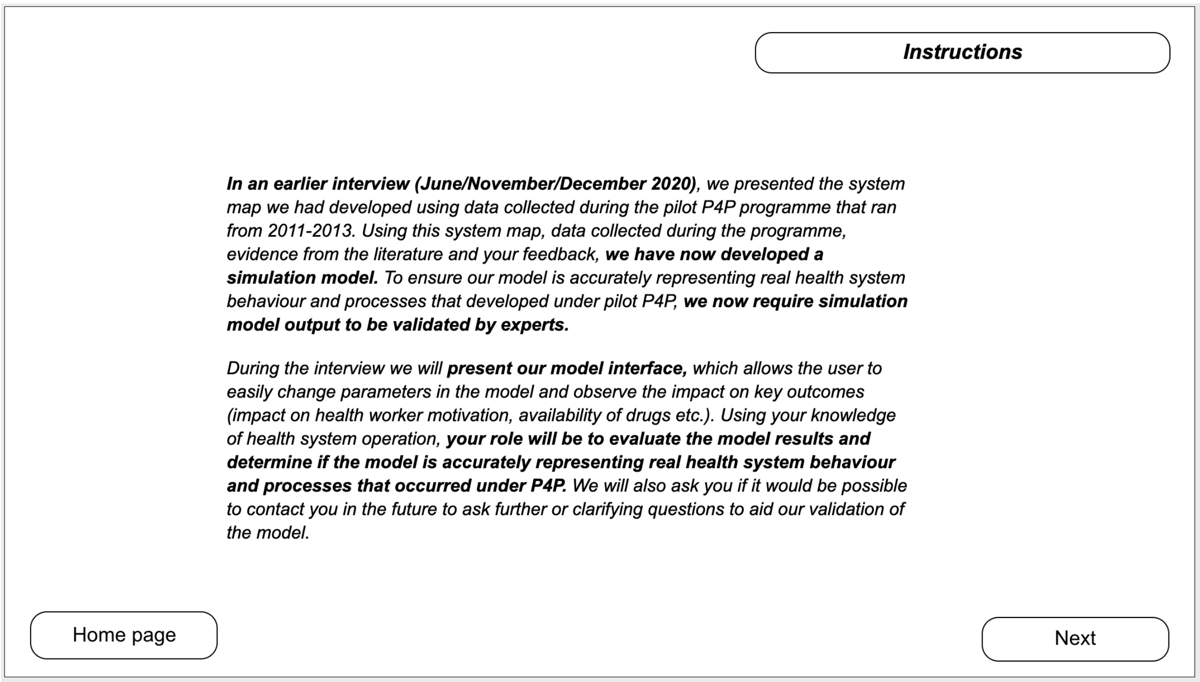
Slide 1: Introduction slide.



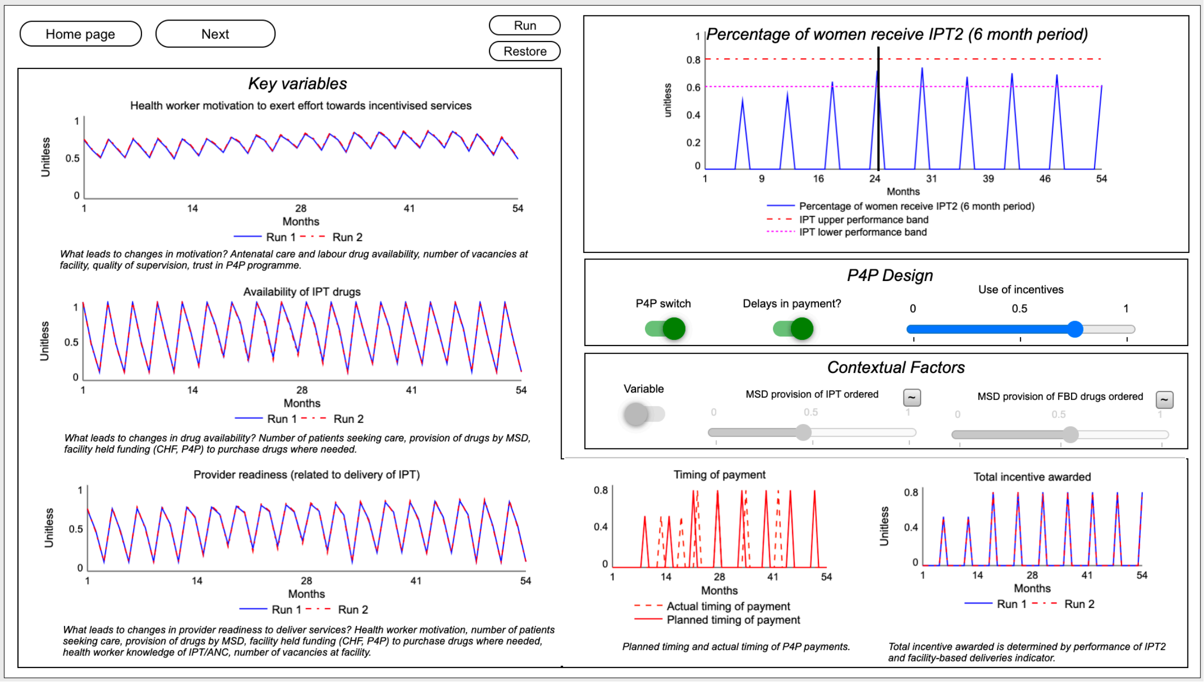
Slide 2: Background (1) slide.



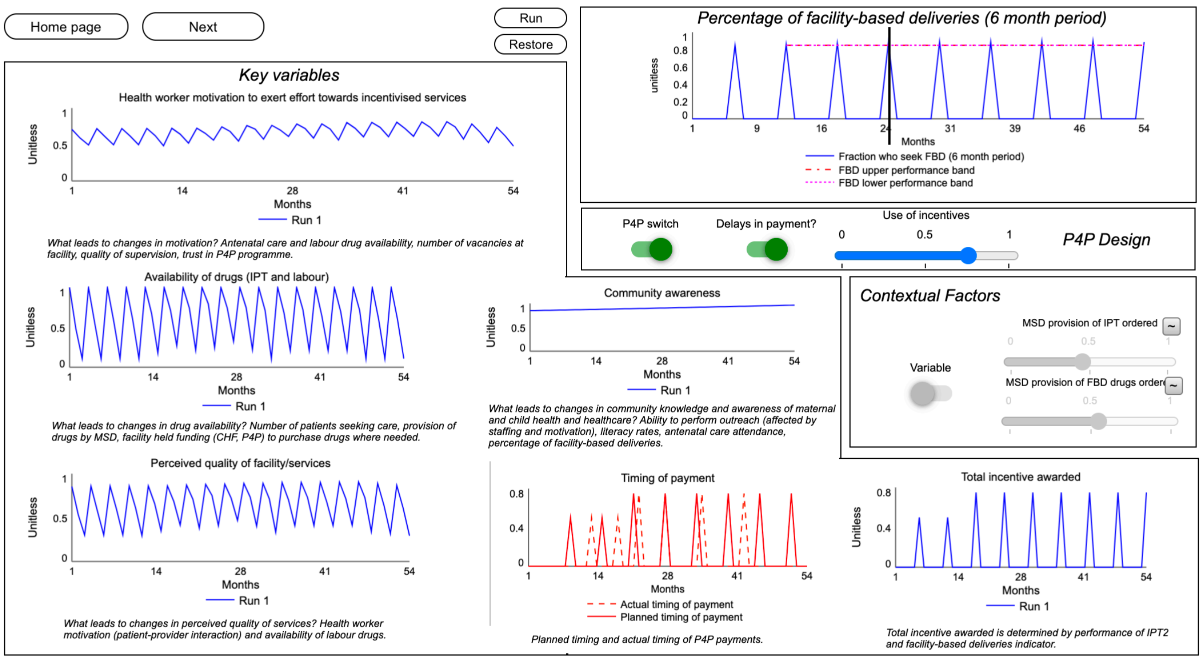
Slide 3: Background (2) slide.



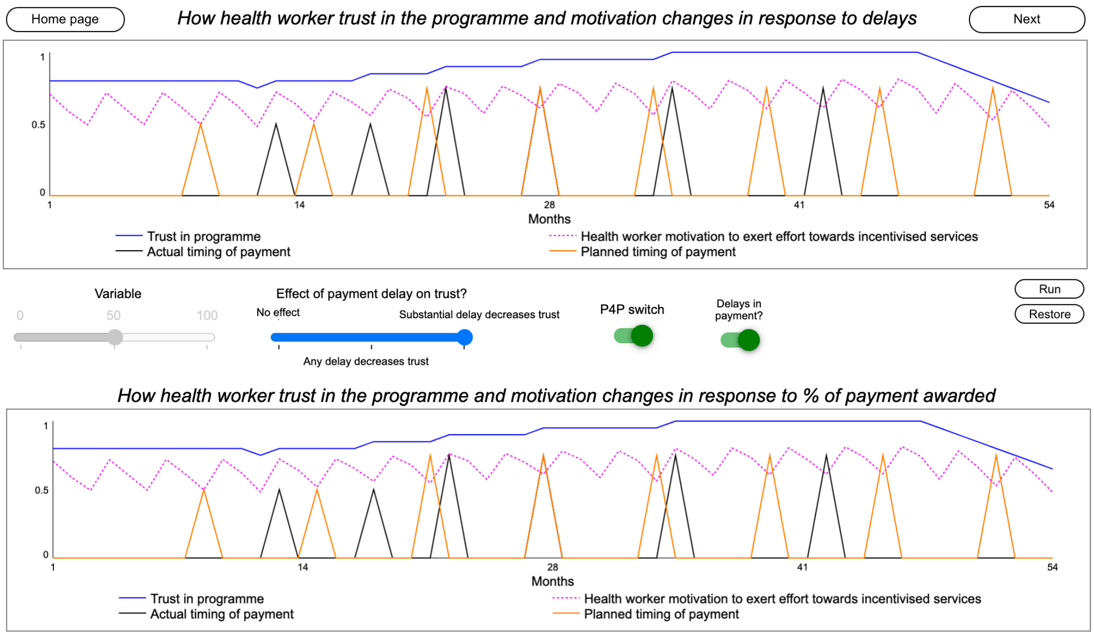
Slide 4: Instructions slide.



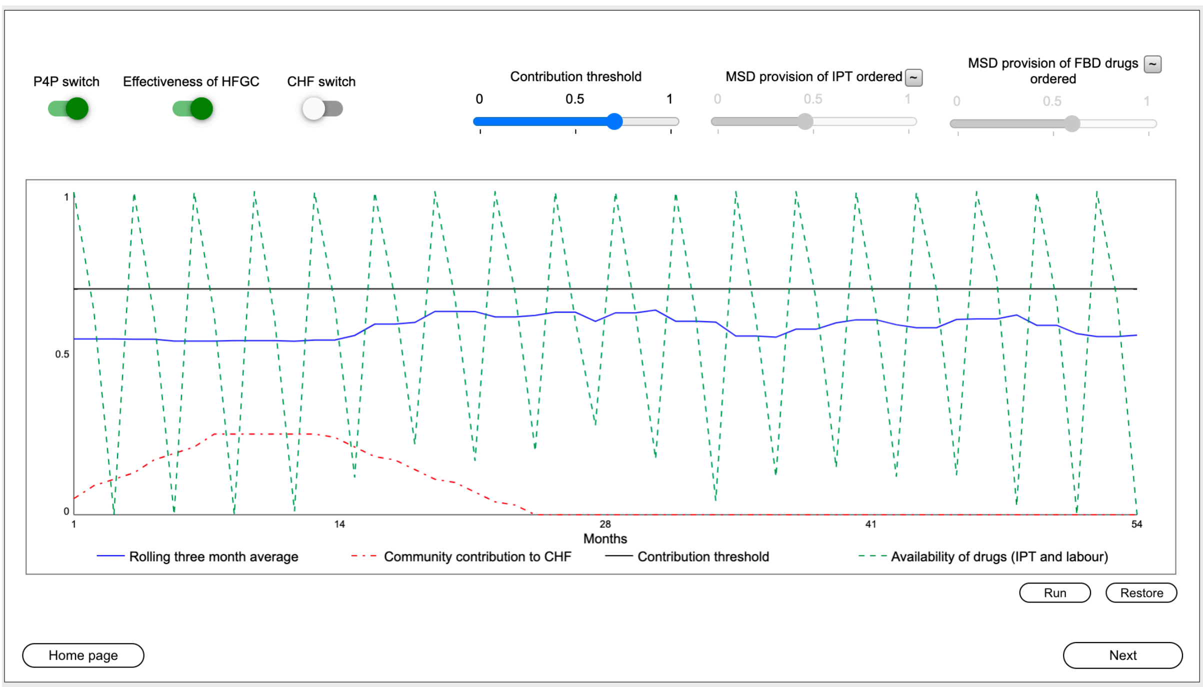
Slide 5: Interactive slide to discuss dynamics around percentage of women who receive at least two doses of IPT during ANC.



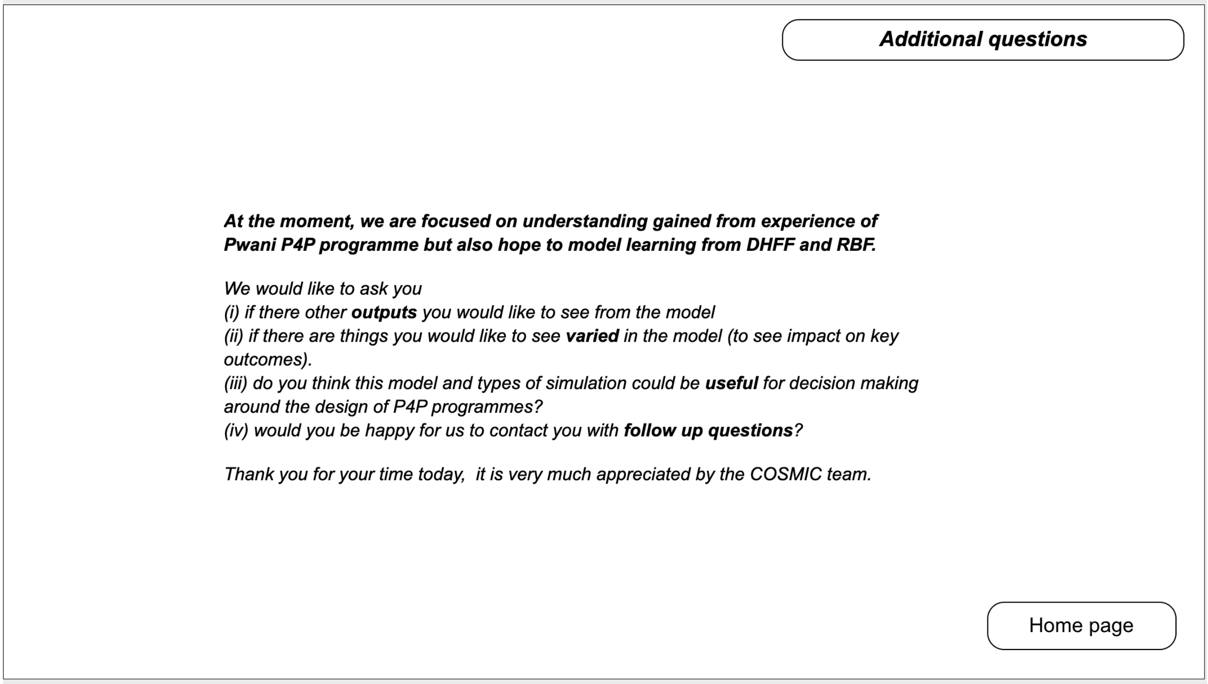
Slide 6: Interactive slide to discuss dynamics around percentage of women who seek facility-based deliveries.



Slide 7: Interactive slide to discuss how health worker trust and motivation responds to changes in timing of payments and amount of payment.



Slide 8: Interactive slide to discuss how drug availability, perceived threshold of acceptable drug availability and effectiveness of Health Facility Governing Committee (FHGC) affect community payment into the Community Health Fund (CHF).



Slide 9: Additional questions and conclusion slide.

2f: Information sheet for simulation model validation interviews

Novel methods for optimising health systems payment for performance interventions to improve maternal and child health in low-resource settings.

Participants Information Sheet

Introduction

We would like to interview you as part of our study. Participating in this interview is entirely up to you. Before you decide, you need to understand why we need your expertise and what it would involve from you. We have made this information sheet available prior to the interview in order for you have time to digest the information presented in this document and allow time for you to ask any questions. Please ask the research team (contact details given at the end of this document) if anything you read is not clear or you would like more information.

What is the purpose of the study?

The Ifakara Health Institute (IHI) and London School of Hygiene and Tropical Medicine (LSHTM) are conducting research to improve maternal and child health (MCH) in low-resource settings. We are hoping to maximise the impact of health systems payment for performance (P4P) interventions through the development and use of computer models of the Tanzanian health system and the relationships between elements within the system, such as the behaviour of health workers and patients, and how they respond to the P4P programme.

In an earlier interview (November/December 2020), we presented the system map we had developed using data collected during the P4P programme in Pwani that ran from 2011-2013. Using this system map, data collected during the programme, evidence from the literature and your feedback, we have now developed a computer model. In the model, we focus on the impact of P4P on the two programme targets where improvements were documented: provision of two doses of intermittent preventive treatment and facility-based deliveries. To ensure our model is accurately representing the behaviour of the Tanzanian health system and its response to the Pwani P4P scheme, we now require model output to be validated by experts. We also intend to further develop the model to examine health system response to more recent health financing programmes:
(i) The national Results-Based Finance programme (RBF, 2016-2020)
(ii) Direct Health Facility Financing programme (DHFF, 2019-Present)

Why have I been asked to take part?

You have been invited to take part in this interview because of your expertise and experience of health system operation and/or (i) the P4P programme in Pwani, (2) the national RBF programme, (3) DHFF programme. We will be interviewing approximately 10 individuals, which includes representatives from the Ministry of Health (MOHCDGEC), President's Office Regional Administration and Local Government (PORALG), Regional Health Management Team (RHMT), Development partners (e.g., World Bank, USAID, and WHO) and programme evaluation team members. Based on the feedback we receive from participants during interviews, we will refine the model to ensure it is accurately representing the impact of these programmes on the Tanzania health system.

info@ihi.or.tz | www.ihi.or.tz

Ifakara Branch
Off Mlabani Passage
P.o. Box 53 Ifakara
Phone: +255232931572

Dar es Salaam Office
Plot 463, Kiko Avenue, Mikocheni
P.o. Box 78,373 Dar es Salaam
Phone: +255222774756

Bagamoyo Branch
Inside District Hospital
P.o. Box 74 Bagamoyo
Phone: +255232625164

Do I have to take part?
No, your participation in this interview is voluntary. We have made this information sheet available prior to the interview in order for you to have time to digest the information presented in this document and ask any questions you might have. If you agree to take part, we will ask you to sign a consent form. If you change your mind about wanting to be interviewed, you can let us know using the contact details at the bottom of the form. You can also stop the interview at any time.

What will I have to do?
During the interview we will present our model interface, which allows the user to easily change parameters in the model and observe the impact on key outcomes (impact on health worker motivation, availability of medical commodities etc.). Using your knowledge of health system operation, you will then be asked to evaluate the model results and determine if the model is accurately representing real health system behaviour and processes that occurred under P4P. We will also ask you if it would be possible to contact you in the future to ask further questions to aid our validation of the model.

What data is collected in this study?
Our researchers will make written notes and may also take audio-recordings of feedback to ensure we have a comprehensive record of your feedback. If you would rather not have your feedback audio-recorded we will only take written notes. Please indicate on this consent form if you would prefer not to have your feedback audio-recorded.

How data is collected in this study?
Due to COVID-19, there are now restrictions in place on international travel and requirements to observe a social distance (which restricts face to face interviews). We will therefore conduct interviews over Zoom (or alternative web call service) with measures in place to ensure the security of the web call. Our researchers will make written notes during the interview and, with your permission, use a handheld audio recording device to record the interview.

What will happen to the data collected in this study?
All feedback collected from this interview will be anonymised. We will assign participants a code which will be used to refer to a participant's feedback in analysis or published written work so they cannot be identified. The participants code will confer what their respondent group is (e.g., donor, government, health care provider) but they will not be individually identifiable. Only the research team will have access to the secure password-protected folder containing the original feedback (either written or audio-recordings) obtained in interviews, protecting the anonymity of participants. The feedback collected in these interviews will be made available in published work, academic presentations, PhD thesis and study reports but you will not be identified as having participated and feedback will be anonymised.

What will happen to the results of this study?
The information we get from the study will aid our knowledge and understanding of the health system impact of provider payment programmes in Tanzania. The study results will be published in a health systems journal, and will be presented at conferences, within a PhD thesis and policy briefs so that others can learn from our findings.

What are the possible risks and disadvantages?
Participants will not be exposed to any physical risks and the study is unlikely to pose any risk greater than the risk encountered in daily life. The study team will also be particularly cautious about protecting respondents' identity, and, where relevant, that of their organisations, when reporting findings.

Who is funding this study?
Our study is funded by the Health Systems Research Initiative, jointly funded by Department of International Development (DFID), the Economic and Social Research Council (ESRC), the MRC and the Wellcome Trust.

Who has checked this study?
All research involving human participants is looked at by an independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by Ifakara Health Institute institutional review board (IHI/IRB/No:15 -2019), the Tanzania National Institute for Medical Research (NIMR/HQ/R.8a/ Vol. IX/3154), and the London School of Hygiene and Tropical Medicine Research Ethics Committee (ref: 16139) has also approved this study.

What if I have a question/ concern?
If you have a question or concern about any aspect of this study, please feel free to contact the local principal investigator on this study: Dr Peter John Binyaruka (+255655363361, pbinyaruka@ihi.or.tz) from IHI. This research has been approved by board of research ethics of IHI-IRB and NIMR. For ethical issues please contact Mr. Fakih Bakari (fbakari@ihi.or.tz) representative of the IHI IRB, P.O. Box 78373 Dar es salaam, (+255 23 2625164/ +255 22 2774714); and Ms. Sia Malekia (smalekia@nimr.or.tz), representative of NIMR, P. O. Box 9653, 11101 Dar es Salaam, Tanzania, Tel: +255-22-2121400, Fax: +255-22-2121360. You can contact also the London School of Hygiene and Tropical Medicine ethics committee by telephone at +4420 7927 2221 or by email at ethics@lshtm.ac.uk.

2g: Consent form for simulation model validation interviews

Novel methods for optimising health systems payment for performance interventions to improve maternal and child health in low-resource settings.

CONSENT FORM

Study Purpose:
The Ifakara Health Institute (IHI) and London School of Hygiene and Tropical Medicine (LSHTM) are conducting research to improve maternal and child health (MCH) in low-resource settings. We are hoping to maximise the impact of health systems payment for performance (P4P) interventions through the development and use of computer models of the Tanzanian health system and the relationships between elements within the system, such as the behaviour of health workers and patients, and how they respond to the P4P programme.

In an earlier interview (November/December 2020), we presented the system map we had developed using data collected during the P4P programme in Pwani that ran from 2011-2013. Using this system map, data collected during the programme, evidence from the literature and your feedback, we have now developed a computer model. In the model, we focus on the impact of P4P on the two programme targets where improvements were documented: provision of two doses of intermittent preventive treatment and facility-based deliveries. To ensure our model is accurately representing the behaviour of the Tanzanian health system and its response to the Pwani P4P scheme, we now require model output to be validated by experts. We also intend to further develop the model to examine health system response to more recent health financing programmes:
(i) The national Results-Based Finance programme (RBF, 2016-2020)
(ii) Direct Health Facility Financing programme (DHFF, 2019-Present)

Participant Information Sheet:
Before reading and completing this consent form, please review the participant information sheet for further details on what we would like to ask during the interview, the data that we collect, how this data will be used, study funding and ethics approval, and study benefits and risks.

Rights:
If you have a question or concern about any aspect of this study, please feel free to contact the local principal investigator on this study: Dr Peter John Binyaruka (+255655363361, pbinyaruka@ihi.or.tz) from IHI. This research has been approved by board of research ethics of IHI-IRB and NIMR. For ethical issues please contact Mr. Fakih Bakari (fbakari@ihi.or.tz) representative of the IHI IRB, P.O. Box 78373 Dar es salaam, (+255 23 2625164/ +255 22 2774714); and Ms. Sia Malekia (smalekia@nimr.or.tz), representative of NIMR, P. O. Box 9653, 11101 Dar es Salaam, Tanzania, Tel: +255-22-2121400, Fax: +255-22-2121360. You can contact also the London School of Hygiene and Tropical Medicine ethics committee by telephone at +4420 7927 2221 or by email at ethics@lshtm.ac.uk.

Consent:

Please read the following statements and initial the boxes to provide your consent.

Statement	Please initial
I understand that my participation is voluntary.	
I understand that I can withdraw from the interview at any time.	
I understand that my feedback during the interview will be recorded via written notes and audio-recorded by a member of the research team (using either encrypted web-based software or a handheld recording device).	
I understand that my feedback collected in this interview will be kept confidential and be used for research purpose only. Also, I understand that my responses will be only shared with/by authorised individuals in the research team from LSHTM and IHI and any information included in the report, academic presentation or in published work I will not be identified as the respondent.	
I confirm that I am happy for the research team to contact me in the future to ask follow up questions.	
I agree to take part in the above-named study.	

Printed name of participant	Signature of participant	Date

Printed name of person obtaining consent	Signature of person obtaining consent	Date

info@ihi.or.tz | www.ihi.or.tz

Ifakara Branch
Off Mlabani Passage
P.o. Box 53 Ifakara
Phone: +255232931572

Dar es Salaam Office
Plot 463, Kiko Avenue, Mikocheni
P.o. Box 78,373 Dar es Salaam
Phone: +255222774756

Bagamoyo Branch
Inside District Hospital
P.o. Box 74 Bagamoyo
Phone: +255232625164

Appendix 3: Evidence of retention of copyright or use of published materials in this thesis

Appendix 3 contains evidence of retention of copyright or use of published materials in this thesis for the three published papers:

- 3a: Evidence of author copyright retention for Paper 1, 'Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models', published in BMC Health Services Research in November 2019.
- 3b: License agreement for paper use in thesis for Paper 2, 'Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach', published in Social Science & Medicine in September 2021.
- 3c: License agreement for paper use in thesis for Paper 3, 'How to do (or not to do)…using causal loop diagrams for health system research in low and middle-income settings', published in Health Policy and Planning in August 2022.

3a: Evidence of author copyright retention for Paper 1, 'Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models', published in BMC Health Services Research in November 2019.

BMC Health Services Research

Copyright

- Copyright on any open access article in a journal published by BioMed Central is retained by the author(s).
- Authors grant BioMed Central a license to publish the article and identify itself as the original publisher.
- Authors also grant any third party the right to use the article freely as long as its integrity is maintained and its original authors, citation details and publisher are identified.
- The Creative Commons Attribution License 4.0 formalizes these and other terms

and conditions of publishing articles.

In addition to BioMed Central's copyright policy, some journals also follow an Open Data policy and the Creative Commons CC0 1.0 Public Domain Dedication waiver applies to all published data in these journals. Further information can be found on the individual journals pages.

Where an author is prevented from being the copyright holder (for instance in the case of US government employees or those of Commonwealth governments), minor variations may be required. In such cases the copyright line and license statement in individual articles will be adjusted, for example to state '© 2016 Crown copyright'. Authors requiring a variation of this type should inform BioMed Central during or immediately after submission of their article. Changes to the copyright line cannot be made after publication of an article.

Exceptions to copyright policy

Our policy pages provide details concerning copyright and licensing for articles which were previously published under policies that are different from the above. For instance, occasionally BioMed Central may co-publish articles jointly with other publishers, and different licensing conditions may then apply. In all such cases, however, access to these articles is free from fees or any other access restrictions.

Information specifically regarding permissions and reprints can be found here. Please contact us if there are questions.

BMC Series

Submit manuscript

3b: License agreement for paper use in thesis for Paper 2, 'Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach', published in Social Science & Medicine in September 2021.

Marketplace

This is a License Agreement between Rachel Cassidy ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Order General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Order General Terms and Conditions below.

Order Date	24-Nov-2022	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1293469-1	**Publisher**	PERGAMON
ISSN	1873-5347	**Portion**	Chapter/article

LICENSED CONTENT

Publication Title	Social science & medicine	**Rightsholder**	Elsevier Science & Technology Journals
Article Title	Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach	**Publication Type**	e-Journal
		Start Page	114277
		Volume	285
		URL	http://www.sciencedirect.com/science/journal/02779536
Date	01/01/1982		
Language	English		
Country	United Kingdom of Great Britain and Northern Ireland		

REQUEST DETAILS

Portion Type	Chapter/article	**Rights Requested**	Main product
Page Range(s)	1-14	**Distribution**	Worldwide
Total Number of Pages	14	**Translation**	Original language of publication
Format (select all that apply)	Electronic	**Copies for the Disabled?**	No
Who Will Republish the Content?	Academic institution	**Minor Editing Privileges?**	Yes
		Incidental Promotional Use?	No
Duration of Use	Life of current edition		
Lifetime Unit Quantity	Up to 499	**Currency**	GBP

NEW WORK DETAILS

Title	Using systems thinking to optimise health system interventions for improved maternal and child health in low-resource settings	**Institution Name**	The London School of Hygiene and Tropical Medicine
		Expected Presentation Date	2023-03-03
Instructor Name	Rachel Cassidy		

ADDITIONAL DETAILS

Order Reference Number	N/A	**The Requesting Person/Organization to Appear on the License**	Rachel Cassidy

REUSE CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach	**Title of the Article/Chapter the Portion Is From**	Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach
Editor of Portion(s)	Cassidy, Rachel; Tomoaia-Cotisel, Andrada; Semwanga, Agnes Rwashana; Binyaruka, Peter; Chalabi, Zaid; Blanchet, Karl; Singh, Neha S.; Maiba, John; Borghi, Josephine	**Author of Portion(s)**	Cassidy, Rachel; Tomoaia-Cotisel, Andrada; Semwanga, Agnes Rwashana; Binyaruka, Peter; Chalabi, Zaid; Blanchet, Karl; Singh, Neha S.; Maiba, John; Borghi, Josephine
Volume of Serial or Monograph	285	**Issue, if Republishing an Article From a Serial**	N/A
Page or Page Range of Portion	114277	**Publication Date of Portion**	2021-09-01

RIGHTSHOLDER TERMS AND CONDITIONS

Elsevier publishes Open Access articles in both its Open Access journals and via its Open Access articles option in subscription journals, for which an author selects a user license permitting certain types of reuse without permission. Before proceeding please check if the article is Open Access on http://www.sciencedirect.com and refer to the user license for the individual article. Any reuse not included in the user license terms will require permission. You must always fully and appropriately credit the author and source. If any part of the material to be used (for example, figures) has appeared in the Elsevier publication for which you are seeking permission, with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder. Please contact permissions@elsevier.com with any queries.

Marketplace Order General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

“Terms” means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

“User” or “you” is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User’s behalf, such person shall be deemed jointly a User for purposes of such Terms.

“Work(s)” are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC’s Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User’s use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

9) **Copyright Notice.** Use of proper copyright notice for a Work is required as a condition of any License granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Used with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either on a cover page or in another location that any person, upon gaining access to the material which is the subject of a permission, shall see, or in the case of republication Licenses, immediately adjacent to the Work as used (for example, as part of a by-line or footnote) or in the place where substantially all other credits or

notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

10) **Indemnity.** User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs, and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein and in the Order Confirmation, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy, or other tangible or intangible property.

11) **Limitation of Liability.** UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OR BOTH OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for the relevant License. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors, and assigns.

12) **Limited Warranties.** THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS." CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS, OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

13) **Effect of Breach.** Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the License set forth in the Order Confirmation and/or the Terms, shall be a material breach of such License. Any breach not cured within 10 days of written notice thereof shall result in immediate termination of such License without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

14) **Additional Terms for Specific Products and Services.** If a User is making one of the uses described in this Section 14, the additional terms and conditions apply:

a) ***Print Uses of Academic Course Content and Materials (photocopies for academic coursepacks or classroom handouts).*** For photocopies for academic coursepacks or classroom handouts the following additional terms apply:

i) The copies and anthologies created under this License may be made and assembled by faculty members individually or at their request by on-campus bookstores or copy centers, or by off-campus copy shops and other similar entities.

ii) No License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied) (ii) permit "publishing ventures" where any particular anthology would be systematically marketed at multiple institutions.

iii) Subject to any Publisher Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the academic pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to no more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular anthology, whether photocopied or electronic, at more than one institution of learning;

E) in the case of a photocopy permission, no materials may be entered into electronic memory by User except in order to produce an identical copy of a Work before or during the academic term (or analogous period) as to which any particular permission is granted. In the event that User shall choose to retain materials that are the subject of a photocopy permission in electronic memory for purposes of producing identical copies more than one day after such retention (but still within the scope of any permission granted), User must notify CCC of such fact in the applicable permission request and such retention shall constitute one copy actually sold for purposes of calculating permission fees due; and

F) any permission granted shall expire at the end of the class. No permission granted shall in any way include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied).

iv) Books and Records; Right to Audit. As to each permission granted under the academic pay-per-use Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any photocopies sold or by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this License for any reason.

b) ***Digital Pay-Per-Uses of Academic Course Content and Materials (e-coursepacks, electronic reserves, learning management systems, academic institution intranets).*** For uses in e-coursepacks, posts in electronic reserves, posts in learning management systems, or posts on academic institution intranets, the following additional terms apply:

i) The pay-per-uses subject to this Section 14(b) include:

A) **Posting e-reserves, course management systems, e-coursepacks for text-based content,** which grants authorizations to import requested material in electronic format, and allows electronic access to this material to members of a designated college or university class, under the direction of an instructor designated by the college or university, accessible only under appropriate electronic controls (e.g., password);

B) **Posting e-reserves, course management systems, e-coursepacks for material consisting of photographs or other still images not embedded in text,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorization: to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above, including any necessary resizing, reformatting or modification of the resolution of such requested material (provided that such modification does not alter the underlying editorial content or meaning of the requested material, and provided that the resulting modified content is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms), but not including any other form of manipulation, alteration or editing of the requested material;

C) **Posting e-reserves, course management systems, e-coursepacks or other academic distribution for audiovisual content,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorizations: (i) to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above; (ii) to display and perform the requested material to such members of such class in the physical classroom or remotely by means of streaming media or other video formats; and (iii) to "clip" or reformat the requested material for purposes of time or content management or ease of delivery, provided that such “clipping” or reformatting does not alter the underlying editorial content or meaning of the requested material and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms. Unless expressly set forth in the relevant Order Conformation, the License does not authorize any other form of manipulation, alteration or editing of the requested material.

ii) Unless expressly set forth in the relevant Order Confirmation, no License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied or, in the case of Works subject to Sections 14(b)(1)(B) or (C) above, as described in such Sections) (ii) permit "publishing ventures" where any particular course materials would be systematically marketed at multiple institutions.

iii) Subject to any further limitations determined in the Rightsholder Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the electronic course content pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to not more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular materials, whether photocopied or electronic, at more than one institution of learning;

E) electronic access to material which is the subject of an electronic-use permission must be limited by means of electronic password, student identification or other control permitting access solely to students and instructors in the class;

F) User must ensure (through use of an electronic cover page or other appropriate means) that any person, upon gaining electronic access to the material, which is the subject of a permission, shall see:

- a proper copyright notice, identifying the Rightsholder in whose name CCC has granted permission,
- a statement to the effect that such copy was made pursuant to permission,
- a statement identifying the class to which the material applies and notifying the reader that the material has been made available electronically solely for use in the class, and
- a statement to the effect that the material may not be further distributed to any person outside the class, whether by copying or by transmission and whether electronically or in paper form, and User must also ensure that such cover page or other means will print out in the event that the person accessing the material chooses to print out the material or any part thereof.

G) any permission granted shall expire at the end of the class and, absent some other form of authorization, User is thereupon required to delete the applicable material from any electronic storage or to block electronic access to the applicable material.

iv) Uses of separate portions of a Work, even if they are to be included in the same course material or the same university or college class, require separate permissions under the electronic course content pay-per-use Service. Unless otherwise provided in the Order Confirmation, any grant of rights to User is limited to use completed no later than the end of the academic term (or analogous period) as to which any particular permission is granted.

v) Books and Records; Right to Audit. As to each permission granted under the electronic course content Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any electronic copies used by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this license for any reason.

c) ***Pay-Per-Use Permissions for Certain Reproductions (Academic photocopies for library reserves and interlibrary loan reporting) (Non-academic internal/external business uses and commercial document delivery).*** The License expressly excludes the uses listed in Section (c)(i)-(v) below (which must be subject to separate license from the applicable Rightsholder) for: academic photocopies for library reserves and interlibrary loan reporting; and non-academic internal/external business uses and commercial document delivery.

i) electronic storage of any reproduction (whether in plain-text, PDF, or any other format) other than on a transitory basis;

ii) the input of Works or reproductions thereof into any computerized database;

iii) reproduction of an entire Work (cover-to-cover copying) except where the Work is a single article;

iv) reproduction for resale to anyone other than a specific customer of User;

v) republication in any different form. Please obtain authorizations for these uses through other CCC services or directly from the rightsholder.

Any license granted is further limited as set forth in any restrictions included in the Order Confirmation and/or in these Terms.

d) ***Electronic Reproductions in Online Environments (Non-Academic-email, intranet, internet and extranet).*** For "electronic reproductions", which generally includes e-mail use (including instant messaging or other electronic transmission to a defined group of recipients) or posting on an intranet, extranet or Intranet site (including any display or performance incidental thereto), the following additional terms apply:

i) Unless otherwise set forth in the Order Confirmation, the License is limited to use completed within 30 days for any use on the Internet, 60 days for any use on an intranet or extranet and one year for any other use, all as measured from the "republication date" as identified in the Order Confirmation, if any, and otherwise from the date of the Order Confirmation.

ii) User may not make or permit any alterations to the Work, unless expressly set forth in the Order Confirmation (after request by User and approval by Rightsholder); provided, however, that a Work consisting of photographs or other still images not embedded in text may, if necessary, be resized, reformatted or have its resolution modified without additional express permission, and a Work consisting of audiovisual content may, if necessary, be "clipped" or reformatted for purposes of time or content management or ease of delivery (provided that any such resizing, reformatting, resolution modification or “clipping” does not alter the underlying editorial content or meaning of the Work used, and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular License described in the Order Confirmation and the Terms.

15) **Miscellaneous.**

a) User acknowledges that CCC may, from time to time, make changes or additions to the Service or to the Terms, and that Rightsholder may make changes or additions to the Rightsholder Terms. Such updated Terms will replace the prior terms and conditions in the order workflow and shall be effective as to any subsequent Licenses but shall not apply to Licenses already granted and paid for under a prior set of terms.

b) Use of User-related information collected through the Service is governed by CCC's privacy policy, available online at www.copyright.com/about/privacy-policy/.

c) The License is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the License or any rights granted thereunder; provided, however, that, where applicable, User may assign such License in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in any new material which includes the Work(s) licensed under this Service.

d) No amendment or waiver of any Terms is binding unless set forth in writing and signed by the appropriate parties, including, where applicable, the Rightsholder. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by or on behalf of the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the License described in the Order Confirmation, which terms are in any way inconsistent with any Terms set forth in the Order Confirmation, and/or in CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

e) The License described in the Order Confirmation shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such License shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.

3c: License agreement for paper use in thesis for Paper 3, 'How to do (or not to do) …using causal loop diagrams for health system research in low and middle-income settings', published in Health Policy and Planning in August 2022.

Marketplace

Order Date	24-Nov-2022	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1293468-1	**Publisher**	OXFORD UNIVERSITY PRESS
ISSN	1460-2237	**Portion**	Chapter/article

LICENSED CONTENT

Publication Title	Health policy and planning	**Publication Type**	e-Journal
Article Title	How to do (or not to do)... Using Causal Loop Diagrams for Health System Research in Low- and Middle-Income Settings.	**Start Page**	1328
		End Page	1336
		Issue	10
		Volume	37
		URL	http://heapol.oupjournals.org/
Author/Editor	Oxford University Press.		
Date	01/01/1986		
Language	English		
Country	United Kingdom of Great Britain and Northern Ireland		
Rightsholder	Oxford University Press - Journals		

REQUEST DETAILS

Portion Type	Chapter/article	**Rights Requested**	Main product
Page Range(s)	1-9	**Distribution**	Worldwide
Total Number of Pages	9	**Translation**	Original language of publication
Format (select all that apply)	Electronic	**Copies for the Disabled?**	No
Who Will Republish the Content?	Academic institution	**Minor Editing Privileges?**	Yes
		Incidental Promotional Use?	No
Duration of Use	Life of current edition		
Lifetime Unit Quantity	Up to 499	**Currency**	GBP

NEW WORK DETAILS

Title	Using systems thinking to optimise health system interventions for improved maternal and child health in low-resource settings	**Institution Name**	The London School of Hygiene and Tropical Medicine
		Expected Presentation Date	2023-03-03
Instructor Name	Rachel Cassidy		

ADDITIONAL DETAILS

Order Reference Number	N/A	**The Requesting Person/Organization to Appear on the License**	Rachel Cassidy

REUSE CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	How to do (or not to do)… Using Causal Loop Diagrams for Health System Research in Low- and Middle-Income Settings.	**Title of the Article/Chapter the Portion Is From**	How to do (or not to do)… Using Causal Loop Diagrams for Health System Research in Low- and Middle-Income Settings.
Editor of Portion(s)	Cassidy, Rachel; Borghi, Josephine; Rwashana Semwanga, Agnes; Binyaruka, Peter; Singh, Neha S; Blanchet, Karl	**Author of Portion(s)**	Cassidy, Rachel; Borghi, Josephine; Rwashana Semwanga, Agnes; Binyaruka, Peter; Singh, Neha S; Blanchet, Karl
Volume of Serial or Monograph	37	**Issue, if Republishing an Article From a Serial**	10
Page or Page Range of Portion	1328-1336	**Publication Date of Portion**	2022-12-01

Marketplace Order General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the

Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User's use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

9) **Copyright Notice.** Use of proper copyright notice for a Work is required as a condition of any License granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Used with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either on a cover page or in another location that any person, upon gaining access to the material which is the subject of a permission, shall see, or in the case of republication Licenses, immediately adjacent to the Work as used (for example, as part of a by-line or footnote) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

10) **Indemnity.** User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs, and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein and in the Order Confirmation, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy, or other tangible or intangible property.

11) **Limitation of Liability.** UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OR BOTH OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for the relevant License. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors, and assigns.

12) **Limited Warranties.** THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS." CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS, OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

13) **Effect of Breach.** Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the License set forth in the Order Confirmation and/or the Terms, shall be a material breach of such License. Any breach not cured within 10 days of written notice thereof shall result in immediate termination of such License without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

14) **Additional Terms for Specific Products and Services.** If a User is making one of the uses described in this Section 14, the additional terms and conditions apply:

a) ***Print Uses of Academic Course Content and Materials (photocopies for academic coursepacks or classroom handouts).*** For photocopies for academic coursepacks or classroom handouts the following additional terms apply:

i) The copies and anthologies created under this License may be made and assembled by faculty members individually or at their request by on-campus bookstores or copy centers, or by off-campus copy shops and other similar entities.

ii) No License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied) (ii) permit "publishing ventures" where any particular anthology would be systematically marketed at multiple institutions.

iii) Subject to any Publisher Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the academic pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to no more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular anthology, whether photocopied or electronic, at more than one institution of learning;

E) in the case of a photocopy permission, no materials may be entered into electronic memory by User except in order to produce an identical copy of a Work before or during the academic term (or analogous period) as to which any particular permission is granted. In the event that User shall choose to retain materials that are the subject of a photocopy permission in electronic memory for purposes of producing identical copies more than one day after such retention (but still within the scope of any permission granted), User must notify CCC of such fact in the applicable permission request and such retention shall constitute one copy actually sold for purposes of calculating permission fees due; and

F) any permission granted shall expire at the end of the class. No permission granted shall in any way include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied).

iv) Books and Records; Right to Audit. As to each permission granted under the academic pay-per-use Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any photocopies sold or by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this License for any reason.

b) ***Digital Pay-Per-Uses of Academic Course Content and Materials (e-coursepacks, electronic reserves, learning management systems, academic institution intranets).*** For uses in e-coursepacks, posts in electronic reserves, posts in learning management systems, or posts on academic institution intranets, the following additional terms apply:

i) The pay-per-uses subject to this Section 14(b) include:

A) **Posting e-reserves, course management systems, e-coursepacks for text-based content,** which grants authorizations to import requested material in electronic format, and allows electronic access to this material to members of a designated college or university class, under the direction of an instructor designated by the college or university, accessible only under appropriate electronic controls (e.g., password);

B) **Posting e-reserves, course management systems, e-coursepacks for material consisting of photographs or other still images not embedded in text,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorization: to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above, including any necessary resizing, reformatting or modification of the resolution of such requested material (provided that such modification does not alter the underlying editorial content or meaning of the requested material, and provided that the resulting modified content is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms), but not including any other form of manipulation, alteration or editing of the requested material;

C) **Posting e-reserves, course management systems, e-coursepacks or other academic distribution for audiovisual content,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorizations: (i) to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above; (ii) to display and perform the requested material to such members of such class in the physical classroom or remotely by means of streaming media or other video formats; and (iii) to "clip" or reformat the requested material for purposes of time or content management or ease of delivery, provided that such “clipping” or reformatting does not alter the underlying editorial content or meaning of the requested material and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms. Unless expressly set forth in the relevant Order Conformation, the License does not authorize any other form of manipulation, alteration or editing of the requested material.

ii) Unless expressly set forth in the relevant Order Confirmation, no License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied or, in the case of Works subject to Sections 14(b)(1)(B) or (C) above, as described in such Sections) (ii) permit "publishing ventures" where any particular course materials would be systematically marketed at multiple institutions.

iii) Subject to any further limitations determined in the Rightsholder Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the electronic course content pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to not more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular materials, whether photocopied or electronic, at more than one institution of learning;

E) electronic access to material which is the subject of an electronic-use permission must be limited by means of electronic password, student identification or other control permitting access solely to students and instructors in the class;

F) User must ensure (through use of an electronic cover page or other appropriate means) that any person, upon gaining electronic access to the material, which is the subject of a permission, shall see:

- a proper copyright notice, identifying the Rightsholder in whose name CCC has granted permission,
- a statement to the effect that such copy was made pursuant to permission,
- a statement identifying the class to which the material applies and notifying the reader that the material has been made available electronically solely for use in the class, and
- a statement to the effect that the material may not be further distributed to any person outside the class, whether by copying or by transmission and whether electronically or in paper form, and User must also ensure that such cover page or other means will print out in the event that the person accessing the material chooses to print out the material or any part thereof.

G) any permission granted shall expire at the end of the class and, absent some other form of authorization, User is thereupon required to delete the applicable material from any electronic storage or to block electronic access to the applicable material.

iv) Uses of separate portions of a Work, even if they are to be included in the same course material or the same university or college class, require separate permissions under the electronic course content pay-per-use Service. Unless otherwise provided in the Order Confirmation, any grant of rights to User is limited to use completed no later than the end of the academic term (or analogous period) as to which any particular permission is granted.

v) Books and Records; Right to Audit. As to each permission granted under the electronic course content Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any electronic copies used by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this license for any reason.

c) ***Pay-Per-Use Permissions for Certain Reproductions (Academic photocopies for library reserves and interlibrary loan reporting) (Non-academic internal/external business uses and commercial document delivery).*** The License expressly excludes the uses listed in Section (c)(i)-(v) below (which must be subject to separate license from the applicable Rightsholder) for: academic photocopies for library reserves and interlibrary loan reporting; and non-academic internal/external business uses and commercial document delivery.

i) electronic storage of any reproduction (whether in plain-text, PDF, or any other format) other than on a transitory basis;

ii) the input of Works or reproductions thereof into any computerized database;

iii) reproduction of an entire Work (cover-to-cover copying) except where the Work is a single article;

iv) reproduction for resale to anyone other than a specific customer of User;

v) republication in any different form. Please obtain authorizations for these uses through other CCC services or directly from the rightsholder.

Any license granted is further limited as set forth in any restrictions included in the Order Confirmation and/or in these Terms.

d) ***Electronic Reproductions in Online Environments (Non-Academic-email, intranet, internet and extranet).*** For "electronic reproductions", which generally includes e-mail use (including instant messaging or other electronic transmission to a defined group of recipients) or posting on an intranet, extranet or Intranet site (including any display or performance incidental thereto), the following additional terms apply:

i) Unless otherwise set forth in the Order Confirmation, the License is limited to use completed within 30 days for any use on the Internet, 60 days for any use on an intranet or extranet and one year for any other use, all as measured from the "republication date" as identified in the Order Confirmation, if any, and otherwise from the date of the Order Confirmation.

ii) User may not make or permit any alterations to the Work, unless expressly set forth in the Order Confirmation (after request by User and approval by Rightsholder); provided, however, that a Work consisting of photographs or other still images not embedded in text may, if necessary, be resized, reformatted or have its resolution modified without additional express permission, and a Work consisting of audiovisual content may, if necessary, be "clipped" or reformatted for purposes of time or content management or ease of delivery (provided that any such resizing, reformatting, resolution modification or “clipping” does not alter the underlying editorial content or meaning of the Work used, and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular License described in the Order Confirmation and the Terms.

15) **Miscellaneous.**

a) User acknowledges that CCC may, from time to time, make changes or additions to the Service or to the Terms, and that Rightsholder may make changes or additions to the Rightsholder Terms. Such updated Terms will replace the prior terms and conditions in the order workflow and shall be effective as to any subsequent Licenses but shall not apply to Licenses already granted and paid for under a prior set of terms.

b) Use of User-related information collected through the Service is governed by CCC's privacy policy, available online at www.copyright.com/about/privacy-policy/.

c) The License is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the License or any rights granted thereunder; provided, however, that, where applicable, User may assign such License in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in any new material which includes the Work(s) licensed under this Service.

d) No amendment or waiver of any Terms is binding unless set forth in writing and signed by the appropriate parties, including, where applicable, the Rightsholder. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by or on behalf of the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the License described in the Order Confirmation, which terms are in any way inconsistent with any Terms set forth in the Order Confirmation, and/or in CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

e) The License described in the Order Confirmation shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such License shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.

Appendix 4: Initial shared CLD

Appendix 4 contains views of the shared CLD at pre- and post-validation stages.

- Figure A4.1: Initial shared CLD (pre-validation).
- Figure A4.2: Updated shared CLD (post-validation).
- Figure A4.3: Refined updated shared CLD.
- Figure A4.4: Final shared CLD (post-validation).

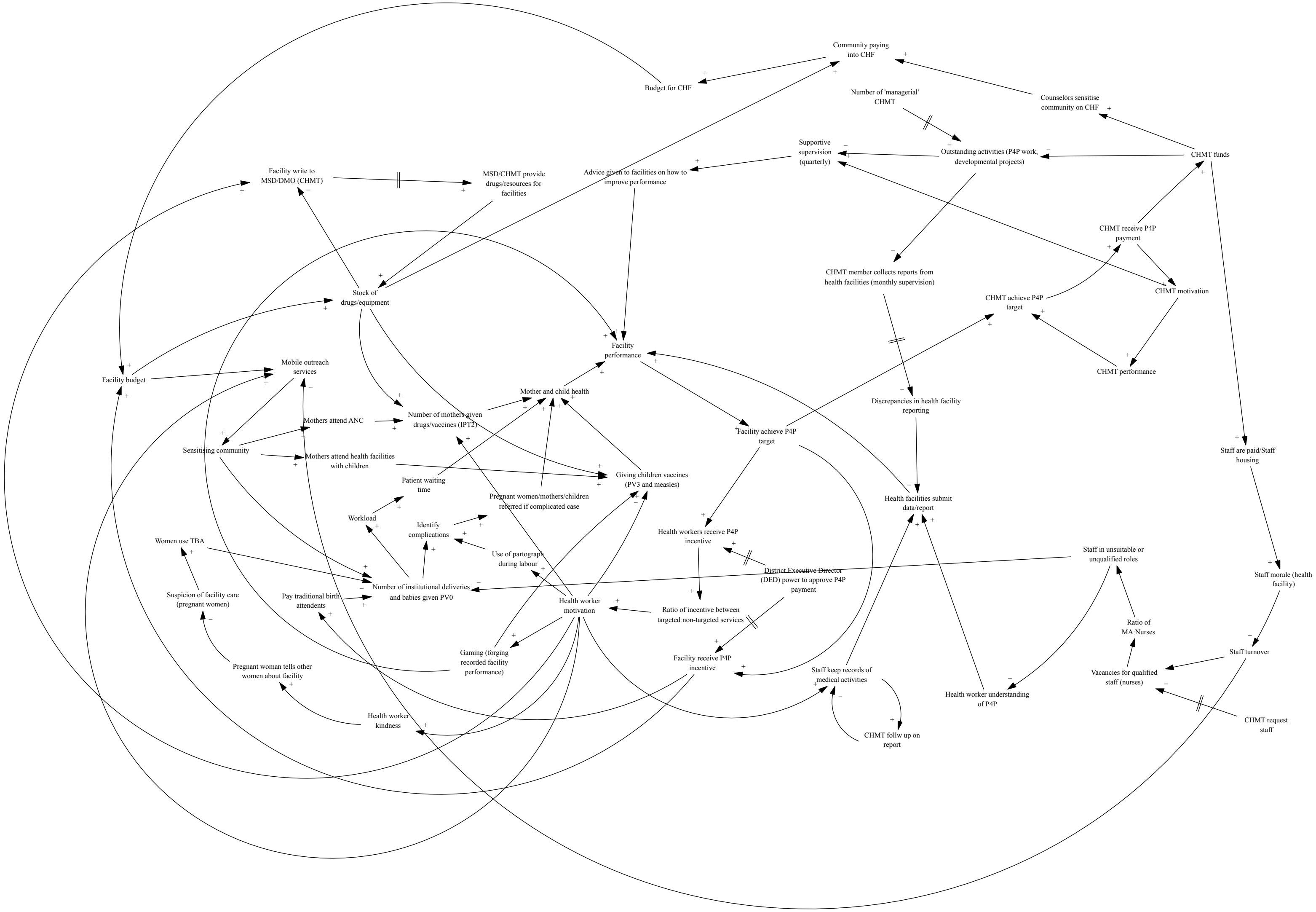


Figure A4.1: Initial shared CLD (pre-validation).

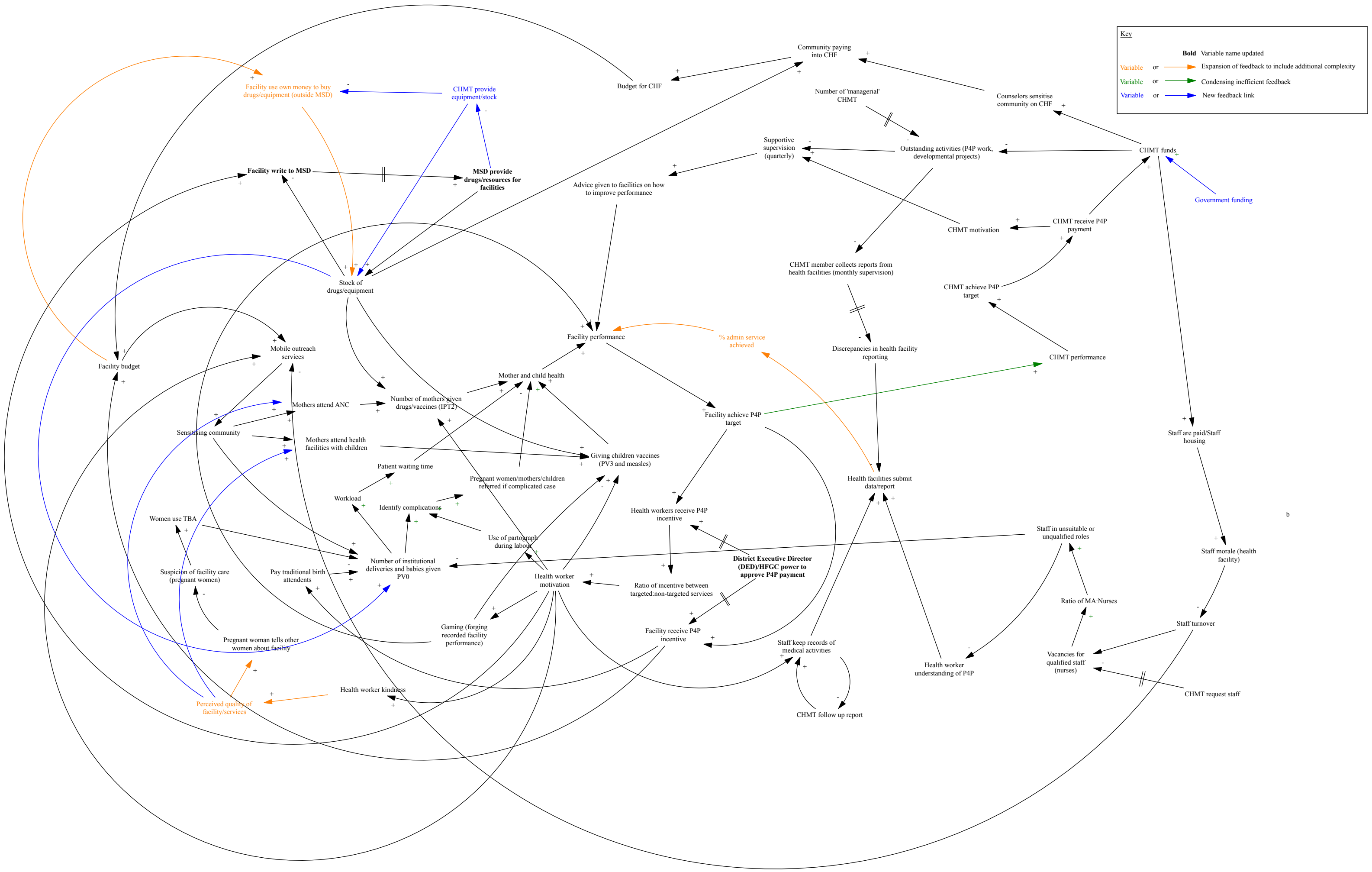


Figure A4.2: Updated shared CLD (post-validation stage 1).

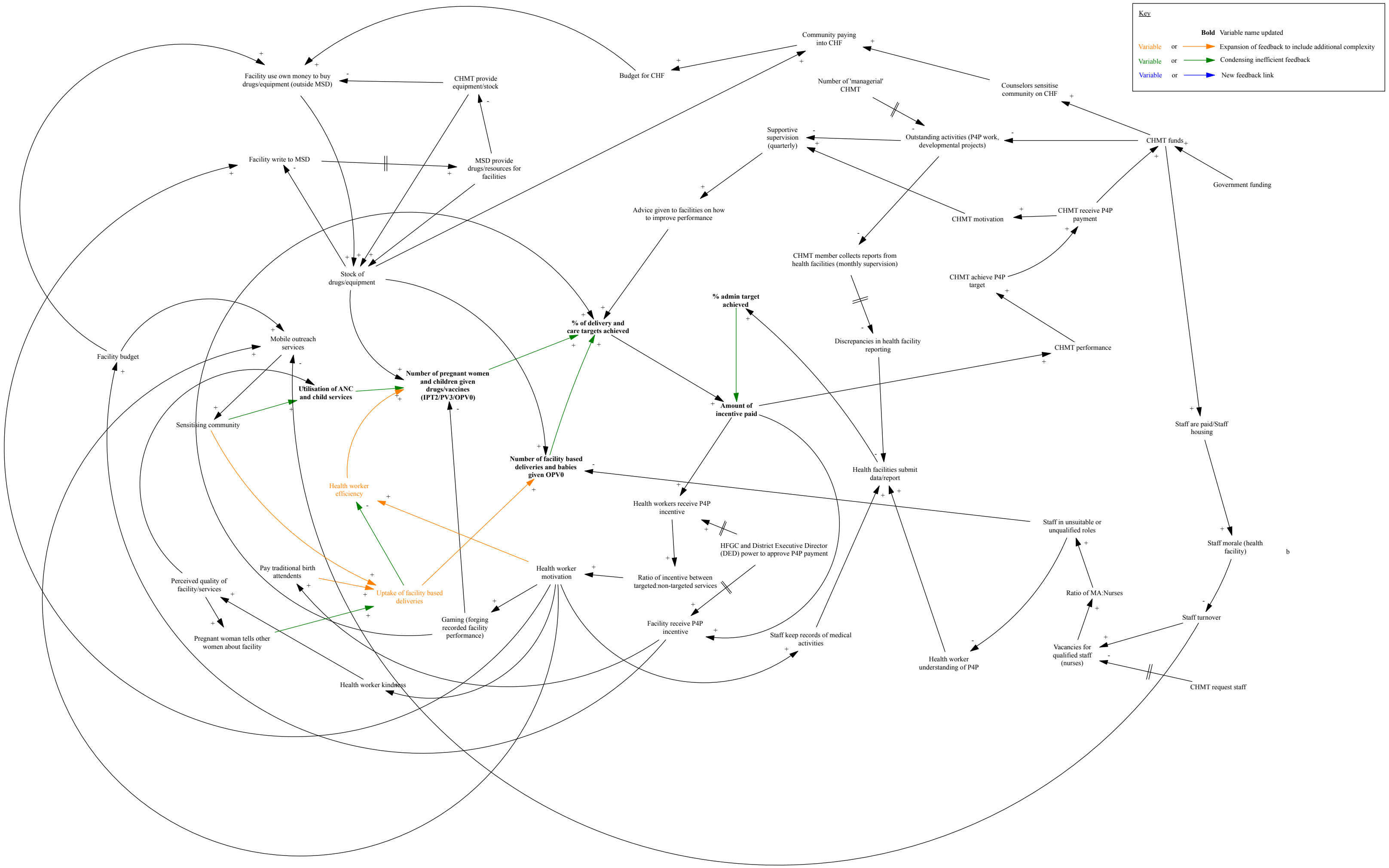


Figure A4.3: Refined updated shared CLD.

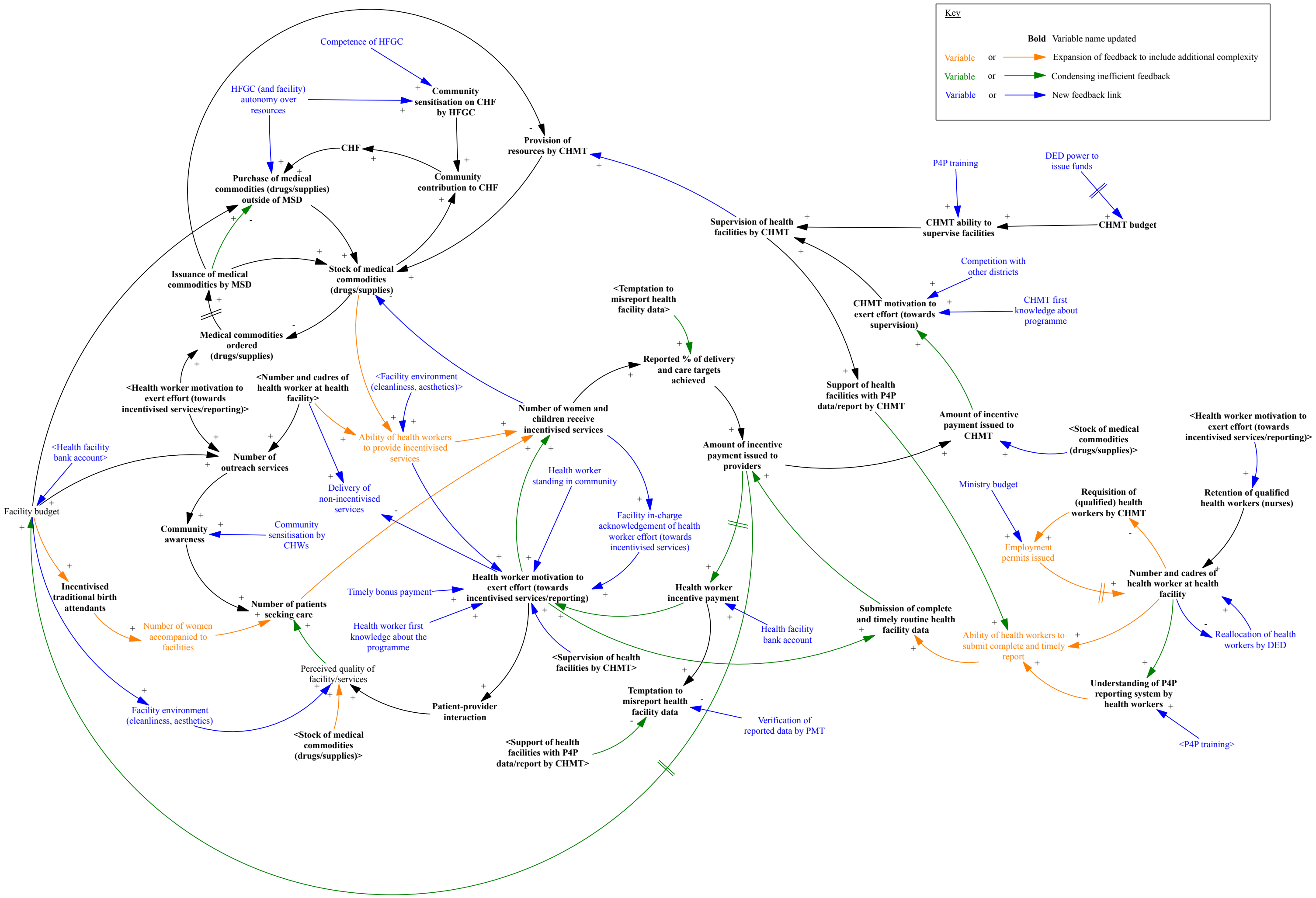


Figure A4.4: Updated shared CLD (post-validation stage 2).

Appendix 5: Publications and conference presentations relevant to this dissertation

Peer reviewed publications (published)

Cassidy R, Borghi J, Semwanga AR et al. 2022. How to do (or not to do)…using causal loop diagrams for health system research in low and middle-income settings. Health Policy Plan, DOI: 10.1093/heapol/czac064.

Singh NS, Kovacs RJ, **Cassidy R** et al. 2021. A realist review to assess for whom, under what conditions and how pay for performance programmes work in low- and middle-income countries. Soc Sci Med 270: 113624.

Cassidy R, Tomoaia-Cotisel A, Semwanga AR et al. 2021. Understanding the maternal and child health system response to payment for performance in Tanzania using a causal loop diagram approach. Soc Sci Med 285: 114277.

Cassidy R, Singh NS, Schiratti P-R et al. 2019. Mathematical modelling for health systems research: a systematic review of system dynamics and agent-based models. BMC Health Serv Res 19: 845.

Publications prepared for submission

Cassidy R, Semwanga AR, Binyaruka, P et al. 2023. Using mathematical modelling to identify the active ingredients in payment for performance programmes. Soc Sci Med.

Conference oral presentations **Joint first author*

Maiba J, Binyaruka P, Singh N, **Cassidy R**, Mziray N, Foss A, Mtei G and Borghi J. Strategic purchasing through direct health facility financing mechanisms in Tanzania: Examining the purchaser-provider relationship, 7th Global Symposium on Health Systems Research, Colombia (November 2022)

Foss A, Mziray N, Binyaruka P, Alibrahim A, Maiba J, **Cassidy R**, Chalabi Z, Borghi J. Using agent-based modelling with data to better understand the impact mechanisms affecting health system performance through a financial incentives scheme for healthcare providers in Tanzania, 7th Global Symposium on Health Systems Research, Colombia (November 2022)

Cassidy R, Tomoaia-Cotisel A, Semwanga A, Binyaruka P, Chalabi Z, Blanchet K, Singh N, Maiba J, Borghi J. Analysing the Health System Response to Payment for Performance in Tanzania Using Systems Thinking Methodology, 2021 iHEA Congress, virtual (July 2021).

Foss A, Mziray N, Binyaruka P, Chalabi Z, **Cassidy R**, Borghi J. Using Agent-Based Modelling As an Ex-Post Evaluation Tool to Better Understand the Impact Mechanisms of a Financial Incentives Scheme for Healthcare Providers, 2021 iHEA Congress, virtual (July 2021).

Ismail S*, **Cassidy R***, Tomoaia-Cotisel A, Singh N, Blanchet K, Borghi J. A critical reflection on approaches to problem articulation for system dynamics modelling in health policy analysis projects, The 38th International Conference of The System Dynamics Society, Norway (virtual) (July 2020).

Conference poster presentations

Cassidy R, Rwashana Semwanga A, Binyaruka P, Maiba J, Blanchet K, Singh NS, Borghi J. Using system dynamics to optimise the design of payment for performance programmes to improve delivery of maternal and child health services in Tanzania, 7th Global Symposium on Health Systems Research, Colombia (November 2022)

Sachingongu N, Chitalu C-C, **Cassidy R**, Semwanga R. A, Borghi J, Generalization of the maternal and child health system response to payment for performance to Zambia: A causal loop diagram approach, 7th Global Symposium on Health Systems Research, Colombia (November 2022)

Mziray N, Maiba J, Binyaruka P, **Cassidy R**, Alibrahim A, Chalabi Z, Borghi J, Foss A. Stakeholders' engagement in Interactive Agent-Based Model Building, 7th Global Symposium on Health Systems Research, Colombia (November 2022)

Cassidy R, Tomoaia-Cotisel A, Semwanga A, Singh N, Binyaruka P, Maiba J, Mziray N, Chalabi Z, Blanchet K, Borghi J. Using system dynamics to optimise health system interventions in low-resource settings, 6th Global Symposium on Health Systems Research, United Arab Emirates (virtual) (November 2020).

Cassidy R, Tomoaia-Cotisel A, Semwanga A, Singh N, Binyaruka P, Mziray N, Maiba J, Chalabi Z, Blanchet K, Borghi J. Analysing the health system response to payment for performance interventions in Tanzania, The 38th International Conference of The System Dynamics Society, Norway (virtual) (July 2020).

Other presentations

Presented CLD and SDM methods, results and reflections as part of the COSMIC Project dissemination workshop for in country stakeholders including representatives from the

Ministries of Health for Tanzania and Zambia and Chief Executive Director for Ifakara Health Institute (November 2022).

Presented CLD methods, results and reflections to the SYSTAC (Systems Thinking Accelerator) support forum as part of a series aimed at introduction and application of systems thinking methods for health research, hosted at the Swiss Tropical and Public Health Institute. (in person) (September 2022).

Presented CLD methods and results, and SDM work in progress to the Thanzi La Onse modelling group, with researchers from University of Malawi, Imperial College London, University College London and University of York, as part of a joint learning session between institutes. (virtual) (June 2022).

Presented CLD methods and results, and SDM work in progress as part of a seminar titled 'Systems thinking approaches applied to health – an overview of the COSMIC project', hosted by International Institute for Applied Systems Analysis, Austria. (virtual) (March 2022).

Presented CLD methods and results, and SDM work in progress as part of 'UNICEF Speaker Series: Understanding the impact mechanisms of a financial incentives scheme for healthcare providers working in maternal and child health in Tanzania'. (virtual) (January 2022).

Presented CLD methods and results at the Complex Adaptive Systems and Global Health Workshop organised by LSHTM, to researchers from the global complexity science community including members from UCL (UK), Graduate Institute of Geneva (Switzerland), Fridtjof Nansen Institute (Norway), Institute for Applied Systems Analysis (Austria), Copernicus Institute of Sustainable Development (Netherlands), Stellenbosch (South Africa), Leeds University (UK), Open University (UK), University of Newcastle (Australia), University of Ghana (Ghana) and University of Queensland (Australia). (virtual) (March 2021).

Presented CLD methods and results at the Complexity Science Workshop organised internally at LSHTM, to complexity science researchers from the Faculty of Public Health and Policy and Faculty of Public Health, Environments and Society. (virtual) (March 2021).

Presented CLD work in progress at an Economic Evaluation meeting held by Centre for Health Economics in London (CHIL), LSHTM. (in person) (February 2020).