

Health Technology Assessment

Volume 25 • Issue 57 • September 2021

ISSN 1366-5278

Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update

Kathryn Skivington, Lynsay Matthews, Sharon Anne Simpson, Peter Craig, Janis Baird, Jane M Blazeby, Kathleen Anne Boyd, Neil Craig, David P French, Emma McIntosh, Mark Petticrew, Jo Rycroft-Malone, Martin White and Laurence Moore



Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update

Kathryn Skivington^{1*} , Lynsay Matthews¹ ,
Sharon Anne Simpson¹ , Peter Craig¹ , Janis Baird² ,
Jane M Blazeby³ , Kathleen Anne Boyd⁴ ,
Neil Craig⁵ , David P French⁶ , Emma McIntosh⁴ ,
Mark Petticrew⁷ , Jo Rycroft-Malone⁸ ,
Martin White⁹  and Laurence Moore^{1*} 

¹Medical Research Council/Chief Scientist Office Social and Public Health Sciences Unit, Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK

²Medical Research Council Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK

³Medical Research Council ConDuCT-II Hub for Trials Methodology Research and Bristol Biomedical Research Centre, University of Bristol, Bristol, UK

⁴Health Economics and Health Technology Assessment Unit, Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK

⁵Public Health Scotland, Glasgow, UK

⁶Manchester Centre for Health Psychology, University of Manchester, Manchester, UK

⁷London School of Hygiene and Tropical Medicine, London, UK

⁸Faculty of Health and Medicine, Lancaster University, Lancaster, UK

⁹Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, UK

*Corresponding authors

Declared competing interests of authors: Additional time on the study was funded by grants from the Medical Research Council (MRC) for Kathryn Skivington (MC_UU_12017/11, MC_UU_00022/3), Lynsay Matthews, Sharon Anne Simpson, Laurence Moore (MC_UU_12017/14, MC_UU_00022/1), and Peter Craig (MC_UU_12017/15, MC_UU_00022/2). Additional time on the study was also funded by grants from the Chief Scientist Office of the Scottish Government Health Directorates for Kathryn Skivington (SPHSU11), Lynsay Matthews, Sharon Anne Simpson and Laurence Moore (SPHSU14), and Peter Craig (SPHSU13 and SPHSU15). Kathryn Skivington and Sharon Anne Simpson were also supported by a MRC Strategic Award (MC_PC_13027). Sharon Anne Simpson was a member of the National Institute for Health Research (NIHR) Clinical Evaluation and Trials Programme Panel (November 2016 to November 2020) and the Chief Scientist Office Health Improvement, Protection and Services Committee (2018 to present) at the time of commissioning and during the project, and was a member of the NIHR Policy Research Programme during the latter parts of the project (November 2019 to present). Janis Baird is a member of the NIHR Public Health Research (PHR) Programme Funding Committee (May 2019 to present). She is also a core member (since 2016) and vice chairperson (since 2018) of the National Institute for Health and Care Excellence (NICE) Public Health Advisory Committee C. Jane M Blazeby is a former member of the NIHR Clinical Trials Unit Standing Advisory Committee (2015–19). Jane M Blazeby is funded in part by the NIHR Biomedical

Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. David P French is a former member of the NIHR PHR programme research funding board (2015–19) and a former member of the MRC–NIHR Methodology Research Panel (2014–18). David P French is funded in part by the NIHR Manchester Biomedical Research Centre (IS-BRC-1215-20007) and NIHR Applied Research Collaboration – Greater Manchester (NIHR200174). Emma McIntosh is a member of the NIHR PHR funding board panel. Jo Rycroft-Malone is programme director and chairperson of the NIHR’s Health Services Delivery Research programme (2014 to present) and is a member of the NIHR Strategy Board (2014 to present). Mark Petticrew is funded in part as director of the NIHR’s Public Health Policy Research Unit, and has funding as part of the MRC-funded TRIPLE C project (grant reference MR/S014632/1) and the NIHR School for Public Health Research at London School of Hygiene & Tropical Medicine. Martin White was funded in part as director of NIHR’s PHR programme (July 2014 to June 2020), which contributed funding to the project via the NIHR/MRC Methodology Research Programme. He is a member of the MRC’s Population Health Sciences Strategic Committee (July 2014 to June 2020). Laurence Moore was a member of the MRC–NIHR Methodology Research Programme Panel (2015–19) and MRC Population Health Sciences Group (2015–20) at the time of commissioning and early stages of the project.

Published September 2021

DOI: 10.3310/hta25570

This report should be referenced as follows:

Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, *et al.* Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update. *Health Technol Assess* 2021;**25**(57).

Health Technology Assessment is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.014

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, the Cochrane Library and Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme or, commissioned/managed through the Methodology research programme (MRP), and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

This issue of the Health Technology Assessment journal series contains a project commissioned by the Medical Research Council's (MRC) Population Health Sciences Group (PHSG). Jointly funded by the MRC and NIHR, the work refreshed the previous version of the Medical Research Council framework for development and evaluation of complex interventions: A comprehensive guidance (2006).

PHSG is responsible for developing the MRC's strategy for research to improve population health. NIHR's mission is to improve the health and wealth of the nation through research. As population level interventions in community and clinical settings become more important, and as science advances and innovates, both funding partners agreed that updating the existing framework was timely and needed.

The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. This report has been published following a shortened production process and, therefore, did not undergo the usual number of proof stages and opportunities for correction. The Health Technology Assessment (HTA) programme editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research. The views and opinions expressed in this publication are those of the authors and do not necessarily reflect those of the MRC, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health and Social Care.

Copyright © 2021 Skivington *et al.* This work was produced by Skivington *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

NIHR Journals Library Editor-in-Chief

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Professor of Digital Health Care, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Senior Scientific Adviser (Evidence Use), Wessex Institute, University of Southampton, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Emeritus Professor of Wellbeing Research, University of Winchester, UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update

Kathryn Skivington^{1*}, Lynsay Matthews¹, Sharon Anne Simpson¹, Peter Craig¹, Janis Baird², Jane M Blazeby³, Kathleen Anne Boyd⁴, Neil Craig⁵, David P French⁶, Emma McIntosh⁴, Mark Petticrew⁷, Jo Rycroft-Malone⁸, Martin White⁹ and Laurence Moore^{1*}

¹Medical Research Council/Chief Scientist Office Social and Public Health Sciences Unit, Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK

²Medical Research Council Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK

³Medical Research Council ConDuCT-II Hub for Trials Methodology Research and Bristol Biomedical Research Centre, University of Bristol, Bristol, UK

⁴Health Economics and Health Technology Assessment Unit, Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK

⁵Public Health Scotland, Glasgow, UK

⁶Manchester Centre for Health Psychology, University of Manchester, Manchester, UK

⁷London School of Hygiene and Tropical Medicine, London, UK

⁸Faculty of Health and Medicine, Lancaster University, Lancaster, UK

⁹Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, UK

*Corresponding authors Kathryn.skivington@glasgow.ac.uk and Laurence.moore@glasgow.ac.uk

Background: The Medical Research Council published the second edition of its framework in 2006 on developing and evaluating complex interventions. Since then, there have been considerable developments in the field of complex intervention research. The objective of this project was to update the framework in the light of these developments. The framework aims to help research teams prioritise research questions and design, and conduct research with an appropriate choice of methods, rather than to provide detailed guidance on the use of specific methods.

Methods: There were four stages to the update: (1) gap analysis to identify developments in the methods and practice since the previous framework was published; (2) an expert workshop of 36 participants to discuss the topics identified in the gap analysis; (3) an open consultation process to seek comments on a first draft of the new framework; and (4) findings from the previous stages were used to redraft the framework, and final expert review was obtained. The process was overseen by a Scientific Advisory Group representing the range of relevant National Institute for Health Research and Medical Research Council research investments.

Results: Key changes to the previous framework include (1) an updated definition of complex interventions, highlighting the dynamic relationship between the intervention and its context; (2) an emphasis on the use of diverse research perspectives: efficacy, effectiveness, theory-based and systems perspectives; (3) a focus on the usefulness of evidence as the basis for determining research perspective and questions; (4) an increased focus on interventions developed outside research teams, for example changes in policy or health services delivery; and (5) the identification of six 'core elements' that should

guide all phases of complex intervention research: consider context; develop, refine and test programme theory; engage stakeholders; identify key uncertainties; refine the intervention; and economic considerations. We divide the research process into four phases: development, feasibility, evaluation and implementation. For each phase we provide a concise summary of recent developments, key points to address and signposts to further reading. We also present case studies to illustrate the points being made throughout.

Limitations: The framework aims to help research teams prioritise research questions and design and conduct research with an appropriate choice of methods, rather than to provide detailed guidance on the use of specific methods. In many of the areas of innovation that we highlight, such as the use of systems approaches, there are still only a few practical examples. We refer to more specific and detailed guidance where available and note where promising approaches require further development.

Conclusions: This new framework incorporates developments in complex intervention research published since the previous edition was written in 2006. As well as taking account of established practice and recent refinements, we draw attention to new approaches and place greater emphasis on economic considerations in complex intervention research. We have introduced a new emphasis on the importance of context and the value of understanding interventions as ‘events in systems’ that produce effects through interactions with features of the contexts in which they are implemented. The framework adopts a pluralist approach, encouraging researchers and research funders to adopt diverse research perspectives and to select research questions and methods pragmatically, with the aim of providing evidence that is useful to decision-makers.

Future work: We call for further work to develop relevant methods and provide examples in practice. The use of this framework should be monitored and the move should be made to a more fluid resource in the future, for example a web-based format that can be frequently updated to incorporate new material and links to emerging resources.

Funding: This project was jointly funded by the Medical Research Council (MRC) and the National Institute for Health Research (Department of Health and Social Care 73514).

Contents

List of tables	xi
List of figures	xiii
List of boxes	xv
List of key points	xvii
List of signposts for further reading	xix
Glossary	xxi
List of abbreviations	xxvii
Plain English summary	xxix
Scientific summary	xxxix
Chapter 1 Development of the framework	1
Introduction	1
Methods and results	2
<i>Stage 1: gap analysis</i>	2
<i>Stage 2: expert workshop</i>	5
<i>Stage 3: open consultation</i>	9
<i>Stage 4: writing the updated framework</i>	13
Conclusion	13
<i>Recommendations</i>	15
Chapter 2 The new framework	17
Introduction	17
<i>Aims of the framework</i>	17
<i>Structure of the framework</i>	17
<i>How to use the framework</i>	17
What is a 'complex intervention'?	18
<i>Sources of complexity</i>	18
<i>Complexity and the research perspective</i>	21
Framework for the main phases and core elements of complex intervention research	25
<i>Overview</i>	25
<i>Core elements</i>	27
Phases of research	39
<i>Developing or identifying the intervention</i>	39
<i>Feasibility</i>	45
<i>Evaluation</i>	50
<i>Implementation</i>	61
Case studies	66
<i>Case study 1: travel through the main phases of complex intervention research</i>	67
<i>Case study 2: considering context</i>	67
<i>Case study 3: developing, refining and testing programme theory</i>	69

CONTENTS

<i>Case study 4: example of a complex logic model</i>	69
<i>Case study 5: example of a dark logic model</i>	71
<i>Case study 6: example of a realist matrix-presenting programme theory</i>	71
<i>Case study 7: system map to develop programme theory</i>	73
<i>Case study 8: engaging stakeholders (service users)</i>	75
<i>Case study 9: refining the intervention</i>	75
<i>Case study 10: economic considerations</i>	76
<i>Case study 11: intervention development: example systems map</i>	77
<i>Case study 12: feasibility trial</i>	77
<i>Case study 13: feasibility, evaluability assessment</i>	77
<i>Case study 14: evaluation (effectiveness perspective: cluster randomised controlled trial with theory-informed process evaluation)</i>	79
<i>Case study 15: evaluation (efficacy)</i>	80
<i>Case study 16: evaluation (effectiveness perspective: individual randomised controlled trial)</i>	80
<i>Case study 17: evaluation (effectiveness perspective: stepped-wedged cluster randomised design)</i>	81
<i>Case study 18: evaluation (effectiveness perspective: synthetic control)</i>	81
<i>Case study 19: development and evaluation – effectiveness perspective (drawing on a systems perspective for intervention development and evaluation)</i>	82
<i>Case study 20: evaluation (theory-based perspective, alongside effectiveness)</i>	82
<i>Case study 21: evaluation (theory-based perspective with realist approach)</i>	83
<i>Case study 22: evaluation (theory-based approach)</i>	84
<i>Case study 23: evaluation (systems perspective)</i>	84
<i>Case study 24: evaluation (systems perspective with process evaluation)</i>	85
<i>Case study 25: evaluation (systems perspective with qualitative comparative analysis)</i>	86
<i>Case study 26: intervention development – discrete choice experiments</i>	86
<i>Case study 27: implementation in all phases</i>	87
Acknowledgements	89
References	93
Appendix 1 Oversight of the project	119
Appendix 2 Updating the Medical Research Council guidance on complex interventions: an outline proposal for developing an integrated body of guidance	121
Appendix 3 Advisors/consultees	125
Appendix 4 Online consultation questionnaire	127
Appendix 5 Approaching complexity in evidence synthesis	129
Appendix 6 Checklist for developing and evaluating complex interventions	131

List of tables

TABLE 1	Summary of the key points from the expert workshop in 2018	7
TABLE 2	Properties of complex adaptive systems, with examples	20
TABLE 3	Framework of actions for intervention development	41
TABLE 4	Core elements for developing or identifying the intervention	43
TABLE 5	Core elements for the feasibility phase	48
TABLE 6	Research perspectives and evaluation	51
TABLE 7	Core elements for the evaluation phase	59
TABLE 8	Core elements of the implementation phase	64
TABLE 9	Case studies referred to in the report	66
TABLE 10	Matrix of context, mechanism and outcomes for the supply component of the community health worker programme	73

List of figures

FIGURE 1 Key elements for developing and evaluating complex interventions (consultation version)	11
FIGURE 2 Framework for addressing complexity within evaluation (consultation version)	11
FIGURE 3 Main phases and core elements of complex intervention research	26
FIGURE 4 Two pathways of evidence generation	40
FIGURE 5 Case study demonstrating how the FFIT programme moved through the research process	68
FIGURE 6 Logic model of a complex intervention to improve payment of salaries to health workers	70
FIGURE 7 Dark logic model for Communities in Charge of Alcohol	72
FIGURE 8 Conceptual system map for the UK SDIL	74
FIGURE 9 Physical activity system map with interventions	78

List of boxes

BOX 1 Considering context: 'Lucky Iron Fish' example	19
BOX 2 Continuum of research perspectives	21
BOX 3 Different approaches to taking a theory-based perspective	23
BOX 4 Examples of complex intervention studies in smoking prevention research taking different research perspectives	24
BOX 5 Risk of underestimating the importance of context: example	27
BOX 6 When is evidence sufficient?	34
BOX 7 Frameworks for economic evaluation	58

List of key points

KEY POINTS 1	Intervention complexity and research perspectives	25
KEY POINTS 2	Context	28
KEY POINTS 3	Programme theory	30
KEY POINTS 4	Stakeholders	33
KEY POINTS 5	Uncertainties	36
KEY POINTS 6	Intervention refinement	37
KEY POINTS 7	Economic considerations	39
KEY POINTS 8	Developing or identifying an intervention	44
KEY POINTS 9	Feasibility	49
KEY POINTS 10	Evaluation	60
KEY POINTS 11	Implementation	65

List of signposts for further reading

SIGNPOSTS TO FURTHER READING 1	Systems thinking in health research	20
SIGNPOSTS TO FURTHER READING 2	Context	28
SIGNPOSTS TO FURTHER READING 3	Programme theory	31
SIGNPOSTS TO FURTHER READING 4	Stakeholders	33
SIGNPOSTS TO FURTHER READING 5	Uncertainties	36
SIGNPOSTS TO FURTHER READING 6	Intervention refinement	37
SIGNPOSTS TO FURTHER READING 7	Economic considerations	39
SIGNPOSTS TO FURTHER READING 8	Developing or identifying an intervention	45
SIGNPOSTS TO FURTHER READING 9	Feasibility	50
SIGNPOSTS TO FURTHER READING 10	Evaluation	60
SIGNPOSTS TO FURTHER READING 11	Implementation	65
SIGNPOSTS TO FURTHER READING 12	Complexity in systematic reviews	129

Glossary

Adaptation Where an existing intervention is modified or tailored to adapt it to a new population, setting or context (note that this is distinct from ‘adaptation’ as a property of complex adaptive systems: a change of system behaviour in response to an intervention).

Agent ‘An agent is any entity whose behaviour is autonomous yet complying with implicit or explicit behavioural rules’ (Signal LN, Walton MD, Ni Mhurchu C, Maddison R, Bowers SG, Carter KN, *et al.* Tackling ‘wicked’ health promotion problems: a New Zealand case study. *Health Promot Int* 2012;**28**:84–94).

Complex intervention An intervention is conceived to be complex either (1) because of the characteristics of the intervention itself, for example multiple components or mechanisms of change, and/or (2) because how the intervention generates outcomes is dependent on exogenous factors, including the characteristics of recipients, and/or the context or system within which it is implemented.

Complexity-informed research Research that attends to complexity rather than focusing on controlling for it. This involves being aware of system properties and how they may play a role in how the intervention affects change.

Context ‘Any feature of the circumstances in which an intervention is conceived, developed, implemented and evaluated’, for example social, political, economic and geographical context {Craig P, Di Ruggiero E, Frolich KL, Mykhalovskiy E, White M, on behalf of the Canadian Institutes of Health Research (CIHR)–National Institute for Health Research (NIHR) Context Guidance Authors Group. *Taking Account of Context in Population Health Intervention Research: Guidance for Producers, Users and Funders of Research*. Southampton: National Institute for Health Research; 2018 [contains information licensed under the Non-Commercial Government Licence v2.0]}.

Cost–benefit analysis A form of economic evaluation that expresses all gains and costs in monetary terms, allowing a judgement to be made about whether or not or to what extent an objective should be pursued.

Cost–consequences analysis A form of economic evaluation in which the whole array of outputs is presented alongside the costs, without any attempt to aggregate the outputs.

Cost-effectiveness analysis A form of economic evaluation that compares the relative costs and outcomes (effects) of different courses of action. Outcomes are measured in natural units.

Cost–utility analysis A form of cost-effectiveness analysis in which benefits are measured in terms of a utility measure, such as the quality-adjusted life-year.

Critical realism A philosophy of science associated with realist methods of evaluation. Critical realism asserts that causes can be understood in terms of mechanisms, rather than just inferred from sequences of events [Archer M, Decoteau C, Gorski P, Little D, Porpora D, Rutzou T, *et al.* What is critical realism? *Perspectives: American Sociological Association Theory* 2016;**38**:4–9. URL: www.asatheory.org/uploads/4/3/3/7/43371669/perspectives_38_2_fall_2016_final.pdf (accessed 1 March 2021); Bhaskar R. *A Realist Theory of Science*. New York, NY: Routledge; 2013].

Decision-makers The people who use our research evidence to make decisions on what and how interventions are implemented in practice. The term ‘research end-user’ is sometimes used in other documents to mean the same thing. Examples of decision-makers that are relevant here include policy-makers (local and national), health-care professionals and service managers.

Difference-in-differences A statistical technique using observational data to mimic an experimental research design, by studying the differential effect of a treatment on a 'treatment group' versus a 'control group'.

Disability-adjusted life-year An outcome measure that adjusts the change in life expectancy produced by an intervention according to changes in the level of disability (see also *Quality-adjusted life-year*).

Economic evaluation Research to determine whether or not an intervention is an efficient use of resources; 'the comparative analysis of alternative courses of action in terms of both their costs and consequences' (Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press; 2005).

Effectiveness The degree to which the intervention produces the intended effects in routine practice rather than in tightly controlled, experimental circumstances.

Efficacy Efficacy studies focus on whether or not an intervention can produce an intended outcome in an idealised experimental setting.

Efficiency Maximising the benefit to any resource expenditure or minimising the cost of any achieved benefit.

Equality Equal distribution of resources or benefits among different individuals or groups.

Equity Fair distribution of resources or benefits among different individuals or groups. Fairness is closely related to whether or not departures from equality are avoidable and how the groups are defined.

Evaluability assessment A systematic approach to planning evaluation projects.

Evaluation A study carried out to enable judgements to be made about the value of an intervention. Evaluation studies usually seek to determine whether or not an intervention 'works' in the sense of achieving its intended outcome. Other aims include identifying unanticipated or adverse impacts, theorising how the intervention works, exploring how it interacts with the context in which it is implemented, exploring how it contributes to system change and exploring how evaluation evidence can be used to support real-world decision-making.

Evidence synthesis Comprehensive, systematic and transparent interpretation of a combination of multiple sources of evidence.

External validity The extent to which the results of a study can be applied to a wider population of interest or to some other context.

Feasibility study Research carried out before a main study to answer the question 'Can this study be done?' (Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, *et al.* CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ* 2016;**355**:i5239).

Fidelity The degree to which an intervention or programme is delivered as intended.

Health economics The study of how scarce resources are allocated among alternative uses for the care of sickness and the promotion, maintenance and improvement of health, including the study of how health-care and health-related services, their costs and benefits, and health itself are distributed among individuals and groups in society.

Implementation Delivery of an intervention as part of routine practice. Implementation research refers to 'deliberate efforts to increase the impact of successfully tested health innovations so as to benefit more people and to foster policy and programme development on a lasting basis' [Reproduced from World Health Organization (WHO). *Nine Steps for Developing a Scaling-up Strategy*. Geneva: WHO; 2010].

Incremental cost-effectiveness ratio A relative measure of cost-effectiveness. This is obtained by dividing the difference between the costs of the two interventions by the difference in the outcomes (i.e. the extra cost per extra unit of effect).

Internal validity The degree to which a study provides unbiased estimates of the causal relationships under investigation.

Intervention An 'action or programme that aims to bring about identifiable outcomes' (Rychetnik L, Hawe P, Waters E, Barratt A, Frommer M. A glossary for evidence based public health. *J Epidemiol Comm Health* 2004;58:538–45). This term is used for everything from medical treatment to changes in policy; it could be something developed and implemented by the research team, for example a surgical procedure, or something beyond the researchers' control, for example a change in welfare policy, such as the implementation of universal credit. In some interventions, the main aim is not necessarily to bring about health outcomes, but the intervention may indirectly lead to changes in health. Interventions can be thought of as interruptions in or changes to a system.

Intervention development The term 'development' is used here for the whole process of designing and planning an intervention from initial conception through to feasibility, pilot or evaluation study.

Key uncertainties Identifying the key uncertainties that exist given what is already known and what the programme theory, research team and stakeholders identify as being the most important to find out. These judgements inform the framing of research questions, which in turn govern the choice of research perspective.

Logic model A method of representing some elements of the programme theory, typically presented in a linear pathway. Simple logic models may include only observable inputs, outputs and intended outcomes; more detailed logic (causal) models will include other components of the programme theory including mechanisms by which an intervention is thought to generate outcomes, and the contextual dependencies and other preconditions on which that is based.

Mechanism A causal link between an exposure (e.g. to some feature of an intervention) and an outcome.

Mixed-methods evaluation The use and integration of quantitative and qualitative approaches to collect, analyse and interpret data to answer evaluation questions.

Natural experiment 'Events, interventions or policies which are not under the control of researchers, but which are amenable to research which uses the variation in exposure that they generate to analyse their impact' (Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, *et al.* *Using Natural Experiments to Evaluate Population Health Interventions: Guidance for Producers and Users of Evidence*. London: Medical Research Council; 2012) (reproduced with permission). A natural experimental study is the methodological approach to evaluating the impact of a natural experiment on health or other outcomes.

Process evaluation 'A study which aims to understand the functioning of an intervention, by examining implementation, mechanisms of impact, and contextual factors. Process evaluation is complementary to, but not a substitute for, high quality outcome evaluation' (Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, *et al.* *Process Evaluation of Complex Interventions. Medical Research Council Guidance*. London: Medical Research Council Population Health Science Research Network; 2014) (reproduced with permission).

Programme theory Describes how an intervention is expected to lead to its effects and under what conditions. It articulates (1) the key components of the intervention and how they interact; (2) the mechanisms of the intervention; (3) the features of the context that are expected to influence those mechanisms; and (4) how those mechanisms may influence the context (Funnell SC, Rogers PJ. *Purposeful Program Theory. Effective Use of Theories of Change and Logic Models*. San Francisco, CA: Jossey-Bass; 2011). In the economic literature, conceptual modelling is often used to refer to the process of developing programme theory.

Quality-adjusted life-years An outcome incorporating both quality of life and life expectancy. Calculated by adjusting the estimated number of years of life an individual is expected to gain from an intervention for the expected quality of life in those years. The quality-of-life score will range between 0 for death and 1 for perfect health, with negative scores being allowed for states considered worse than death.

Realist evaluation An approach to evaluation in which researchers develop mid-range theories to take account of how context and mechanisms of interventions interact to produce outcomes, that is causal mechanisms of an intervention are context specific. Realist evaluation sets out to answer the question: 'what works in which circumstances and for whom?' (Pawson R, Tilley N. *Realistic Evaluation*. London: SAGE Publications Ltd; 1997).

Refinement The process of 'fine tuning' or making changes to the intervention once a preliminary version (prototype) has been developed (O'Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, *et al*. Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. *Pilot Feasibility Stud* 2019;5:41).

Resources Things that are used up in the production of output. Money gives a command over resources but is not a resource per se.

Sensitivity analysis A process through which the robustness of an economic or statistical model is assessed by examining the changes in results of the analysis when key variables are varied over a specified range. Sensitivity analysis is used to determine how far conclusions rely on particular assumptions.

Soft system methodologies A set of methods used to engage stakeholders and create conceptual models about issues or problems to consider 'real-world' solutions.

Stakeholders Those who are targeted by the intervention or policy, those who are involved in its development or delivery or, more broadly, those whose personal or professional interests are affected (i.e. who have a stake in the topic). This includes patients and members of the public as well as those linked in a professional capacity.

System A set of things that are interconnected in such a way that they produce their own pattern of behaviour over time (Meadows DH. *Thinking in Systems: A Primer*. White River Junction, VT: Chelsea Green Publishing; 2008).

System boundary A boundary used to define the set of elements that comprise a system of interest to make research on that system tractable. It is not always feasible, necessary or desirable to consider the whole system in evaluative research. Nonetheless, it is important to be aware of the wider system and to justify the choice of boundaries. Boundaries are dependent on the problem and research question, and can be set in different ways, for example based on geography or concept.

System map A conceptual visual representation of the components within a system. Creating a system map should involve discussion between a multidisciplinary team to identify the components in the system and how they interact.

Systems thinking A holistic way of thinking about complex systems, focusing on the interactions between entities that comprise a system and between those entities and their environment, rather than assuming that a system can be understood by breaking it down into its individual entities and studying each part separately.

Theory Beliefs or assumptions underlying action and explanations of the phenomena of interest. A distinction can be drawn between grand (high-level generalisations that can be applied across domains), mid-range (limited to a specific area) and programme theory (specific to an individual intervention).

Theory of the problem Clarification of the problem with stakeholders using the existing research evidence. Once the problem is defined it should be established how the problem is socially and spatially distributed, including who is most/least likely to benefit from an intervention. In addition, to understand the immediate and underlying influences that give rise to the specified problem (Wight D, Wimbush E, Jepson R, Doi L. Six steps in quality intervention development (6SQUID). *J Epidemiol Comm Health* 2016;**70**:520–5). This is relevant not only in intervention development, but in evaluation of natural experiments, for which it is important to understand the problem that the policy change/intervention was attempting to alter.

Utility A measure of the 'satisfaction' (benefit) obtained from consuming goods and services.

Value-of-information analysis A quantitative analysis that puts a monetary value on further research based on the expected gain from reducing uncertainty surrounding the current cost-effectiveness decision through additional research activities.

List of abbreviations

ABM	agent-based modelling	NICE	National Institute for Health and Care Excellence
ASSIST	A Stop Smoking in Schools Trial	NIHR	National Institute for Health Research
CBA	cost-benefit analysis	NIM	New Orleans Intervention Model
CCA	cost-consequence analysis	PARADES	Psychoeducation, Anxiety, Relapse, Advance Directive Evaluation and Suicidality
CEA	cost-effectiveness analysis	PHSG	Population Health Sciences Group
CEEBIT	Continuous Evaluation of Evolving Behavioural Intervention Technologies	PPI	patient and public involvement
COMET	Core Outcome Measures in Effectiveness Trials	PRECIS	PRagmatic Explanatory Continuum Indicator Summary
CONSORT	Consolidated Standards of Reporting Trials	QALY	quality-adjusted life-year
CUA	cost-utility analysis	QCA	qualitative comparative analysis
DELTA2	Difference ELicitation in TriAls	RCT	randomised controlled trial
FFIT	Football Fans in Training	SAG	Scientific Advisory Group
GP	general practitioner	SDIL	Soft Drinks Industry Levy
HIV	human immunodeficiency virus	SMART	Sequential Multiple Assignment Randomised Trial
ICAMO	intervention-context-actor-mechanism-outcome	VOI	value of information
MCDA	multi-criteria decision analysis	WHO	World Health Organization
MRC	Medical Research Council		
MRP	Methodology Research Programme		

Plain English summary

Interventions are actions taken to make a change, for example heart surgery, an exercise programme or a smoking ban in public. Interventions are described as complex if they comprise several stages or parts or if the context in which they are delivered is complex.

A framework on how to develop and evaluate complex interventions was last published by the Medical Research Council in 2006 (Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. *Developing and Evaluating Complex Interventions*. London: Medical Research Council; 2006). This document describes how the framework has been updated to include advances in research methods and thinking and presents the new framework document.

The updating process had four stages: (1) review of the literature to identify areas requiring update; (2) workshop of experts to discuss topics to update; (3) open consultation on a draft of the framework; and (4) writing the framework.

The updated framework divides the research process into four phases: development, feasibility, evaluation and implementation. Key updates include:

1. the definition of a complex intervention was changed to include both the content of the intervention and the context in which it is conducted
2. addition of systems thinking methods: an approach that considers the broader system an intervention sits within
3. more emphasis on interventions that are not developed by researchers (e.g. policy changes and health services delivery)
4. emphasis on the usefulness of evidence as the key goal of complex intervention research
5. identification of six elements to be addressed throughout the research process: context; theory refinement and testing; stakeholder involvement; identification of key uncertainties; intervention refinement; and economic considerations.

The updated framework is intended to help those involved in funding and designing research to consider a range of approaches, questions and methods and to choose the most appropriate. It also aims to help researchers conduct and report research that is as useful as possible to users of research.

Scientific summary

Background and rationale

In 2006, the Medical Research Council published guidance for developing and evaluating complex interventions (Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. *Developing and Evaluating Complex Interventions*. London: Medical Research Council; 2006), building on a framework that was published in 2000 (Medical Research Council. *A Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health*. London: Medical Research Council; 2000). The aim was to help researchers and research funders recognise and adopt appropriate methods to improve the quality of research to develop and evaluate complex interventions and, thereby, maximise its impact.

Since the 2006 edition of the guidance was published, there have been considerable developments in the field of complex intervention research. In some areas, the accumulation of experience and knowledge in the application of approaches and methods has led to the publication of detailed, focused guidance on the conduct and/or reporting of research. In other areas, new challenges in complex intervention research have been identified and the reliance on traditional approaches and methods promoted in previous guidance has been challenged.

There was, therefore, a need to update the 2006 guidance and, given the greater amount of detailed guidance now available and the need for inclusion of a wider range of approaches, the aim was to provide a less prescriptive and more flexible guide. The new framework aims to improve the design and conduct of complex intervention research to increase its utility, efficiency and impact. Consistent with the principles of increasing the value of research and minimising research waste, the framework (1) emphasises the use of diverse research perspectives and the inclusion of research users, clinicians, patients and the public in research teams, and (2) aims to help research teams prioritise research questions and choose and implement appropriate methods.

Methods

The framework was developed in four stages.

Stage 1

The 'gap analysis' aimed to identify and summarise aspects of the previous guidance that required updating. An initial list of topics to explore further was developed by the project team and the Scientific Advisory Group. A brief scoping review of existing literature was undertaken for each topic. This was then used to develop questions to explore in more detail with experts in stage 2. This initial stage set the agenda for the update.

Stage 2

The expert workshop was facilitated to obtain views and record discussions on topics that should be newly covered or updated. Participants were identified by the project team and Scientific Advisory Group, and we aimed to have at least two experts for each of the topics identified in the gap analysis. These topics were then the starting point for the facilitated breakout discussions. Thirty-six experts from many disciplines relevant to complex intervention development and evaluation, such as public health, health services research, clinical research and health economics, participated in the workshop. Discussions were audio-recorded and these recordings were reviewed and thematically organised and used to make decisions on the content of the first draft of the updated framework.

Stage 3

The open consultation (April 2019) aimed to collect views on the first full draft of the framework document. Consultees were targeted via e-mail, but also through public social media channels. Consultation responses ($n = 52$) were organised thematically and the project team, along with the Scientific Advisory Group, decided how to revise the framework.

Stage 4

The project team, in discussion with the Scientific Advisory Group and with the support of all authors, drafted the final document. We then sought further review from the Medical Research Council and National Institute for Health Research (NIHR) Methodology Research Programme Advisory Group and Population Health Sciences Group, plus another eight experts in May/June 2020, and made final edits accordingly.

Results

The findings from each stage of the process described above supported decisions taken on the content of the new framework.

Key changes from the previous guidance:

1. We have extended the definition of a complex intervention to include complexity that arises from the interaction between an intervention and the context in which it is implemented, as well as complexity arising from the structure of the intervention.
2. We draw on recent developments in thinking and methods to suggest ways in which researchers can take account of complexity, and we encourage consideration and use of diverse research perspectives and the pragmatic and pluralist choice of research questions and methods. We define four overlapping research perspectives that can be drawn on to answer different types of research questions:
 - i. Efficacy perspective – to what extent does the intervention produce the intended outcome(s) in experimental or ideal settings?
 - ii. Effectiveness perspective – to what extent does the intervention produce the intended outcome(s) in real-world settings?
 - iii. Theory-based perspective – what works in which circumstances and how?
 - iv. Systems perspective – how do the system and intervention adapt to one another?

We encourage consideration of the different research perspectives, particularly how they could be used to conceptualise and evaluate the intervention, the intervention's place in the wider systems in which it is implemented and the interdependence between the two over time. The aim of this is to encourage appropriate consideration of the potential sources of complexity.

3. We identify six core elements that should be addressed throughout the research process:
 - i. Context – we have introduced a new emphasis on the importance of context and the value of understanding interventions as 'events in systems' that produce effects through interactions with features of the contexts in which they are implemented.
 - ii. Refinement and testing of programme theory – programme theory describes how an intervention is expected to lead to its effects and under what conditions. The programme theory should be tested and refined at all stages and used to guide the identification of uncertainties, research questions and evaluation outcomes. The development of programme theory is a collaborative and iterative process.
 - iii. Stakeholders – genuine engagement of appropriate stakeholders throughout the research process and working with them as partners to jointly develop, identify or prioritise interventions for research and agree programme theories, research perspectives, key uncertainties and research questions.

- iv. Uncertainties – identifying the key uncertainties that exist given what is already known and what the programme theory, research team and stakeholders identify as being the most important. These judgements inform the framing and prioritisation of research questions and should yield more useful evidence for decision-making.
- v. Intervention refinement – the process of ‘fine tuning’ or making changes to the intervention once a preliminary version (prototype) has been developed. We suggest that ongoing refinement of an intervention, consistent with the programme theory, can improve the potential implementability of the intervention.
- vi. Economic considerations – the comparative resource and outcome consequences of the interventions for those people and organisations affected. More emphasis is placed on these than in the previous guidance because we see them as a key element at all phases of a research project, rather than simply a set of methods for assessing cost-effectiveness.

We divide the research process into four phases: development, feasibility, evaluation and implementation. For each, we provide a concise summary of recent developments, key points to address and signposts to further reading.

1. Development

- i. Developing an intervention: we draw heavily on the comprehensive guidance provided by the INDEX study (O’Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, *et al.* Guidance on how to develop complex interventions to improve health and healthcare. *BMJ Open* 2019;**9**:e029954). The steps are adequate planning; involving stakeholders; bringing together an intervention development team with relevant expertise; review related existing literature and theories; develop, test and refine programme theory; draw on primary data; take efforts to understand the context of the intervention; consider future implementation; refine the intervention as appropriate; and end the development phase, including writing up the process. In addition to these steps, we have added four sets of issues that may be useful in the development of an intervention: adaptation of interventions to new contexts, consideration of wider system factors, modelling or simulation of the intervention and consideration of economic factors.
- ii. Identifying an intervention: researchers are not always involved in the design phase of intervention development, for example where interventions are designed and driven in policy settings. This presents slightly different challenges for this phase of research. For example, it may not be entirely clear what the theoretical basis for the intervention is, or there may not be explicit programme theory. At this phase of research, the key step is to develop programme theory, which will be an important basis for considerations around the evaluation design.

2. Feasibility

This phase is designed to explore the uncertainties that have been identified at the development phase. ‘Progression criteria’, ideally developed with input from multiple relevant stakeholders, should be used to guide the decision on whether or not to proceed to the next stage of evaluation, conduct more feasibility work, return to intervention development or terminate the research. Criteria for progression could relate to the evaluation design, for which key aspects are feasibility of recruitment and retention of participants; capacity to achieve appropriate sample size; feasibility of the outcomes of the evaluation (e.g. data collection tools and follow-up duration); the capacity and time to collect and analyse the data; and consideration of the unintended outcomes. Criteria for progression could also relate to uncertainties around the intervention itself, for example optimal and acceptable intervention content and delivery; adherence to the intervention; likelihood of cost-effectiveness; and capacity of those providing the intervention to deliver as planned. This is important even if the intervention has been shown to be efficacious, particularly where contextual or implementation factors may influence the intervention’s effectiveness. A key point here is that further work may be required to refine the intervention before embarking on full-scale evaluation, and the programme theory itself should be refined in an ongoing manner. Economic modelling can

be used to assess whether or not expected benefits from an intervention justify its costs and, thus, is useful for decisions on whether or not it is worth proceeding to full-scale evaluation. Evaluability assessment is a complementary approach to a feasibility study. This is a collaborative approach to support the development and evaluation of interventions, aiming to determine whether or not an intervention can usefully be evaluated, and the most appropriate methods for doing so.

3. Evaluation

At this phase it is important to consider how to maximise the 'usefulness' of information for decision-making. To do this, there needs to be an understanding of the key uncertainties surrounding the intervention, and research questions should be developed that are relevant to resolving these uncertainties. As noted above, there are a range of research perspectives that should be considered, rather than necessarily focusing on effectiveness alone. There are numerous study designs and the decision on which to use should be made based on the research questions posed, which should be defined based on the uncertainties that have been identified. There is no 'best' study design: the approach needs to be tailored. A crucial aspect is the choice of outcome measures or evidence of change. Again, these should be chosen based on which outcomes or change are important for stakeholders and on a good theoretical understanding of the intervention. Outcome measures do not need to be reserved to capturing changes in individuals, but can also assess changes in the system, for example changes to policy or to social norms, or by exploring how changing the dynamics in one part of the system alters behaviour in other parts, for example the displacement of smoking into the home after a public smoking ban.

4. Implementation

This phase is about taking deliberate effort to increase the impact of interventions. Even where interventions are shown to be effective, they are seldom implemented successfully. Implementation science is a developing research field, with one focus being on maximising the effective implementation of interventions that have already demonstrated effectiveness in a study. In implementation studies, the main outcomes of interest are measures of the effectiveness of implementation, for example reach, policy impact and uptake of a service. Implementation should be considered in all phases of complex intervention research. Early consideration of implementation increases the potential of developing interventions that can be adopted and maintained in real-world settings.

We present case studies, which are referred to in relevant parts of the text, to illustrate a variety of approaches and demonstrate particular points.

Throughout the phases of complex intervention research, we encourage a regular return to 'check in' with the core elements mentioned above. We provide a checklist to support and document use of this framework, in particular the attention to the core elements and the key considerations for each research phase, in the preparation of funding applications and journal articles.

Complex intervention research is an interdisciplinary, rapidly developing field in which key terms are used in varying ways. The glossary sets out our working definitions of these terms. The framework document provides brief guidance on how complexity should be approached in evidence synthesis.

Conclusions

The framework adopts a pluralist approach and encourages consideration and use of diverse research perspectives. We acknowledge that to generate the most useful evidence for decision-making will often require a trade-off between precise, unbiased answers to narrowly defined questions and more uncertain answers to broader, more complex questions. Earlier editions of this framework were underpinned by an assumption that unbiased estimates of effectiveness are the cardinal goal of evaluation. We have emphasised that improving theories and understanding how and in what circumstances interventions contribute to change are also an important goal for complex intervention research.

For many complex intervention research problems an efficacy or effectiveness perspective will be the optimal approach, for which a randomised controlled trial will probably provide the best design to achieve an unbiased estimate. For other problems, this will not be the case and alternative perspectives and designs will be more likely to generate useful new knowledge to help reduce decision-maker uncertainty. What is important for the future is that the scope of intervention research commissioned by funders and undertaken by researchers is not constrained to a limited set of perspectives and approaches that may be less risky to commission and more likely to produce a clear and unbiased answer to a specific question. In many cases, what is needed is a bolder approach, including some methods and perspectives for which experience is still quite limited, but for which there is an urgent need to make progress by mainstreaming new methods that are not yet widely used, as well as undertaking methodological innovation.

It is recommended that future updates of this framework continue to adopt a broad, pluralist perspective, which means that given the widening scope of the framework and the rich, diverse and constantly evolving body of detailed methods guidance that is now available on specific methods and topics, it will most usefully be in the form of a high-level framework with signposting, published in a fluid, web-based format, which will ideally be frequently updated to incorporate new material, through both updating of text and the addition of new links to updated and emerging key resources.

Funding

This project was jointly funded by the Medical Research Council (MRC) and NIHR (Department of Health and Social Care 73514).

Chapter 1 Development of the framework

Introduction

In 2006, the Medical Research Council (MRC) published guidance for developing and evaluating complex interventions,¹ building on a framework that was published in 2000.² The aim was to help researchers and research funders recognise and adopt appropriate methods to improve the quality of research to develop and evaluate complex interventions and, thereby, maximise its impact. The guidance documents have been highly influential, and the accompanying papers published in the *British Medical Journal* are widely cited.^{3,4}

Since the 2006 edition of the guidance was published, there have been considerable developments in the field of complex intervention research. In some areas, the accumulation of experience and knowledge in the application of approaches and methods that were novel or undeveloped at the time of writing the previous guidance has led to the publication of detailed, focused guidance on the conduct and/or reporting of research, including the MRC guidance on 'process evaluation'.⁵ In other areas, new challenges in complex intervention research have been identified and the reliance on traditional approaches and methods promoted in previous guidance has been challenged.⁶⁻⁸ The guidance has also been complemented by MRC guidance on 'natural experiments',⁹ an important area of development in methods and practice. Given that complex intervention research is a broader and more active field, this new framework provides a less prescriptive and more flexible guide. The framework aims to improve the design and conduct of complex intervention research to increase its utility, efficiency and impact. Consistent with the principles of increasing the value of research and minimising research waste, the framework (1) emphasises the use of diverse research perspectives and the inclusion of research users, clinicians, patients and the public in research teams; and (2) aims to help research teams prioritise research questions and choose and implement appropriate methods. This aims to provide pragmatic recommendations aimed at an audience from multiple disciplines, and we have taken a pluralist approach.

Updating the framework was a pragmatic, staged process, in which the findings from one stage fed into the next. The next section, therefore, provides the methods followed by the results for each stage (gap analysis, expert workshop, open consultation and writing the new framework). We then provide concluding remarks and suggestions for moving forward. The resulting framework is presented in *Chapter 2*.

This project was overseen by a Scientific Advisory Group (SAG) that comprised representatives of each of the National Institute for Health Research (NIHR) programmes, the MRC-NIHR Methodology Research Panel, key MRC population health research investments and authors of the 2006 guidance (see *Appendix 1*). A prospectively agreed protocol, outlining the work plan, was agreed with the MRC and NIHR and signed off by the SAG (see *Appendix 2*). At various points throughout the writing process, we consulted with other researchers, evidence users, journal editors and funders (see *Appendix 3*).

As terminology can often be ambiguous, and there are often terms used interchangeably, we have provided a *Glossary* of key terms.

Methods and results

The framework was updated using multiple methods over several stages:

- stage 1 – a gap analysis of the evidence base for complex interventions
- stage 2 – a workshop that collated insight from current experts in the field
- stage 3 – an open consultation
- stage 4 – drafting the updated framework and a final feedback exercise.

Various stakeholders, for example researchers, research users (patients, public, policy-influencers and NHS), funders and journal editors, were engaged at different stages of the drafting process. The methods and findings from each of the stages are described in the following sections.

Stage 1: gap analysis

Methods for stage 1: gap analysis

The aim of the gap analysis was to identify and summarise aspects of the previous guidance that required updating. We used these gaps as a starting point for discussion within the project team, SAG (for a list of members, see *Appendix 3*) and identified experts. It was, therefore, a method of agenda setting and, thus, did not aim to be comprehensive. The intention was that issues could be added as the work progressed.

Our first step was a brief horizon scanning review that focused on new approaches/progress since the previous guidance, criticisms of existing guidance and other gaps. Based on initial reading of the literature and the experience of the project team, the SAG were provided with a list of topics for update. This was discussed at the initial SAG meeting (24 November 2017) and the list of topics was updated for more in-depth exploration of the literature.

Separate literature searches were conducted for each of the identified topics using keywords (the topic of interest plus variations of 'complex intervention') in Web of Science, restricted to English language with the date limited to those published since 2008. Where there were existing guidance documents relevant to the development, implementation or evaluation of complex interventions, we used these as our starting point and limited our literature review to documents published after these existing guidance documents. For example, guidance existed for natural experiments,⁹ process evaluation⁵ and context,¹⁰ which we drew heavily on. We also discussed this new framework with those involved in developing other guidance at the time, for example for intervention development,¹¹ exploratory studies¹² and systems thinking.^{13,14} We excluded guidance that did not provide substantive information on methodological issues. Criteria for including other publications were broadly that they provided relevant thinking that could be used to progress the work. A summary of the findings from each topic search was created and used to identify focal points for the expert workshop.

Findings from stage 1: gap analysis

Several limitations of the 2006 guidance were highlighted. These mainly related to (1) the focus on effectiveness; (2) considering randomised controlled trials (RCTs) as the gold standard research design; (3) the lack of detail on economic evaluation; (4) the lack of attention to mechanisms that deliver change; (5) the lack of acknowledgement of complex organisational systems or complexity theory; and (6) omission of the importance of policy context, including the impossibility of standardising context. In addition to these issues, there were several areas that had progressed since 2006, which were

considered to be important to explore in more detail. Here we provide a brief summary of the gaps that were identified in 2017 and what we initially proposed to be discussed with experts at the workshop:

- Complex intervention definition –
 - Key issue for the update – definition is too narrow.
 - There are different dimensions of complexity. How can we improve the definition of complex intervention to better acknowledge contextual complexity and the system-level properties that add to this complexity?
 - Does the term ‘complex intervention’ make sense when complexity arises from the intervention context and the interplay between intervention and context as well as the intervention itself?
- Intervention development –
 - Key issue for update – little practical guidance on this phase and subsequent literature provides more detail on certain aspects, for example using a theory-driven approach;^{15,16} identifying and delivering the mechanism of change;¹⁷ co-production and prototyping of public health interventions;¹⁸ and optimisation of the intervention.¹⁹
 - There is guidance under way on intervention development (INDEX study¹¹) that is about identifying and assessing different approaches to developing complex interventions. Given that this is current and there are clear overlaps, are there any issues that are not covered in the INDEX guidelines that we should consider adding in this document?
- Pre-evaluation phase: appraisal of evaluation options or exploratory work –
 - Key issue for update – previous complex intervention guidance highlighted the importance of preparatory work, with the focus being on conducting a pilot study; however, detail on feasibility issues and how to develop the feasibility stage is required.
 - Similar to the development phase, there is work in progress to create guidance for Exploratory Studies of Complex Public Health Interventions (GuEst study¹²), which includes a lot of relevant information. Should the current update include anything further and how can we make it relevant beyond public health?
- Context –
 - Key issue to update – although the previous complex intervention guidance states the importance of taking account of context, this is mostly about how context affects outcomes at the evaluation phase and how interventions may require adaptation for implementation in different contexts. There is little attention to the consideration of context throughout the research phases or guidance on how to take it into account.
 - How do ‘context’ and ‘system’ differ/overlap?
 - Context is a critical construct; how do we ensure that we refer to it throughout the research process?
 - Do we want to go further than the recently published guidance on taking account of context in population health intervention research?¹⁰ What are the key points for considering context in complex intervention research more broadly?
- Ideas from complex systems science –
 - Key issue to update – this is an area that has received increasing attention in the last decade, and for this reason the previous complex intervention guidance did not draw on it.
 - Examples of using complex systems thinking in public health research have been limited to describing and modelling systems; this has not yet been taken further and been used to develop and evaluate interventions.⁷

- When is it critical to embrace a complexity perspective (and when it is not necessary: simple and complicated questions and approaches have their merits) and how can such a perspective be implemented methodologically?
- How can a complex systems approach guide each phase of complex intervention research?
- Programme theory –
 - Key issue to update – the previous complex intervention guidance provided brief information on causal mechanisms and on developing a theoretical understanding of the process of change; however, this lacks the required level of information to guide researchers in developing programme theory from the outset.
 - Further detail is needed to illustrate the steps required to undertake a robust planning phase, including (1) identifying appropriate theories of change, (2) considering potential mechanisms of change, (3) anticipating important contextual factors that could influence the change mechanism and outcomes and (4) mapping appropriate methods to operationalise the chosen theory into practice.
- Implementation research –
 - Key issue to update – the previous complex intervention guidance has limited information on the practical implementation process and needs to understand and account for dynamic contextual factors.
 - Successful implementation is critical to the scaling up of interventions and the new framework should reflect this by emphasising implementation throughout the research process.
 - When do you stop doing effectiveness studies and start doing implementation studies?
 - How can we include the wider aspects of implementation that may enable or constrain desired change? For example, how much guidance do we provide on addressing intervention context and addressing future implementation on a greater scale?
 - How do we make the information palatable for decision-makers?
- Economic evaluation and priority setting –
 - Key issue to update – the previous complex intervention guidance did not go into any detail on how standard economic evaluation methods need to be adapted to deal with particularly complex interventions.
 - Issues around timeline – outcomes are likely to extend beyond the lifetime of an evaluation – can economists work with proxies to system change?
 - How do we best guide on issues for existing economic evaluation methods where interventions aim to change the properties of complex systems? That is, it is not appropriate to evaluate health outcomes only at the individual level if a component of the intervention is to effect change to the system; outcomes are broader than individual health and costs (is a societal rather than a health-care perspective required?).
 - (How) should we include equity issues and economic evaluation analytical approaches, which are growing and complicated methodological areas?
 - How can we guide on economic evaluation for priority setting? That is, what is the most efficient use of resources (to determine whether or not the additional cost of a research project or particular study design is justified)? Are decision-modelling and value-of-information analysis (VOI) practical propositions?
- Systematic reviews of complex interventions –
 - Key issue to update – the previous complex intervention guidance did not address issues related to the inclusion of complex intervention studies in systematic reviews, much beyond acknowledging that they can be problematic. Should we add more?

- Systematic review methods may differ from standard methods and extra consideration is necessary where the systematic review includes complex interventions (if the review is about complexity), for example in defining the research questions, developing the protocol, the use of theory, searching for relevant evidence, and assessing complexity and quality of evidence (how to identify key components of complex interventions; how to assess study quality).
- What should be the end point of a systemic review of complex interventions? For example, effect size, decision model, improved theory or supporting policy decisions?
- Patient and public involvement (PPI) and co-production –
 - Key issue to update – previous complex intervention guidance mentioned that stakeholders should be consulted at various points, but did not emphasise the need to engage relevant stakeholders throughout the research process or provide any guidance on how to do this.
 - How do we guide on effective engagement of stakeholders throughout?
- Evaluation –
 - Key issue to update – the previous complex intervention guidance focused on designing evaluations to minimise bias (i.e. with high internal validity) and, in doing so, did not consider how to maximise the usefulness of evidence for decision-making. These are not mutually exclusive concerns: could both be considered?
 - Should we take an approach that promotes ‘usefulness of evidence’ rather than hierarchy of evaluation study design?
 - Should we present evaluation options that go beyond individual-level primary health outcomes? For example, taking account of system change.
 - Evaluation study designs – what should be added to reflect development in this area? For example, *n*-of-1, adaptive trials. How much information should we present on individual study design?

These topic areas and questions were intended to be a foundation for discussion and further consideration, rather than an exhaustive or definitive list.

Stage 2: expert workshop

Methods for stage 2: expert workshop

A 1-day expert workshop was convened in London in May 2018. A list of those who attended the workshop is given in *Appendix 3*. The aim of the workshop was to obtain views and record discussions on topics that should be newly covered or updated. Participants were identified by the project team and SAG. We aimed to have at least two experts for each of the identified topics and include a range of people from across the UK, plus international representation as far as budget allowed.

In advance of the workshop, the participants were asked to provide two key points, each with one sentence of explanation, that they felt should be taken into account in the update. These key points, alongside findings from the stage 1 gap analysis and discussion with the SAG, were used to inform the agenda and develop content for an interactive, multidisciplinary expert workshop.

After an introductory presentation by the project team, participants were split into five groups (of seven or eight) for the morning session round-table discussion.

The topics covered for all groups (presented in a different order) were:

- the definition of complexity
- the overall framing and scope

- potential impact of the new framework
- the main diagram of the framework (key elements of the development and evaluation process)
- complex systems thinking.

For each of the two afternoon sessions, participants were split into five 'expert groups' aligned with their topic areas of expertise. Topics covered in these smaller specialised groups included:

- options for study design
- the previous guidance's emphasis on 'effectiveness'
- choice of outcomes
- considerations for economic evaluation
- pre-evaluation and development phases
- considerations for implementation
- key focus areas to improve applications for funding
- evidence synthesis of complex interventions
- considerations for digital health
- programme theory.

Each session was facilitated by a member of the project team and was supported by a colleague from the MRC/Chief Scientist Office Social and Public Health Sciences Unit, University of Glasgow. Colleagues assisted the facilitators by taking notes of key points during each discussion, clarifying main points with attendees and producing a written summary of each discussion after the workshop. SAG members were also present in each discussion. Round-table discussions were audio-recorded. Throughout the day, participants were asked to provide their thoughts on key points, case study examples and key references on Post-it® notes (3M, Saint Paul, MN, USA) on dedicated noticeboards.

Data from each of the 15 workshop discussions and post-it points were thematically coded, and summaries drawing on all of the data were created for each theme. These workshop summaries were sent to workshop participants by e-mail as a follow-up consultation to ensure that the thematic summaries that we created from the workshops were accurate overviews of the discussions in which they were involved. Final summaries were discussed in detail with the SAG to support the decision-making on the content of the document.

Findings from stage 2: expert workshop

Seventy experts were invited to the workshop (with the aim of facilitating a workshop of around 40 participants). In total, 37 experts confirmed their attendance; one who accepted the invitation did not attend (owing to sickness), three people did not respond to our invitation and 30 people could not attend for various reasons, some of whom recommended others who did attend. In total, 36 participants attended the workshop. Key issues that were identified are summarised in *Table 1*.

Decisions taken following the expert workshop

There was considerable agreement across the workshop discussions; however, as seen in *Table 1*, there were some issues for which consensus was not reached or for which competing points were made in different break-out discussions. The main example of controversy was the purpose of evaluation (theory as an end point, the need for primary outcome). In addition, some of the points that were made were very specialised, for example related to particular methods or specialties. Along with the SAG, the project team determined which focus areas to incorporate in the document, keeping them high level rather than getting into specific detail. With respect to the issues for which views diverged (primarily related to effectiveness and the purpose of evaluation), we consider the document as a 'thinking tool' to provide recommendations to arrive at the most appropriate approach for each piece of complex intervention research (with no 'one size fits all' approach, instead determined by the problem that is being addressed and taking a 'usefulness of evidence' approach).

TABLE 1 Summary of the key points from the expert workshop in 2018

Topic	Key points
Definition of complex intervention	<ul style="list-style-type: none"> • Requires clarity regarding the use of the terms simple and complex, and the use of the term intervention • Needs to include context, systems thinking, feedback loops and the fundamental aspects of complexity • Thinking about complex interventions in terms of how 'simple' or 'complicated' they are is not very helpful, as there are many aspects that make an intervention complex
Framing and scope	<ul style="list-style-type: none"> • Reaching a new audience is important: as well as researchers, the framework needs to reach 'practitioner researchers', Health Technology Assessment and policy-makers. A common language is needed and it is, therefore, important to include different stakeholders in this process. Provide clarification of terminology • A more iterative and fluid research process needs to be emphasised. More options should be given to avoid saying there's 'one way of doing this'
Intervention development	<ul style="list-style-type: none"> • More emphasis on determining the problem and establishing the research questions • Acknowledge that there are different starting points to the research process, for example the process does not necessarily start with development, researchers may not be involved in the intervention's development (e.g. a national policy) and researchers could join/start at any phase
Study design	<ul style="list-style-type: none"> • Agreement regarding the need to choose a design that is best suited to answer the research question in a given context (i.e. there is no novel design that caters for 'complex interventions') • It is important to emphasise that there are more (and often more suitable) options than RCTs • Guide people to think about the function of the intervention in choosing the study design. Give examples of what researchers have undertaken in different contexts and include case studies • Usefulness of evidence is a good approach, rather than the previous hierarchy of evidence approach. This requires thinking of the right research questions for the intended use of the research • Acknowledge that rapidly changing fields require faster routes of evaluation before the overall context changes (e.g. in digital health)
Systems thinking	<ul style="list-style-type: none"> • General agreement regarding including systems thinking and encouraging researchers to think with a 'systems lens' • However, there was also agreement that this field is developing and there are limitations to how much guidance we can currently give • Introduce three levels of systems thinking: (1) conceptualisation of the system (what does it currently look like), (2) what parts of the system can be influenced (and what are the boundaries) and (3) what is happening outside the boundary (providing richer context) • In relation to this, context is often poorly articulated in funding applications and guidance is required
Implementation science	<ul style="list-style-type: none"> • Clarity is required on ongoing modification at the implementation phase • Clarity on terminology required (delivery of intervention/implementation) • Implementation should be emphasised throughout (e.g. from development stages)
Programme theory	<ul style="list-style-type: none"> • Clarity on terminology is required (e.g. people are typically unclear on what theory is; there do not seem to be consistent definitions of terms, such as 'logic model' and 'programme theory'; and other terms, such as mechanism of action, etc., need to be defined) • Include information on theory (e.g. thinking about what the problem is at the start and what should be changed; articulating what theory is and is not; considering how theory may be influenced by context; and encouraging adaptive and iterative theory development) • Researchers need to be encouraged to articulate their theory in full, not solely in a visual model that will miss some of the important detail
Economic evaluation	<ul style="list-style-type: none"> • There was general agreement that new methods were not needed, but that standard methods could be adapted to more effectively explore complex interventions • Signpost recent developments since the previous guidance and link them to the existing guidance • Emphasise the need to measure aspects of programme theory and resource use, rather than just effectiveness of outcomes • Mention cost-consequence analysis to help highlight links between processes and outcomes • Emphasise the need for a broader range of outcomes • End point: it should be about revealing the resource cost/outcome trade-offs and causal connection, as much as it is about producing a ratio or a number. We need the trade-offs for each resource use

continued

TABLE 1 Summary of the key points from the expert workshop in 2018 (continued)

Topic	Key points
Effectiveness	<ul style="list-style-type: none"> • There were diverse views on whether or not traditional effectiveness is an appropriate end point for an evaluation, for example some were keen to abolish effect sizes altogether, look at other things and answer different questions (e.g. what happened?). This would include having theory as an end point in itself. Others disagreed that theory should be an end point and felt strongly that evaluations have to answer the 'does it work?' question; however, there was general agreement that this (does it work?) in itself is not enough • In relation to this, there was debate about whether or not evaluations should have primary (health) outcomes, with some of the opinion that there needs to be a prespecified outcome for the intervention that is being tested, and without a primary outcome there is the risk of 'cherry picking' the most improved outcome to make the intervention appear effective. Others disagreed and felt that evaluation should explore the impact on multiple effects, including system change (not individual effect sizes), with the goal of theory development
Stakeholders	<ul style="list-style-type: none"> • There was more emphasis required than in the previous guidance • Place higher priority on co-production and non-researcher-led interventions • Articulation of the problem needs to come from shared space; genuine co-creation of interventions from this starting point
Evidence synthesis	<ul style="list-style-type: none"> • Acknowledge that methods still need to be developed • Context: <ul style="list-style-type: none"> ○ The value of an evidence synthesis does not lie only in obtaining the most unbiased estimate (in terms of effect size) or in meta-analysis; there needs to be a focus on exploring heterogeneity across contexts and identifying the mechanisms that drive variation ○ There will never be a complete suite of studies of complex interventions in all of the relevant contexts, so there needs to be a way of extrapolating from what we have ○ Provide pointers to things that would help decision-makers know if they could use the evidence and make judgements about transferability, how their context differs and what might be the things that facilitate change ○ Decision-makers could be encouraged to think about socially significant differences in context • One challenge is that, with some exceptions, evidence synthesis organisations focus on trials. The new framework should seek to widen the range of evidence included in syntheses used for decision-making and ensure the inclusion of mixed-methods research. It should also clarify what we mean by 'evidence' and acknowledge that this includes theory as well as information about outcomes • General agreement that an improvement in primary studies that follows from the recommendations will have a positive impact on evidence synthesis in time (there will be more appropriate studies to synthesise)

Some examples of the decisions taken are as follows:

- Clarity in terminology – include a comprehensive glossary.
- Include a series of case studies as an appendix, highlighting particular aspects of each phase and core elements of the research process.
- Highlight the distinctive methods of evidence generation, emphasising that the research can begin at any stage of the intervention and that there may be different approaches for researchers not involved in intervention development.
- Not to be prescriptive but rather provide options for approaching the research, which should be chosen by taking the problem as the starting point and working out what is most useful in terms of decision-making for policy and practice going forward.
- Update the diagram included in the 2006 guidance that showed 'Key elements of the development and evaluation process',¹ particularly to include context.
- Include a greater focus on programme theory, but one that encourages its consideration and refining throughout.

- Systems thinking – not to provide detailed guidance on systems thinking and methods because this is beyond the scope; rather it will be a starting point for encouraging people to consider how a systems perspective could help develop and evaluate complex interventions, with methodological development to follow.
- Evidence synthesis – following the expert workshop, information that others were developing guidance in this area and discussion with the SAG we took the decision to focus on primary studies; therefore, we did not include a section in the main document on evidence synthesis. It is hoped that an improvement in primary studies, brought about by this new framework, would in time have a positive impact on evidence synthesis. We added an appendix to highlight some of the main considerations for evidence synthesis (see *Appendix 5*).

Further decisions were taken regarding the need to obtain further expertise in drafting the document. We approached three health economists for a follow-up meeting to discuss further issues related to economic considerations for complex intervention research; following this, they agreed to take on the responsibility of drafting sections that related to economic considerations and became co-authors. We also approached experts in systems thinking to discuss some of the emerging ideas on taking a systems perspective to complex intervention research. We convened a meeting in December 2018 in London with a group of researchers with such expertise (individually acknowledged in this monograph). Similarly, we convened a meeting in January 2019 with researchers who were creating guidelines on intervention development (individually acknowledged in this monograph) to discuss the overlap and the use of the INDEX guidance within the current document.

Stage 3: open consultation

Methods for stage 3: open consultation

The first draft of the updated document was made available for open consultation from Friday 22 March to Friday 5 April 2019.

Potential respondents were targeted, as follows:

- those invited to the expert workshop
- other experts identified from the suggestions of workshop participants, with greater focus on international experts
- early and mid-career researchers (identified via e-mail groups)
- journal editors
- funders
- service users/public
- policy-influencers/-makers.

We e-mailed potential respondents with advance notice of the consultation dates and a link to register their interest in participating, and sent a further message when the consultation opened. Two reminder e-mails were also sent. As well as targeted promotion, we used social media to publicise the consultation and encouraged others to pass on the link.

Consultees were informed that they were responding about an early draft of the revised framework and that their involvement was an important part of the process for its final development. We asked them to relate topics in the draft to a project that they had recently worked on and to provide feedback on its usability.

The online consultation was guided by a questionnaire that was developed by the project team (the questions that all consultees were asked to complete are presented in *Appendix 4*). Responses were anonymous.

Findings from stage 3: open consultation

We received 52 individual responses, plus some follow-up e-mail comments. This amounted to 25,000 words of response. The majority of responses were from researchers, but some identified as funders ($n = 3$), journal editors ($n = 7$), NHS ($n = 7$), policy-influencers ($n = 3$) and service users (patient or public, $n = 5$). Most of the respondents said that their main field of expertise was public health ($n = 21$) or health services research ($n = 20$), with others stating clinical medicine ($n = 6$), implementing policy ($n = 3$), systems-based research ($n = 4$), patient or public involvement ($n = 4$) and other ($n = 7$: statistics, sociology, health economics and triallist) as their main field of expertise. A summary of the consultation suggestions is provided below; however, it is important to note that there were conflicting views on some aspects, which we have noted.

Overall

- Overall layout: extra sections are required – an executive summary and a preface chapter that details how this is related to previous guidance and that this document is a standalone framework that does not require reference back to the 2006 version. Consider placing more emphasis on development in the earlier sections of the document rather than delve straight into evaluation.
- Definition of complexity: the distinction made between complicated and complex interventions was said to be unclear. Respondents stressed that a clear definition of complex intervention and a more accessible account of how complexity affects the research process are required.
- Key elements for developing and evaluating complex interventions (*Figure 1*): respondents felt that the ‘overarching considerations’ should all be highlighted as central to the research process and that some text detail should be added to each phase box to provide more information on what each means.
- Evaluation perspectives (shown in the x-axis of *Figure 2*): many respondents felt that the perspectives that we presented were shown to be mutually exclusive and hierarchical (which was not the intention). There was significant pushback on using the term ‘realist’ as an evaluation perspective. Respondents questioned whether or not we were advocating for evaluations that do not measure effectiveness, with some conflicting views on whether or not this was a positive change. It was felt that there was not enough focus on how the perspectives relate to intervention development or to the development of research questions.
- Framework for addressing complexity within an evaluation (see *Figure 2*): although some liked this framework, on balance respondents did not feel that this figure complemented the text or was very clear. Complexity does not increase in a linear fashion based on intervention components and perspective taken. Context and system were missing from the diagram despite being a large focus of the text.

Research phases (shown *Figure 1*, plus a section of text was also dedicated to each phase)

- Developing and identifying complex interventions: suggestions included that we consider minimising detail in this section and signpost to the MRC-funded INDEX guidance;¹¹ clarify the different circumstances in which development versus identification of interventions is appropriate; and consider including something specific on digital interventions.
- Feasibility: make sure that the definition of feasibility is clear, for example in line with other standard definitions. There was a call for additional detail on the role of context in determining uncertainties for feasibility testing.
- Evaluation: as in the expert workshop, there was conflicting feedback from respondents on how to provide guidance on evaluation. Suggestions included highlighting that evaluations must focus on effectiveness, with additions (not replacements) relating to theory and systems perspectives, but also to include better examples of evaluations focusing on systemic questions. Many respondents felt that there was too much focus on realist evaluation and little mention of theory-based evaluation approaches. The section on study design needs to be clearer, particularly on why some designs are included but others are not.
- Implementation: suggestions were made to differentiate between clinical and implementation interventions; add EPOC (Effective Practice and Organisation of Care) criteria²⁰ and diagnostic approaches to implementation; and clarify the time and stage of modification in relation to implementation.

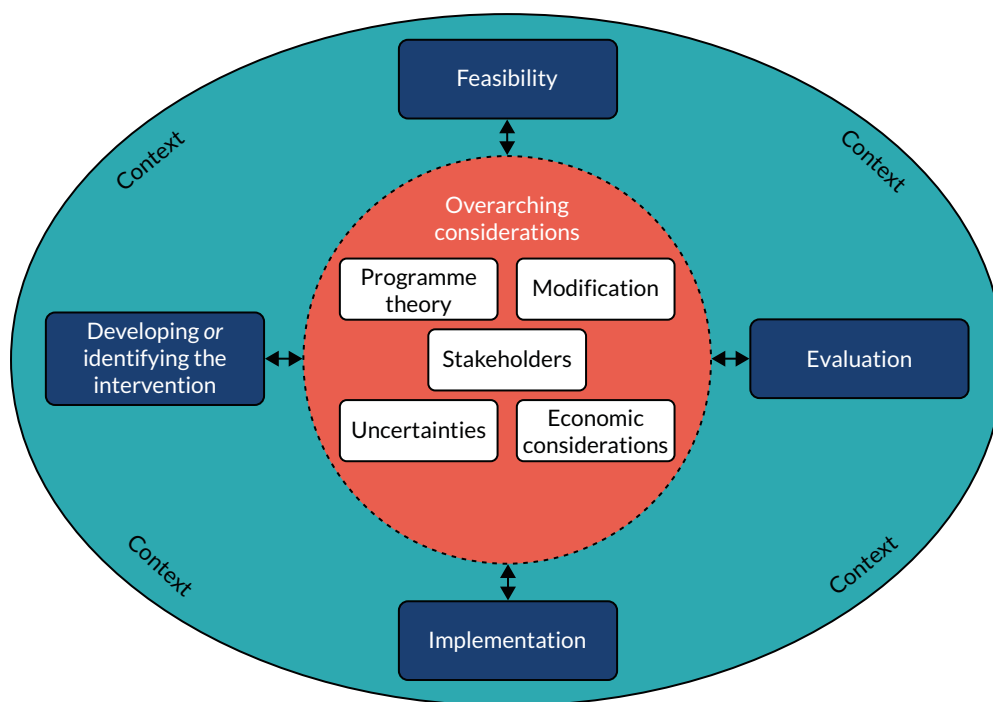


FIGURE 1 Key elements for developing and evaluating complex interventions (consultation version).

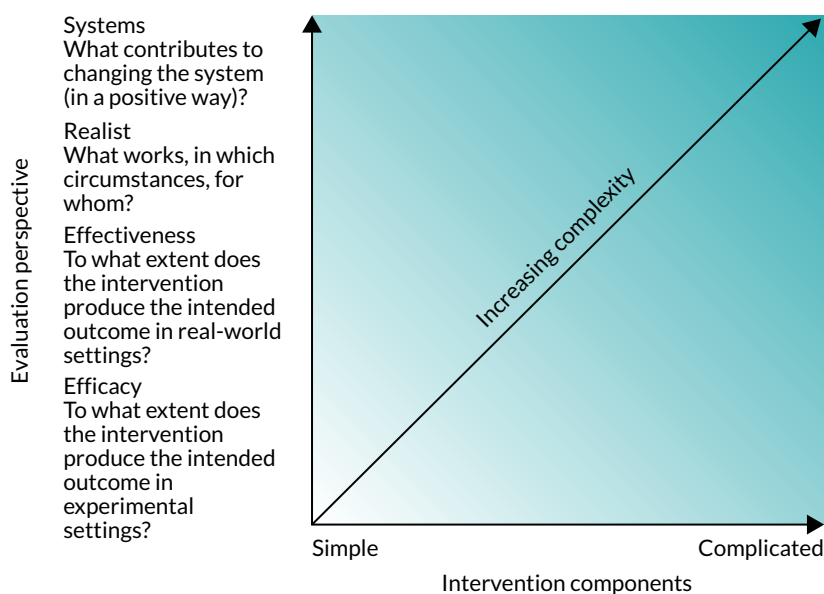


FIGURE 2 Framework for addressing complexity within evaluation (consultation version).

Overarching considerations (see Figure 1)

- Programme theory: it was suggested that we provide greater detail on theory-led research, with a balance of signposting to appropriate resources; address how theory-based content is written and presented for readers who come from non-theory-based disciplines, to avoid alienating people; and clarify the terminology relating to 'mechanism', 'programme theory' and 'logic models'.
- 'Modification' (changed to refinement in the final version): there were conflicting opinions on the use of terminology, particularly with the (MRC-/NIHR-funded) adaptation guidance (in preparation²¹). Respondents called for guidance on where/when to perform modifications and how to agree acceptable boundaries, as well as examples to help readers understand the different approaches to modification.

- Stakeholders: respondents encouraged a greater focus on PPI, more consideration of the challenges of 'stakeholder engagement', as well as practical examples of how to engage stakeholders.
- Economic considerations: respondents suggested that we make sure that sensitivity analysis is discussed in relation to statistical models as well as economic models; that we give mention to the tensions between equity and efficiency in evaluating complex interventions; and that we provide more detail on generalisability and context dependency of cost-effectiveness as well as effectiveness, and the possibility of using programme theory to achieve this.

The majority of respondents were positive about the document overall, albeit with constructive criticism that required the project team to reconsider various aspects. Examples of changes that were made following consultation (note that this is not an exhaustive list of changes) were:

- Figures –
 - the framework for addressing complexity in evaluation was removed (see *Figure 2*)
 - *Figure 1* was updated to place 'context' in the centre with the other 'overarching considerations' (re-termed 'core elements') and short descriptors added to each of the phases.
- Evaluation perspectives (see *Figure 2*) – a rewrite of this section was required (now termed 'research perspectives'). We changed 'realist perspective' to 'theory-based perspective' to take account of other approaches to evaluation that aim to explore how and why interventions bring about change.
- Terminology –
 - the definition of 'complex intervention' was updated
 - 'modification' was changed to 'refinement'
 - programme theory/logic model – a decision was taken to use 'logic model' for the visualisation of the 'programme theory', with programme theory detailed with text.
- Programme theory: we have clarified terminology in the text and *Glossary*.
- Modification/refinement: we changed the term from 'intervention modification' to 'intervention refinement', and differentiated from 'adaptation'. We have added detail to this section on when you would expect interventions to be refined and why, including a separate section on rapid refinement of digital interventions.
- Stakeholders: we separated the section on stakeholders into PPI and professional stakeholders, and added text to highlight the challenges in engaging stakeholders.
- Economic considerations: we have edited and moved some of this section to other parts of the document to avoid repetition. We add detail on the potential trade-off between equity and efficiency.
- Developing and identifying interventions: we removed text and used the INDEX guidance as the basis for this section, adding three more points that were not highlighted in that guidance but were felt to be important throughout the process of developing this document.
- Feasibility: we further clarify what we mean by 'feasibility'. We have re-ordered this section to improve readability. We have added a section on 'efficacy signals' to further show the potential of feasibility studies.
- Evaluation: we have added detail on how the research perspectives are related to evaluation, as well as more case studies to illustrate the main points. We have emphasised the need for qualitative study in an evaluation and have added detail on process evaluation. We have added detail on the strengths and limitations of each type of economic evaluation.
- Implementation: we have considered separately in this section (1) implementation science research, which focuses specifically on the development and evaluation of interventions to maximise effective implementation; and (2) the need to emphasise implementation considerations in earlier phases, including hybrid effectiveness/implementation designs. In the earlier phases and in the core elements, we have highlighted context, stakeholder input and the need for a broader programme theory, all of which contribute to increased consideration of implementation factors.

Stage 4: writing the updated framework

Methods for stage 4: writing the updated framework

The writing of the framework was led by the project team and was supported by co-authors in the writing group and the SAG. Feedback was received at various stages throughout the writing process from members of the MRC's Population Health Sciences Group (PHSG) and the MRC-NIHR Methodology Research Programme (MRP) Advisory Group.

Given that the document had changed substantially from the open consultation draft, we asked a further set of external individuals to provide comments on the near-final draft. We received feedback from eight people in May/June 2020. The final draft was then sent to all co-authors for approval.

Findings from stage 4: final approval and sign-off

The final draft was approved by the MRC's PHSG in March 2020.

Patient and public involvement

This project was methodological; views of patients and the public were included at the open consultation stage of the update. The open consultation, involving access to an initial draft, was promoted to our networks via e-mail and via digital channels, such as our unit Twitter account (Twitter, Inc., San Francisco, CA, USA; www.twitter.com). We received five responses from people who identified as 'service users' (rather than researchers or professionals in a relevant capacity). Their input included helpful feedback on the main complexity diagram, the different research perspectives, the challenge of moving interventions between different contexts and overall readability and accessibility of the document. Several respondents also highlighted useful signposts to include for readers.

In relation to broader PPI, the resulting updated framework (see *Chapter 2*) highlights the need to include PPI at every phase of developing and evaluating complex interventions. We have drawn on and referred to numerous sources that provide further detail or guidance in how to do so.

Limitations

There was a huge amount to cover in developing this document. We have not provided detailed methodological guidance where that is covered elsewhere because we have tried to focus on the main areas of change and novelty. In many of these areas of novelty, methods and experience are in some parts quite limited. In addition, we have foregrounded the very important concept of 'uncertainties' and, although there are methods of doing this through, for example, decision-modelling and more qualitative soft system methodologies, this area is limited and specific guidance on how to determine uncertainties in a formal way may seem unclear. We recommend that due consideration is given to this concept and call for further work to develop methods and provide examples in practice. Invariably we may have missed something in our writing and, furthermore, the fields will inevitably move on at pace following publication of this document.

Conclusion

Parts of this text have been reproduced with permission from Skivington *et al.*²⁶ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <https://creativecommons.org/licenses/by/4.0/>. The text below includes minor additions and formatting changes to the original text.

In this document, we have incorporated developments in complex intervention research that were published since the previous edition was written in 2006. We have retained the basic structure of the research process as comprising four phases – development, feasibility, evaluation and implementation –

but we emphasise that a programme of research may begin at any of these points, depending on what is already known. We have emphasised that complex intervention research will not always involve the development of new researcher-led interventions, but will often involve the evaluation of interventions that are not in the control of the researcher, but instead led by policy-makers or service managers, or are the adaptation of interventions from another context. We have highlighted the importance of engaging stakeholders throughout the research process, including patients, the public, practitioners and decision-makers. We emphasise the value of working with them as partners in research teams to jointly identify or prioritise research questions; develop, identify or prioritise interventions; and agree programme theories, research perspectives, key uncertainties and research questions.

As with earlier editions, we stress the importance of thorough development and feasibility testing prior to large-scale evaluation studies. As well as taking account of established practice and recent refinements in the methodology of intervention development, feasibility and pilot studies, we draw attention to new approaches, such as evaluability assessment, that can be used to engage stakeholders in collaborative ways of planning and conducting research. We place greater emphasis than in the previous edition on economic considerations in complex intervention research. We see these as a vital to all phases of a research project, rather than simply a set of methods for assessing cost-effectiveness.

We have introduced a new emphasis on the importance of context and the value of understanding interventions as 'events in systems' that produce effects through interactions with features of the contexts in which they are implemented. We adopt a pluralist approach and encourage consideration and use of diverse research perspectives, namely efficacy, effectiveness, theory-based and systems perspectives, and the pragmatic choice of research questions and methods that are selected to optimally address the key uncertainties that remain. We acknowledge that to generate the most useful evidence for decision-making will often require a trade-off between precise, unbiased answers to narrowly defined questions and less certain answers to broader, more complex questions.

Although we have not explicitly discussed epistemology, we have challenged the position established in earlier editions that unbiased estimates of effectiveness are the cardinal goal of evaluation, and we have emphasised that improving theories and understanding of how and in what circumstances interventions contribute to change is also an important goal for complex intervention research. For many complex intervention research problems, an efficacy or effectiveness perspective will be the optimal approach, for which a RCT will probably provide the best design to achieve an unbiased estimate. For other problems this will not be the case, and alternative perspectives and designs will be more likely to generate useful new knowledge to help reduce decision-maker uncertainty. What is important for the future is that the scope of intervention research commissioned by funders and undertaken by researchers is not constrained to a limited set of perspectives and approaches that may be less risky to commission and more likely to produce a clear and unbiased answer to a specific question. What is needed is a bolder approach, including some methods and perspectives for which experience is still quite limited, where we (supported by our workshop participants and respondents to our consultations) believe that there is an urgent need to make progress by mainstreaming new methods that are not yet widely used, as well as undertaking methodological innovation.

We have emphasised the importance of continued deliberation by the research team of what the key uncertainties are that are relevant to that stage of research, and then defining research questions and selecting research perspectives and methods that will reduce that uncertainty. We reiterate that our recommendation is not to undervalue research principally designed to minimise bias in the estimation of effects; rather, we encourage the use of a wider range of perspectives and methods, augmenting the available toolbox and, thus, increasing the scope of complex intervention research and maximising its utility for decision-makers. This more deliberative, flexible approach is intended to reduce research waste and increase the efficiency with which complex intervention research generates knowledge that contributes to health improvement.

We acknowledge that some readers may prefer more detailed guidance on the design and conduct of any specific complex intervention research project. The approach taken is to help researchers identify the key issues that ideally need to be considered at each stage of the research process, to help research teams choose research perspectives and prioritise research questions, and to design and conduct research with an appropriate choice of methods. We have not provided detailed methodological guidance, primarily because that is well covered elsewhere. We have been fortunate to be able to draw on and refer to many other guidance documents that address specific and vitally important aspects of the complex intervention research process and specific aspects of research design, conduct and reporting. We encourage researchers to consult these sources, which provide more detail than we were able to here. We have provided more emphasis and detail in areas of change and novelty introduced in this edition. However, in many of these areas there is an urgent need for further methods development and guidance for their application and reporting in complex health intervention research. These include more formal methods to quantify or consider uncertainty, for example decision-modelling approaches, Bayesian approaches, uncertainty quantification or more qualitative soft systems methodologies, and methods suited to a systems perspective including simulation approaches and qualitative comparative analysis methods.

Recommendations

The recommendations of this work are given in *Chapter 2*. At the end of each research phase section (see *Chapter 2, Phases of research*) we include a table of elements that we recommend should be considered at that phase. The overall recommendation, therefore, is that people use the tables at the end of each phase when developing research questions and use the checklist in *Appendix 6* as a tool to record where/how the recommendations have been followed.

Monitoring the use of this framework and evaluating its acceptability and impact is warranted: this has been lacking in the past. We encourage research funders and journal editors to support the diversity of research perspectives and methods that are advocated and to seek evidence that the key considerations are attended to in research design and conduct. The use of the checklist that we provide to support the preparation of funding applications, research protocols and journal publications (see *Appendix 6*) offers one way to monitor impact of the framework on researchers, funders and journal editors. Further refinement of the checklist is likely to be helpful.

We recommend that future updates of this framework continue to adopt a broad, pluralist perspective. Given the widening scope and the rich, diverse and constantly evolving body of detailed methods guidance that is now available on specific methods and topics, the framework will most usefully be in the form of a high-level framework with signposting, published in a fluid, web-based format, which will ideally be frequently updated to incorporate new material, both through updates to the text and case studies and through the addition of new links to updated and emerging key resources.

Chapter 2 The new framework

Introduction

Aims of the framework

The framework aims to improve the design and conduct of complex intervention research to increase its utility, efficiency and impact. Consistent with the principles of increasing the value of research and minimising research waste,²² the framework (1) emphasises the use of diverse research perspectives and the inclusion of research users, clinicians, patients and the public in research teams, and (2) aims to help research teams prioritise research questions and choose and implement appropriate methods.

Structure of the framework

The framework is presented as follows:

- In *What is a 'complex intervention'?*, we provide an updated definition of 'complex intervention' informed by a broader understanding of complexity, and introduce how different research perspectives can be employed in complex intervention research.
- In *Framework for the main phases and core elements of complex intervention research*, we set out the revised framework for developing and evaluating complex interventions and provide an overview of the key core elements that we recommend are considered repeatedly throughout the phases of complex intervention research.
- In *Phases of research*, we go into further detail about each of the key phases in complex intervention research.
- In *Case studies*, we present illustrative case studies to help exemplify aspects of the framework using a variety of study designs, from a range of disciplines, undertaken in a range of settings.

How to use the framework

In the 2000² and 2006¹ versions of this document, there was a key explanatory diagram in which complex intervention research was divided into a number of key phases. In this edition, we have developed a revised explanatory diagram (see *Figure 3*). This does not offer a linear or even cyclical pathway through the phases of development and evaluation of interventions, but emphasises that at any phase key core elements should be considered to guide decisions as to whether or not the research should proceed to the next phase, return to a previous phase, repeat a phase or be aborted. The framework encourages teams to consider the research perspective(s) most suited to the research challenge that they are aiming to address and to use the six core elements to guide the choices that they make at each research phase. A programme of intervention research may begin at any one of the four phases, depending on the key uncertainties that are associated with the intervention and taking into account existing evidence and/or experience. Users can, therefore, choose which sections of the recommendations are most appropriate to their own research plans. To aid navigation through the document, we provide extensive hyperlinks to cross-references between sections. Throughout the document, we refer to existing detailed guidance and provide key points and signposts to further reading. At the end of the sections on each phase of research (see *Tables 4–8*) we include a table that lists the core considerations relevant to that phase, which we recommend to be used as a tool in research planning.

Given that this is a pragmatic framework aimed at an audience from multiple disciplines, we have taken a pluralist approach. Terminology that is related to the study of complex interventions is not used consistently across disciplines; our use of terms is detailed in the *Glossary*. Note that we focus on evaluation rather than *ex ante* appraisal,²³ and specifically on evaluations that contribute to the scientific evidence base, rather than those that are conducted primarily for monitoring and accountability purposes.

Previous guidance has focused on statistical and qualitative considerations; we refer to this as well as highlight the importance of statistical and qualitative considerations throughout the research process. Economic issues in the context of developing and evaluating complex interventions have been given less attention; however, their consideration is also important from the early phases of research in relation to intervention development, evaluation and future implementation. We, therefore, provide more detail in this section of the framework.

We focus on primary research but note that other types of research, in particular evidence synthesis, must consider how to approach complex interventions as well. In *Appendix 5*, we provide some brief points and signposts to further reading on approaching complexity for producers and users of evidence syntheses. In *Appendix 6*, we provide a checklist to support and document the use of this framework in the preparation of funding applications and journal articles.

What is a ‘complex intervention’?

Few interventions are truly simple. Complexity arises from the properties of the intervention itself, the context in which an intervention is delivered and the interaction between the two.^{8,24,25}

Sources of complexity

Complexity owing to characteristics of the intervention

Parts of this text have been reproduced with permission from Skivington *et al.*²⁶ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <https://creativecommons.org/licenses/by/4.0/>. The text below includes minor additions and formatting changes to the original text.

Interventions become more complex in line with increasing the:¹

- number of intervention components and the interactions between them
- range of behaviours, expertise and skills (e.g. particular techniques and communication) required by those delivering or receiving the intervention
- number of groups, organisational levels or settings that are targeted by the intervention
- level of flexibility or tailoring of the intervention or its components that is permitted (i.e. how dynamic or adaptive the intervention is).

For example, the Links Worker Programme²⁷ was an intervention in primary care in Glasgow, Scotland, that aimed to link people with community resources that could help them to ‘live well’ in their communities. It targeted individual, primary care [general practitioner (GP) surgery] and community levels; the intervention was flexible in that it could differ between primary care GP surgeries. In addition, there was no single health or well-being issue that the Link Workers specifically supported: bereavement, substance use, employment and learning difficulties were all things that could be included.²⁷ The inherent complexity of this intervention had implications for many aspects of its evaluation, such as the choice of appropriate outcomes.

Complexity arising from context

Context: . . . any feature of the circumstances in which an intervention is conceived, developed, implemented and evaluated.

Reproduced with permission from Craig et al.¹⁰ Contains information licensed under the Non-Commercial Government Licence v2.0

Examples of features include social, political, economic and geographical contexts. Whether or not and how an intervention generates outcomes can be dependent on a wide range of contextual factors. These contextual factors will be intervention specific and may be difficult to anticipate. An example of an intervention that may seem simple, until context is considered, is the 'Lucky Iron Fish' (Box 1).¹⁰

It is important to consider what features of context may be important in determining how an intervention achieves outcomes and under what circumstances an intervention may be more or less effective. This is important for all phases, that is for developing a new intervention, adapting or translating an existing intervention to a new context, evaluation and implementation.

Interventions as 'events in systems'

System: a set of things that are interconnected in such a way that they produce their own pattern of behaviour over time.

Meadows and Wright³¹

Systems thinking can help us to understand the interaction between an intervention and the context in which the intervention is implemented in a more dynamic way. Systems can be thought of as complex and adaptive,³² in that they are defined by system-level properties, such as feedback, emergence, adaptation and self-organisation (Table 2). We can theorise interventions as 'events in systems'²⁴ and can conceive an intervention's outcomes as being generated through the interdependence of the intervention and a dynamic system context.

Change in complex systems may be unpredictable. In a social system, people interact with each other and other parts of the system in non-linear and interconnected ways so that the actions of one person alter the context for others.³⁵ For example, removing hospital car-parking charges has clear beneficiaries. However, by encouraging people to drive, the policy may reduce demand for public transport, leading to a reduction or withdrawal of services. The net effect may be to reduce accessibility to those without a car. Demand for car parking will increase, possibly beyond capacity, and alternative options will have diminished. What might appear at first sight to be a positive intervention may have adverse effects on health and serve to widen inequalities.

BOX 1 Considering context: 'Lucky Iron Fish' example

A small fish-shaped iron ingot placed in a pot while cooking or boiling drinking water. The ingots have been shown to be an effective way of reducing iron deficiency anaemia in women in some communities in rural Cambodia.²⁸ The intervention was carefully developed to be effective and sustainable in this setting and population. The ingots could be produced locally, at lower cost than conventional nutritional supplements, and making them fish-shaped encouraged uptake because fish are considered to be lucky by Cambodian villagers. Even so, an earlier trial found that short-term improvements in iron status were not sustained, because seasonal changes in water supply reduced the dietary availability of iron from the cooking water.²⁹ A subsequent trial in a different region of Cambodia also found little benefit, because anaemia in this region was primarily due to inherited problems with haemoglobin production, rather than dietary iron deficiency.³⁰

Example taken from *Taking Account of Context in Population Health Intervention Research: Guidance for Producers, Users and Funders of Research*.¹⁰ Reproduced with permission from Craig *et al.*¹⁰ Contains information licensed under the Non-Commercial Government Licence v2.0.

TABLE 2 Properties of complex adaptive systems, with examples

Properties of complex adaptive systems	Example
Emergence	
Complex systems have emergent properties that are a feature of the system as a whole. Emergent properties are often unanticipated, arising without intention	Group-based interventions that target at-risk young people may be undermined by the emergence of new social relationships among the group that increase exposure to and reinforce risk behaviours, while reducing their contact with mainstream youth culture where risk-taking is less tolerated ³³ Enhanced recovery pathways are introduced to optimise early discharge and improve patient outcomes. They involve changes across pre-operative care where patient expectations are managed; changes in theatre by minimising the length of surgical incisions and the use of surgical drains; and changes in postoperative care on surgical wards (e.g. the use of physiotherapy). As enhanced recovery pathways are introduced for specific patients it influences management of other patients; thus, the whole culture of surgical practice within a hospital changes towards 'enhanced recovery'
Feedback	
Where one change reinforces, promotes, balances or diminishes another	A smoking ban in public places reduces the visibility and convenience of smoking. Fewer young people start smoking because of its reduced appeal, thus further reducing its visibility and so on in a reinforcing loop ⁷
Adaptation	
A change of system behaviour in response to an intervention	Retailers adapted to the ban on multibuy discounts for alcohol by placing discounts on individual alcohol products, offering them at the same price individually as they would have been if part of a multibuy offer ³⁴
Self-organisation	
Organisation and order achieved as the product of spontaneous local interaction rather than of a preconceived plan or external control	Individually focused treatment for people who misuse alcohol did not address some social aspects of alcohol dependency; as a result, recovery groups were self-organised in a collective effort and Alcoholics Anonymous (New York, NY, USA) was formed
Reproduced with permission from Skivington <i>et al.</i> ²⁶ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: https://creativecommons.org/licenses/by/4.0/ . The table includes minor additions and formatting changes to the original table.	

SIGNPOSTS TO FURTHER READING 1 Systems thinking in health research

- For a more detailed accessible introduction to applying systems thinking to public health evaluation, see Egan *et al.*^{13,14}
- The Health Foundation's overarching description and evidence scan.³²
- Understanding complexity in health systems: international perspectives. A series of academic papers on the topic.⁸
- The Magenta Book³⁶ is from the UK Government and provides guidance on the evaluation of government interventions, with recommendations for the planning, conduct and management of the evaluation. It provides relevant guidance on complex systems-informed evaluation and a supplementary guide that specifically focuses on handling complexity in policy evaluation. This highlights the challenges of complexity to policy evaluation and the importance of its consideration in commissioning and managing interventions, including guidance on the approaches available to support such projects.³⁶
- A report published by the Centre for Complexity Across the Nexus and commissioned by the Department for Environment, Food and Rural Affairs provides a framework for evaluation, specifically to support evaluations of government policy to consider the implications of complexity theory.³⁷

Complexity and the research perspective

We aim to encourage wider awareness, understanding and use of ‘complexity-informed’ research,⁸ by which we mean research that gives sufficient and appropriate consideration to all of the sources of complexity outlined in the previous section. There are several overlapping perspectives that can be employed in complex intervention research (Box 2), each associated with different types of research questions. Examples of complex intervention studies taking different research perspectives are given at the end of this subsection (see Box 4).

Efficacy perspective

Research taking an efficacy or effectiveness perspective is principally concerned with obtaining unbiased estimates of the average effect of interventions on predetermined outcomes. Studies from an efficacy perspective aim to test hypotheses about the mechanisms of action of interventions.³⁸ Such research aims for high internal validity, taking an explanatory approach to test causal hypotheses about the outcome(s) generated by the intervention. This contrasts with effectiveness studies, which take a more pragmatic approach. The PRagmatic Explanatory Continuum Indicator Summary (PRECIS)-2 tool³⁹ identifies nine domains on which study design decisions may vary according to the perspective or approach. Few studies exhibit all of the characteristics of a purely efficacy or a purely effectiveness perspective; there is not necessarily a dichotomy between efficacy and effectiveness studies, rather a continuum.

Studies from an efficacy perspective will typically evaluate interventions in idealised, controlled conditions among a homogeneous group of highly selected participants using a proximal outcome, such as disease activity, with the intervention delivered in a standardised manner with high fidelity to a protocol or manual, by highly trained practitioners, without the flexibility and variability that is likely to occur in real-world practice. Evidence from an efficacy study indicates whether or not an intervention can work in idealised conditions. Typically, efficacy studies control for contextual variation, so do not usually help to illuminate context dependence. An efficacy perspective could be taken to the development, feasibility and evaluation phases of intervention research. However, if found efficacious, this would not provide sufficient evidence for implementation. For example, testing whether or not a vaccine is efficacious in preventing infection under optimal conditions would be undertaken prior to developing a delivery programme, which would then need to be tested for effectiveness in practice. Another example is testing a psychosocial intervention for efficacy in optimal conditions, using experienced practitioners and carefully selected patients, a precursor to a further research cycle to refine and test the intervention for effectiveness in real-world settings.

BOX 2 Continuum of research perspectives

Research perspectives that are used in the development and evaluation of interventions are best not thought of as mutually exclusive. The types of questions that these perspectives can be used to answer include:

- Efficacy perspective: to what extent does the intervention produce the intended outcome(s) in experimental or ideal settings?
- Effectiveness perspective: to what extent does the intervention produce the intended outcome(s) in real-world settings?
- Theory-based perspectives: what works in which circumstances and how?
- Systems perspective: how do the system and intervention adapt to one another?

Adapted with permission from Skivington *et al.*²⁶ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <https://creativecommons.org/licenses/by/4.0/>. The box includes minor additions and formatting changes to the original box.

Effectiveness perspective

Research from an effectiveness perspective seeks to answer pragmatic questions about the effects that are produced by interventions in routine practice to directly inform the choice between intervention options.³⁸ Effectiveness studies aim to test an intervention in samples and settings representative of those in which the intervention would be implemented in everyday practice, usually with a health or health service outcome. Flexibility in intervention delivery and adherence may be permitted to allow for variation in how, where and by whom interventions are delivered and received. Standardisation of interventions may relate more to the underlying process and functions of the intervention than on the specific form of components delivered.⁴⁰ For example, the INCLUSIVE trial⁴¹ assessed the effectiveness of the Learning Together programme of restorative justice to reduce bullying and aggression in schools. Although key intervention inputs were provided to all intervention schools with the aim that key functions were delivered to trigger the theorised mechanisms of change, each school was encouraged to ensure that the form of local implementation was appropriate for their students and context, with scope for locally decided actions.⁴¹

Theory-based perspective

The primary aim of adopting a theory-based perspective is to provide evidence on the processes through which interventions lead to change in outcomes and what prerequisites may be required for this change to take place, thus exploring how and why they bring about change. This differs from developing or evaluating interventions using an effectiveness perspective, which focuses on identifying whether or not they 'work' based primarily on average estimates of effect. It also differs from an efficacy perspective, as theory-based perspectives explore interventions in practice, taking account of context, and often explore more than one single theoretical account of how the intervention may work. Such approaches to evaluation aim to broaden the scope of the evaluation to understand how an intervention works and how this may vary across different contexts or for different individuals.⁴² In research taking a theory-based perspective, interventions are developed and evaluated through a continuous process of developing, testing and updating programme theory. Research from this perspective can generate an understanding of how mechanisms and context interact, providing evidence that can be applied in other contexts. For example, there are numerous mechanisms by which group-based weight loss interventions may bring about behaviour change. Change may be motivated by participants' relationship with the facilitator or by interaction and sharing of experiences among members of the group, as well as by the specific content of the intervention. Whether or not and how such mechanisms generate outcomes will depend on the context in which the intervention is being applied.⁴³ Thus, the impacts of interventions cannot confidently be determined in the absence of knowledge of the context in which they have been implemented.⁴⁴

There are several approaches that take a theory-based perspective (Box 3 shows a selection of examples). They are 'methods neutral' in the sense that they draw on both quantitative and qualitative study designs to test and refine programme theories.

Systems perspective

A systems perspective suggests that interventions can be better understood with an examination of the system(s) in which they are embedded or the systems that they set out to change. A systems perspective treats interventions as events within, or disruptions to, systems.²⁴ The properties of a system cannot be fully explained by understanding only each of the system's individual parts.⁴⁹ This perspective is concerned with an awareness and understanding of the whole system:

The essential point is that the theory driving the intervention is about the dynamics of the context or system, not the psyche or attributes of the individuals within it.

Hawe et al.²⁴

BOX 3 Different approaches to taking a theory-based perspective

Fourth Generation Evaluation⁴⁵

This approach to evaluation was developed in the 1980s; it aims to move beyond 'just getting the facts—to include the myriad human, political, social, cultural and contextual elements that are involved'.⁴⁵ To do this, the involvement of stakeholders and their 'claims, concerns and issues'⁴⁵ are essential.

Theory of change approach^{46,47}

'A systematic and cumulative study of the links between activities, outcomes, and contexts of the initiative'.⁴⁶ This involves developing 'plausible, doable, and testable'⁴⁶ programme theories in collaboration with stakeholders to determine the intended outcomes of the intervention, the activities required to achieve those outcomes and the potential influence of contextual factors. The programme theory determines which outcome and interim measures should be collected in evaluation and which contextual factors should be considered. If activities, context and outcomes occur as expected in the prespecified theory of change, then the outcomes can be attributed to the intervention.

Realist Evaluation⁴⁸

Realist evaluation sets out to answer 'what works in which circumstances and for whom?'.⁴⁸ The important aspect of realist evaluation is that the intervention's explanatory components – context(s), mechanisms and outcomes – are identified, articulated, tested and refined. The development of context-mechanism-outcome configurations provides plausible explanations for the observed patterns of outcomes, and a key purpose is to test and refine programme theory based on the evaluation findings.

Key to a systems perspective is considering the relationships between intervention and context, engaging with multiple perspectives, and identifying and reflecting on system boundaries.⁵⁰ A systems perspective encourages researchers to consider how the intervention may be influenced by and impact on many elements of the system, and over an extended period of time. Rather than focusing on a narrow and predetermined set of individuals and outcomes within a fixed time period, a systems perspective will aim to consider multiple ways by which an intervention may contribute to system change. These may occur through multiple, often indirect, routes over an extended time period and through spillover and diffusion processes, some of which may be unintended. It is not usually possible to take account of a whole, often open, system; therefore, to make an evaluation tractable it will be necessary to limit its scope by determining a system boundary and restricting the range of potential mechanisms or explanations that are investigated.

Summary

Interventions rarely achieve effects evenly across individuals, populations and contexts. For complex intervention research to be useful to decision-makers it needs to take into account the complexity that arises both from the intervention's components and from the intervention's interaction with context. Intervention effects can be dependent on, can influence or change contexts, and this interdependence can vary over time. Complex intervention research should be conducted with an awareness of these multiple potential sources of complexity, with the design of any specific research study adopting the research perspective that is most suited to the research question that it aims to address (see *Box 4* for examples). The preponderance of complex health intervention research to date has taken an efficacy or effectiveness perspective, and there will continue to be many situations in which research questions for which these perspectives are most suited will continue to be prioritised. However, for many interventions, the most critical research questions and the needs of decision-makers who use research evidence are not met by research that is restricted to questions of efficacy and effectiveness;

Efficacy perspective

Initial trials of NRT for smoking cessation focused on establishing the efficacy of different forms of NRT. One example is for nasal nicotine spray, for which the efficacy study included a highly selected group of participants (in that they were patients attending a smokers' clinic). It suggested that the intervention was efficacious but that it would require further testing for generalisability to other smokers and settings, and to compare it with other forms of NRT.⁵¹

Effectiveness perspective

Two school smoking education programmes were evaluated and found to be efficacious in terms of delaying the onset and reducing the uptake of smoking in young people, and were subsequently evaluated for effectiveness under normal classroom conditions using a cluster RCT.⁵² The research found no significant differences in uptake of smoking between intervention and control groups. The authors suggest that the experimental conditions in which the interventions were originally tested may be associated with success, in contrast to this study where the programmes were taught under typical classroom conditions, by usual classroom teachers. It was recommended that further work be carried out to develop interventions that are effective in practice and that it is important to formally 'field test' under usual conditions before widespread dissemination.

Theory-based perspective

Although there has been promising evidence about smoking cessation programmes, they do not necessarily work for everyone in every context. Further investigation using a theory-based approach is, therefore, appropriate to provide a better understanding of how such interventions work, for whom and why.⁵³ An example is research that explores the perspectives of smoking and non-smoking pregnant women with regard to smoking in pregnancy, and relating this to anti-smoking interventions, identifying why standard cessation efforts may not be successful for some women. For example, rather than pregnant smokers being ignorant of the facts of smoking in pregnancy (the key issue that mass-media interventions target), they may be aware of the facts but do not place credibility in them in the same way as non-smoking pregnant women do, instead favouring information from family and friends.⁵⁴

Systems perspective

The strategy to evaluate the legislation to ban smoking in enclosed public places in Scotland took a comprehensive approach that involved seven linked studies.⁵⁵ The primary aim of the legislation was to reduce the negative health impact of exposure to environmental tobacco smoke, which has been evaluated.⁵⁶⁻⁵⁸ In addition, included in the evaluation strategy was research to determine whether or not the legislation had an impact on smoking quit attempts and successful quitting;⁵⁹ those at particular risk, such as bar workers;⁶⁰ and the performance of the hospitality sector (that has potential impacts for unemployment in certain areas and a plausible impact on health).⁶¹ Studies were also commissioned to explore the potential effect of the legislation on the context, for example smoking culture and social norms about smoking, and the potential harmful effects, including the risk of increased exposure to environmental tobacco smoke in the home among children.⁵⁷ Although not explicitly theorised as taking a systems perspective, the evaluation was informed by a programme theory that hypothesised multiple effects across a range of outcomes, settings and population groups. In retrospect, a systems perspective can be used to understand how and why these effects occurred in an even broader context, by theorising smoke-free legislation as an event that occurred at a critical time in the history of the system and activated further system change including, for example, the growing adoption of smoke-free homes and the greater advocacy and public support for further tobacco control.⁶²

NRT, nicotine replacement therapy.

therefore, a wider range of research perspectives and methods needs to be considered and used by researchers, and supported by funders. This may particularly be the case in health service delivery and public health research, in which organisational-level and population-level interventions and outcomes are research priorities that are not well served by traditional methods.^{6,7} Key questions for evaluation include identifying generalisable determinants of beneficial outcomes and assessing how an intervention contributes to reshaping a system in favourable ways, rather than the dominant focus on the binary question of effectiveness in terms of individual-level outcomes. In the next section, we set out the revised framework that aims to help research teams consider the research perspective(s) most suited to the research challenge that they are aiming to address, to prioritise research questions, and choose and implement appropriate methods.

KEY POINTS 1 Intervention complexity and research perspectives

- Complexity arises from properties of the intervention, from the context in which it is delivered and from the interaction between the two.
- Consider different research perspectives and how they could be used to conceptualise and evaluate the intervention, its place in the wider system in which it is implemented and the interdependence between the two over time.
- The key to choosing a research perspective is to consider the available approaches and use those that are most helpful, given your research questions.
- Complexity-informed research with an awareness of system(s) can encourage (1) researchers to develop research questions that take into account the wider contextual factors that influence and/or are influenced by an intervention over time, and (2) researchers, funders, practitioners and policy-makers to develop, evaluate and implement interventions using the most appropriate tools and methods.

Framework for the main phases and core elements of complex intervention research

Overview

This section presents an overview of the main phases and core elements of complex intervention research.

Figure 3 shows the four main phases of intervention research: development or identification, feasibility, evaluation and implementation. Each phase is connected to a set of core elements considering context, developing and refining programme theory, engaging stakeholders, identifying key uncertainties, refining the intervention and economic considerations. We recommend that these elements are revisited continually throughout the research process and especially before moving between phases (e.g. between feasibility assessment and evaluation). It is vital that stakeholder input, contextual considerations and economic considerations are included early and continually throughout the research process, given that there is little point in going through the other phases if ultimately the intervention is unlikely to ever be implementable in real-world practice. *Case study 1* provides an example of a complex intervention throughout its journey from development to implementation and its adaptation to different contexts.

A complex intervention research programme may begin at any phase, depending on what the key uncertainties are about the intervention in question. The research may involve numerous iterative stages. Repeating of phases is preferable to automatic progression if uncertainties remain unresolved. As with drug development research, for which only a tiny minority of promising drug discoveries make it through to implementation,⁶³ it is likely that many new intervention ideas will prove not to be worthwhile. If the feasibility phase shows that the initial intervention idea is not deliverable or acceptable, a decision is needed on whether to stop the research altogether or to do further

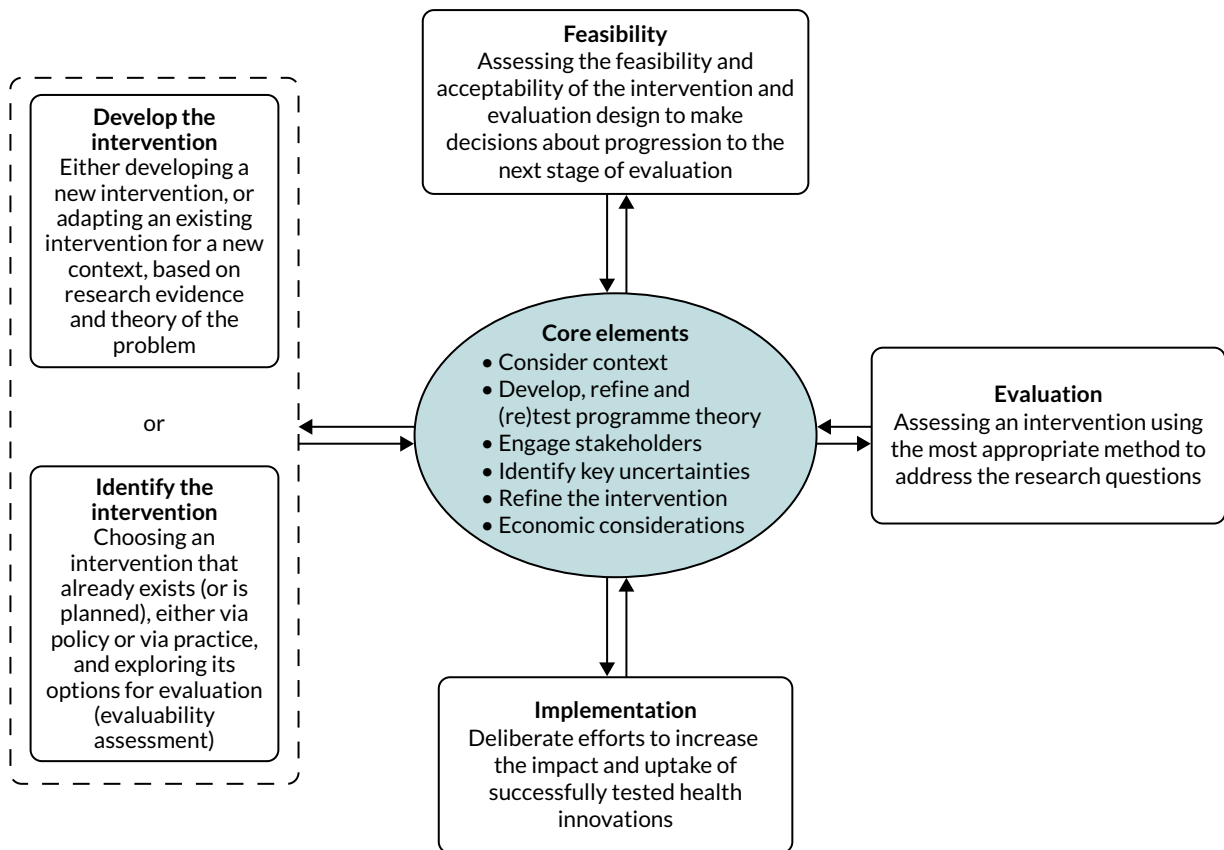


FIGURE 3 Main phases and core elements of complex intervention research. The diagram illustrates the non-linear, iterative pathway through the research process. The core elements should be repeatedly considered and may lead to further work at the same phase or movement to another phase. Reproduced with permission from Skivington *et al.*²⁶ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <https://creativecommons.org/licenses/by/4.0/>. The figure includes minor additions and formatting changes to the original figure.

Definitions of terms used in *Figure 3*:

- **Context:** any feature of the circumstances in which an intervention is conceived, developed, evaluated and implemented.
- **Programme theory:** how an intervention is expected to lead to its effects and under what conditions. The programme theory should be tested and refined at all stages and used to guide the identification of uncertainties and research questions.
- **Stakeholders:** those who are targeted by the intervention or policy, involved in its development or delivery, or more broadly those whose personal or professional interests are affected (i.e. who have a stake in the topic). This includes patients and members of the public, as well as those linked in a professional capacity.
- **Refinement:** the process of 'fine tuning' or making changes to the intervention once a preliminary version (prototype) has been developed.
- **Uncertainties:** identifying the key uncertainties that exist given what is already known and what the programme theory, research team and stakeholders identify as being most important to find out. These judgements inform the framing of research questions that, in turn, govern the choice of research perspective.
- **Economic considerations:** what are the comparative resource and outcome consequences of the interventions for those people and organisations affected?

development work, rather than press ahead with an effectiveness study. In the case of policies or programmes, for which stopping the intervention is not within the researchers' control, an evaluability assessment (see *Evaluability assessment*) may show that the intervention cannot be usefully evaluated in its current form and that further development is needed before an informative evaluation can be undertaken.

Core elements

Parts of this text have been reproduced with permission from Skivington *et al.*²⁶ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <https://creativecommons.org/licenses/by/4.0/>. The text below includes minor additions and formatting changes to the original text.

We have identified six core elements that require consideration throughout all of the phases of complex intervention research. These are (1) context; (2) development and ongoing iteration of programme theory; (3) involvement of relevant stakeholders; (4) identification of key uncertainties to define and prioritise research questions; (5) ongoing intervention refinement; and (6) economic considerations. This section provides an overview of these core elements, with the following sections linking them to the four research phases (development, feasibility, evaluation and implementation).

Context

Context: ... any feature of the circumstances in which an intervention is conceived, developed, implemented and evaluated.

Reproduced with permission from Craig et al.¹⁰ Contains information licensed under the Non-Commercial Government Licence v2.0

Context is dynamic and multidimensional. Key dimensions include physical, spatial, social, cultural, digital, political or economic features of the circumstances in which interventions are implemented. Context 'interacts, influences, modifies and facilitates or constrains the intervention'.⁶⁴ How an intervention works may vary across context and what works in one context may not be as effective, or may even be harmful, in different contexts (see *Case study 2*). The impact of a new intervention will depend on what provision already exists. Interventions may need to be explicitly adapted to fit different contexts (*Box 5*). As well as being influenced by context, many 'interventions work to change the context within which health is produced and become part of the context for future health behaviours, interventions and outcomes' (reproduced with permission from Craig *et al.*¹⁰

BOX 5 Risk of underestimating the importance of context: example

Underestimation of the importance of context, or a failure to take account of all of the relevant contextual dimensions, can lead to difficulties with implementation or a lack of effectiveness.¹⁰ The Nurse–Family Partnership was shown in a number of US trials^{65,66} and one European trial⁶⁷ to be effective in improving a range of pregnancy, parenting, maternal health and social outcomes before it was implemented in the UK as the Family Nurse Partnership. A large, well-conducted UK effectiveness trial of the Family Nurse Partnership found that, despite successful implementation, there was no improvement across four primary outcomes. The trial report concluded that benefits could not be assumed in 'different health-care settings and service populations, even when objective programme fidelity from US trials was maintained'.⁶⁸ Compared with the US settings in which the intervention had first been tested, the UK has well-organised health visiting services to which the Family Nurse Partnership may not have added a great deal.⁶⁹ The intervention needed further development to better meet the requirements of a UK context [see www.dartington.org.uk/fnpadapt (accessed 1 March 2021)].

Contains information licensed under the Non-Commercial Government Licence v2.0). Theorising interventions as interruptions in systems highlights the dynamic interdependence of intervention and context over time. Taking context into account is crucial throughout the research process:

- in theorising how the intervention will generate effects, intended and unintended
- in anticipating and understanding barriers to and facilitators of the implementation, and scaling up and scaling out of interventions
- in the development phase to make it more likely that the intervention is effective and implementable among the target population and setting
- at the evaluation phase, to understand how interventions 'work' in terms of how they interact with their context
- in the reporting of research, to help decision-makers understand whether or not the intervention will work and for whom in the context for which they are responsible.¹⁰

Circumstances surrounding the intervention may change after the study has begun, for example through the introduction of a new policy or service or, in the case of evaluations with long-term follow-up, more fundamental changes in the social or economic context. The context may also interact with the intervention and interventions may be designed to be adaptive and responsive to context.

It is important to develop a clear understanding of what aspects of context may be important in a broader consideration of intervention outcomes and processes, and which of these are particularly relevant to the research questions to be addressed. Considering these wider contextual factors at the earliest stage can help to ensure that the research improves understanding of how an intervention works and how it might operate in or respond to different circumstances. It allows interventions to be developed, refined and adapted so that they fit with the contexts in which they are implemented.

KEY POINTS 2 Context

- The context (see *Context*) in which an intervention is being delivered has a role in its success, or otherwise.
- Theorising an intervention as an event in a system highlights the dynamic interdependence of intervention and context over time.
- Context is multidimensional and it is important to be clear about which aspects of context are likely to influence the delivery, effectiveness, reach and sustainability of the intervention.
- An understanding of the interdependence between context, intervention and outcomes should be developed at the outset and continuously revisited and documented throughout the research process.
- Clear understanding, description and explanation of context enables judgements to be made about transferability of the intervention and better enables intervention adaptation and implementation.

SIGNPOSTS TO FURTHER READING 2 Context

- *Taking Account of Context in Population Health Intervention Research: Guidance for Producers, Users and Funders of Research.*¹⁰
- Pfadenhauer *et al.*'s⁷⁰ the Context and Implementation of Complex Interventions (CICI) framework.
- A realist conceptualisation of context.⁷¹
- Viewing intervention context in a more dynamic way through a systems lens provides many challenges to complex intervention research.⁶²

Programme theory

Programme theory describes how an intervention is expected to lead to a set of specified outcomes and under what conditions. It articulates (1) the key components of the intervention and how they interact; (2) the mechanisms of the intervention; (3) the features of the context that are expected to influence those mechanisms; and (4) how those mechanisms may influence the context.⁷²

Case study 3 provides an example of how programme theory has been developed and used. For interventions designed to be adaptable across contexts, the programme theory should articulate the key functions that must be consistently delivered and indicate how far interventions can be responsive to different and changing contexts.^{40,73} A programme theory may include aspects of mid-range or grand theories, but its aim is more specific: to articulate how and why an intervention is expected to work.

Developing programme theory

Programme theory can be used to promote a shared understanding of the intervention and how it works among diverse stakeholders involved in intervention implementation and evaluation, as well as among those consuming and using the research. It can also be used to identify key uncertainties and research questions. Programme theory should be 'adaptive', continually updated, developed and refined as understanding of the intervention and its interaction with the context and system changes over time. The best practice is to develop programme theory at the beginning of the research project based on the evidence base and theory from relevant fields with involvement of diverse stakeholders, and to refine it during successive stages of the research to produce an updated programme theory at project completion. A refined programme theory is an important evaluation outcome and is the principal aim for which a theory-based perspective is taken. Improved theory will help inform transferability of interventions across settings and help produce evidence and understanding that is useful to decision-makers. It may demonstrate that an intervention is less likely to be effective when transferred to a given context; this in itself is a useful outcome.

Where researchers are involved in evaluating an intervention (such as a policy) developed by others, they still need to theorise the intervention before attempting to evaluate it.⁷⁴ This involves defining the content and purpose of the policy clearly, and articulating and refining a programme theory. As with researcher-developed interventions, involvement of stakeholders is critical. Key documents can also be drawn on (e.g. policy documents or existing funding applications) to refine programme theory. Developing a programme theory is part of completing an evaluability assessment of an existing or proposed policy intervention,⁷⁵ discussed further in *Evaluability assessment*.

For further guidance, see *Signposts to further reading 3: programme theory*.

Representing programme theory

Programme theory should always be clearly articulated. It can be helpful to provide a visual representation of aspects of the programme theory. This is particularly useful for shared understanding among stakeholders of the intervention and the potentially multiple ways that it may generate outcomes. It is important to note that any single visual representation will often not be sufficient to fully articulate the programme theory. There are various ways of depicting a programme theory, for example logic models, realist matrices and systems maps. Each method shows something slightly different and is, therefore, useful in different circumstances.

Logic models The terms logic model and programme theory have been used inconsistently in the past, with logic model often being used interchangeably with programme or intervention theory. Given that logic models (a method of visually representing some elements of the programme theory) have

varying levels of detail, typically showing the inputs, outputs and intended outcomes but not always considering the intervention mechanisms or context dependencies, we distinguish the terms to highlight the importance of clearly articulating the full programme theory. Nonetheless, logic models can be very useful alongside programme theory to clarify thinking, to gain a shared definition and understanding of the intervention among diverse stakeholders, and to guide and communicate with others about the outcome evaluation. More complex forms of logic model are increasingly used. They include causal logic models that show mechanisms and pre-conditions (see *Case study 4*) and 'dark logic models' that depict unintended and potentially negative consequences of the intervention (see *Case study 5*).⁷⁶

Realist approaches Programme theory is at the core of realist evaluation. There are a variety of methods for developing and presenting realist programme theory, such as creating multiple draft theories for different aspects of the intervention to inform the overall programme theory⁷⁷ and the use of realist matrices.⁷² A realist matrix is a table addressing aspects of the intervention, such as agency, context, resources, mechanisms and outcomes (see *Case study 6*). Realist matrices present the hypothesised causal mechanisms of the intervention and attempt to address the question 'what works in which circumstances and for whom?'

System maps If taking a systems perspective, it may be appropriate to create a system map that embeds the intervention in the wider dynamic context. Here, the intervention is viewed as an event within a complex system, in which different components of the system and relationships between them will react in various ways, with the potential to affect different outcomes and generate system change. A system map is a conceptual visual map that is created to describe the components of the system and the relationships between them. Group model building and other soft systems methodologies can be used to create system maps using inputs from multiple perspectives to gain a holistic and shared understanding.⁷⁸⁻⁸¹ An initial system map, created with the input of various stakeholders, may help to refine the programme theory and to define the scope of the evaluation. A systems perspective supported by systems mapping will support the inclusion of longer-term, spill-over feedback and potential emergent effects of interventions that would potentially be missed by other research perspectives that tend to focus on more proximal, linear effects. For example, an evaluation of the UK Soft Drinks Industry Levy (SDIL) used system mapping, involving multiple stakeholders, as a starting point for developing programme theory (see *Case study 7*). System maps can also be complementary to the realist approaches described above.⁸²

KEY POINTS 3 Programme theory

- Programme theory describes how an intervention is expected to lead to a set of specified outcomes, and under what conditions. It should be developed at the outset as a collaborative process involving researchers and stakeholders.
- The development of programme theory is an iterative process. An updated programme theory should be produced at the end of each phase of development and evaluation to capture learning and guide next steps.
- Use the programme theory to choose appropriate evaluation outcomes, including measurement and testing of the theory.
- Visual representation of the programme theory can be useful for summarising programme theory and communicating with stakeholders. However, a comprehensive programme theory should always be clearly articulated and not just represented in a logic model(s) or other simplified visual representation.

SIGNPOSTS TO FURTHER READING 3 Programme theory

- Overview of the various types of theory.⁸³
- The use of programme theory within intervention development.¹¹
- The use of programme theory for evaluating policy-led processes and outcomes.⁷⁴
- The 6SQuID model for intervention development provides detail on the use of theory.¹⁷
- Intervention mapping is a theory-led approach to developing interventions and their evaluation.⁸⁴
- Extending logic models with a complexity perspective.⁸⁵
- *What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation of complex population health interventions.*⁸⁶
- *Purposeful Program Theory. Effective Use of Theories of Change and Logic Models.*⁷²
- Realist reviews to inform programme theories.⁸⁷

6SQuID, 6 Steps in Quality Intervention Development.

Stakeholders

Stakeholder engagement is required throughout the research process to maximise the likelihood that the research addresses relevant questions and leads to changes in practice or policy:

Stakeholders: Those who are targeted by the intervention or policy, involved in its development or delivery, or more broadly those whose personal or professional interests are affected i.e. who have a stake in the topic. This includes patients and members of the public as well as those linked in a professional capacity.

Stakeholders should be involved early in the research process and remain engaged throughout. Reciprocity is vital, and the goal should be the development of collaborative relationships that continue throughout the phases of development, evaluation and implementation.

The purpose of stakeholder engagement will differ depending on the context and phase of the research, but could include drawing on stakeholders' practical experience; obtaining stakeholders' perspectives on the problem being addressed by the intervention; gaining stakeholders' input to and shared understanding of the programme theory, including understanding of the context and system properties; identifying potential areas for change; choosing appropriate outcomes; and ensuring that all 'voices' are heard. It can also help with prioritising research questions, choosing the most useful evaluation perspective, building consensus and fostering relationships to make evaluation more straightforward, for example by accessing gatekeepers, supporting intervention refinement and considering issues around transferability and implementation. Where a systems perspective is taken, stakeholder engagement can help to define the system boundaries and, thus, the scope of the evaluation.

The risks of not engaging relevant stakeholders in complex intervention research include the intervention being theorised or considered from a narrow perspective only; the intervention not being readily implementable or supported by existing system context; the intervention being infeasible or unacceptable; and the chosen outcomes not being important to the patients, public, policy or practice. For example, outcomes of surgical interventions for severe and complex obesity depend on patient engagement with new lifestyle and diet, and work better with multidisciplinary care from psychologists and dieticians in addition to the traditional surgical teams.⁸⁸ The development and evaluation of a surgical intervention like this example, therefore, requires multidisciplinary stakeholder engagement, not least with the patients who may receive the intervention. Broad thinking is needed to identify a diverse range of appropriate stakeholders beyond the usual suspects.

Stakeholders from professional groups

Engaging stakeholders from professional groups can be difficult for practical reasons: because they have other priorities or because they have competing or vested interests. For example, social security policy-makers may attach low priority to the health impact of welfare reform because their chief concern is the effects on labour market participation; therefore, engaging them in research to evaluate health outcomes is difficult. There may be drawbacks from aligning research too closely with stakeholder priorities. It is important to elicit their research priorities, but also to consider why they are priorities and what decisions they are likely to influence.

These issues may also apply to researchers, particularly those who have been involved in intervention development and who may have a stake in the success of the intervention.⁸⁹ Studies have shown that early positive results are often not replicated in independent evaluation.⁹⁰ Systematic bias has also been evidenced in drug trials, whereby the products of the company funding the research are shown to be favoured over others.⁹¹ Thorough consideration of the governance issues, conflicts of interest, risks and benefits is particularly important for research partnerships with commercial organisations. Unhealthy commodity industries have been shown to engage in numerous tactics when engaging with research,⁹² for example to divert attention away from population-level to individual-level interventions, despite the former having greater potential to improve population health.⁹³

Researchers need to be aware of these sensitive issues. There should be thorough consideration of the appropriate involvement of identified stakeholders with integrity and transparency. An element of critical (and self-critical) distance is needed without losing the benefits of working effectively with stakeholders (see *Signposts to further reading 4: stakeholders*).

Co-production of interventions with effective collaboration between a group of stakeholders can support the move from interventions being driven in a closed manner by researchers, to interventions being developed in partnership with a range of stakeholders.⁹⁴ As noted above for stakeholder engagement, however, partnership and co-production must be conducted openly and in ways that do not distort research priorities or compromise research integrity.

Patient and public involvement

One group of stakeholders that is particularly important to involve in research is the potential recipients of an intervention. PPI in research is defined as ‘research being carried out “with” or “by” members of the public rather than “to”, “about” or “for” them’ (reproduced with permission from INVOLVE).⁹⁵ As well as collaboratively developing research questions and interventions that are relevant and appropriate to participants, and identifying outcomes that are important to them, appropriate PPI is likely to improve the recruitment and retention in evaluation studies.⁹⁶ In particular, people with lived experience of the condition or issue should be seen as having expertise that others cannot bring to the team.⁹⁷ (*Case study 8* illustrates this ‘patient-as-partner approach’.)

It is critical that PPI is meaningful and robustly conducted, rather than tokenistic.⁹⁸ The type of engagement should also be appropriate to the project, for example to the particular condition or issue of interest, which could be sensitive or stigmatised.⁹⁹ There are numerous challenges to overcome to engage patients and the public effectively in research. They include power imbalances between researchers and patients/public, competing priorities, engagement difficulties and creating effective partnerships. Researchers should consider following a formal framework for engaging, monitoring and evaluating stakeholder involvement (see *Signposts to further reading 4: stakeholders*). To improve PPI, research should also include some evaluation of the engagement process,¹⁰⁰ which should be published where possible, to enable shared learning.¹⁰¹

KEY POINTS 4 Stakeholders

- ‘Stakeholder’ is a wide-reaching term for anyone who has a stake in the intervention and/or its evaluation or implementation; therefore, careful consideration at the outset of the relevant stakeholders (for different phases of the research) is important.
- It is particularly important to involve patients and the public in research. Consider following a predefined framework for PPI in your research.
- Stakeholder engagement needs to be genuine rather than to satisfy criteria for funding, publication, etc. and should include the prioritisation and design of the research rather than merely its delivery.
- Involvement of appropriate stakeholders is crucial to each phase of research, and ultimately to the prospects of the intervention being effective and the research leading to changes in policy or practice. At the same time, an element of critical distance between researchers and stakeholders is required, and researchers need to beware of sensitive issues that involving certain stakeholders may invoke. The nature of stakeholder involvement should be transparent, with upfront and agreed terms of references and processes for declaring conflicts of interest.

SIGNPOSTS TO FURTHER READING 4 Stakeholders

- INVOLVE [www.invo.org.uk/resource-centre/resource-for-researchers/ (accessed 1 March 2020)] is a national advisory group funded by NIHR to support active public and patient involvement in NHS, public health and social care research. There is an online resource to support researchers to involve the public in how research is identified, prioritised, designed, conducted and disseminated. INVOLVE’s Values and Principles Framework¹⁰² aims to support the building of collaborative partnerships in research and to collaboratively produce national standards for public involvement.
- See the National Standards for Public Involvement in Research website for updates.¹⁰³
- NIHR handbook for researchers on PPI in health and social care research.¹⁰⁴
- Different frameworks exist for supporting PPI in research; for a critical overview see Greenhalgh *et al.*¹⁰⁵
- Recommendations for effective engagement with stakeholders for research prioritisation, alongside a checklist of practices.¹⁰⁶
- Reporting of PPI in research: primary research guidelines¹⁰⁷ and systematic review guidelines.¹⁰⁸
- Guidance on collaboration with industry:
 - MRC *Industry Charter*.¹⁰⁹
 - NIHR *Model Industry Collaborative Research Agreement (mICRA Guidance)*¹¹⁰ and MRC *Industry Collaboration Agreement (MICA)*.¹¹¹
 - MRC *UKPRP: Principles for Working with Industry*.¹¹²
 - The Lambert Toolkit for universities and companies that wish to undertake collaborative research projects.¹¹³
- Researchers can draw on different tools to engage stakeholders, for example:
 - For systems methodology – *Systems Concepts in Action: A Practitioner’s Toolkit*,¹¹⁴ and *Soft Systems Methodology in Action*.¹¹⁵ Examples of soft systems methodology include rich pictures, context diagrams, multiperspective diagrams, group model building, laddering, concept maps, cognitive maps and causal loop diagrams.⁸¹
 - The Network for Transdisciplinary Research provides a series of methods and tools for collaboration between experts and stakeholders (e.g. ideas for clarifying expectations and dealing with power issues in the multidisciplinary stakeholder group).^{116,117}

Key uncertainties

A major challenge for complex intervention research is that, at any phase, there are multiple research questions that could be usefully addressed and alternative research perspectives that may seem equally plausible. Key to maximising the efficiency and value of research is to focus on addressing the key uncertainties that exist, taking into account what is already known and what the programme theory, research team and stakeholders identify as being the most important to find out. These judgements inform the framing of research questions that, in turn, govern the choice of research perspective. Even if the fullest possible assessment of an intervention depends on understanding its impact on the system in which it is embedded, that understanding may need to be developed through a series of studies that adopt a range of different perspectives, chosen at each phase to be the most relevant to addressing the key uncertainty at that phase. For example, a common approach that fits well with the linear stages of research that were proposed in the 2000 complex intervention framework,³ is to focus on establishing whether or not the intervention ‘works’ by obtaining an unbiased estimate of the average effect. This is undertaken initially through an efficacy perspective that is applied within tightly controlled conditions, and then through effectiveness research that is conducted in more realistic conditions. For interventions where the programme theory suggests that there are homogeneous effects across population subgroups, with a low degree of system and context dependence, efficacy and effectiveness are likely to be the key initial uncertainties, with implementation and transferability being the focus of later stages of research. However, for interventions that are designed to be flexible and adaptable or that interact with contexts and systems in a dynamic and variable way, the key initial uncertainties may relate less to efficacy and more to acceptability and compatibility with the wider systems in which they are embedded, with the assessment of impact on key outcomes taking place at a later evaluation phase.

At any given phase of research there may be more than one key uncertainty, which can be addressed through separate research questions and parallel paths of inquiry. For example, modelling the effectiveness and economic implications of scaling up an intervention may be worthwhile alongside a development or feasibility study, to reduce the risk of developing and testing an intervention that has little real prospect of being implemented in practice (Box 6).

Prioritising research questions and choosing a research perspective will be influenced both through the identification of key uncertainties and through the assessment of the likelihood of achieving an unequivocal answer to that question. There is a risk that research questions, research perspectives and research designs that are favoured by existing guidance and research funding mechanisms are those that have a greater

BOX 6 When is evidence sufficient?¹¹⁸

Decision analysis can help decision-makers to weigh up the consequences of adopting a new intervention, taking into account evidence of therapeutic effectiveness for both the new and the existing treatments, but also wider costs, such as side effects of treatment, the productivity losses of leaving the illness untreated and the consequences of adopting the intervention if the effectiveness evidence turns out to be false.

Value-of-information analysis can be used to answer the question of whether or not it is worth gathering extra evidence to reduce uncertainty in the adoption decision. Gathering extra evidence is more likely to be worthwhile if a new intervention is expected to be much better than the one currently thought of as optimal, and if the cost of gathering the evidence is small relative to the benefit of making the right adoption decision.

probability of finding a certain answer, even though the question may be of less importance. One of the motivations for updating this framework was to answer calls for a change in research priorities, towards allocating greater effort and funding to research that can have the greatest impact on health-care and population health outcomes. This may require a shift away from, for example, efficacy trials of relatively uncomplicated interventions in tightly controlled conditions, for which research questions are answered with great certainty but for which the translation of this evidence into diverse, real-world settings is often highly problematic,¹¹⁹ requiring heroic assumptions about the potential for generalisation and implementation. In many areas of health intervention research, it may be more useful to give higher priority to an evaluation that is sensitive to complexity, from which improvements in understanding and important implications for decision-makers can be identified, albeit with caveats, assumptions and limitations.⁷ Thus, theory-based or systems evaluations of complex population-level interventions and mixed-methods evaluations that emphasise implementation, context and system fit may provide more useful knowledge to address key uncertainties for decision-makers than evaluations that seek to provide unequivocal answers to tightly specified questions of less importance. For example, decision-makers may place greater value on research from a systems perspective that assesses an intervention in terms of its potential to contribute to long-term impacts on multiple outcomes, through multiple and often indirect pathways that will vary over place and time, than on research from an effectiveness perspective that measures short-term effects on a single prespecified outcome.

Identifying key uncertainties, priority research questions and optimal research perspectives will ideally involve the active participation of a multidisciplinary research team and a diversity of stakeholders, and ideally with independent oversight, perhaps from a funders' steering group. Different stakeholders will have different views on what is the most important to find out. Some may not be in equipoise about the intervention's effectiveness because they are satisfied with what is already known and may, therefore, be more focused on questions of reach, context dependence and issues of implementation; others may feel that greater certainty of effectiveness is required owing to limitations and potential bias in the existing knowledge base. Choice of outcome may also vary according to stakeholder perspective. For example, a local authority evaluating a suite of policies to increase physical activity may prioritise a natural experiment using routinely collected data on pre- and post-intervention use of leisure centres and green space in localities with and without the policy. Researchers may place higher priority on validated measures of physical activity in exposed and unexposed groups to be convinced of the effect of the intervention. An evaluability assessment would be a useful tool in this case to help identify the most worthwhile design and focus of evaluation (see *Evaluability assessment*).

Key uncertainties should be continually revisited as data accumulate and the programme theory and intervention are updated and refined. Formal methods, such as decision analysis, VOI analysis (see *Box 6*) or agent-based modelling (ABM) with uncertainty quantification, may be useful in prioritising research questions and optimising research design.¹²⁰⁻¹²² However, in many situations this will be largely a deliberative process underpinned by the purpose of the evaluation and the needs of research users. Traditionally, research funders and researchers have prioritised research questions and their associated designs and methods that have a good chance of being answered with a high degree of certainty and are strongly influenced by hierarchies of evidence. For many complex interventions, the needs of research users and the complexity of the evaluation may favour decision-making based on a greater diversity of evidence, thus valuing a wider range of research perspectives and methods, and on 'balance of probabilities' rather than 'evidence beyond reasonable doubt'. By involving a wider range of stakeholders in the research team, a consideration of uncertainty in prioritising the next stage of research may lead to a greater focus on reducing uncertainty around important questions than on answering less important questions with certainty; the findings may be equivocal but will nevertheless be of greater value.

KEY POINTS 5 Uncertainties

- Identify key uncertainties that the research aims to address, given what is already known and the intervention programme theory. What are the key uncertainties for decision-makers?
- Formulate and prioritise research questions accordingly.
- Uncertainties will change as evidence accumulates, so should be reviewed at each phase of the research process.
- Engage stakeholders to develop a detailed understanding of where uncertainty lies for those funding, delivering or receiving the intervention.
- Consider both the importance of the uncertainty and the likelihood of reducing that uncertainty when prioritising questions and perspectives.
- Identifying which outcomes are important and/or measurable may help prioritise the focus of research.
- In the case of interventions that are occurring anyway outside the control of the research team, such as changes in policy or health-care delivery, evaluability assessment can support the identification of key uncertainties.

SIGNPOSTS TO FURTHER READING 5 Uncertainties

- Uncertainty and the evaluation of health and social interventions.^{118,123}
- Decision-theoretic approaches for appraising evidence.¹²⁴
- Two discussions of the uncertainty of evidence surrounding face masks for the protection of the public during COVID-19.^{125,126}

Intervention refinement

Within each phase of complex intervention research and on transition from one phase to another, there may be good reason, based on the data collected and/or the development of programme theory, to refine or optimise the intervention:

Refinement: The process of 'fine tuning' or making changes to the intervention once a preliminary version (prototype) has been developed.

O'Cathain et al.¹¹

Intervention refinement (see *Case study 9*) in the development, feasibility and implementation phases can be relatively straightforward, provided that the changes are consistent with the intervention's programme theory. Such refinements are covered by the argument that the form of an intervention can be flexible as long as the key functions specified in the programme theory are delivered.⁴⁰ However, where the changes are substantial and imply a significant change to the programme theory, such adaptation may require a more fundamental rethinking of the research process. This more fundamental kind of adaptation is distinct from refinement, in that it refers to an existing intervention being adapted to a new population, setting, target outcome or context, and is discussed in *Developing an intervention*.

Intervention refinement and flexibility are discouraged in an efficacy evaluation, in which standardisation and consistency of intervention delivery are prioritised to maximise internal validity. However, from a systems perspective, adaptability and responsiveness are likely to be desirable features of the intervention and a key focus of the research. The extent to which variation in programme delivery is desirable in an effectiveness or theory-based evaluation will depend on the extent to which a flexible form of delivery is allowed, while preserving the delivery of the intervention's key functions.⁴⁰

In instances of an 'identified intervention' (i.e. one that is driven by policy or practice, rather than researcher led; see *Identifying an intervention*), limits on refinement may be set by the policy or practice context. Refinement of policy- or practice-led interventions will require negotiation with those co-ordinating the intervention and all other relevant stakeholders, but may be valuably informed by evaluation results.

Rapid refinement of digital interventions

There are specific challenges that are associated with evaluating digital interventions. These interventions offer the potential for wide reach and high impact at a relatively low cost. However, such interventions may require a rapid route to implementation and frequent refinements to avoid becoming 'out of date'.¹²⁷ Considerations specific to evaluating digital interventions have been summarised by Murray *et al.*,¹²⁸ who suggest that traditional evaluation methods, such as RCTs, could be combined with a more iterative approach to intervention development and refinement. To allow for multiple versions of web-based or mobile interventions to be deployed simultaneously, Continuous Evaluation of Evolving Behavioural Intervention Technologies (CEEBIT) trials have been developed.¹²⁹

Transparent reporting of changes

The level of refinement that is considered acceptable at different points in the research process will differ depending on the programme theory and research perspective. However, if refinements are made to an intervention, they should be reported transparently and with justification.

KEY POINTS 6 Intervention refinement

- Refinements to an intervention should retain the key features of the programme theory. Therefore, acceptable boundaries of refinement should be agreed beforehand.
- It benefits all stakeholders for the optimal version of an intervention to be evaluated and/or implemented.
- Ongoing refinement can improve the potential implementability of the intervention.

SIGNPOSTS TO FURTHER READING 6 Intervention refinement

- The INDEX study: guidance on developing complex interventions to improve health and health care.¹¹
- *Normalisation process theory: a framework for developing, evaluating and implementing complex interventions* [see also www.normalizationprocess.org/ (accessed on 1 March 2021)].¹³⁰
- *A conceptual framework for adaptive preventive interventions*.¹³¹
- *Developing and evaluating digital interventions to promote behaviour change in health and health care: recommendations resulting from an international workshop*.¹²⁷
- CEEBIT trials: *Continuous evaluation of evolving behavioural intervention technologies*.¹²⁹

Economic considerations

Economic considerations: *What are the comparative resource and outcome consequences of the interventions for those people and organisations affected?*

Nearly all complex interventions will be costly to implement and will impose costs on a range of populations or organisations. Economic evaluation, the comparative analysis of alternative courses of action in terms of both costs (resource use) and consequences (outcomes and effects), should be a

core component of all phases of intervention research. Early identification of the potential costs and outcomes can help with the setting of research questions and the design of an appropriate economic evaluation (see *Case study 10*). Methods for the economic evaluation of health interventions are well developed.¹³² Here, we concentrate on aspects that are particularly important for complex interventions.

Early engagement of economic expertise will help identify which perspective to adopt to answer questions that matter most to decision-makers.¹³³ Recent National Institute for Health and Care Excellence (NICE) guidance¹³⁴ recognises that complex interventions are likely to have costs and outcomes across different sectors, and recommends the adoption of a broad (e.g. societal) rather than a narrow (e.g. health service) perspective. Use of an impact inventory is a good way of making choices of perspective and of which costs and outcomes to include, explicit, transparent and comprehensive.¹³⁵ Another important initial consideration is what economic evaluation framework should be adopted. Again, the NICE guidance¹³⁴ advocates broad cost-consequence analysis (CCA) or cost-benefit analysis (CBA) approaches because narrower approaches, such as cost-effectiveness analysis (CEA) or cost-utility analysis (CUA), are unlikely to reflect the full range of costs and consequences. These approaches and the choice between them are discussed more fully in *Economic considerations*.

Central to all forms of economic evaluation is the identification, measurement and valuation of the resources and outcomes of both the intervention being evaluated and the appropriate comparators. What to include may not be obvious to evaluators given the variety of resources required to deliver a complex intervention and the range of intended and unintended consequences for those directly and indirectly affected.¹³⁶ Engaging stakeholders (organisations and individuals involved in commissioning and delivering the intervention or expected to experience additional costs, savings, benefits or negative impacts) in the identification process is essential. Economic considerations are often very important to decision-makers because many of their decisions concern the optimal allocation of scarce resources. A decision-modelling framework may be useful to guide and synthesise complex intervention research.

This understanding of resources (costs) and outcomes, together with evidence from existing studies, such as economic evaluations of comparable interventions, should be incorporated in a programme theory. Programme theories can be used to identify the cost and outcome variables that should be taken forward into the economic evaluation. Interactions between an intervention and the context in which the intervention is delivered may lead to variation in costs and outcomes, underlining the importance of unpicking how causal pathways might work differently for different groups and incorporating this understanding in an explicit programme theory.¹³⁷ Use of programme theories in economic evaluation is considered further in *Developing an intervention*.

A further consideration in designing economic evaluations of complex interventions is the trade-off between equity and efficiency; this is important where differential impacts are likely. The costs and the benefits of complex interventions may vary across groups in the population, such that the most efficient interventions in terms of population health gain may not be the most equitable in terms of reducing (or at least not increasing) health inequalities. Considering who the beneficiaries might be and how costs and impacts might differ between them is a crucial part of planning an economic evaluation of a complex intervention, and methods have been developed for conducting economic evaluation in ways that take account of trade-offs between equity and efficiency.^{138,139}

Economic considerations that should be taken into account at particular phases in the process of developing and evaluating a complex intervention are addressed in the relevant sections below. Ways in which economic evaluation of a complex intervention might differ from standard approaches are identified and new methods highlighted.

KEY POINTS 7 Economic considerations

- Include economists in project teams from the early stages of complex intervention research to ensure that economic considerations are fully incorporated in all important decisions about intervention and evaluation design.
- Inclusion of economic considerations in the programme theory will help with understanding of the problem and shape the design of future feasibility and evaluation studies.
- Depending on the research perspective, it will often be appropriate in complex intervention research to adopt a broad (e.g. societal) rather than a narrow (e.g. health service) perspective to the scope of economic considerations included.
- Equity–efficiency trade-offs should be considered when the intervention is likely to have different costs and benefits for different groups, or where the case for intervention is driven by health equity concerns.

SIGNPOSTS TO FURTHER READING 7 Economic considerations

- Guidance on the development of programme theories and conceptual modelling for economic aspects of complex intervention development.¹⁴⁰
- An introduction to the principles of health economics and economic evaluation relevant to the specific challenges of public health and its aims of improving population health.¹⁴¹
- Guidance for conducting economic evaluations alongside natural experiments.¹⁴²
- Textbook on *Methods for the Economic Evaluation of Health Care Programmes*.¹³²
- NICE methods guidance and reference case for undertaking economic evaluations.¹³⁴
- *Exploring uncertainty in cost-effectiveness analysis*.¹²¹
- *Using Realist Approaches to Explain the Costs and Cost-effectiveness of Programmes*.¹³⁷
- *Conceptualizing a model: a report of the ISPOR-SMDM modelling good research practices Task Force-2*.¹⁴⁰

Phases of research

Parts of this text have been reproduced with permission from Skivington *et al.*²⁶ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <https://creativecommons.org/licenses/by/4.0/>. The text below includes minor additions and formatting changes to the original text.

The following section outlines four phases of research: (1) developing or identifying the intervention, (2) feasibility testing, (3) evaluation and (4) implementation. As illustrated in *Overview*, the phases may overlap and a research programme may begin at any phase, depending on the key uncertainties about the intervention in question. Each phase is discussed in detail below. After each phase, we present a table that addresses the core elements (see *Core elements*) and relevant issues to consider for that phase, followed by a 'key points' box and a final box that provides signposts to further guidance and resources.

Developing or identifying the intervention

Intervention development: *The term 'development' is used here for the whole process of designing and planning an intervention from initial conception through to feasibility, pilot or evaluation study.*

This phase of the research process was referred to as intervention development in the previous version of this framework.¹ However, this did not address the evaluation of interventions where researchers are not involved in their development and do not have experimental control of implementation, which is often

the case in policy- or practice-led interventions. Intervention identification and intervention development represent two distinct pathways of evidence generation to inform policy or practice, characterised by the two pathways shown in *Figure 4*.¹⁴³ In both cases, a key requirement in this phase is to develop a programme theory for the intervention, identify key uncertainties and formulate research questions to decide which research phase to move onto next and what further research to prioritise.

Developing an intervention

The MRC-/NIHR-funded INDEX study has developed comprehensive guidance on intervention development.¹¹ Rather than replicate the INDEX guidance here, we provide an overview of the key steps and encourage readers to access the main INDEX document (see *Signposts to further reading 8: developing or identifying an intervention*).

The INDEX guidance recommends that a robust process of intervention development should involve the 11 steps outlined in *Table 3*.

In practice, intervention development approaches will need to be tailored to the capacity of the team, context and resources. Some of the steps in *Table 3* may overlap or need to be repeated. It may not be possible or necessary to address all of them, but each should be considered for relevance and importance. Teams involved in developing interventions are encouraged to publish their development work so that learning can be shared.

Four additional sets of issues that may need to be addressed are (1) the adaptation of existing interventions to a new context; (2) wider system factors that may need to be taken into account; (3) the consideration of economic factors; and (4) mathematical modelling or simulation of the intervention.

Adapting an existing intervention

Although the INDEX guidance focuses on the development of de novo interventions, it may often be the case that the need for a new intervention is responded to through the adaptation of one or more existing interventions. The term adaptation is distinct from that of refinement, in that it refers to one or more existing interventions being adapted to a new population, setting or context.^{144,145} Similarly, an intervention found to be effective in improving one particular outcome can be adapted to target other

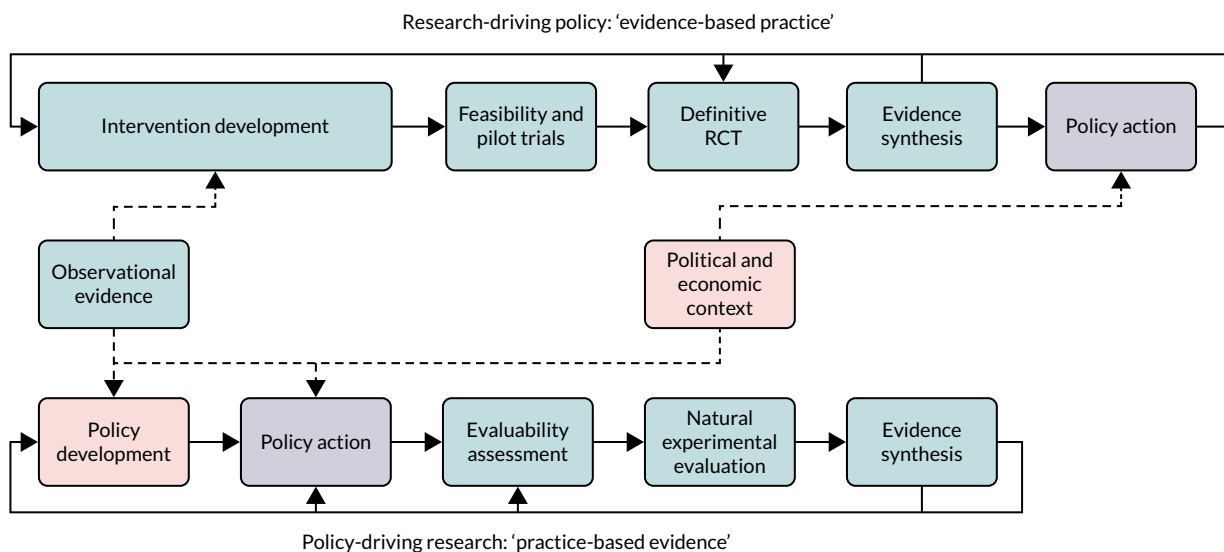


FIGURE 4 Two pathways of evidence generation. Reproduced with permission from Ogilvie *et al.*¹⁴³ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>. The figure includes minor additions and formatting changes to the original figure.

TABLE 3 Framework of actions for intervention development

Action	Consider the relevance and importance of the following:
1. Plan the development process	<ul style="list-style-type: none"> Identify the problem to be targeted and refine the understanding of it throughout the process Assess whether or not the problem is a priority Consider which aspects of the problem are amenable to change Ask whether or not a new intervention is really needed and if the potential benefit of the new intervention justifies the cost of development (and consider the likely short- and long-term cost-effectiveness) Determine the time needed to undertake intervention development Obtain sufficient resources/funding for the intervention development study Draw on one or more of the many published intervention development approaches, recognising that there is no evidence about which approach is best, and apply flexibly depending on the problem and context Involve stakeholders during the planning process (see next table row) Produce a protocol detailing the processes to be undertaken to develop the intervention
2. Involve stakeholders, including those who will deliver, use and benefit from the intervention	<ul style="list-style-type: none"> Work closely with relevant stakeholders throughout the development process: patients, the public, the target population, service providers, those who pay for health and social services or interventions, policy-makers and intervention design specialists Develop a plan at the start of the process to integrate public and patient involvement into the intervention development process Identify the best ways of working with each type of stakeholder from consultation to co-production, acknowledging that different ways may be relevant for different stakeholders at different times Use creative activities within team meetings to work with stakeholders to understand the problem and generate ideas for the intervention
3. Bring together a team and establish decision-making processes	<ul style="list-style-type: none"> Include within the development team individuals with relevant expertise: in the problem to be addressed by the intervention include those with personal experience of the problem, in behaviour change when the intervention aims to change behaviour, in maximising engagement of stakeholders and with a strong track record in designing complex interventions It may be hard to make final decisions about the content, format and delivery of the intervention, so only some team members may do this. There is no consensus about the size or constituency of the team that makes these final decisions, but it is important early on to agree a process for making decisions within the team
4. Review published research evidence	<ul style="list-style-type: none"> Review published research evidence before starting to develop the intervention and throughout the development process (e.g. to identify existing interventions and to understand the evidence base for each proposed substantive intervention component) Look for, and take into account, evidence that the proposed intervention may not work in the way intended
5. Draw on existing theories	<ul style="list-style-type: none"> Identify an existing theory or framework of theories to inform the intervention at the start of the process (e.g. behaviour change or implementation theory) Where relevant, draw on more than one existing theory or framework of theories (e.g. both psychological and organisational theories)
6. Articulate programme theory	<ul style="list-style-type: none"> Develop a programme theory. The programme theory may draw on existing theories. Aspects of the programme theory can be represented by a logic model or set of models. Programme theory can also help the development of economic research questions developed in parallel with the intervention Test and refine the programme theory throughout the development process
7. Undertake primary data collection	<ul style="list-style-type: none"> Use a wide range of research methods throughout (e.g. qualitative research to understand the context in which the intervention will operate and quantitative methods to measure change in intermediate outcomes)

continued

TABLE 3 Framework of actions for intervention development (*continued*)

Action	Consider the relevance and importance of the following:
8. Understand context	<ul style="list-style-type: none"> Understand the context in which the intervention will be implemented. Context may include population and individuals; physical location or geographical setting; social, economic, cultural and political influences; and factors affecting implementation (e.g. organisation, funding and policy)
9. Pay attention to future implementation of the intervention in the real world	<ul style="list-style-type: none"> From the start, understand facilitators of and barriers to reaching the relevant population, future use of the intervention, and 'scale up' and sustainability in real-world contexts
10. Design and refine the intervention	<ul style="list-style-type: none"> Generate ideas about content, format and delivery with stakeholders Once an early version or prototype of the intervention is available, refine or optimise it using a series of iterations. Each iteration includes an assessment of how acceptable, feasible and engaging the intervention is, including potential harms and unintended consequences, resulting in refinements to the intervention. Repeat the process until uncertainties are resolved Check that the proposed mechanisms of action are supported by early testing
11. End the development phase	<ul style="list-style-type: none"> There are no established criteria for stopping the intensive development phase and moving on to the feasibility/pilot or evaluation phases. The concepts of data saturation and information power may be useful when assessment of later iterations of the intervention produces few changes Describe the intervention to facilitate transferability of an intervention outside the original team and location in which it was developed Write up the intervention development process so that judgements can be made about the quality of the process, links can be made in the future between intervention development processes and the subsequent success of interventions, and others can learn how it can be done

Reproduced with permission from O'Cathain *et al.*¹¹ This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>. The table includes minor additions and formatting changes to the original table.

outcomes, for example the A Stop Smoking in Schools Trial (ASSIST) adolescent smoking prevention intervention being adapted to tackle substance misuse and sexual health.¹⁴⁶⁻¹⁴⁸ A well-developed programme theory can help identify what features of the antecedent intervention(s) need to be changed for the intervention to work in its new application.¹⁴⁴

Consider the wider system

Acknowledging that an intervention takes place within a wider system guides the intervention development team to consider the ways in which the intervention and system influence one another, for example via feedback loops and adaptivity (see *Table 2*). Mapping the system or parts of the system can help facilitate this process. See *Case study 11* for an example of how system mapping has been used to support intervention development.

Economic considerations

Consideration of economic aspects at the development phase can help frame the research questions that will underpin the design of an economic evaluation. The review undertaken to inform intervention development (see *Table 3*) should include economic evidence to help decide what economic issues need to be addressed and how economic ideas, such as insights from behavioural economics,¹⁴⁹ can contribute to the programme theory.

Modelling

The development and justification for an intervention, the potential for an existing intervention to be adapted to a different context and/or the identification of uncertainties to be addressed in its

evaluation can often be usefully supported by a mathematical model or simulation of the intervention in silico. This may be particularly useful for upstream interventions and policies, or where a systems perspective is being taken to the research. A range of methods can be used (discussed further in *Types of evaluation design and methods*), including deterministic models, microsimulation models (e.g. to assess tobacco control policy¹⁵⁰) or agent-based models (e.g. to assess the impact of a potential educational campaign on fruit and vegetable consumption¹⁵¹). Assessing and/or further refining these models as data accumulate through all research phases may be an important focus of evaluation activities. For the economic evaluation, the programme theory needs to be translated into a credible conceptual and mathematical structure^{152,153} so that an early-stage decision-analytic model can be formulated that compares the intervention with one or more relevant comparators. This will help to establish the potential cost-effectiveness of the planned intervention and highlight areas of uncertainty that should be explored. VOI analyses can then be undertaken to estimate a monetary value on further research activities, based on the expected gain from reducing uncertainty surrounding the current cost-effectiveness decision.^{154,155} Expected value of perfect parameter information can also be usefully undertaken to identify parameters that drive the uncertainty in the economic model. The evaluation can then focus on those parameters with the most uncertainty.

Identifying an intervention

Often, interventions are policy-driven and/or designed in clinical practice. In such cases, researchers are not usually involved in intervention development, but evidence that they have generated may have informed this process. In some instances, implementation of policy- or practice-led intervention may have taken place without a prior evaluation, in which case researchers may need to plan a retrospective evaluation.

Even if researchers are not involved in the design phase, they should still consider the six core elements (*Table 4*). In particular, they should develop a programme theory and use this to identify key uncertainties. Some policy interventions are based on explicit theory, but many are not. In these cases, identifying the implicit theoretical basis of an intervention is essential to working out how it might be evaluated. This is an important step, even if the programme theory is developed after the intervention has been rolled out. This helps identify mechanisms, important contextual factors and relevant outcome measures.

TABLE 4 Core elements for developing or identifying the intervention

Core elements	Issues to consider	Risk of not considering them
Context	<ul style="list-style-type: none"> • What are the immediate contextual influences on the intervention? • What is the environment in which the evaluation will take place? • Where is the system boundary? 	<ul style="list-style-type: none"> • Lack of understanding of context may result in an intervention that works in one setting, but may be ineffective, or even harmful, when delivered elsewhere
Programme theory	<ul style="list-style-type: none"> • What is the underlying theory or combination of theories? • What are the proposed mechanisms of change? • What are the proposed short-, medium- and long-term outcomes? • What are the wider impacts of the intervention on the system? • Explore the current evidence base, in particular, for information to suggest that the proposed intervention may not work as intended • Consider context (see row above) • The programme theory process is an important step in identifying priority research questions and helping choose an evaluation perspective 	<ul style="list-style-type: none"> • An intervention that has little chance of working even in idealised settings • An intervention that is unlikely to be sensitive to different contexts 'in the real world' • Failure to measure impacts of the intervention that might be important • Unable to advance theory

continued

TABLE 4 Core elements for developing or identifying the intervention (continued)

Core elements	Issues to consider	Risk of not considering them
Stakeholders	<ul style="list-style-type: none"> • Include relevant stakeholders in the development process, in particular for developing theory of the problem; considering system and context; developing programme theory; providing feedback to help refine and improve the intervention; and identifying potential issues with future implementation • Choose different approaches to engage different stakeholders (e.g. evaluability assessment, interviews/focus groups and soft system methodologies) 	<ul style="list-style-type: none"> • Resulting intervention that fails to achieve its proposed effect • Effective intervention that fails to be adopted and implemented • Resulting evaluation that fails to generate useful data on the processes, mechanisms and/or outcomes of the intervention • Reduced chance of impact on policy and practice
Uncertainty	<ul style="list-style-type: none"> • Where are the key areas of uncertainty to be addressed in any evaluation of this intervention? • What are the priority research questions and optimal research perspective to address them? • What is the proposed research design to address these questions and what is the next stage of research? • Is the intervention, the programme theory and the proposed evaluation design sufficiently developed for the next stage of the research to be undertaken? 	<ul style="list-style-type: none"> • Failure to agree and optimise the research perspective and design • Embark prematurely on underspecified research that fails to address the key questions of interest
Intervention refinement	<ul style="list-style-type: none"> • What aspects of the intervention could be improved prior to delivery? 	<ul style="list-style-type: none"> • Delivery of a suboptimal intervention and/or evaluation • Failure to effectively gather meaningful data on potential mechanisms and outcomes of the intervention • Waste of resources
Economic considerations	<ul style="list-style-type: none"> • What is the existing economic evidence? • What is the appropriate economic question: taking account of the system and programme theory/economic conceptual model developed? • What perspective(s) is/are appropriate? • What type of economic evaluation is most appropriate to address the research question? • Should spill-over impacts be included? 	<ul style="list-style-type: none"> • Remaining uncertainty not identified and/or studies duplicate existing evidence • Inappropriate question with limited focus • Economic results that underestimate/overestimate potential cost-effectiveness • Study does not answer questions of interest • Study does not capture the full range of costs and outcomes

KEY POINTS 8 Developing or identifying an intervention

- Specific considerations for intervention development:
 - plan the development process
 - involve stakeholders, including PPI
 - bring together a team and establish decision-making processes
 - review published research evidence
 - draw on existing theories
 - articulate programme theory
 - undertake primary data collection
 - understand context
 - pay attention to future implementation of the intervention in the real world
 - design and refine the intervention
 - consider the wider system within which the intervention operates
 - consider the future evaluation design (that may affect some of the things that you do in the intervention development phase).

KEY POINTS 8 Developing or identifying an intervention (*continued*)

- Specific considerations for intervention identification, in addition to the relevant key points above:
 - identify your starting point within the research process most applicable to the current status of the intervention (i.e. evaluating an existing intervention)
 - consider which natural experimental methods could be used to evaluate the intervention
 - develop and refine programme theory if one does not exist.

SIGNPOSTS TO FURTHER READING 8 Developing or identifying an intervention

- The INDEX study: guidance on developing complex interventions to improve health and health care¹⁵⁶ and the associated reporting guidance GUIDED.¹⁵⁷
- *Six steps in quality intervention development (6SQuID)*.¹⁷
- Intervention mapping is a theory-led approach to developing interventions and their evaluation.⁸⁴
- *Normalisation process theory: a framework for developing, evaluating and implementing complex interventions*.¹³⁰
- Evaluating the public health impact of health promotion interventions: the RE-AIM framework.¹⁵⁸
- *Taking Account of Context in Population Health Intervention Research: Guidance for Producers, Users and Funders of Research*.¹⁰
- *Natural experiments: an overview of methods, approaches, and contributions to public health intervention research*.¹⁵⁹
- Using natural experimental studies to guide public health action.¹⁴³
- Guidance on conducting economic evaluations alongside natural experiments.¹⁴²
- Economic evaluation of public health interventions.¹⁴¹
- Reporting of behaviour change techniques.¹⁶⁰
- A checklist has been produced for intervention description and replication (TIDieR)¹⁶¹ and has been extended to population health and policy interventions (TIDieR Population Health and Policy).¹⁶²

TIDieR, Template for Intervention Description and Replication.

Feasibility

We follow the Consolidated Standards of Reporting Trials (CONSORT) extension for pilot and feasibility trials,¹⁶³ and use the term feasibility to cover both the feasibility of the intervention and the feasibility of the evaluation design. A feasibility study asks whether or not something can be undertaken, should we proceed with it and, if so, how:

Feasibility studies: *Pieces of research done before a main study in order to answer the question ‘Can this study be done?’*

*Eldridge et al.*¹⁶³

Why is a feasibility phase needed?

The feasibility phase explores the feasibility of the intervention and/or the evaluation design to underpin decisions about whether or not and how to progress to evaluation. Although feasibility studies are most commonly undertaken prior to effectiveness evaluation, they can also be necessary prior to an efficacy, a theory-based or a systems perspective evaluation study. This vital preparatory work is sometimes overlooked or rushed,¹⁶³ which leads to evaluation studies that are undermined by problems related to either the intervention or the evaluation design that could have been identified and overcome through appropriate feasibility research. On the other hand, feasibility studies can

sometimes be overspecified and take more time and resources than are necessary to address the key uncertainties with the planned intervention or evaluation design.

A feasibility study is designed to explore the uncertainties that have been identified at the development phase (see *Case study 12*). The aim is then to address these uncertainties to ascertain whether or not it is appropriate to move onto the evaluation phase and optimise study design.

Progression criteria

A feasibility study should be designed to assess progression criteria that relate to the feasibility and acceptability of the intervention and/or the evaluation design. Progression criteria should ideally be developed with input from an independent steering committee and relevant stakeholders, and be agreed and signed off prior to the collection of data.

Progression criteria should be used to guide the decision on whether to proceed to the next stage of evaluation or to undertake further feasibility work, or to return to the development phase and/or terminate the research. Assessment of progression criteria may require both qualitative and quantitative data. Examples include (1) qualitative data, such as the acceptability of the intervention content and mode of delivery for participants and providers, and (2) quantitative data, such as recruitment and retention rates. See the HelpMeDolt! study for an example of progression criteria.¹⁶⁴

Progression criteria should be worded in a way that provides scope for considered discussion among the study team and stakeholders, ideally with an independent steering committee making the final recommendation. The use of a traffic light system for each progression criterion is a useful approach (green = go, where there are no issues identified; red = stop, where issues identified cannot be resolved; and amber = amend, where issues identified can potentially be resolved).

In many cases, the key uncertainties to be addressed prior to the evaluation phase relate solely to aspects of evaluation design. This may be where the development phase has involved substantial empirical assessment of the intervention, where the intervention is not in the control of the researcher and so cannot be changed, or where the intervention has already been tested for efficacy. In these cases, questions of feasibility may relate to evaluation design only, given that there may be no reason to expect that further changes will need to be made to the intervention to successfully conduct the evaluation. If this is the case, feasibility work may take the form of a pilot study: a small-scale version of some or all of the future evaluation study. To expedite the research process, it may be advantageous to commission a full evaluation in which an initial pilot stage is included, rather than a separate feasibility phase incorporating a pilot study. All being well, the pilot stage could then be continued and expanded into a full evaluation, eliminating a substantial delay between the pilot and the evaluation. If there were problems that needed to be addressed in the pilot, such as problems with recruitment, retention or measurement of study participants, the pilot would be stopped and the research design or methods reviewed.

Feasibility of the evaluation design

In terms of evaluation design, a feasibility study should assess key aspects of research design and conduct, which typically include:

- recruitment – time frame, willingness of participants to be randomised, capacity of those involved (e.g. clinicians) to recruit participants and effective strategies
- retention – follow-up rates and reasons for attrition
- sample size – capacity to achieve an appropriate sample size
- outcomes – choice of outcome measures, data collection tools and processes (e.g. missing data on questionnaires or accessibility of routine clinical data), and duration of follow-ups
- analysis – capacity to collect and analyse data within a time frame, and floor or ceiling effects
- unintended outcomes – for example, potential harms or the impact of the intervention on inequalities.

Feasibility of the intervention design

Feasibility studies provide an important opportunity to implement the intervention on a small scale and assess any uncertainties regarding the intervention prior to embarking on a full-scale evaluation. Where prior stages of research have not tested the implementation of the intervention in the same kind of circumstances that will exist in the evaluation study, the feasibility phase is the first and an important opportunity to address questions related to optimal intervention content and mode of delivery; the acceptability of the intervention to participants and providers; adherence; the likelihood of cost-effectiveness; and the capacity of providers to deliver the intervention and fit within the setting. Where the intervention programme theory suggests that contextual or implementation factors may influence the effectiveness of the intervention, these questions should be addressed at the feasibility phase, even if the intervention has already been tested for efficacy.

Many intervention-related questions that might be assessed in a feasibility study relate to process evaluation, which aims to understand the functioning of an intervention by examining implementation, mechanisms of impact and contextual factors.^{5,165}

Depending on the results of the feasibility study, further work may be required to progressively refine the intervention before embarking on a full-scale evaluation. As discussed in *Intervention refinement*, any refinement should have agreed limits and be reported transparently. Ideally, incremental refinements should each be tested separately; however, in practice refinements can be made simultaneously if sufficiently rich data are collected to enable judgements about which refinements are helpful and which are not.

The programme theory should be refined in an ongoing way regularly throughout the feasibility study and indeed all phases of the research process.

Efficacy signals

To justify the expense of a full-scale evaluation from an effectiveness, theory of change or systems perspective, as well as evidence of feasibility it is also helpful to have convincing empirical or theoretical evidence to suggest that the intervention has the potential to be effective. Such evidence may come from prior research conducted from an efficacy perspective that has demonstrated that the intervention can work in ideal, controlled conditions. Such 'efficacy signals' or evidence of potential effectiveness may also be drawn from evidence of the effectiveness of similar interventions for which the programme theory suggests that the effects may be transferable to different populations, settings or outcomes. Feasibility studies should generally not aim to identify efficacy signals or estimate effect sizes for use in sample size calculations, given that these will be very imprecise. However, in some cases, a feasibility study may provide an opportunity to collect data to support the case for potential effectiveness, for example evidence of short-term impacts on proximal, sensitive and intermediate outcomes.

Economic considerations for the feasibility phase

The aim of an economic evaluation conducted within a feasibility study is to explore the likelihood that the intervention will be worthwhile, identify key areas of uncertainty and understand whether or not the data required for a full economic evaluation can be captured. Economic considerations that should be addressed during the feasibility phase include refinement of the programme theory that will underpin the full economic evaluation, development and testing of primary data collection tools, and identification of relevant sources of routinely collected resource use and outcome data. Economic modelling can be undertaken to assess the likelihood that the expected benefits of the intervention justify the costs and to help decision-makers decide whether or not it is worthwhile to proceed to a full-scale evaluation. A VOI study (see *Developing an intervention*) should be considered if one was not included in the intervention development phase. If a VOI study was included then it should be updated at this stage. It should be noted that such analyses may increase rather than reduce uncertainty by identifying new sources of variation in costs and outcomes. This can be helpful when a policy-maker believes that an intervention is cost-effective, despite a lack of evidence.

Case study 10 demonstrates how economic considerations at the design and feasibility phases were used to build the case for a full-scale trial of a complex intervention.

Evaluability assessment

An alternative or complementary approach to feasibility is evaluability assessment, a systematic, collaborative approach that is often used with policy-driven or practice-driven (identified) interventions that can equally be used to plan the development and evaluation of a researcher-developed intervention. It aims to determine whether or not an intervention can usefully be evaluated and, if so, which methods are the most appropriate. Its purpose is to ensure that evaluation resources are used efficiently and to minimise research waste on evaluations of poorly designed and/or implemented interventions, or those for which the available data or the method of implementation do not permit outcome evaluation. Evaluability assessment involves collaboration with stakeholders to reach agreement on what are the expected outcomes of the intervention, what data could be collected to assess processes and outcomes and what are the options for designing the evaluation.⁷⁵ The end result is a recommendation on whether or not an evaluation is feasible, should be undertaken and can be carried out at a reasonable cost, and by which methods this can be undertaken.¹⁶⁶ Five questions to assess evaluability have been suggested by Ogilvie *et al.*:¹⁶⁶

1. Where is a particular intervention situated in the evolutionary flow chart of an overall intervention programme?
2. How will an evaluative study of this intervention affect policy decisions?
3. What are the plausible sizes and distribution of the intervention's hypothesised impacts?
4. How will the findings of an evaluative study add value to the existing scientific evidence?
5. Is it practicable to evaluate the intervention in the time available?

Case study 13 provides an example of using evaluability assessment of existing community interventions to identify practice-based evidence. Table 5 summarises the core elements for the feasibility phase.

TABLE 5 Core elements for the feasibility phase

Core elements	Issues to consider	Risk of not considering them
Context	<ul style="list-style-type: none"> • Determine the contextual dependencies for the intervention (e.g. factors affecting delivery and receipt) and evaluation (e.g. factors affecting collection of outcome data) • Consider how these contextual dependencies may change for a future evaluation or implementation within different settings • Incorporate aspects of process evaluation 	<ul style="list-style-type: none"> • An intervention and/or evaluation that does not demonstrate feasibility and/or acceptability, with no understanding of why • An intervention and/or evaluation that fails to meet progression criteria
Programme theory	<ul style="list-style-type: none"> • Begin with a proposed programme theory • Use programme theory to identify measurable outcomes • Refine the programme theory as data are gathered to inform how the intervention operates • Incorporate aspects of process evaluation to explore potential mechanisms of impact • Produce a revised programme theory by the end of the phase 	<ul style="list-style-type: none"> • Key contextual dependencies are neglected resulting in an intervention that fails to operate as designed
Stakeholders	<ul style="list-style-type: none"> • Input is critical to ensure that relevant data are collected to inform future delivery, evaluation, implementation and impact • Input is critical to programme theory and key uncertainties 	<ul style="list-style-type: none"> • Lack of useful and convincing data to inform progression to the next phase of evaluation and/or implementation • Planning and progression to the next phase will not meet stakeholders' needs or benefit from their critical insights • Risk of proceeding with a full-scale definitive trial for an intervention that is highly unlikely to be implemented in practice

TABLE 5 Core elements for the feasibility phase (continued)

Core elements	Issues to consider	Risk of not considering them
Uncertainty	<ul style="list-style-type: none"> • What are the key questions that need to be addressed to progress to an evaluation study? • Translate these into progression criteria and use these progression criteria as the principal questions to be addressed in the feasibility study • Prior to evaluation, the feasibility study may assess the acceptability of intervention and evaluation to participants, providers and the research setting • Consider undertaking an evaluability assessment to determine whether or not and how an evaluation should be undertaken 	<ul style="list-style-type: none"> • Progression to a large-scale, potentially expensive evaluation study that then fails because the evaluation design and methods and/or the intervention design and implementation have not been sufficiently tested for feasibility for acceptability • Contextual and implementation factors undermine the intervention in ways that were not experienced or anticipated in prior development or efficacy research
Intervention refinement	<ul style="list-style-type: none"> • Agree boundaries and limits on how much the intervention or evaluation design can be refined • Report all refinement transparently 	<ul style="list-style-type: none"> • Risk of proceeding to the next phase of evaluation with a suboptimal intervention and/or evaluation design
Economic considerations	<ul style="list-style-type: none"> • Identify and measure data on resource use across a range of cost categories consistent with the perspective of the study • Identify relevant unit cost sources for the valuation of resource use • Consider logistical issues with collecting resource use data (e.g. practical and logical) • Develop economic data collection tools and systems. Test the acceptability of data collection tools • Consider whether or not qualitative and process evaluation can feed valuable information into the economic evaluation alongside the feasibility study or economic model • Identify and measure suitable economic outcomes (preferably preference based) • Consider which economic evaluation frameworks are suited to the research question (possibly consider using more than one framework) • Consider the use of VOI: potentially useful for identifying areas of particular uncertainty and further research, and identifying a monetary value limit for the costs of any further research activities 	<ul style="list-style-type: none"> • Underestimation of time and resources required to collect resource use and outcome data: could ultimately lead to 'failure' of the trial or 'ineffective intervention' owing to poor data collection • Data collection tools or systems that are unacceptable to patients/clients or those delivering the intervention • Inappropriately designed economic evaluation for a definitive trial (e.g. full range of resources and outcomes not correctly identified and measured) • Limited economic evaluation frameworks considered • Expense of proceeding with a full-scale definitive trial for an intervention that is highly unlikely to be cost-effective or implementable in practice

KEY POINTS 9 Feasibility

- Assess the feasibility of the intervention design and/or the evaluation design.
- Use qualitative and quantitative methods to explore and address uncertainties.
- Involve stakeholders to ensure that relevant data are collected for future decision-making.
- If needed, refine the intervention and/or evaluation design (within agreed boundaries) and report changes transparently.
- Use predefined progression criteria to guide decisions about proceeding to evaluation.
- Identify and measure key resource use and outcomes for the purposes of designing a full economic evaluation.
- Consider whether or not it is appropriate and relevant to undertake early economic modelling and VOI analyses.
- Consider conducting an evaluability assessment to determine whether or not and how a full-scale evaluation should be undertaken.

- More detailed guidance on feasibility studies is in preparation by the MRC/NIHR GUEST project team.^{167,168}
- *Taking Account of Context in Population Health Intervention Research: Guidance for Producers, Users and Funders of Research.*¹⁰
- *Process Evaluation of Complex Interventions: Medical Research Council Guidance.*⁵
- *Applied Methods of Cost-Benefit Analysis in Health Care.*¹⁶⁹
- *Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement.*¹⁷⁰
- *Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report.*¹⁷¹
- *Evaluability Assessment: A Systematic Approach to Deciding Whether and How to Evaluate Programmes and Policies. Working Paper.*⁷⁵
- *Assessing the evaluability of complex public health interventions: five questions for researchers, funders, and policymakers.*¹⁶⁶

Evaluation

Evaluation: . . . enables judgments to be made about the value of an intervention. It includes whether an intervention ‘works’ in the sense of achieving its intended outcome identifying what other impacts it has, theorising how it works, taking account of how it interacts with the context in which it is implemented, how it contributes to system change, and how the evidence can be used to support real-world decision-making.

In *Purpose of evaluation*, we discuss the purpose of evaluation and suggest that the primary focus should be providing useful evidence rather than minimising bias. *Research perspectives and evaluation* considers how research perspectives should inform design choices. The remaining sections consider choice of outcomes (see *Choice of evaluation outcomes*), provide information on selection of evaluation designs (see *Types of evaluation design and methods*), consider understanding processes (see *Understanding processes*), consider economic considerations (see *Economic considerations*) and consider good practice in reporting evaluations (see *Reporting evaluations*).

Purpose of evaluation

A key change in the field of evaluation has been a shift towards assessing the ‘usefulness’ of information for decision-making, in contrast to focusing exclusively on obtaining unbiased estimates of effectiveness.¹⁷² In a trial, randomisation, coupled with other elements of trial design (such as concealment of the allocation sequence and blinding of participants and researchers), can minimise bias. A statistical model of the process that determines exposure to the intervention serves the same purpose in a natural experimental study (see *Natural experimental designs*); however, this may not be as effective as randomisation at creating balanced (or ‘exchangeable’) intervention and control groups. However, to maximise the usefulness of the evidence for decision-making, other elements are required, such as the use of programme theory and consultation with stakeholders to identify research questions that matter and a process evaluation to assess elements of implementation, context and mechanisms of impact.⁵ We emphasise that here we are advocating for additions to the toolkit of methods, rather than replacements for traditional methods, when they are appropriate. A more fundamental challenge to the conventional idea that precise and unbiased estimates of effectiveness are the principal goal of evaluation is the proposal that improving theories about how to intervene should be the ambition. In this view, effect estimates are inherently context bound, so that average effects are not a useful guide to decision-makers working in a particular context. More useful may be contextualised understandings of how an intervention works and details on the most important enablers and constraints on its delivery across a range of settings.

Often, questions about the delivery, reach and maintenance of interventions are assigned to 'implementation research' and are explored separately from evaluation questions. Some frameworks, such as RE-AIM,¹⁵⁸ try to pull them together. A key recommendation here is that implementation questions should be considered alongside evaluation questions from the outset. Moving away from a narrow 'effectiveness' focus should increase the relevance of the evidence produced by evaluation studies and increase the speed with which interventions that make a useful contribution can be implemented within policy and practice. This is about explicitly considering from the outset how the research will make a difference, who will use it and who could benefit from it. For example, who needs to know the answer to the research questions? One component of a research proposal should be a 'Pathways to Impact' statement or an implementation plan; these are often required by funders,¹⁷³ but even if not it is a good idea to have given it due consideration.

Research perspectives and evaluation

Table 6 sets out the main research perspectives used in evaluation and relates them to study designs: (1) efficacy, (2) effectiveness, (3) theory based and (4) systems. Although some designs and perspectives tend to be closely associated with one another, the mapping is not fixed and designs can be combined to answer a range of research questions about any given intervention. Efficacy and effectiveness evaluations do not represent a dichotomy but a continuum,³⁹ and it is 'likely impossible to perform a pure efficacy study or pure effectiveness study'.¹⁸⁹ An effectiveness evaluation can draw on a theory-based perspective to help to explore the impact of context and mechanisms on the outcomes being assessed. For example, a theory-informed process evaluation was conducted alongside an effectiveness trial of the 3D intervention¹⁹⁰ (to improve the management of multimorbidity in primary care, see *Case study 14*) to understand how and why the intervention was effective or not, as well as to understand the effect of context. An effectiveness evaluation can also draw on a systems perspective to help determine which outcomes should be evaluated. For example, in determining an evaluation strategy to explore the impacts of the SDIL, researchers drew on a systems perspective and methods (see *Case study 7*).¹⁹¹ Theory-based and systems perspectives to evaluation can also be used in conjunction; for example, methods developed using systems thinking could be used to support and focus the involvement of stakeholders in a realist evaluation.⁸²

TABLE 6 Research perspectives and evaluation

Perspective	What the research perspective means for evaluation	Example of study design used
Efficacy: can this work in ideal circumstances?	Answers research questions about whether or not the intervention works in experimental settings. It aims to assess whether or not targeting the hypothesised mechanisms of action of an intervention produces changes in the expected outcome, often a mediating measure (e.g. disease activity), rather than a health or health-care outcome. The aim is to obtain a precise, unbiased estimate of the average effect of the intervention in a homogeneous context. In an efficacy study, the experimental treatment is compared with a placebo or with no intervention	<i>Case study 15</i> : efficacy RCT for carotid endarterectomy ¹⁷⁴

continued

TABLE 6 Research perspectives and evaluation (continued)

Perspective	What the research perspective means for evaluation	Example of study design used
Effectiveness: what works in the real world?	Answers research questions about whether or not the intervention works in pragmatic real-world settings. The principal focus is to identify an unbiased estimate of the average effect of the intervention, in a heterogeneous context that is representative of the scenarios in which the intervention will be applied in practice. In an effectiveness study, the comparison is with 'treatment as usual' or sometimes with an optimised version of treatment as usual	<p><i>Case study 16:</i> individual RCT to evaluate an intervention for preventing repeat falls in people with Parkinson's disease (PDSAFE)¹⁷⁵</p> <p><i>Case study 14:</i> cluster RCT, with embedded process evaluation, to evaluate an approach to support the management of multimorbidity^{176,177}</p> <p><i>Case study 17:</i> stepped-wedge cluster RCT to evaluate the World Health Organization surgical checklist¹⁷⁸</p> <p><i>Case study 18:</i> synthetic controls to evaluate pay or performance on hospital mortality¹⁷⁹</p> <p><i>Case study 19:</i> whole systems trial for prevention of obesity¹⁸⁰</p>
Theory based: what works in which circumstances and how?	This perspective emphasises the importance of understanding how intervention mechanisms may generate one or more outcomes that may vary in different contexts and for different people. This approach places less value on the average effect estimate and more value on understanding the interplay of mechanism and context, how change is being brought about and how this may vary across different contexts and recipients	<p><i>Case study 20:</i> realist process evaluation to evaluate robotic-assisted surgery^{181,182}</p> <p><i>Case study 21:</i> realist process evaluation using case study to evaluate an intervention to support adherence to antiretroviral treatment and retention in care¹⁸³</p> <p><i>Case study 22:</i> mixed-methods evaluation of the New Deal for Communities¹⁸⁴</p>
Systems: how do the system and intervention adapt to one another?	Treats the intervention as a change or disruption to a complex system and/or to relationships within the system. For example, it focuses on the way that a system's constituent parts inter-relate and how systems work over time and within the context of larger systems. It recognises the dynamic nature of systems and specific properties of that dynamic interdependence, such as emergence and feedback. Examples given here include modelling studies, which are not evaluations in the traditional sense, but simulations of what might happen under a given set of circumstances	<p><i>Case study 23:</i> social network analysis (and other systems methods) to evaluate a developmental asset approach to evaluate adolescent well-being¹⁸⁵</p> <p><i>Case study 24:</i> process evaluation, taking a systems perspective, of a local alcohol availability intervention (Reducing the Strength)^{186,187}</p> <p><i>Case study 25:</i> QCA to evaluate factors associated with narrowing health inequalities¹⁸⁸</p>

QCA, qualitative comparative analysis.

Links to case studies are included in *Table 6* to highlight how different study designs may be used to answer different types of questions. A short description of a selection of these methods is given in *Types of evaluation design and methods*; however, before deciding on study design, the range of outcomes that an evaluation may include should be considered, which is described in the next section.

Choice of evaluation outcomes

A crucial aspect of the design of an evaluation is the choice of outcome measures or evidence of change. Evaluators need to assess which outcomes are the most important for stakeholders (e.g. patients, public and decision-makers) and how to deal with multiple outcomes in the analysis.

A sharp distinction between a single primary outcome and a number of secondary outcomes is not necessarily appropriate, particularly where the programme theory identifies impacts across a range of domains. A good theoretical understanding of the intervention, derived from careful development work, is key to choosing suitable outcome measures. Rather than always choosing one health outcome and judging the effectiveness based on that only, effects on intermediate outcomes, such as health behaviour, health care or a determinant of health or a range of outcomes, should be explored and theoretically justified. These outcomes should be identified prospectively during the process of developing the programme theory and be incorporated alongside an a priori analysis plan in a published study protocol and reported comprehensively in the final evaluation. This approach allows a more broad-based judgement on effectiveness to be made.

Inconsistencies in measured outcomes makes synthesis of evidence more difficult; therefore, the use of core outcome sets is recommended where appropriate.¹⁹² The Core Outcome Measures in Effectiveness Trials (COMET) initiative supports the development of core outcome sets, largely for clinical trials, although includes some resources that may be more widely applicable.¹⁹³

Outcome measures may capture changes to a system rather than changes in individuals. Examples of system-level outcomes include changes in relationships within an organisation, the introduction of policies or new ways of working and changes in social norms or normalisation of practice. They also include how changing the dynamics of one part of a system alters behaviours in other parts, such as the displacement of smoking into the home after a public smoking ban. It is possible to use traditional evaluation designs to answer questions relevant to a systems perspective (see *Case study 19*). It is unlikely that all aspects of complexity can be investigated in an evaluation;⁶² however, a broad awareness of the system should be considered at the development phase and system boundaries determined, with consideration of the usefulness of evidence that will be generated.

Pre-identified sources of variation in outcomes are important, and prespecified subgroup analyses should be carried out and reported. In the case of interventions that are expected to affect inequalities in health, analyses stratified by socioeconomic position, gender, ethnicity or geography may be appropriate. Even where such analyses are underpowered they should be included in the protocol and reported, as they may be useful as input into subsequent meta-analyses or for developing hypotheses for testing in further research.

A priori sample size calculations are made to ensure that there is enough statistical power to detect a meaningful target difference in the outcome between the intervention and the comparison groups. This calculation is often based on a single primary outcome, although different candidate outcomes can be considered in turn. The final decision for sample size will depend on practicality and usefulness of evidence, and ideally will be sufficient to detect a meaningful change in all specified candidate outcomes. The Difference ELicitation in TriAls (DELTA2) guidance¹⁹⁴ provides guidance on choosing target differences in outcomes and on associated sample size calculations.

Another outcome of evaluation is refined or developed theory. Improved theory helps us to understand the possibility of transferability across settings and to produce evidence and understanding that is more applicable to informing future decisions by policy-makers. Theory is used to inform evaluations, but the purpose of evaluation should also be the gradual incremental testing and refinement of theory. Interventions are likely to require refinement or adaptation to be successful in different contexts, whereas the theory produced from the study can be generalisable. It is important that results from experimental trials are not overinterpreted; unbiased estimates may not provide evidence suited to the application of the intervention in different contexts.¹⁷² Decision-makers operating in a context different from that of the research may be able to learn more from the theory that is developed from the intervention than from an effect estimate with little information on context or transferability.

Types of evaluation design and methods

There are many study designs to choose from and different designs are each optimally suited to different research questions and different circumstances.¹⁹⁵ As we have emphasised, there is no one-to-one mapping of research perspectives to evaluation designs. A purely quantitative approach using an experimental design with no additional elements, such as a process evaluation, is rarely adequate. This is particularly true in complexity-informed research using a theory-based or systems perspective; in such cases, qualitative and mixed-methods designs are essential for evaluation to answer questions beyond effectiveness.

Studies that use both qualitative and quantitative methods often analyse the data for each method separately and then consider them together in discussion of the findings. There are techniques for integrating qualitative and quantitative data and findings;¹⁹⁶⁻¹⁹⁸ however, this is not common practice in the context of RCTs for health interventions, and effort is ongoing to provide future guidance (see Richards *et al.*¹⁹⁹ for some examples).

A design may rarely be used in a particular field, but that does not mean it cannot be used. You should make your choice on the basis of the specific needs of your evaluation, in particular the research questions that you aim to answer. Awareness of the range of experimental and non-experimental approaches should lead to more appropriate methodological choices. In the following sections we elaborate on some of the evaluation designs from the case studies referred to in *Table 6* and in *Case studies*. We also include social network analysis and modelling, which are not designs as such but provide an overview of further methods that can be used to evaluate complex interventions. We focus on experimental and theory-based designs that have come to the fore since the previous version of this document¹ and the natural experiments guidance²⁰⁰ were published. The examples that we give of systems methodologies are those most commonly discussed in relation to systems methods for evaluation of interventions within health research, not necessarily an exhaustive list of systems methodologies.

Stepped-wedge designs

This is a variant of the cluster RCT design. It can be used to address practical or ethics objections to experimentally evaluating an intervention for which there is some initial evidence of effectiveness or that cannot be made available to the whole population at once. It allows a trial to be conducted without delaying roll-out of the intervention. Eventually, the whole population receives the intervention, but in a phased implementation. Allocation of the intervention can be undertaken in a randomised or non-randomised manner.²⁰¹ This design is increasingly being used and provides an alternative to regular cluster RCTs, for which there are pragmatic constraints on randomisation or allocation of the intervention to groups.²⁰² Given that the roll-out of the intervention in a stepped-wedge design has a temporal element, particular consideration should be given to any changes to the context that occur concurrently. Stepped-wedge designs should be used with caution given that they introduce many complexities and there are ethics and resource implications of delivering to all study participants what the study may find to be an ineffective or even harmful intervention.

Adaptive trial designs

This trial design involves advanced planning of modification to the evaluative design throughout the study period based on interim data.²⁰³ Adaptations to evaluations could include, for example, the decision to drop inferior trial arms, reassessment of sample size requirements, reassessment of eligibility criteria or changes to analytical methods. Adaptation of the evaluation is based on careful planning, simulation of potential scenarios and resulting decision rules. Interim data are evaluated in accordance with predefined decision criteria and any changes that are made require maintaining integrity and validity of the study.²⁰⁴ The aims of adaptive evaluations are to address uncertainties in the planning process and to identify promising interventions by eliminating those with little likelihood of benefit and reduce the probability of failure early on. One example of an adaptive trial design is Sequential Multiple Assignment Randomised Trial (SMART),^{205,206} in which participants are re-randomised to alternative intervention conditions depending on their responsiveness to the first

intervention sequence. Typically, unresponsive participants are re-allocated to alternative interventions. SMART is also being used to refine interventions to optimise the delivery, duration or frequency.²⁰⁵

Natural experimental designs

Natural experimental studies exploit a process or event that occurs independently of the researcher and divides a population into exposed and unexposed groups, between which outcomes can be compared to identify the effect of the intervention. There is increasing recognition of the value of using these designs and considerable methodological development.^{159,207,208} Sometimes this process is effectively random, either by design²⁰⁹ or by some imperfection in the way that the intervention is implemented; usually a statistical model is required to create comparable or 'conditionally exchangeable' groups. Alongside well-established approaches, such as interrupted time series, difference-in-difference and regression discontinuity studies,²¹⁰⁻²¹⁴ there are some newer methods, such as synthetic controls, for which experience is accumulating rapidly.²¹⁵ The synthetic control method involves creating a counterfactual from a pool of control units such that the pre-intervention trend in the outcome is similar in the exposed unit (usually a geographical area) and the synthetic control. Mixed-methods evaluations of natural experiments, as with other effectiveness studies, can provide further insight into intervention mechanisms and processes.²¹⁶ Qualitative evaluations of natural experiments can also provide more in-depth information on the mechanisms by which outcomes are achieved.²¹⁷ A natural experimental approach may be appropriate when a RCT is not practical or ethical; where a significant health impact is expected but there is uncertainty about the nature and size of the effect; or where the intervention has potential for being replicated and scaled up.¹⁵⁹ Other advantages of natural experimental approaches are that they can be applied retrospectively to interventions that were implemented without any prospective evaluation using routinely collected data; produce evidence relatively quickly and cheaply; generate evidence with high external validity while minimising threats to internal validity; and identify effects on long-term and non-health outcomes.²¹⁸ Further detail is provided by the MRC's guidance on natural experiments²⁰⁰ and in a series of papers on quasi-experimental studies in the *Journal of Clinical Epidemiology*.²¹⁹

n-of-1 trials

An *n*-of-1 trial is a randomised controlled crossover trial with a single unit (e.g. person, hospital and school) that receives different intervention options and, therefore, acts as its own control.²²⁰ The unit of randomisation is not the individual but the intervention order within the cycle allocated to the individual. This allows the evaluation of how intervention effects vary within individuals, answering research questions about individual cases.²²¹ A traditional trial would provide the average effect of an intervention for the group that received it, but would not necessarily provide information on differences in effect between individuals; *n*-of-1 trials are particularly useful for testing interventions and theory at the individual level, potentially providing evidence that allows for personalisation of interventions. In addition, an aggregated series of *n*-of-1 trials can be used to determine the effect of the intervention, including heterogeneity of effects, at the population level. However, challenges include the difficulty of calculating statistical power and sample size, and the fact that statistical approaches are as-yet underdeveloped/tested for this type of study design. For examples of *n*-of-1 trial designs being used in health behaviour research see McDonald *et al.*'s²²¹ systematic review, and for medical intervention see Gabler *et al.*'s²²² systematic review.

Network analysis

Network analysis is an approach that can be applied within the other study designs already mentioned, rather than being a design in itself. It is a tool to understand the structures of systems by studying changing relationships within a system, for example of individuals or organisations. It is useful either where the outcome is connections within a system structure, for example an evaluation of an intervention to tackle social isolation and improve social networks, or where these connections are part of the process, for example Rosas and Knight¹⁸⁵ show change in actor networks between year 1 and year 2 of a community intervention designed to address adolescent health and well-being (see *Case study 23*).¹⁸⁵ This method can also provide information on the agents that are central to the

network or on who may be ‘influencers’ within the network,²²³ as well as on understanding how the intervention works.

Modelling

This approach is distinct from traditional evaluation methods in that it is used to predict intervention effects by simulating outcomes. It can, therefore, be useful in intervention development and justification, as well as being used in the evaluation phase as data are collected to assess the model predictions and/or to further improve the model. Microsimulation modelling predicts the effects of an intervention using known relationships between exposures and outcomes to model the effect of a change in exposure. Agent-based and system dynamics modelling methods use computational simulations that draw on objectively measured and experiential data from a range of perspectives. The starting point for these methods is often other soft systems methodologies. These models can be used to evaluate different hypothetical scenarios, for example different variants of an intervention, or to explore spill-over or longer-term effects on health or economic outcomes, which may occur outside the scope or time frame of an empirical evaluation. For example, in the Evaluation of the health impacts of the UK Treasury SDIL (see *Case study 7*) outcomes are measured at 2 years, with additional modelling to estimate effects for a wider range of health outcomes (e.g. morbidity and mortality, as well as health service costs) over 5–20 years.²²⁴

Models may, thus, be used alone or to complement other designs and can be used throughout research phases, including development and implementation as well as evaluation. These methods are just emerging in health-related evaluation research, with few examples as yet. ABM involves virtual representation of real-world situations that models individuals and their interactions with their environment, which allows for system properties, such as feedback and emergence.²²⁵ See Tracy *et al.*²²⁶ for an overview of ABM in public health, with examples of application and discussion of future direction (see also Silverman *et al.*¹²²). System dynamics modelling is an approach that aims to predict the behaviour of complex systems over time, where the starting point is the structures that trigger the behaviours of the system itself, rather than the entities within it (as in agent-based models). As yet, less work has been carried out using system dynamic modelling in evaluative health research.

Other modelling methods also use synthetic populations and simulate individual trajectories under different scenarios. For example, the effectiveness of the NHS health check programme in Liverpool was modelled from 2017 to 2040 under five scenarios, including continuation of the implementation of the programme as it is and targeting implementation to the most deprived areas, with comparison with no NHS health check.²²⁷ The study showed that under current implementation the intervention was unlikely to be cost-effective or equitable, but that implementation under alternative scenarios would be more positive. Additional examples that use this (IMPACT_{NCD} model²²⁸) and other models have explored salt policies in England; the potential of trans fats policies to reduce socioeconomic inequalities in coronary heart disease mortality in England;²²⁹ and the impact of pandemic influenza on the UK.²³⁰ The Sheffield Alcohol Policy Model was used to predict the effect of minimum unit pricing for alcohol in Scotland, including differential impacts on heavy and light drinkers and by socioeconomic group; this model was key to the development of the policy and in countering legal challenges to its implementation and will continue to be used in the evaluation of the policy as data accumulate.²³¹

Qualitative comparative analysis

Qualitative comparative analysis (QCA) is a mixed method for establishing causation based on systematic qualitative and quantitative comparison across a number of cases, for example intervention sites and exploring how outcomes were achieved.^{232–234} Qualitative attention to each case is followed by quantitative exploration of each case by a number of attributes determined to be important. These attributes are defined through careful review of existing and new data, for which stakeholder engagement and theoretical arguments are crucial (see *Case study 25*).¹⁸⁸ The QCA approach does not specify a single causal model but follows the principle of equifinality, which means that the same outcomes can be achieved by different means, that is there can be more than one causal set.

Understanding processes

Process evaluation is a critically important component of the evaluation of complex interventions and is an essential adjunct to the study designs presented so far (see the case study examples set out in *Table 6*). Process evaluations can use qualitative and/or quantitative process data to answer questions around fidelity and quality of implementation (e.g. what is implemented and how?), mechanisms of change (e.g. how does the delivered intervention produce change?) and context (e.g. how does context affect implementation and outcomes?).¹⁶⁵

Within an efficacy study, a process evaluation would have relatively narrow questions. The key idea of efficacy studies is that there is a test of causal mechanism, often through mediation analysis. Such analyses are usually quantitative, but could use qualitative methods to assess how research participants understood the intervention and made use of intervention components. Because the focus of efficacy studies is to maximise internal validity, process analyses could usefully examine fidelity to assess whether or not the intervention was delivered with high fidelity to the theorised mechanism of action. Consideration of context would not be the focus of a process analysis from an efficacy perspective.

Process evaluations in an effectiveness perspective would be used to inform findings, for example to provide insight into why an intervention fails unexpectedly or has unanticipated consequences, or why a successful intervention works and how it can be optimised. This includes exploring the fidelity and quality of the intervention, clarifying causal mechanisms and identifying contextual factors associated with variation in outcomes (see *Case study 14*). Details on the type of measures to consider are provided in MRC guidance on process evaluation of complex interventions.⁵

Process evaluations are key components of theory-based and systems evaluations, to the extent that they seek to answer research questions about how the intervention produces change in a generalisable way, taking the study further than assessing solely whether or not it works in a specific setting. An emerging approach is to conduct a realist process evaluation, specifically drawing on realist evaluation.⁴⁸ If undertaken alongside an effectiveness evaluation, a realist process evaluation can help to understand the effects of intervention components rather than only the intervention as a whole; examine mechanisms of change; explore the effect of intervention context; collect and analyse qualitative and quantitative data; and aim to build and validate intervention programme theories.²³⁵ There are also increasing examples of taking a systems approach to process evaluations, with a framework developed by McGill *et al.*²³⁶ offering some guidance.

Process evaluations should be conducted to the same high methodological standards and reported as thoroughly as evaluation of outcomes. The MRC guidance on process evaluation for complex interventions⁵ provides further information on how to conduct a process evaluation, including consideration of relevant theoretical approaches.

Economic considerations

As noted in *Economic considerations*, there are a range of economic evaluation frameworks (*Box 7*).

Each framework has different strengths and limitations. CEA measures outcomes in natural units that relate to, but do not measure directly, the health outcomes of interest; for example, reduced blood pressure is associated with reduced incidence of stroke, but it does not directly measure the incidence of stroke or capture the associated range of health improvements. This limits the usefulness of an incremental cost-effectiveness ratio when evaluating complex interventions with multiple impacts. CUA uses measures, such as quality-adjusted life-years (QALYs), to compare interventions with a range of physical and mental health outcomes in terms of their impact on length and quality of life. However, complex interventions often have outcomes that are difficult to capture fully using QALYs. Ascribing monetary values to all of the dimensions of benefit in a CBA is methodologically challenging. CCA is more straightforward and can be used to present evidence on a wide range of costs and outcomes; however, without valuation of these outcomes (e.g. preference weighting in QALYs or monetary

BOX 7 Frameworks for economic evaluation

- CEA: comparison of costs and outcomes measured in ‘natural units’, such as life-years gained, cancers detected and strokes prevented.
- CUA: comparison of costs and outcomes measured in terms of life expectancy adjusted for ‘utility’, such as QALYs or disability-adjusted life-years.
- CBA: comparison of costs and health and non-health benefits valued in monetary terms. This can involve the measurement of non-health benefits across different sectors.
- CCA: comparison of costs and health and non-health benefits across different sectors, measured in units appropriate to the benefit being considered and reported in disaggregated form.

valuation in CBA) decision-making requires judgement. Rare but costly outcomes may acquire too much influence if their frequency is not taken into account. Furthermore, if some outcomes improve more under some interventions and other outcomes fare better under others, the policy implications of a CCA may be ambiguous. The importance or ‘weight’ of differing consequences is likely to vary across stakeholders. A separate exercise, such as a discrete choice experiment (see *Case study 26*), could add additional evidence to the value of different consequences to different stakeholders.

Given that complex interventions typically have effects across a wide and diverse range of outcomes, recent guidance emphasises the use of a broad framework, such as CCA or CBA.¹³⁴ CCA has the intuitive appeal of providing data across a range of ‘consequences’ (outcomes) with which decision-makers are likely to be concerned when making a policy decision. This allows the analyst to adopt a broad public sector or societal perspective that is appropriate to many complex interventions. It is important to note, however, that the different frameworks are not mutually exclusive: CBA and CEA can be carried out and embedded in a wider CCA that provides data on a more comprehensive inventory of outcomes. The CCA can be useful for identifying the range of impacts likely to be of interest to researchers and policy-makers prior to deciding which outcomes will be measured and how. For this reason, it can be a helpful first step towards a CBA.

In addition, there are a number of emerging methodologies that can lend themselves to economic evaluations of complex interventions. They include Social Return on Investment²³⁷ and multi-criteria decision analysis (MCDA).²³⁸ These are not usually considered to be standard tools for economic evaluation owing to theoretical weaknesses and their handling of outcomes, but they are becoming more commonly used in the area of complex interventions. MCDA is particularly helpful following a CCA because it involves a structured, explicit approach to decisions involving multiple criteria.²³⁸

Decision analytic modelling^{120,239} is a useful way of combining data from an economic evaluation with data from other sources to extrapolate costs and outcomes beyond the study time frame. The use of advanced modelling techniques, such as microsimulation modelling, network analysis, discrete event simulation and ABM,^{153,240,241} is particularly useful for complex and public health interventions, although simpler model types can also be appropriate.^{242,243}

Reporting evaluations

Evaluations should always be reported in accordance with established guidelines because this will help to ensure that the key information is available for replication studies, evidence syntheses and guideline development. The Enhancing the QUALity and Transparency Of health Research (EQUATOR) Network maintains a comprehensive database of reporting guidelines [www.equator-network.org/ (accessed 2 March 2021)].

Relevant guidelines for reporting are as follows:

- The CONSORT statement is a well-established standard for reporting randomised trials. Relevant extensions are published for –
 - social and psychological interventions (CONSORT-SPI 2018)²⁴⁴
 - randomised pilot and feasibility trials¹⁶³
 - cluster RCTs²⁴⁵
 - pragmatic trials²⁴⁶
 - *n*-of-1 trials²⁴⁷
 - cluster stepped-wedge trials²⁴⁸
 - non-pharmacological interventions²⁴⁹
 - patient-reported outcomes (CONSORT PRO)²⁵⁰
 - adaptive designs²⁵¹
 - and the SPIRIT-PRO extension for patient-reported outcomes.²⁵²
- Reporting of statistical analysis plans.²⁵³
- Reporting process.²⁵⁴
- RAMESES II for realist evaluation.²⁵⁵
- SPIRIT 2013 statement provides evidence-based recommendations for the minimum content of a clinical trial protocol.²⁵⁶
- Standards for reporting qualitative research.²⁵⁷
- Strengthening the Reporting of OBservational studies in Epidemiology (STROBE).²⁵⁸
- Consolidated Health Economic Evaluation Reporting Standards (CHEERS).¹⁷⁰

Table 7 provides a summary of the consideration of each of the core elements at the evaluation phase.

TABLE 7 Core elements for the evaluation phase

Core elements	Issues to consider	Risk of not considering them
Context	<ul style="list-style-type: none"> • Consider the wider contextual factors that may influence the evaluation in relation to influences on either the 'intervention' or the 'study processes' 	<ul style="list-style-type: none"> • Questions related to external validity will not be answered
Programme theory	<ul style="list-style-type: none"> • Has the programme theory changed from the development and/or feasibility phase? • What aspects of the wider system have influenced the desired outcomes? 	<ul style="list-style-type: none"> • Lack of potential learning and understanding owing to either lack of programme theory or underdeveloped programme theory
Stakeholders	<ul style="list-style-type: none"> • Gather appropriate input from a range of relevant stakeholders to ensure that evaluation methods and outcomes meet their requirements 	<ul style="list-style-type: none"> • Not collecting the correct data to inform changes in policy and practice could negatively impact the utility of the evaluation and future implementation of the intervention
Uncertainty	<ul style="list-style-type: none"> • The evaluation should respond to key uncertainties identified by engaging with stakeholders, existing evidence and relevant theory 	<ul style="list-style-type: none"> • Answering research questions that do not address important areas of uncertainty, so that the utility of the evaluation will be limited
Intervention refinement	<ul style="list-style-type: none"> • Decide beforehand what type of refinement is appropriate to allow for a robust evaluation 	<ul style="list-style-type: none"> • Not allowing for appropriate changes to the intervention or study processes may lead to outcomes that are not useful or meaningful. Validity of the evaluation may be compromised by failing to consider what kind of intervention refinement is acceptable

continued

TABLE 7 Core elements for the evaluation phase (continued)

Core elements	Issues to consider	Risk of not considering them
Economic considerations	<ul style="list-style-type: none"> • Importance of (broad) perspective • Sensitivity analysis and subgroup analyses will be crucial • Identify preferred economic evaluation framework (other frameworks can be embedded) • Consider both a short-term and a long-term time horizon for economic analyses 	<ul style="list-style-type: none"> • Too narrow a perspective will not reveal the full economic impact • Cost-effectiveness of alternative scenarios will not be identified • Incorrect frameworks may not identify all relevant costs and impacts • Ignoring long-term costs and outcomes may result in an incorrect cost-effectiveness conclusion

KEY POINTS 10 Evaluation

- Where appropriate, involve stakeholders in defining research questions and the choice of evaluation design and outcomes to ensure that meaningful and useful data are produced for relevant stakeholders.
- Determine the key uncertainties that your evaluation will resolve and frame appropriate research questions.
- Consider a range of 'useful' evaluation perspectives and outcomes rather than necessarily focusing on effectiveness.
- Choose from a range of experimental or non-experimental designs that are focused on answering the research questions.
- Include methods of evaluation to explore the change process (i.e. mechanisms, process, context and theory).
- Conduct economic evaluation and consider the range of impacts that an intervention may have.
- Clearly report the evaluation, including updated programme theory, following recognised reporting standards if available.

SIGNPOSTS TO FURTHER READING 10 Evaluation

- For detail on different experimental designs, see the 2006 MRC guidance.¹
- PRagmatic Explanatory Continuum Indicator Summary (PRECIS and PRECIS-2) as a tool to help determine whether or not a trial should be/is on the pragmatic-explanatory continuum.^{39,259}
- For guidance on sample size decisions see the DELTA2 guidance¹⁹⁴ and a series of papers in *Trials* (2018, issue 19).²⁶⁰
- The COMET database provides a searchable set of developed core outcome sets for various health areas and target populations.¹⁹³
- MRC guidance on natural experiments.⁹
- Synthetic controls.²¹⁵
- Stepped-wedge cluster randomised trials.^{202,248}
- Commentary and hypothetical example on how theories of change and realistic evaluation differ.⁴⁴
- Qualitative comparative analysis for evaluation.²³³
- Discussion and examples of treating qualitative and quantitative data from a RCT interdependently.¹⁹⁹

Economic evaluation

- Further reading and advances in the field overall.^{132,141,242,243,261,262}
- NICE guidance.¹³⁴ The NICE guidance, updated in 2017,¹³⁴ allows wider perspectives (e.g. NHS and Personal Social Services, local government and societal where appropriate), inclusion of alternative non-health outcomes (such as the ICECAP²⁶³ measurement of capabilities²⁶⁴ and 'broader' approaches such as CCA and CBA, as appropriate).⁹²
- Alongside observational studies and natural experiments,^{265,266} economic modelling²⁴³ and conceptual modelling frameworks,¹⁴⁰ applied health economic methods, public health practice and research,¹⁴¹ and reporting guidelines.¹⁷⁰

Implementation

'Implementation' is a term used inconsistently within and across disciplines. Here, we use the World Health Organization (WHO) definition:

Implementation: *Deliberate efforts to increase the impact of successfully tested health innovations so as to benefit more people and to foster policy and programme development on a lasting basis.*

Reproduced from WHO²⁶⁷

Successful implementation of interventions, even after they have been shown to be effective, remains relatively rare.²⁶⁸ There are numerous reasons for this: some relate to how intervention research is conducted and others focus more on how research findings translate into policy and practice. In this section, we focus on how the conduct and reporting of intervention research can be undertaken to maximise its utility and impact. First, we briefly discuss implementation studies that focus specifically on interventions and strategies to maximise the translation of research into practice or policy.

Implementation science

Within health care, implementation science is an important and developing field of research that undertakes studies that develop and evaluate strategies for implementing effective interventions that address local realities, can be implemented at scale and are potentially sustainable.²⁶⁹ Implementation science has been defined as 'the scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practice, and, hence, to improve the quality and effectiveness of health services';²⁷⁰ however, a wider definition would not be restricted to health care or to the limited range of designs and methods typically used in that context. With this wider definition, implementation science, therefore, includes interventions and strategies to increase the uptake and impact of research in policy, including knowledge mobilisation across a range of potentially relevant organisations, sectors and professions, as exemplified by the range of case studies of implementation research in a global health context.²⁷¹

A large part of implementation science research, therefore, involves the development and evaluation of complex interventions to maximise effective implementation in practice and/or the policy of interventions that have already demonstrated efficacy or effectiveness. The whole of this framework, therefore, applies to those implementation studies just as it does to other forms of intervention. The main specific characteristic of such implementation studies is that the main outcome measures used will not be health outcomes but measures of implementation effectiveness, such as practitioner or patient behaviour, uptake of services, reach and policy impact. For example, a cluster randomised trial of strategies to improve nurses' hand-washing behaviour used hand-hygiene compliance and adherence as outcome measures rather than infection rates, with a nested process evaluation identifying which components of the implementation strategy and context particularly supported compliance.²⁷² The field of implementation science has many theories and frameworks that can importantly inform the development of implementation interventions and associated programmes (see *Signposts to further reading 11: implementation*). There are also specific standards for reporting implementation studies and other specific considerations for implementation-focused intervention research that complement this framework.²⁷³

Implementation in all phases

A key motivation for this update was to maximise the usefulness of complex intervention evidence for decision-makers. For the evidence to be useful, it is important that interventions are developed with implementation considerations in the foreground at all phases (see *Case study 27*). A key part of this is involving stakeholders and being mindful of the realities of the intervention context at all phases/from the beginning and throughout. Early consideration of implementation increases the potential to develop an intervention that can be widely adopted and maintained in real-world settings.²⁴⁵ For example, this may encourage the development of 'very brief interventions' that fit with routine practice and are effective and cost-effective when implemented at scale.²⁷⁴ Such interventions may have more

impact than very ambitious researcher-led interventions that demonstrate effectiveness and cost-effectiveness in a research setting, but are not 'implementable' for a variety of reasons, such as lack of fit to health-care practitioners' usual practice.¹³⁰ The consequence of neglect is that issues, such as acceptability to the public or to those who deliver the intervention, may undermine implementation later on.²⁷⁵ Other benefits of including stakeholders in all phases are to gain better understanding of the time and resource constraints of real-world evidence users and to engage with, understand and influence organisational, professional and social systems that may be resistant to change.²⁷⁶⁻²⁷⁹

As well as anticipating implementation factors in intervention development and design, research questions in the feasibility and evaluation phases can help to understand and inform implementation. The intervention programme theory and process and outcome data collected in these phases can help inform dissemination of the findings and preparation of an implementation manual or plan to maximise the effective roll-out of the intervention in real-world settings. In addition to the publication of findings, sufficient details of the intervention and its underlying theory should be reported.²⁷³ This improves the synthesis of evidence that may be more useful to decision-makers (e.g. theoretical outputs compared with effect sizes) and improves their understanding of the potential transferability of the intervention to their specific context (see *The role of context in implementation*).

Details of the intervention and key information related to effective delivery identified in feasibility and evaluation phases need to be communicated clearly in dissemination via methods that are accessible and convincing to implementers.²⁸⁰ These may include resource requirements for prerequisites and delivery (e.g. equipment, staff training, skills and time); barriers to and facilitators of effective implementation; and potential flexibility in implementation while maintaining programme fidelity. This includes making recommendations as specific as possible.²⁸¹

Implementation and evaluation phases combined

In a natural experimental study, evaluation takes place within the implementation of the intervention in a real-world context. It may be possible to ascertain both the principal health outcomes during implementation to demonstrate effectiveness and/or the more proximal or process measures available through readily available data sources collected as part of the implementation. This would allow the assessment of which implementation or contextual factors support, amplify or undermine the uptake, (cost-) effectiveness or reach of the intervention.

In the case of highly pragmatic effectiveness trials or specific hybrid effectiveness-implementation designs, effectiveness and implementation outcomes can be combined in one study, with the aim of reducing time for translation of research on effectiveness into routine practice. Different hybrid designs have been proposed based on their focus on effectiveness versus implementation: conducting an effectiveness study while exploring and gathering data on implementation; dual focus on effectiveness and implementation; and conducting a study with a primary focus on implementation, while gathering data on effectiveness.^{282,283} Green *et al.*²⁸⁴ describe three studies to show how these approaches differ in practice.

The role of context in implementation

Understanding interdependence with the context of a given intervention is critical to the success or failure of implementation. Developing an understanding of context dependence throughout the development, feasibility and evaluation phases can, therefore, importantly inform what prerequisites need to be in place for the effects identified in the evaluation phase to be successfully reproduced in implementation.

Understanding context is also central to the transferability of an intervention into different contexts: an important aspect of long-term implementation. Developing and refining programme theory and considering wide-ranging contextual factors can help to identify issues related to transferability, for example that the intervention is unlikely to be effective when transferred to a given context.²⁸⁵ Stakeholder involvement is then a helpful process to identify potential solutions to overcome identified issues.

The 'Context and Implementation in Complex Interventions' framework that integrates context, implementation and setting can be used to guide discussions between researchers and stakeholders on the transferability of interventions.⁷⁰

Economic considerations of implementation

Economic issues at the implementation phase have not traditionally formed part of the guidance on economic evaluation, but they can and should be considered (1) at the early stages of intervention and study development and (2) in the reporting of economic results.

How the results of economic analyses are reported and presented to decision-makers can affect whether or not and how they act on the results. A key issue is avoiding and anticipating misconceptions in reporting the results of economic analyses. Interventions described as 'cost saving' may well be potentially cost saving, but savings may be difficult to realise in financial terms during implementation if there is an element of fixed costs or if the reduced demand for a service withdrawn on economic grounds is simply replaced with demand for services to meet previously unmet needs. The 'cost saving' may refer not to a 'financial' saving but to a release of resources for other potential uses, such as a freeing up of beds or the shifting of resource use from the NHS to another sector, such as social services (e.g. palliative care).²⁸⁶

Another important issue that is relevant for implementation is how to deal explicitly with multisectoral issues. The payers of costs and the receivers of benefits of a multisector intervention may not be the same organisation, which can reduce the incentive to implement an intervention even if it is shown to be beneficial relative to its cost. Early identification and engagement with appropriate stakeholders and decision-makers to identify such cross-sectoral impacts will help frame appropriate research questions and pre-empt any implementation challenges that might arise. Recent work in this area has suggested methods, such as co-financing between and across sectors,²⁶¹ to deal with these issues.

Studies to assess the cost of implementation can also be undertaken post evaluation to help aid the implementation process. This could be particularly important where an intervention is proven to be effective and cost-effective, yet implementation requires additional investment.

Other considerations specific to the implementation phase

At the implementation phase, observational or experimental studies are often useful even though there may not be any specific implementation strategies being assessed, as would be the case in implementation science studies (see *Implementation science*) or hybrid implementation/effectiveness designs (see *Implementation and evaluation phases combined*). Questions that can be assessed at the implementation stage include assessment of intervention receipt, 'dose' received, programme completion and the characteristics of those receiving the intervention and any inequalities in reach or completion. Similarly, barriers to and facilitators of effective implementation, including contextual, system or resource constraints, can be investigated. A key issue in understanding implementation is the extent to which the intervention has been delivered with fidelity,²⁸⁷ so that problems with implementation can be distinguished from a lack of effectiveness. Where an intervention is delivered with low fidelity, it is difficult to determine whether or not a lack of impact in implementation reflects an intervention that is genuinely ineffective, ineffective in that context or potentially effective but not observable in this instance owing to low-fidelity implementation. On the other hand, some flexibility in intervention implementation may support intervention effectiveness across contexts, provided that the key functions of the programme are maintained.⁴⁰ Where adaptations have been made to an intervention in a particular context, it is important to understand exactly what has been adapted.¹⁴⁵ Further detail on implementation related to fidelity can be found in the MRC's guidance on process evaluation of complex interventions.⁵

Few evaluation studies are powered to detect rare events, while unanticipated or adverse outcomes and potentially important longer-term system effects of interventions, such as spill over or diffusion of

effects through networks, may also not be feasible to measure in an evaluation study within a specific sample and time frame. These can potentially be measured within the implementation phase, as they may have differential reach or effectiveness across population subgroups.²⁸⁸ Effects are likely to be smaller and more variable once the intervention becomes implemented more widely, and unintended consequences may begin to emerge.^{76,136} Long-term follow-up may be required to determine whether or not short-term changes persist, to determine whether or not benefits demonstrated from the original study do in fact occur and to capture any likely cost impacts (e.g. cost savings). Although long-term follow-up of complex interventions is uncommon, such studies can be highly informative. It is worth thinking about how to measure rare or long-term impacts, for example through routine data sources and record linkage or by recontacting study participants. Plans for the collection of appropriate outcome data and obtaining appropriate consents should be built into the study design at the outset. Implementation phase studies may involve randomised or non-randomised control groups or can be undertaken solely among those who receive the intervention.

Table 8 provides a summary of the considerations of the core elements at the implementation phase.

TABLE 8 Core elements of the implementation phase

Core elements	Issues to consider	Risk of not considering them
Context	<ul style="list-style-type: none"> Consider the contextual factors that could influence the intervention in this and other settings (e.g. adoption, reach and maintenance of the intervention) What are the key factors that may determine whether or not evaluation findings are replicated in implementation? 	<ul style="list-style-type: none"> Failure to understand how the intervention interacts and/or interrupts the system in which it is located Interventions implemented in contexts in which they are unlikely to work
Programme theory	<ul style="list-style-type: none"> What are the key functions and components of the intervention that need to be preserved in implementation to maximise the likelihood of achieving effects found in evaluation? Continue to update programme theory throughout any implementation phases What potential harms, spill-over effects or longer-term effects can be measured in implementation? 	<ul style="list-style-type: none"> Implementation failure Failure to learn in the implementation phase about where fidelity, flexibility and refinement can be balanced to maintain intervention effects Failure to identify harms, side effects or potential wider, long-term effects that were not assessed in the evaluation
Stakeholders	<ul style="list-style-type: none"> Careful communication of research findings to stakeholders to support decision-making as to where and how interventions may be transferred and key aspects of programme implementation relevant to their context Identifying potential 'champions' of the intervention to support its implementation 	<ul style="list-style-type: none"> Intervention implemented in contexts in which the evaluation evidence is unlikely to generalise Implementation failure
Uncertainty	<ul style="list-style-type: none"> What are the areas of uncertainty in our implementation plan and theory? Are there key barriers to and facilitators of implementation in real-world contexts that have not been identified or theorised in other phases of research? What are the critical intervention components and functions that need to be maintained in implementation and where can there be flexibility in intervention form while preserving key functions? Are their effects not anticipated or measured in the previous stages that can occur and be assessed in implementation? 	<ul style="list-style-type: none"> Evidence of effectiveness may not translate as the intervention is implemented outside research conditions Failure to continue to learn about how the intervention can best be implemented to maximise effects Failure to optimise implementation across contexts Failure to capture important unanticipated, rare or long-term system effects that were not included in the evaluation design

TABLE 8 Core elements of the implementation phase (continued)

Core elements	Issues to consider	Risk of not considering them
Intervention refinement	<ul style="list-style-type: none"> Refine the intervention, while retaining its core elements and programme theory, in response to implementation and contextual factors 	<ul style="list-style-type: none"> Intervention failure owing to unnecessarily rigid adherence to intervention as evaluated
Economic considerations	<ul style="list-style-type: none"> Highlight and work towards multisectoral evaluation from the outset Establish who payers are and who receives benefits. Co-payment options may be relevant Cost of implementation analyses: overcome logistical barriers Consider requirements for additional sensitivity analysis to explore applicability of economic evaluation evidence to the differing context of large-scale implementation 	<ul style="list-style-type: none"> Misconceptions of reported economic results Decision-makers misunderstanding economic results and what that means for their practice/budgets Cost-effective interventions not implemented in practice Cost-effectiveness of large-scale implementation underestimated or overestimated

KEY POINTS 11 Implementation

- Determine key uncertainties about intervention implementation that need to be resolved to ensure impact, and develop appropriate research questions related to implementation.
- Consider implementation throughout all research phases.
- Involve stakeholders from the early development phase and throughout.
- Incorporate theories, models or frameworks of implementation into intervention design.
- Learn from where others have identified a 'failure of implementation'.
- Incorporate economic considerations alongside implementation issues from outset. These should be considered in the programme theory developed at the development phase.
- Publish and actively disseminate full details of the intervention and the updated programme theory, including a clear indication of the critical functions of the intervention and key context dependencies.

SIGNPOSTS TO FURTHER READING 11 Implementation

- Implementation fidelity: the MRC guidance on process evaluation of complex interventions.⁵
- The Context and Implementation of Complex Interventions (CICI) Framework.⁷⁰
- Development of a framework and coding system for modifications and adaptations of evidence-based interventions.*¹⁴⁵
- Unintended outcomes evaluation approach: a plausible way to evaluate unintended outcomes of social development programmes.*¹³⁶
- Standards for reporting implementation studies (StaRI) statement.*²⁷³
- The Implementation Research ('ImpRes') tool.²⁸⁹
- An introduction to implementation science.²⁹⁰
- Making sense of implementation theories, models and frameworks.*²⁹¹
- Normalisation process theory: a framework for developing, evaluating and implementing complex interventions.*¹³⁰
- Implementation research: new imperatives and opportunities in global health.*²⁷¹
- Diffusion of innovations in service organizations: systematic review and recommendations.*²⁹²
- The quality implementation framework: a synthesis of critical steps in the implementation process.*²⁹³

Case studies

This section includes 27 case studies that have been chosen to highlight key points throughout (Table 9).

TABLE 9 Case studies referred to in the report

Case study	Description
Core elements	
Case study 1	Travel through the main phases of complex intervention research
Case study 2	Considering context
Case study 3	Developing, refining and testing programme theory
Case study 4	Complex logic model
Case study 5	Dark logic model
Case study 6	Realist matrix
Case study 7	System map to develop programme theory
Case study 8	Engaging stakeholders (service users)
Case study 9	Refining the intervention
Case study 10	Economic considerations
Phases of the research process	
Case study 11	Intervention development and example systems map
Case study 12	Feasibility trial
Case study 13	Feasibility: evaluability assessment
Case study 14	Evaluation: effectiveness perspective – cluster RCT with theory-informed process evaluation
Case study 15	Evaluation: efficacy perspective
Case study 16	Evaluation: effectiveness perspective with individual RCT
Case study 17	Evaluation: effectiveness perspective with stepped-wedged cluster randomised design
Case study 18	Evaluation: effectiveness perspective with synthetic control design
Case study 19	Development and evaluation: effectiveness perspective drawing on a systems perspective for intervention development and evaluation
Case study 20	Evaluation: theory-based perspective alongside effectiveness
Case study 21	Evaluation: theory-based perspective with realist approach
Case study 22	Evaluation: theory-based perspective
Case study 23	Evaluation: systems perspective
Case study 24	Evaluation: systems perspective with process evaluation
Case study 25	Evaluation: systems perspective with qualitative comparative analysis
Case study 26	Intervention development and discrete choice experiments
Case study 27	Implementation in all phases

Case study 1: travel through the main phases of complex intervention research

Study title: Football Fans In Training (FFIT).^{294–298}

Summary: many men lack basic information about how to maintain a healthy weight by eating a balanced diet and staying physically active. They often feel uncomfortable about joining NHS or commercial weight management programmes, which they see as being mainly for women. FFIT aimed to help men lose at least 5–10% of their weight and keep it off over the long term. The programme was developed to appeal to Scottish football fans and to help them improve their eating and activity habits.

A RCT that was undertaken in professional football clubs established the effectiveness and cost-effectiveness of the FFIT programme; the mean difference between the intervention and the control groups at 12-month follow-up was 4.94 kg in favour of the intervention group. In addition to weight loss during the 12-week programme, almost 40% of participants maintained $\geq 5\%$ weight loss at 12-month follow-up. This level of maintained weight loss is associated with various health benefits, including lower risk of stroke, heart disease, cancer, diabetes mellitus and other health issues. A CEA demonstrated that FFIT was inexpensive to deliver, making it appeal to decision-makers for local and national health provision.

The journey of the research project through the phases of complex intervention research is illustrated in *Figure 5*.

Key points:

- The study team considered all six core elements during each of the four phases of research.
- Implementation was considered from the outset and the study team engaged with key stakeholders in the development phase to explore how the intervention could be implemented in practice, if proven to be effective.

Case study 2: considering context

Study title: *Why do interventions work in some places and not others: a breastfeeding support group trial.*²⁹⁹

Summary: the Breastfeeding In Groups intervention involved the provision of breastfeeding groups for pregnant and breastfeeding individuals and aimed to increase breastfeeding and maternal satisfaction. Findings from the effectiveness study (pragmatic cluster RCT) showed that breastfeeding rates increased in three intervention localities, declined in three and remained unchanged in one.³⁰⁰ This could not be explained by the quantity of the intervention delivered (number of group meetings or attendances). Qualitative and quantitative data were collected to develop seven case studies to explore differences in context in the different areas. The study found that the history and context of the locality, including structural and organisational factors, professional relationships, leadership, multidisciplinary partnership working and readiness to change, all influenced successful policy implementation.

Key points:

- This study was informed by a realist approach and offered insights into the influence of context on the impact of the intervention.
- The study chose methods in a pragmatic way and used ethnographic in-depth interviews, focus groups, observations and survey data to build case studies.
- Data collection and analysis were conducted before the findings of the outcome evaluation were known to minimise bias in interpretation. Retroductive logic was then applied to explain the outcomes from the effectiveness trial.
- Systems thinking was used to inform the analysis; for example, the localities were interpreted as systems composed of numerous interacting elements.
- Overall, the explanatory model that was constructed provided insights into the necessary attributes for successful policy initiation and sustainability. As well as providing a plausible explanation for the trial results, the model may be useful for other trials of complex interventions.

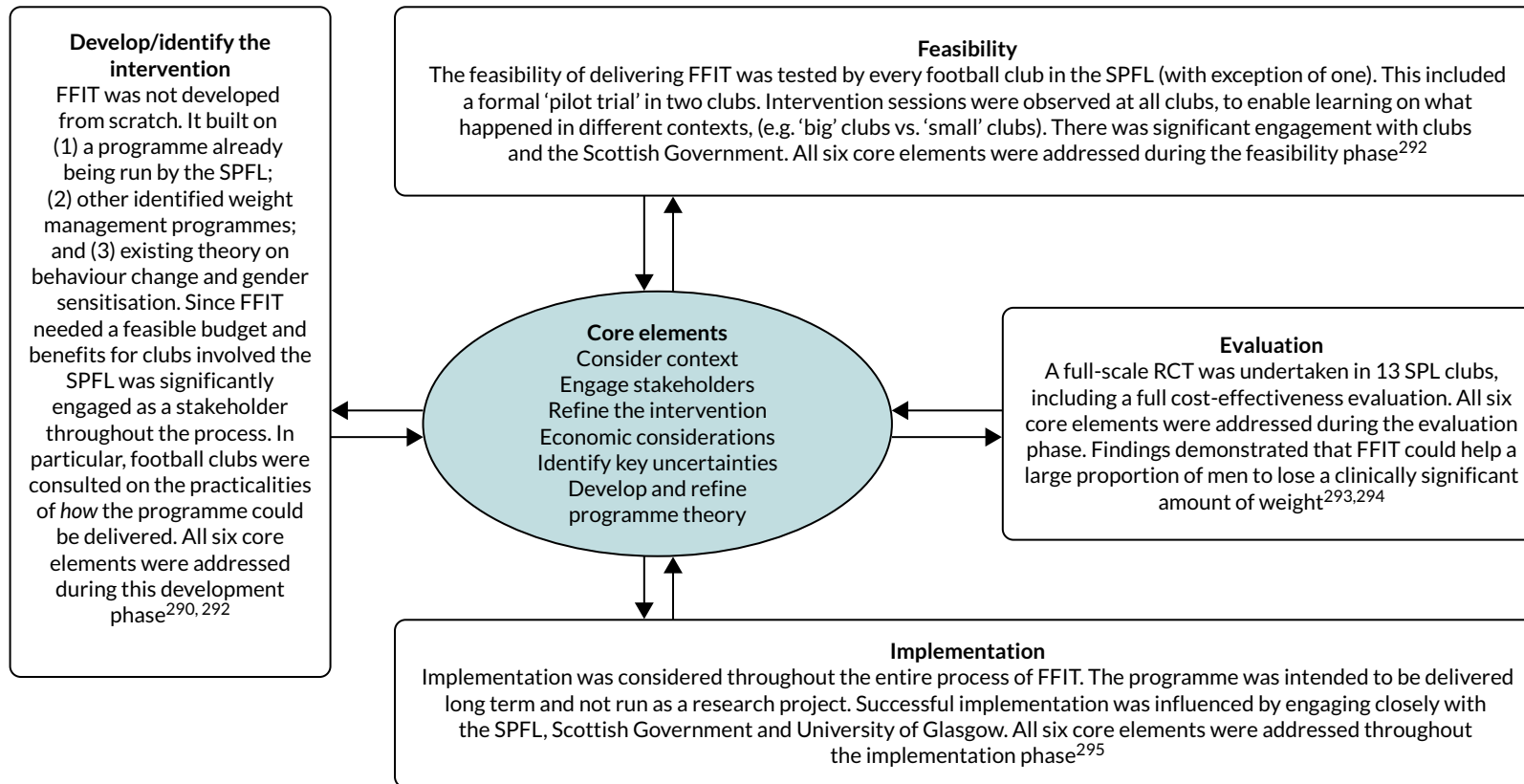


FIGURE 5 Case study demonstrating how the FFIT programme moved through the research process. SPFL, Scottish Premier Football League.

Case study 3: developing, refining and testing programme theory

Study title: *Improving care at scale: process evaluation of a multi-component quality improvement intervention to reduce mortality after emergency abdominal surgery (EPOCH trial)*.³⁰¹

Summary: this stepped-wedge cluster randomised trial tested a large-scale quality improvement programme that aimed to improve 90-day survival rates for patients undergoing emergency abdominal surgery.³⁰² The programme theory was developed prior to evaluating the quality improvement programme. Current evidence and learning from other quality improvement programmes were used to justify inclusion of elements of the programme theory. Contextual factors were considered so that delivery of the programme theory could be tailored to local contexts; those leading the implementation in each hospital were supported to focus on the aspects of the intervention that would work in their local context, for example one consideration of context was the available human and financial resources. Desired outcomes were identified, along with the strategies, individual activities and resources intended to achieve those outcomes. Importantly, data collection was then aligned with elements of the programme theory to ensure that the mixed-methods process evaluation gathered relevant data to allow interpretation of the programme theory. Findings from the process evaluation, therefore, informed reflections on the programme theory. The overall findings suggested that some, but not all, aspects of the EPOCH programme theory were implementable in the context of hospital resources. The programme theory potentially underestimated the influence of social influences (e.g. the time-consuming and challenging nature of building and maintaining social relationships) and system influences (e.g. a major system reorganisation in the NHS) on the quality improvement programme. This helped to identify barriers to improvement, and feeds into the development of future quality improvement programmes.

Key points:

- The programme theory was well articulated throughout the paper.
- Contextual factors were considered allowing tailored programmes to be delivered in different hospitals.
- Data collection methods were aligned with elements of the programme theory to support interpretation of the findings.
- Overall, the development, implementation and post-study reflection of the programme theory resulted in suggested improvements for future implementation of the quality improvement programme.

Case study 4: example of a complex logic model

Study title: *How to and how not to develop a theory of change to evaluate a complex intervention: reflections on an experience in the Democratic Republic of Congo*.³⁰³

Summary: the Democratic Republic of Congo faces many health system challenges, one of which is the flawed wage system for health workers. This has a number of impacts, including reduced motivation and inadequate delivery of health care. The authors outline the process of developing a theory of change to inform the development and evaluation of an intervention that sought to improve government payments to health workers. They developed a logic model in collaboration with key stakeholders that was reviewed before the intervention was implemented and again afterwards. The logic model (*Figure 6*) attempts to demonstrate the complexity of the intervention. Key findings from this work include the importance of ensuring that all relevant stakeholders are included, the challenge of managing power dynamics among stakeholders and the importance of exploring the impact of the intervention on the wider system.

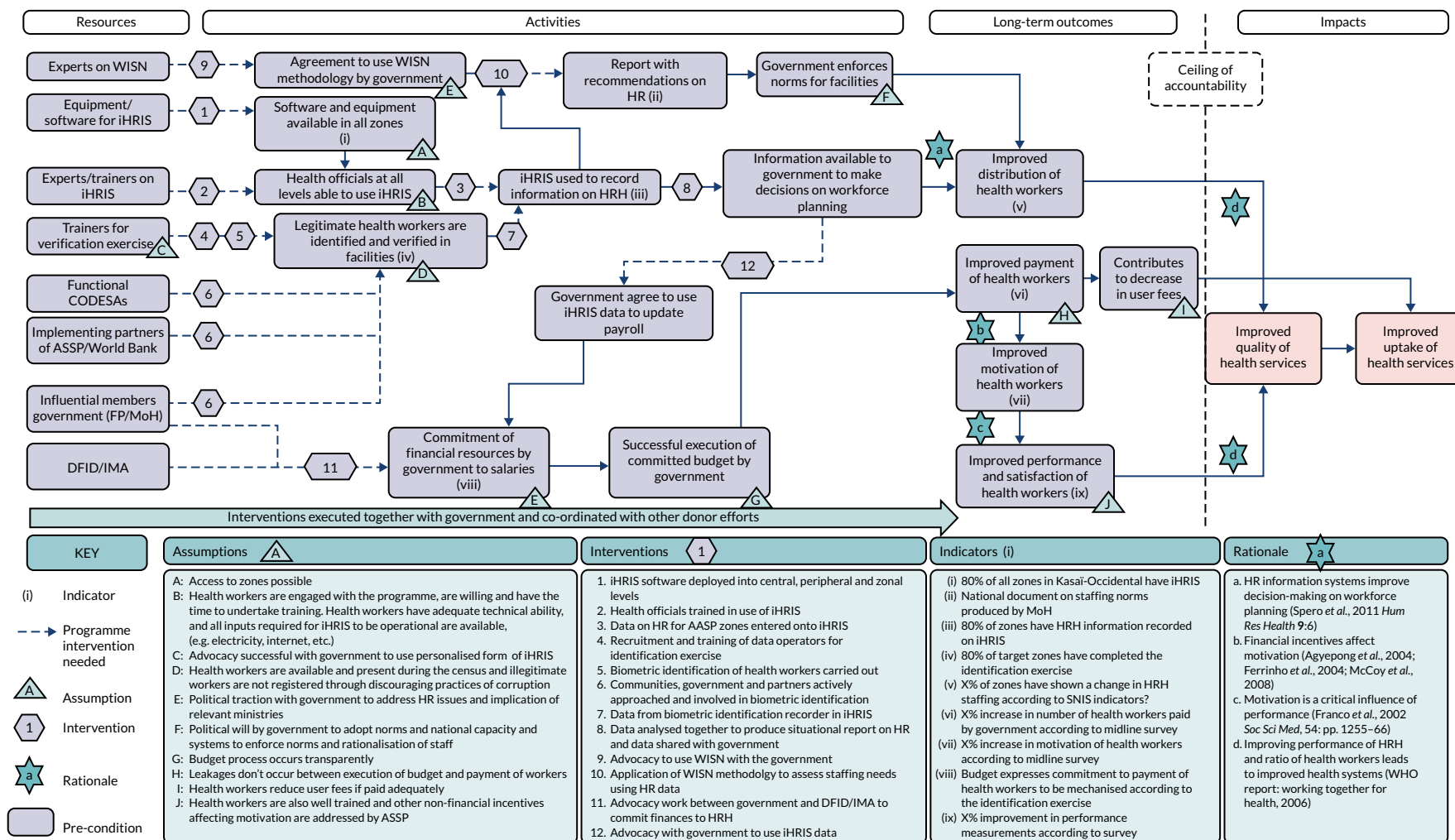


FIGURE 6 Logic model of a complex intervention to improve payment of salaries to health workers. Reproduced with permission from Maini *et al.*³⁰³ This is an Open Access article distributed in accordance with the Creative Commons Attribution Non-Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>. The figure includes minor additions and formatting changes to the original figure. ASSP, Accès Aux Soins de Santé Primaires; CODESA, community health committee; DFID, Department for International Development; FP, family planning; HR, human resources; HRH, human resources for health; iHRIS, IntraHealth International (Chapel Hill, NC, USA); IMA, IMA World Health (Washington, DC, USA); MoH, Ministry of Health; SNIS, single-number indicators; WISN, Workload Indicator of Staffing Need. Sources included as per original.^{304–306}

Key points:

- The logic model shows the resources and activities as well as the outcomes and impacts that you might expect to see in a more traditional logic model; however, it also includes assumptions, interventions, indicators and rationale.
- Importantly, the hypothesised mechanisms of change, including the pathways of each element, are considered.
- Contextual factors and their influence on the intervention processes are also included.

Case study 5: example of a dark logic model

Study title: *Communities in charge of alcohol (CICA): a protocol for a stepped-wedge randomised controlled trial of an alcohol health champions programme.*³⁰⁷

Summary: this study evaluated an asset-based community development approach to reducing alcohol harm. Volunteers were trained to become 'Alcohol Health Champions', to provide advice to individuals and to mobilise community-level activity focused on alcohol availability. As part of the development work for the evaluation, the authors considered the possible unintended harms of the intervention (dark logic). They developed a 'dark' logic model that demonstrates how potential negative consequences could result from the intervention (*Figure 7*).

Key points:

- The development of a dark logic model in advance of the evaluation helps evaluators to include appropriate methods to measure possible negative consequences.
- The dark logic model could inform changes to the intervention and reduce the risk of future harm.

Case study 6: example of a realist matrix-presenting programme theory

Study title: *Dealing with context in logic model development: reflections from a realist evaluation of a community health worker programme in Nigeria.*³⁰⁸

Summary: in this paper, the authors describe the development of a logic model within the context of an 'ongoing realist evaluation which assesses the extent to which and under what circumstances a community health workers programme promotes access to maternity services in Nigeria'.³⁰⁸ The development of a logic model is seen as key to identifying relevant hypotheses regarding context, mechanisms and outcomes, and how interventions lead to change. As part of the process of developing the logic model and to help elucidate the contextual factors influencing the community health workers programme, the authors ran stakeholder workshops and developed matrices of candidate theories as in *Table 10*, which outlines the tentative 'context, mechanism and outcomes' for the supply component of the community health worker programme. This paper helpfully outlines eight lessons learned by the authors of this realist evaluation.

Key points:

- The realist matrix enabled blending of context into the initial working programme theory, which in turn helped to map the relations between programme elements and identify evaluation measures to understand how, why, for whom and under what circumstances the intervention achieved its outcomes.
- Creating a programme theory is not the end – it is important to view programme theory as a flexible and iterative process to understand an intervention.
- Stakeholder engagement is crucial to ensure the development of a 'shared language' that is meaningful to a wider audience, not only researchers.

THE NEW FRAMEWORK

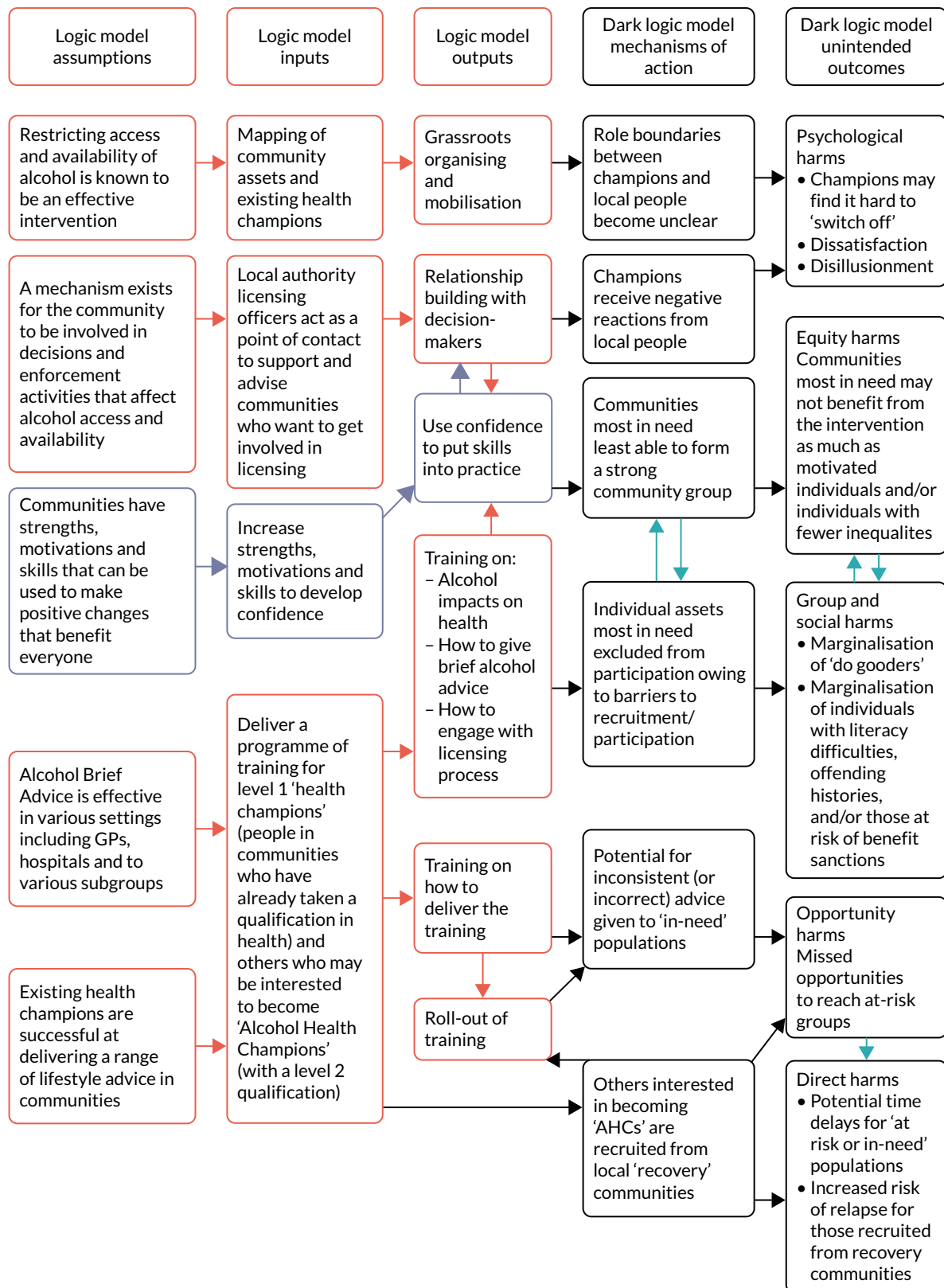


FIGURE 7 Dark logic model for Communities in Charge of Alcohol. Light blue arrows, potential links between harms; orange arrows, links between logic model inputs and outputs; black arrows, dark logic model mechanism; dark blue arrows, main logic model mechanism. AHC, Alcohol Health Champion. Reproduced with permission from Cook *et al.*³⁰⁷ This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated. The figure includes minor additions and formatting changes to the original figure.

TABLE 10 Matrix of context, mechanism and outcomes for the supply component of the community health worker programme

Levels of context	Levels of mechanisms	Levels of outcomes
Individual context:	Individual mechanism:	Individual outcome:
C1 – Non-experienced staff experience	M1 – Availability of health workers and skill mix of MCH staff ensured	O1 – Altruism and increased social responsibility
C2 – Demoralised staff		O2 – Increased staff motivation
C3 – Status and skill mix of MCH staff (CHWs, CHEWs and midwives)		O3 – Increased satisfaction
		O4 – Improved staff performance
Institutional context:	Institutional mechanism:	Institutional outcome:
C3 – Irregular salaries	M2 – Continuous training of staff	O5 – Increased staff retention
C4 – Poorly functioning facilities	M3 – Supportive supervision of staff	O6 – Improved quality of care delivered by facility
	M4 – Collegial working environment	O7 – Increased utilisation of ANC by women
C5 – Strained working relationships between CHEWs and nurses following policy change in PHC facility management	M5 – Regular payments are instituted	O8 – Increased skilled birth attendance
	M6 – Availability of equipment supplies and infrastructure	
Macro context:	Macro mechanism:	Macro outcome:
C6 – New government policy on social protection of vulnerable populations implemented as a pilot	M7 – Availability of SURE-P regulatory oversight	O9 – Reduced maternal mortality rate
		O10 – Reduced infant mortality rate

ANC, antenatal care/clinic; CHEW, community health extension worker; CHW, community health worker; MCH, maternal and child health; PHC, public health care; SURE-P, Subsidy Reinvention and Empowerment Programme.

Notes

From the above, we can start formulating hypotheses such as C1 + M1, M2, M4, M5 = O5, O6.

Candidate theory 1: if different incentives (e.g. regular payments, training and improved working environment) are available in a timely manner, this will lead to improved and sustained health worker motivation, job satisfaction, performance and improved retention of staff in the context of Anambra State that is characterised by irregular salaries and poorly functioning facilities.

Reproduced with permission from Ebenso *et al.*³⁰⁸ This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>. The table includes minor additions and formatting changes to the original table.

Case study 7: system map to develop programme theory

Study title: *Evaluation of the Health Impacts of the UK Treasury Soft Drinks Industry Levy (SDIL) Project Protocol.*²²⁴

Summary: White *et al.*²²⁴ took a systems approach to evaluate the UK's SDIL: a graduated levy on the producers of sugar-sweetened drinks. The levy is intended to encourage the reformulation of drinks, with the aim of reducing sugar consumption and its associated health consequences. The researchers theorised the levy as an 'event in a complex system'²²⁴ that might precipitate reactions among a wide array of stakeholders and affect a range of dietary and health outcomes. Anticipating direct and indirect effects across a range of interacting food and health systems, the researchers developed a bounded conceptual system map to articulate their understanding of how the intervention was expected to work, guide evaluation design and help identify data sources to support evaluation (Figure 8). The evaluation design includes multiple methods, for example qualitative process evaluation, natural experiment study, economic evaluation and simulation modelling of long-term (5–20 years) health and economic outcomes, with synthesis and triangulation of findings across methods planned.

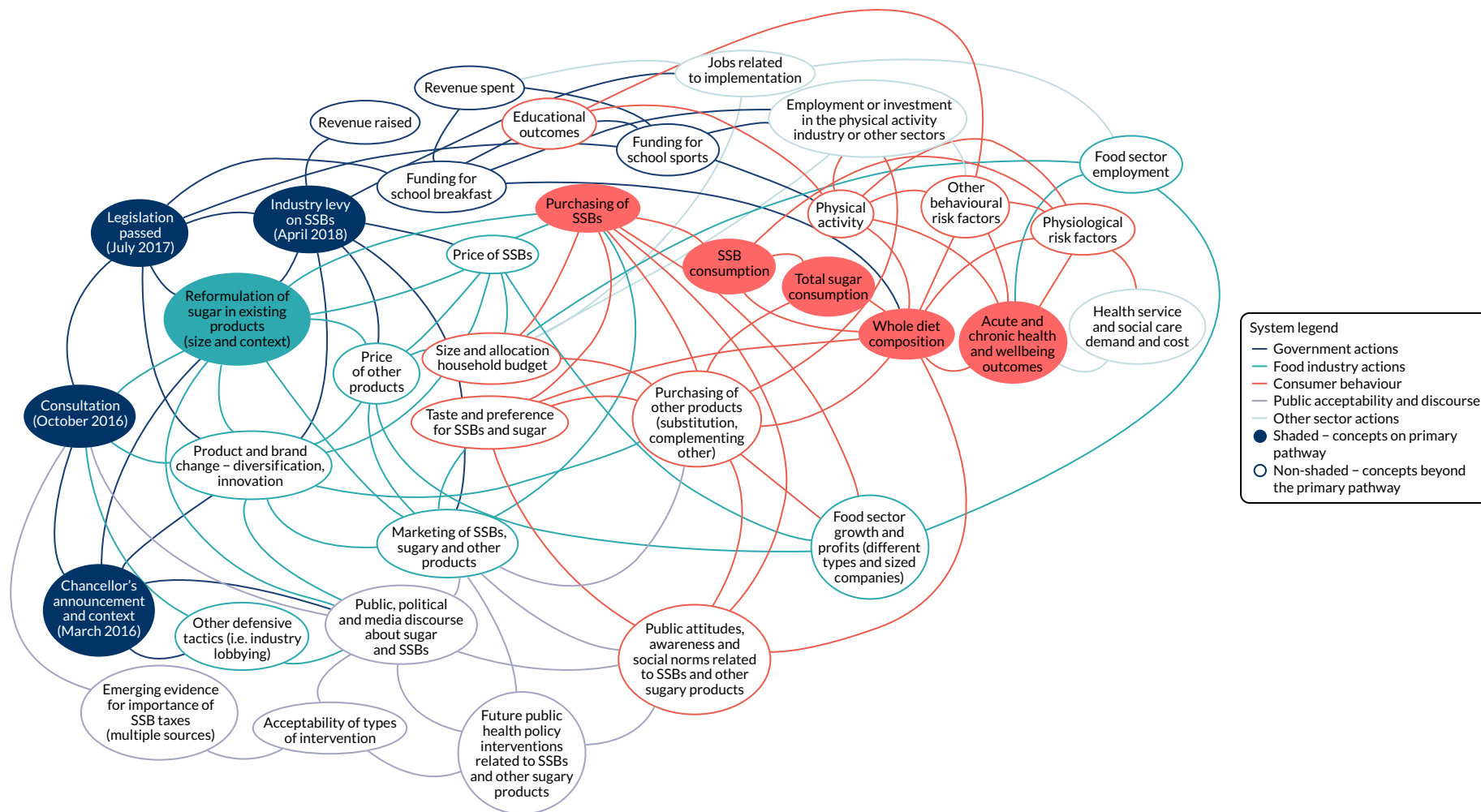


FIGURE 8 Conceptual system map for the UK SDIL. Context: modifying factors (e.g. age, gender, socioeconomic position and household composition) and macro-level factors (e.g. Brexit, economic climate, agricultural policy, global governance and trade). Adapted with permission from White *et al.*²²⁴

Key points:

- A review of published evidence enabled the development of a preliminary conceptual map of the system and a theory for the SDIL.
- An interactive expert workshop refined the system map and verified hypothesised entities and their relationships, as well as the underlying programme theory.
- An online Delphi study built consensus among a wide range of stakeholders from all sectors and groups identified on the map, and enabled refinement of the map and programme theory.
- The final map allowed the generation of specific hypotheses concerning causal pathways central to the programme theory, the identification of potential data sources to measure outcomes associated with each key pathway and the design of a range of qualitative and quantitative studies to assess change in relation to these pathways.
- Synthesis of the findings from these studies will be used to revisit and refine the system map and programme theory, and derive generalisable causal inference.

Case study 8: engaging stakeholders (service users)

Study title: *Reducing relapse and suicide in bipolar disorder: practical clinical approaches to identifying risk, reducing harm and engaging service users in planning and delivery of care – the PARADES (Psychoeducation, Anxiety, Relapse, Advance Directive Evaluation and Suicidality) programme.*³⁰⁹

Summary: this study evaluated approaches to reduce harm and improve outcomes for people with bipolar disorder. Stakeholder engagement, mainly in the form of service user involvement, was central throughout the duration of the Psychoeducation, Anxiety, Relapse, Advance Directive Evaluation and Suicidality (PARADES) programme, from the initial grant application to study close and dissemination. Service users were individuals with lived experience of bipolar disorder. ‘Involvement principles’ were agreed between service users and researchers at the outset of the programme and guided stakeholder engagement. Service users contributed to the PARADES programme in numerous ways, including delivering the intervention; offering insights on the duration and frequency of intervention sessions; practical issues, such as venue choices and accessibility; translating materials into culturally sensitive resources for participants; and helping to develop recruitment strategies. They also contributed to the interpretation and writing up of findings and the dissemination strategy. At the end of the study, they developed a user-friendly guide to the Mental Capacity Act³¹⁰ for people with bipolar disorder.

Key points:

- PPI activities in this study were wide-ranging and central throughout the project.
- Involving service users with lived experiences of bipolar disorder had many benefits for the research. It enhanced the intervention but also improved the evaluation and dissemination methods.
- The benefits were two-way because there were positive outcomes for service users involved in the study, including more settled employment, progression to postgraduate study, fellowship funding and co-authorship of several papers.

Case study 9: refining the intervention

Study title: *The person-based approach to enhancing the acceptability and feasibility of interventions.*³¹¹

Summary: this paper describes using mixed-methods approaches to develop an in-depth understanding of the needs and views of people who will use the intervention. This understanding ‘enables intervention developers to design or modify the intervention to make it more relevant, persuasive, accessible and engaging’.³¹¹ Three illustrative examples are used to demonstrate the iterative process:

- Illustrative example 1 involved developing guiding principles that informed the design of a digital intervention for asthma self-management. Guiding principles were developed based on existing evidence and relevant issues that the intervention would need to address. The process involved

developing key intervention design objectives to address identified issues, and intervention features were developed to address these design objectives.

- Illustrative example 2 used think-aloud interviews, alongside PPI input, to improve the intervention booklet. This involved changing the images and wording of the booklet several times, leading to a more acceptable resource for participants. This refinement process is likely to lead to higher engagement with the booklet.
- Illustrative example 3 also used think-aloud interviews to improve the usability of a physical activity planner for people with diabetes mellitus. In this example, a physical activity planner was found to be difficult to use, resulting in participants overestimating their level of physical activity. Several iterations of the planner were developed based on interviews and observations and this iterative process led to the planner being much more feasible and accurate.

Key points:

- The feasibility and acceptability of these interventions was improved by involving potential users to inform intervention refinements.
- This process of iterative intervention refinement increases the chances that the interventions will be acceptable to participants and feasible to deliver.
- It is also likely that this refinement process will increase the chances of the interventions being found to be effective in a future evaluation.

Case study 10: economic considerations

Study title: Adaptation of the New Orleans Intervention Model (NIM) for infants entering foster care – the Best Services Trial (BeST?) in Scotland, UK.^{265,312–318}

Summary: the New Orleans Intervention Model (NIM) for infants entering foster care seeks to improve the quality of placement decisions while also improving the mental health outcomes for these infants. The intervention is set within a complex context, spanning social care and the NHS, in which the comparator (service as usual) varies in service delivery across the UK. The development and testing of the intervention within a UK context has been conducted in an iterative manner,³¹² with the inclusion of economic considerations from the outset.³¹³

An early-stage exploratory economic model was developed comparing the proposed NIM intervention with the existing local foster care service. A full economic evaluation was deemed inappropriate at this stage, as the existing foster care service was not clearly defined and there was no direct clinical evidence on health outcomes in this or the proposed new system. Instead, a CCA framework was used to map and measure existing local services and compare them with a hypothetical model of the NIM intervention. This allowed decision-makers to compare the potential costs and consequences of these two approaches and form their own view of the importance of the different outcomes.¹⁴¹

Evidence from the early-stage model helped to build a case for a Phase II exploratory trial in Glasgow,^{314–317} including development and feasibility testing of economic resource use data systems. The feasibility study led to a definitive Phase III trial³¹⁸ that included short- and long-term economic analysis from multiple perspectives;²⁶⁵ included NHS and Personal Social Services, public sector and wider societal perspectives; and used a range of frameworks, including CUA and CCA to capture changes in the intersectoral costs and outcomes associated with child maltreatment.

Key points:

- An iterative approach was taken for the design and conduct of the economic evaluation.
- The evaluation took into account outcomes spanning a range of sectors; short-term, intermediate and long-term time horizons; and evidence from a range of sources.
- The use of multiple economic evaluation frameworks will provide decision-makers with a comprehensive, multiperspective guide to the cost-effectiveness of the NIM.

Case study 11: intervention development: example systems map

Study title: *Tackling 'wicked' health promotion problems: a New Zealand case study.*³¹⁹

Summary: this study describes a complex environmental approach to addressing 'wicked' health promotion problems. The starting point for this research was the problem – obesity, poor nutrition and inadequate levels of physical activity among Māori, Pacific and low-income people in New Zealand – rather than a particular intervention. The project involved literature reviews, focus groups, stakeholder workshops and key informant interviews to identify important areas for intervention and provide recommendations for action. The system was mapped (Figure 9) and analysed, identifying 'control parameters', 'elements of a system that have a controlling influence on other system elements, acting to set the boundaries within which the system can operate',³¹⁹⁻³²² that are open to manipulation. Stakeholder workshops explored the options for interventions to affect change on the control parameters, and intervention areas were then prioritised.

Key points:

- The research took a systems perspective that aimed to inform system-level change.
- Mapping of the relevant systems facilitated the identification of a suite of interventions that were likely to be effective, taking into account feasibility, acceptability, sustainability, equity, side effects and costs.
- A collaborative strategy for intervention development was used, with effort taken to include relevant multidisciplinary stakeholders, including community members and policy-makers.
- A mixed-methods approach was taken utilising data from multiple sources.
- Cost-benefit in terms of intervention options was considered from the outset to ensure value for money.

Case study 12: feasibility trial

Study title: *Randomized feasibility trial of directly observed versus unobserved hepatitis C treatment with ledipasvir-sofosbuvir among people who inject drugs.*³²³

Summary: this study explored the feasibility of an intervention to reduce secondary transmission of the hepatitis C virus in people who inject drugs. The uncertainties identified at the start of the study included the feasibility of treating people at high risk for secondary transmission and the optimal approach for delivering treatment. The study also explored the feasibility of recruitment, adherence to the intervention, retention in the study and potential harms. The findings demonstrated that both modes of delivery were feasible and acceptable to participants. It was also feasible to recruit people who inject drugs for this type of treatment, retain them over a 12-week follow-up period and achieve positive viral responses to treatment.

Key points:

- Uncertainties were identified before the study commenced, which guided the choice of intervention and evaluation methods.
- The feasibility and acceptability of the intervention and evaluation methods were tested.
- Useful data were gathered to optimise the design of a future evaluation.
- The findings addressed the initial uncertainties, thereby strengthening the support for future funding and/or evaluation.

Case study 13: feasibility, evaluability assessment

Study title: *Practice to evidence: using evaluability assessment to generate practice-based evidence in rural South Georgia.*³²⁴

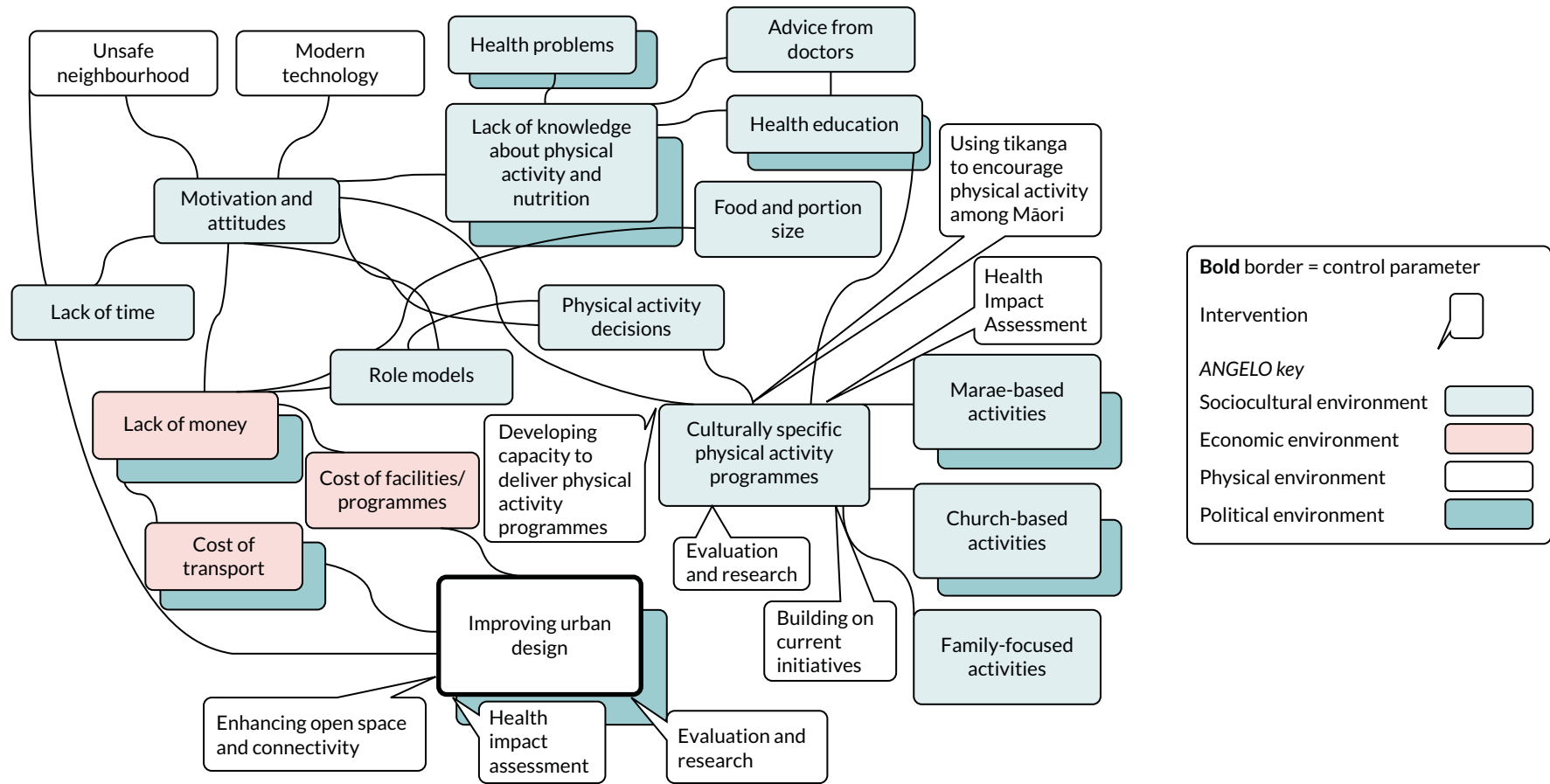


FIGURE 9 Physical activity system map with interventions. Reproduced from Signal LN, Walton MD, Mhurchu CN, et al. Tackling ‘wicked’ health promotion problems: a New Zealand case study, *Health Promotion International* 2013;28(1):84–94,³¹⁹ by permission of Oxford University Press.

Summary: the Emory Prevention Research Centre used evaluability assessment methods as part of an initiative to gather practice-based evidence on cancer prevention from community interventions in south-west Georgia, USA. Of nine interventions that were identified and screened for eligibility, two were selected for a full evaluability assessment. Each evaluability assessment comprised a review of documents related to the intervention, site visits, meetings with stakeholders and reviews of published literature. The aim was to assess the feasibility and usefulness of an evaluation based on the programme theory underpinning the intervention, the availability of data, stakeholder interest and its potential contribution to the evidence base. From the two evaluability assessments, one intervention was identified as eligible for a quasi-experimental outcome evaluation. Data limitations precluded an outcome evaluation of the second intervention and a recommendation was made instead for either a process evaluation of one component of the programme or a descriptive case study.

Key points:

- In the context of scarce evaluation resources, partnerships between research and practice can helpfully identify interventions that can generate useful findings and inform future implementation.
- Interventions were screened from current practice and evaluability assessments used to assess those with the potential for formal evaluation and to recommend an evaluation method.
- Findings from the initiative described were included in national resources for practitioners looking at evidence-based practices and helped the community organisations expand and strengthen their programmes.

Case study 14: evaluation (effectiveness perspective: cluster randomised controlled trial with theory-informed process evaluation)

Study title: *Management of multimorbidity using a patient-centred care model: a pragmatic cluster-randomised trial of the 3D approach.*¹⁷⁷

Summary: this pragmatic cluster RCT is an effectiveness study of the 3D intervention, a primary care approach for patients with multimorbidity, which focuses on quality of life and self-management to achieve goals agreed between the patient and the health-care provider. The intervention aimed to improve health-related quality of life. Process and cost-effectiveness evaluations were also completed. There was no difference in quality of life between intervention and control groups.

Key points:

- Cluster RCTs are useful when an intervention is aimed at communities rather than individuals, and also where there is potential for treatment contamination, for example where control participants may gain access to intervention materials. However, they are more complex to design and analyse and increase the sample size requirements.
- The trial process evaluation sought to understand how and why the intervention was effective (or not) and to identify strategies for implementation that considered the influence of context, as well as practical issues related to adoption, delivery maintenance and reach. Examples of research questions were 'to what extent was the intended intervention delivered to patients?'¹⁷⁷ and 'how did patients perceive the intervention and its delivery?'¹⁷⁷
- The interpretation of findings was enhanced by the nested qualitative process evaluation that followed a prespecified protocol.^{190,325}
- An external pilot of the study was conducted before proceeding to this trial.
- A study protocol was written, published and followed.³²⁶ This included a prespecified analysis plan for primary and secondary outcomes.
- Economic evaluation of cost consequences and cost-effectiveness was conducted alongside the trial.³²⁷

Case study 15: evaluation (efficacy)

Study title: *Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis*.¹⁷⁴

Summary: this RCT in 50 centres tested, among 659 patients with a high risk of stroke (70–99% stenosis of a carotid artery), whether or not the addition of carotid endarterectomy to optimal medical therapy (compared with optimal medical therapy alone) reduces the risk of stroke. The trial results indicated that carotid endarterectomy was beneficial.

Key points

This efficacy trial can be differentiated from an effectiveness trial based on the following characteristics²⁵⁹ –

- The inclusion criteria for the trial were narrow. The main group of interest were those who were at high risk of stroke and those who were most likely to respond to the intervention. The exclusion criteria were wide-ranging; therefore, the intervention was tested on a very narrow group of patients.
- The intervention delivery was tightly constrained with little flexibility, and those delivering the intervention were highly trained. The surgeons taking part had to be approved by an expert panel and had to have completed at least 50 carotid endarterectomies in the last 24 months.
- The control intervention was not treatment as usual but included antiplatelet therapy and the same co-interventions applied to the intervention arm were also given to control patients.
- Follow-up was very intensive at multiple time points.
- Practitioner adherence to the study protocol was monitored.
- In real-world practice, much of the above would not happen; therefore, the impact of the intervention is likely to be different in a more pragmatic setting.

Case study 16: evaluation (effectiveness perspective: individual randomised controlled trial)

Study title: *Exercise- and strategy-based physiotherapy-delivered intervention for preventing repeat falls in people with Parkinson's: the PDSAFE RCT*.¹⁷⁵

Summary: people with Parkinson's disease are at an increased risk of falls. Although there is some evidence that exercise might reduce fall risk, the findings are mixed. This study examined the clinical effectiveness and cost-effectiveness of PDSAFE, an exercise- and strategy-based intervention for fall reduction for people with Parkinson's disease. The study was a multicentre, individual RCT with an embedded economic evaluation and qualitative process evaluation. A published protocol included prespecified subgroup analyses. This was useful because, although the intervention was not effective overall in terms of reducing fall risk, these analyses indicated that it may be of benefit for those with 'moderate' Parkinson's disease, but it may increase falls in those with more severe Parkinson's disease. The authors concluded, therefore, that a different treatment strategy could be used for those with more severe Parkinson's disease and that the benefits indicated for those with 'moderate' Parkinson's disease would need to be further studied in an effectiveness trial for this group before implementation.

Key points:

- This trial tested the intervention in those patients who are representative of those likely to receive the intervention in routine practice.
- The qualitative process evaluation provided information on barriers to and potential facilitators of participation (e.g. existing social support). These aspects could be incorporated to further enhance this intervention for more targeted groups, for whom the intervention is more likely to be effective.
- An economic evaluation was conducted as part of the trial and included subgroup analyses.
- Engagement with stakeholders, including professionals and patients, supported the development of some of the study materials.
- In line with good reporting practice, CONSORT and Template for Intervention Description and Replication (TIDIER) descriptions were provided.

Case study 17: evaluation (effectiveness perspective: stepped-wedged cluster randomised design)

Study title: *Effect of the World Health Organization (WHO) Checklist on Patient Outcomes: a stepped wedge cluster randomized controlled trial.*¹⁷⁸

Summary: this stepped-wedge cluster RCT examined the effects of implementing WHO's surgical safety checklist on in-hospital complications, length of stay and mortality. The stepped-wedge design was used because the surgical safety checklist intervention was thought more likely to do good than harm; therefore, it would be unethical to not deliver or to withdraw the intervention. The surgical safety checklist was randomly rolled out in sequence until five clusters had received the intervention. The different clusters (surgical specialties) switched treatment from control (no checklist) to intervention (surgical safety checklist) at different time points. The clusters crossed over in one direction, that is the checklist was introduced and not withdrawn after implementation. The surgical specialty cohorts acted as their own controls providing data before and after the checklist intervention. The analyses compared all data in the steps before (controls) with the steps after (intervention) the intervention. The results indicated that use of the surgical safety checklist led to important reductions in morbidity and length of hospital stay.

Key points:

- The stepped-wedge cluster randomised design is especially useful where (1) it would be unethical to not deliver or to withdraw the intervention; (2) randomising individuals or procedures would not be possible; (3) contamination is likely to be a major problem; (4) there are practical or financial constraints that make it impossible to deliver the intervention to all participants at the same time.
- Issues with the stepped-wedge design are that it can require a large number of data. In this example, routinely collected data were used to address this issue. It can also be difficult to implement the interventions within the planned time frame because of the practicalities of working within complex health-care organisations.

Case study 18: evaluation (effectiveness perspective: synthetic control)

Study title: *Examination of the synthetic control method for evaluating health policies with multiple treated units.*¹⁷⁹

Summary: The Advancing Quality scheme (a hospital pay-for-performance scheme) was re-evaluated using the synthetic control method and the findings were contrasted to the original differences-in-differences analysis (in which the main assumption is parallel trends). The original differences-in-differences study compared intervention hospitals with hospitals that did not introduce the Advanced Quality scheme, and found that risk-adjusted hospital mortality was reduced for the three incentivised conditions.³²⁸ Concerns were raised about whether or not the parallel-trends assumption was upheld given an inspection of risk-adjusted mortality in the periods before the introduction of the Advanced Quality scheme. The data were, therefore, re-evaluated using the synthetic control method, which allows for estimation of intervention effects by constructing a weighted combination of control units. This represents an estimate of what the intervention group would have experienced in the absence of the intervention; see the full paper for the methods.¹⁷⁹ Results did not align with those from the original differences-in-differences analysis, but rather found that the scheme did not reduce mortality for incentivised conditions and that mortality was significantly increased in non-incentivised conditions. Sensitivity analysis upheld findings from the synthetic control method.

Key points:

- Many public health interventions cannot be evaluated using RCTs and rely on the assessment of observational data.²¹⁵ Building synthetic controls, using a weighted combination of potential control units, is one way of overcoming any difficulties in selecting available, suitable controls.

- Where the validity of the parallel-trends assumption is doubtful, the synthetic control method may be useful. This method allows for the effects of unobserved confounders to change over time by weighting the control group so that it has similar pre-intervention characteristics to the intervention group.³²⁹
- This study adds to concerns that the parallel-trends assumption underlying the differences-in-differences approach may not be plausible in settings where the differences between the comparison groups in pre-intervention outcomes are relatively large, showing the appropriate use of a synthetic control approach.

Case study 19: development and evaluation – effectiveness perspective (drawing on a systems perspective for intervention development and evaluation)

Study title: *Whole of Systems Trial of Prevention Strategies for childhood obesity: WHO STOPS childhood obesity.*¹⁸⁰

Summary: this paper presents the protocol for a stepped-wedge cluster RCT evaluating WHO STOPS, a complex intervention targeting childhood obesity that was developed using a systems perspective. The intervention engages and facilitates community members to agree a system map of causes of childhood obesity for the community; 'identifies intervention opportunities through leveraging the dynamic aspects of the system';¹⁸⁰ and uses these to co-produce systems-oriented action plans that are implemented in a way that allows flexible and iterative development of the intervention over time. Pilot work undertaken with communities using group model building to build causal loop diagrams indicated that participants were able 'to identify multiple systems that impact childhood obesity'¹⁸⁰ and 'design interventions that considered non-linearity, feedback, and complexity for intervention development';¹⁸⁰ see Allender *et al.*³³⁰ for a worked example of a causal loop diagram. The evaluation will test the impact of the intervention on childhood obesity prevalence. Ten communities will be randomised: five to receive the intervention at year 1 and five to receive the intervention at year 3. The evaluation also draws on a systems perspective and includes numerous systems methods to explore the impact of the intervention at a systems level.

Key points:

- The study involves stakeholder engagement and considers implementation and context from the outset.
- Systems boundaries will be predefined and agreed by a range of community stakeholders.
- Systems change will be assessed alongside the primary outcome (childhood obesity) using a variety of methods, for example using social network analysis.
- A dynamic logic model will be created and updated throughout.
- Economic evaluation is included in the trial, using approaches that are suitable to capture systems change as well as equity impacts.
- Qualitative methods, largely drawing on soft systems methodology (e.g. group model-building workshops),³³¹ and quantitative methods will be employed throughout, and the process evaluation will assess the response of the community to the systems intervention.

Case study 20: evaluation (theory-based perspective, alongside effectiveness)

Study title: *A realist process evaluation of robot-assisted surgery: integration into routine practice and impacts on communication, collaboration and decision-making.*¹⁸²

Summary: robot-assisted surgery has been shown to have potential benefits for patients; however, despite the technology being available, this has not yet fully been implemented into routine practice in the NHS. This realist evaluation aimed to explore 'how and under what circumstances robot-assisted surgery is effectively introduced into routine practice, and how and under what circumstances robot-assisted surgery impacts communication, teamwork and decision-making'.¹⁸² This was done alongside an effectiveness trial of robot-assisted surgery.^{332,333}

Phase 1 comprised a literature review that explored stakeholder theories of how robot-assisted surgery can become embedded into routine practice, and particularly how it affects communication, decision-making and teamwork. The theories were further developed using data from stakeholder interviews. In phase 2, four theories were selected from the initial theories developed in phase 1 for testing in a multisite case study across four NHS trusts. This involved multiple methods, such as video recordings of operations, ethnographic observation, post-operation interviews and a questionnaire. In phase 3, interviews with surgical staff from other specialties were completed to explore whether or not the results of phase 2 were generalisable and to refine the theories to include the experience of a wider group of surgical specialties.

Key points:

- A theory-based evaluation can usefully be carried out alongside an effectiveness trial to understand the variations in implementation of the intervention and the influence of context on outcomes.
- Multiple methods can be used to investigate how and under what circumstances an intervention affects processes related to the implementation of an intervention.
- Data collection and analysis took an iterative approach meaning that refinement of theories was ongoing, and further data could be gathered as necessary.
- Theory-based evaluation can provide standalone findings, as well as complement an effectiveness evaluation.
- The study resulted in several practical recommendations to support the implementation of robot-assisted surgery into routine surgical practice. For example, engaging staff at different levels, including surgeons who will not be using robot-assisted surgery, whole-team training and a suitably sized operating theatre.

Case study 21: evaluation (theory-based perspective with realist approach)

Study title: *'Patients are not following the [adherence] club rules anymore': a realist case study of the antiretroviral treatment adherence club, South Africa.*¹⁸³

Summary: effectiveness studies have shown positive results of the adherence club intervention model, an antiretroviral treatment delivery model, in terms of improved retention in care and adherence to antiretroviral treatment for people living with human immunodeficiency virus (HIV). However, understanding how, why and under what health system contexts the intervention could be most effective were not clear. The aim of this realist evaluation was to test the initial programme theory and to 'determine how, why, for whom, and under what health system context the adherence club intervention works (or not) in real-life implementation'.¹⁸³ The study identified certain mitigating circumstances that could inform implementation of the intervention.

Key points:

- The evaluation was transparent, in that a prespecified plan for the research was set out in a published protocol paper.³³⁴
- Mixed methods were used within a case study approach. Qualitative methods were used 'to explore implementation features related to the context and mechanisms, and quantitative methods to describe and classify the outcomes'.¹⁸³
- Data from various participants (health-care providers and patients) were triangulated to verify what was obtained.
- An explanatory tool [Intervention-Context-Actor-Mechanism-Outcome (ICAMO; intervention-context-actor-mechanism-outcome)] was used to construct theories, building on realist context – mechanism – outcome configurations, to explore how the 'outcome (O) is generated by a mechanism (M), being triggered in context (C), through an actor (A), when the intervention is implemented (I)'.¹⁸³

- Links between each element of ICAMO were conceptualised and then compared and contrasted with the initial programme theory. A transparent and systematic process was followed.
- The initial programme theory was modified to reflect findings, providing an updated explanation of how, why, for whom and under what health system conditions the adherence club intervention works (or not).
- Guidelines for reporting realist evaluation studies were followed.²⁵⁵

Case study 22: evaluation (theory-based approach)

Study title: *The impact on health inequalities of approaches to community engagement in the New Deal for Communities regeneration initiative: a mixed-methods evaluation.*¹⁸⁴

Summary: this study drew on a theory of change approach to evaluate the New Deal for Communities, thus exploring how and why the intervention ‘works’. It explored the social and health impacts of different community engagement approaches that were undertaken as part of the New Deal for Communities regeneration initiative in England. There were 39 local New Deal for Communities programmes that were each given a budget to engage residents in planning and implementing initiatives to reduce crime and worklessness, and improve community cohesion, housing, the physical environment, education and health. The study used primary and secondary data, underpinned by theories of engagement and empowerment, to categorise the community engagement approach of each programme into four types and explore community empowerment as a mechanism of change within these approaches. Analysis explored how the four community engagement types related to different health and social outcomes, ultimately looking at whether or not the different community engagement approaches explained any of the differential outcomes of the New Deal for Communities programmes (that were previously identified). Economic analysis was also undertaken to explore the cost-effectiveness of different community engagement approaches. Statistical results were mixed and most were not statistically significant, and evidence of effectiveness was inconclusive because of limited accurate cost data. However, patterns of change were consistent with theories about the relationship between empowerment and health and social outcomes. The theoretical interpretation highlighted which of the different community engagement approaches could be problematic and may lead to unintended negative health and social outcomes, and which could more plausibly have positive impacts and, thus, how community engagement approaches could be made to work better. The findings highlight the need for prudent application of theory to make sense of effectiveness studies, in this case where results are inconclusive.

Key points:

- The typology of community engagement had strong theoretical underpinnings and was sense checked with the New Deal for Communities residents and practitioners, who were also advisors in the study.
- Theory of change was explored by testing plausible programme theories in practice.
- The value of this study lies in providing an understanding of the mechanisms of change and relationship with context, rather than the effect estimate.

Case study 23: evaluation (systems perspective)

Study title: *Evaluating a complex health promotion intervention: case application of three systems methods.*¹⁸⁵

Summary: the Delaware Young Health Program IM40 was an initiative to address the health and well-being of young people in Delaware through a developmental assets approach. It was funded by AstraZeneca plc (Cambridge, UK), a pharmaceutical company. By drawing on systems thinking and complexity science, the study aimed to identify and examine underlying patterns and structures in the network of individuals and their relationships that influenced system-wide behaviours, for example a culture shift in communities to better value young people. IM40 was a complex intervention in the sense that it sought to influence several levels simultaneously (individual, professional, procedural and policy)

and it was flexible and adaptive. Participating local communities were expected to follow a set of principles rather than adopt particular strategies. Each community had a different starting point in terms of relationships between organisations and existing strategies. Three systems science methods were used: group model building and viable systems model assessment to identify underlying patterns and structures, and social network analysis to evaluate change in relationships over time.

Key points:

- IM40 was expected to change underlying system dynamics, structures and conditions; therefore, the evaluation identified systems-oriented research questions and methods.
- The systems-oriented research questions were checked with stakeholders to ensure that they were relevant and useful. The study used mixed methods to gather data.
- A developmental evaluation focus was taken, which was appropriate to the complexity involved and the early stage of the work.
- Group model building was used to collaboratively create a causal model of the intervention with key stakeholders. The viable systems model method was used to model and assess the emergent organisational structure of the intervention. Social network analysis was conducted to explore shifts in the relationships between actors in the system. Strengths and limitations of each method used are reported in the paper.

Case study 24: evaluation (systems perspective with process evaluation)

Study title: *Consequences of removing cheap, super-strength beer and cider: a qualitative study of a UK local alcohol availability intervention.*^{186,187}

Summary: 'Reducing the Strength' is an intervention in which local authorities ask retailers to stop selling 'super-strength' beer and cider. Street and homeless drinkers are the target population for the intervention. The mixed-methods process evaluation took a systems perspective to explore the pathways through which the intervention may lead to intended and unintended consequences. The authors interviewed relevant stakeholders including compliers and non-compliers (shop keepers), the target public (homeless people) and stakeholders from a wider array of relevant services (homeless service providers relevant to alcohol and accommodation provision, and community safety officers). This was carried out to determine whether or not the intervention was developed to fit coherently within the wider system and to look at the consequences of different forms of adaptation and feedback, for example finding that homeless people adapt by changing their method of intoxication or by switching to shops that do not comply with the intervention. The authors conclude that owing to the above issues and the small-scale implementation and limited range of products included in Reducing the Strength, the intervention can make only a modest impact on alcohol harms. However, the intervention may lead to wider system changes as it impacts on services not just individual drinkers.

Key points:

- Reducing the Strength was conceptualised as an event in a complex system, and the way in which multiple people or agents within the system responded to it was explored.
- Multiple perspectives were sought.
- The study drew on complexity theory and developed multilevel theories of change to explain the impacts of the intervention.
- The qualitative work identified how the Reducing the Strength intervention affected components of the system and teased out the mechanisms by which the intervention triggered system changes.
- Qualitative methods may help address barriers to evaluating local health policy innovation, 'which can include small delivery scales, rapid delivery timescales, and a demand from local decision makers for evidence that is sufficiently contextually rich to be recognisable to them as local'.¹⁸⁶

Case study 25: evaluation (systems perspective with qualitative comparative analysis)

Study title: *A qualitative comparative analysis of factors associated with trends in narrowing health inequalities in England.*¹⁸⁸

Summary: targets for reducing health inequalities in England were adopted by the UK Government in 2001. Seventy 'spearhead' local authority areas were identified (those in the bottom fifth for five factors related to mortality, morbidity and the Index of Multiple Deprivation: 28% of the population in England). An example target was that the difference in life expectancy between the English and the spearhead areas average should reduce by 10%. The study aimed to explore variation between areas on their progress in achieving the targets, particularly looking at cardiovascular disease and cancer outcomes. QCA was used to explain why particular 'cases' have particular outcomes. Each spearhead area was a case in the analysis and each case was thought of as a local system, with different organisational attributes and contextual features. Data were gathered about the local conditions of each case (at present time and 3 years previous) using questionnaires completed by relevant professionals in each area. Secondary data were also used, for example hospital admissions and crime rates. The fsQCA software (version 3.0; Ragin and Davey, Irvine, CA, USA) was used to explore the effects of the attributes on the outcome indicators in combination with each other rather than individually. This showed the different configurations of attributes, for example the combination of 'individual commitment and champions' and 'higher spend per head on cancer programmes', that were associated with narrowing/widening inequalities between spearhead areas and English averages. Necessary and sufficient conditions were then identified for each morbidity and mortality outcome; for example, for narrowing the cancer gap, 'championing' was present in all configurations and was absent where the cancer gap was not narrowing. Results of the study were presented at a series of local stakeholder workshops, and to understand more about the sufficient conditions identified, such as individual commitment and championing, the authors planned to have discussions with local stakeholders.

Key points:

- QCA focuses on inter-relations between the system's constituent parts rather than on the individual components.
- Substantive and theoretical reasoning, rather than experimental design, is used to justify causal pathways.
- Anomalous cases are used as triggers for further investigation.
- Stakeholder involvement is key to the QCA process, both for defining and measuring important attributes and for interpreting the results.

Case study 26: intervention development – discrete choice experiments

Study title: *Using discrete choice experiments to inform the design of complex interventions.*²⁶²

Summary: discrete choice experiments can be used to quantify user preferences for different aspects of an intervention and to explore heterogeneity in preferences among users. This study used a discrete choice experiment in combination with a qualitative study to develop a voluntary medical male circumcision intervention for HIV infection prevention in the United Republic of Tanzania. Qualitative data from interviews and participatory group discussions with women, young men (< 20 years) and older men (\geq 20 years) were used to identify barriers to and facilitators of voluntary medical male circumcision, such as the role of female partners, service providers' attitudes and social stigma. A discrete choice experiment ($n = 325$) was used to measure men's preferences for or against characteristics of voluntary medical male circumcision services, such as availability of partner counselling, age-separated waiting areas and female health-care staff. The findings were used to develop an intervention that included community information booths, separate waiting areas and training for providers on client friendliness.

Key points:

- Qualitative and quantitative data were combined to develop an intervention to inform the development of a complex intervention.
- Participant preferences are an important determinant of uptake and adherence to an intervention, such as voluntary medical male circumcision, given the stigma associated with HIV.
- The use of a discrete choice experiment, informed by qualitative information about barriers to and facilitators of, enabled the researchers to design an intervention aligned with user preferences.

Case study 27: implementation in all phases

Study title: *An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial.*¹⁴⁶

Summary: the ASSIST programme identified the most influential students in schools from the 12- to 13-year age group, who were then trained to act as peer supporters during informal interactions outside the classroom to encourage their peers not to smoke. The intervention was based on diffusion of innovation theory and adapted an effective adult-focused sexual health programme. It was developed in collaboration with health and education professionals, and the intervention and evaluation methods were assessed in a feasibility study. A cluster randomised effectiveness trial in 59 schools with nested process and economic evaluations found the intervention to be cost-effective. Key contextual factors, intervention mechanisms and variations in implementation were identified.³³⁵⁻³³⁷

Key points:

- The intervention was designed to have minimal dependence on school resources, and to avoid poor implementation, often identified as a key problem of smoking education in real-world classrooms.
- A feasibility study identified a number of changes to the prototype programme to improve its acceptability to students and schools and to improve the reach of the programme to male students.
- In the evaluation phase cluster randomised trial, the intervention was delivered as closely as possible to a real-world implementation. Process evaluation data and the programme theory were used to identify the critical components of the intervention and key barriers to and facilitators of delivery.
- Programme implementation involved an intervention manual that identified both critical components and adaptable components, allowing for flexible implementation while retaining the key mechanisms of change. A training manual for the trainers and ongoing quality assurance were built into longer-term roll-out, based on the programme theory and process evaluation findings.

Acknowledgements

We would like to thank the following for their contributions to the work:

- Matt Egan, Mike Kelly, Steve Cummins, Vanessa Er, Karen Lock and Harry Rutter for invaluable initial and ongoing discussions about taking a systems approach.
- Matt Egan, Dave Richards, Jeremy Segrott, Graham Moore, Mark Richardson, Mike Robling and Matti Heino for providing feedback on a near final draft.
- Alicia O’Cathain, Graham Moore, Lucy Yardley and Ed Duncan for discussion and input on the development phase of the framework.
- Gavin Malloch and Sam Rowley (MRC) for their input and assistance throughout
- Martin Ashton-Key for his contribution to the SAG.
- Participants of the expert workshop (named in *Appendix 3*) for a full day of in-depth discussion
- All anonymous respondents of the open consultation on the draft document.
- The conference delegates who participated in the workshops that were held at various conferences [UK Society for Behavioural Medicine (UKSBM), 12–13 December 2018, Birmingham, and UKSBM, 13–14 December 2017, Liverpool; Society for Social Medicine and Population Health, 5–7 September 2018, Glasgow].
- Cindy Grey for work on drafting the FFIT case study.
- We are grateful to NIHR and MRC for the funding received to complete this project.
- MRC Population Health Sciences Group and the MRC–NIHR Methodology Research Programme, particularly Hywel Williams and colleagues, for their review and comment.
- Anonymous peer reviewers (in the NIHR Journals Library submission process) for helpful and considered comments.
- Colleagues at Social and Public Health Sciences Unit, specifically: Olga Utkina-Macaskill for administrative support throughout; Enni Pulkkinen for support with diagrams; Valerie Wells and Lorna Dick for help with editing; Andrew Jackson for setting up the online consultation; Jessica Moran, Marcia Gibson, Michele Hilton-Boon, Andrew Baxter and Grace Dibben for proofreading; Vittal Katikireddi and Mark McCann for discussion on drafts; Susie Smillie, Natalie Chalmers, Olga Utkina-Macaskill, Arlene McGarty and Anne Martin for note-taking at the expert workshop; and Alice MacLachlan for assisting with the *Plain English summary*.

Contributions of authors

Kathryn Skivington (<https://orcid.org/0000-0002-3571-1561>) (Research Fellow) was a member of the project team, drafted the gap analysis, co-facilitated the expert workshop in London in 2018, analysed and summarised findings from all stages. She contributed to SAG and project team discussions, drafted and made final edits to much of the monograph.

Lynsay Matthews (<https://orcid.org/0000-0002-5760-1080>) (Research Fellow) was a member of the project team, drafted the gap analysis, co-facilitated the expert workshop in London in 2018 and analysed and summarised findings from all stages. She contributed to SAG and project team discussions and to the drafting of the monograph.

Sharon Anne Simpson (<https://orcid.org/0000-0002-6219-1768>) (Professor of Behavioural Sciences and Health) was a member of the project team and co-facilitated the expert workshop in London in 2018. She provided substantial intellectual input, contributed to SAG and project team discussions and to the drafting and editing of the monograph.

ACKNOWLEDGEMENTS

Peter Craig (<https://orcid.org/0000-0002-7653-5832>) (Professor of Public Health Evaluation) was a member of the project team and co-facilitated the expert workshop in London in 2018. He provided substantial intellectual input, contributed to SAG and project team discussions and to the drafting and editing of the monograph.

Janis Baird (<https://orcid.org/0000-0002-4039-4361>) (Professor of Public Health and Epidemiology, Honorary Consultant in Public Health) contributed to SAG discussions, attended the expert workshop in London (2018) and provided feedback and edits on drafts of the framework, with particular attention to process evaluation, and provided case studies.

Jane M Blazeby (<https://orcid.org/0000-0002-3354-3330>) (Professor of Surgery) contributed to SAG discussions, attended the expert workshop in London (2018), provided feedback and edits on drafts of the framework, with particular attention to clinical research relevance, and provided case studies.

Kathleen Anne Boyd (<https://orcid.org/0000-0002-9764-0113>) (Doctor, Senior Lecturer in Health Economics and Health Technology Assessment) attended the expert workshop in London (2018) and drafted the sections on economic considerations in the framework.

Neil Craig (<https://orcid.org/0000-0002-7370-5541>) (Honorary Professor, Acting Head of the Evaluation Team within Public Health Sciences, Public Health Scotland) attended the expert workshop in London (2018) and drafted the sections on economic considerations in the framework.

David P French (<https://orcid.org/0000-0002-7663-7804>) (Professor of Health Psychology, Chairperson in Health Psychology) contributed to SAG discussions, attended the expert workshop in London (2018) and provided feedback and edits on drafts of the framework. He provided some text for the 'implementation' section of the framework.

Emma McIntosh (<https://orcid.org/0000-0001-6340-3083>) (Professor of Health Economics) drafted the sections on economic considerations.

Mark Petticrew (<https://orcid.org/0000-0002-6378-5517>) (Professor of Public Health Evaluation) contributed to SAG discussions, attended the expert workshop in London (2018), and provided feedback and edits on drafts of the framework, with particular attention to the 'systems perspective' and 'evidence synthesis'. He also contributed to a smaller group meeting on systems thinking for complex interventions in London (2019).

Jo Rycroft-Malone (<https://orcid.org/0000-0003-3858-5625>) (Professor, Dean of the Faculty of Health and Medicine) contributed to SAG discussions, attended the expert workshop in London (2018), provided feedback and edits on drafts of the framework, with particular attention to 'theory-based approaches' and implementation science, and provided some case studies.

Martin White (<https://orcid.org/0000-0002-1861-6757>) (Professor of Population Health Research) chaired the SAG discussions and provided considerable feedback and edits on multiple drafts of the framework, as well as provided case studies.

Laurence Moore (<https://orcid.org/0000-0003-2182-823X>) (Professor of Public Health Improvement and Director of the MRC/CSO Social and Public Health Sciences Unit) was the principal investigator, with overall responsibility for the work. He provided substantial intellectual input, co-facilitated the expert workshop in London in 2018, contributed to SAG and project team discussions and to the drafting and editing of the monograph. He made final revisions on the guidance.

All co-authors reviewed and approved the final version of the document.

Publication

Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby J, *et al.* A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ* 2021;**374**:n2061.

Data-sharing statement

Owing to this project being methodological in nature there is no data that can be shared. For more information please contact the corresponding author.

References

1. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. *Developing and Evaluating Complex Interventions*. London: Medical Research Council; 2006.
2. Medical Research Council (MRC). *A Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health*. London: MRC; 2000.
3. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;**321**:694–6. <https://doi.org/10.1136/bmj.321.7262.694>
4. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;**337**:a1655. <https://doi.org/10.1136/bmj.a1655>
5. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. *Process Evaluation of Complex Interventions. Medical Research Council Guidance*. London: Medical Research Council Population Health Sciences Research Network; 2014.
6. Raine R, Fitzpatrick R, Barratt H, Bevan G, Black N, Boaden R, et al. Challenges, solutions and future directions in the evaluation of service innovations in health care and public health. *Health Serv Deliv Res* 2016;**4**(16). <https://doi.org/10.3310/hsdr04160>
7. Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, et al. The need for a complex systems model of evidence for public health. *Lancet* 2017;**390**:2602–4. [https://doi.org/10.1016/S0140-6736\(17\)31267-9](https://doi.org/10.1016/S0140-6736(17)31267-9)
8. Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Med* 2018;**16**:95. <https://doi.org/10.1186/s12916-018-1089-4>
9. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. *Using Natural Experiments to Evaluate Population Health Interventions: Guidance for Producers and Users of Evidence*. London: Medical Research Council; 2012.
10. Craig P, Di Ruggiero E, Frolich KL, Mykhalovskiy E, White M, on behalf of the Canadian Institutes of Health Research (CIHR)–National Institute for Health Research (NIHR) Context Guidance Authors Group. *Taking Account of Context in Population Health Intervention Research: Guidance for Producers, Users and Funders of Research*. Southampton: NIHR Evaluation, Trials and Studies Coordinating Centre; 2018. <https://doi.org/10.3310/CIHR-NIHR-01>
11. O’Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. *BMJ Open* 2019;**9**:e029954. <https://doi.org/10.1136/bmjopen-2019-029954>
12. University of Glasgow. *Guidance of Explanatory Studies of Complex Public Health Interventions (The GUEST Study)*. URL: www.gla.ac.uk/researchinstitutes/healthwellbeing/research/mrccsocialandpublichealthsciencesunit/programmes/policy/evaluatinghealthypublicpolicies/guest/ (accessed 1 March 2021).
13. Egan M, McGill E, Anderson de Cuevas R, Er V, Lock K, Popay J, et al. *NIHR SPHR Guidance on Systems Approaches to Local Public Health Evaluation. Part 1: Introducing Systems Thinking*. London: National Institute for Health Research School for Public Health Research; 2019.

14. Egan M, McGill E, Penney T, Anderson de Cuevas R, Er V, Orton L, *et al.* *NIHR SPHR Guidance on Systems Approaches to Local Public Health Evaluation. Part 2: What to Consider When Planning a Systems Evaluation.* London: National Institute for Health Research School for Public Health Research; 2019.
15. De Silva MJ, Breuer E, Lee L, Asher L, Chowdhary N, Lund C, Patel V. Theory of change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions. *Trials* 2014;**15**:267. <https://doi.org/10.1186/1745-6215-15-267>
16. French SD, Green SE, O'Connor DA, McKenzie JE, Francis JJ, Michie S, *et al.* Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. *Implement Sci* 2012;**7**:38. <https://doi.org/10.1186/1748-5908-7-38>
17. Wight D, Wimbush E, Jepson R, Doi L. Six steps in quality intervention development (6SQuID). *J Epidemiol Community Health* 2016;**70**:520–5. <https://doi.org/10.1136/jech-2015-205952>
18. Hawkins J, Madden K, Fletcher A, Midgley L, Grant A, Cox G, *et al.* Development of a framework for the co-production and prototyping of public health interventions. *BMC Public Health* 2017;**17**:689. <https://doi.org/10.1186/s12889-017-4695-8>
19. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *Am J Prev Med* 2007;**32**(Suppl. 5):112–18.
20. Effective Practice and Organisation of Care. *Effective Practice and Organisation of Care (EPOC) Taxonomy.* 2015. URL: <https://epoc.cochrane.org/epoc-taxonomy> (accessed 6 June 2019).
21. Campbell M, Moore G, Evans RE, Khodyakov D, Craig P on behalf of the ADAPT Study team. ADAPT study: adaptation of evidence-informed complex population health interventions for implementation and/or re-evaluation in new contexts: protocol for a Delphi consensus exercise to develop guidance. *BMJ Open* 2020;**10**:e038965. <https://doi.org/10.1136/bmjopen-2020-038965>
22. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, *et al.* How to increase value and reduce waste when research priorities are set. *Lancet* 2014;**383**:156–65. [https://doi.org/10.1016/S0140-6736\(13\)62229-1](https://doi.org/10.1016/S0140-6736(13)62229-1)
23. HM Treasury. *The Green Book. Central Government Guidance on Appraisal and Evaluation.* London: HM Treasury; 2018.
24. Hawe P, Shiell A, Riley T. Theorising interventions as events in systems. *Am J Community Psychol* 2009;**43**:267–76. <https://doi.org/10.1007/s10464-009-9229-9>
25. Petticrew M. When are complex interventions 'complex'? When are simple interventions 'simple'? *Europ J Public Health* 2011;**21**:397–8. <https://doi.org/10.1093/eurpub/ckr084>
26. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby J, *et al.* A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ* 2021;**374**:n2061.
27. Mercer S, Wyke S, Fitzpatrick B, McConnachie A, O'Donnell K, Mackenzie M, *et al.* *Evaluation of the Glasgow 'Deep End' Links Worker Programme.* Edinburgh: NHS Health Scotland; 2017.
28. Charles CV, Dewey CE, Hall A, Hak C, Channary S, Summerlee A. A randomized control trial using a fish-shaped iron ingot for the amelioration of iron deficiency anemia in rural Cambodian women. *Trop Med Surg* 2015;**3**:3. <https://doi.org/10.4172/2329-9088.1000195>
29. Charles CV, Dewey CE, Daniell WE, Summerlee AJ. Iron-deficiency anaemia in rural Cambodia: community trial of a novel iron supplementation technique. *Eur J Public Health* 2011;**21**:43–8. <https://doi.org/10.1093/eurpub/ckp237>

30. Press Release: Iron Fish Update on Clinical Trial in Preah Vihear, Cambodia. Press release. URL: <https://luckyironfish.com/blogs/news/press-release-lucky-iron-fish-update-on-clinical-trial-in-preah-vihear-cambodia> (accessed 2 March 2021).
31. Meadows DH, Wright D. *Thinking in Systems: A Primer*. White River Junction, VT: Chelsea Green Publishing; 2008.
32. The Health Foundation. *Evidence Scan: Complex Adaptive Systems*. London: The Health Foundation; 2010.
33. Wiggins M, Bonell C, Sawtell M, Austerberry H, Burchett H, Allen E, Strange V. Health outcomes of youth development programme in England: prospective matched comparison study. *BMJ* 2009;**339**:b2534. <https://doi.org/10.1136/bmj.b2534>
34. Robinson M, Geue C, Lewsey J, Mackay D, McCartney G, Curnock E, Beeston C. Evaluating the impact of the alcohol act on off-trade alcohol sales: a natural experiment in Scotland. *Addiction* 2014;**109**:2035–43. <https://doi.org/10.1111/add.12701>
35. Plsek PE, Greenhalgh T. Complexity science: the challenge of complexity in health care. *BMJ* 2001;**323**:625–8. <https://doi.org/10.1136/bmj.323.7313.625>
36. HM Treasury. *Magenta Book 2020. Supplementary Guide: Handling Complexity in Policy Evaluation*. London: HM Treasury; 2020.
37. Centre for Complexity Across the Nexus (CECAN). *Complexity Evaluation Framework. Recognising Complexity & Key Considerations for Complexity-Appropriate Evaluation in the Department for Environment, Food and Rural Affairs (DEFRA)*. London: CECAN; 2019.
38. Zwarenstein M. 'Pragmatic' and 'explanatory' attitudes to randomised trials. *J R Soc Med* 2017;**110**:208–18. <https://doi.org/10.1177/0141076817706303>
39. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;**350**:h2147. <https://doi.org/10.1136/bmj.h2147>
40. Hawe P, Shiell A, Riley T. Complex interventions: how 'out of control' can a randomised controlled trial be? *BMJ* 2004;**328**:1561–3. <https://doi.org/10.1136/bmj.328.7455.1561>
41. Bonell C, Allen E, Warren E, McGowan J, Bevilacqua L, Jamal F, *et al*. Effects of the Learning Together intervention on bullying and aggression in English secondary schools (INCLUSIVE): a cluster randomised controlled trial. *Lancet* 2018;**392**:2452–64. [https://doi.org/10.1016/S0140-6736\(18\)31782-3](https://doi.org/10.1016/S0140-6736(18)31782-3)
42. Astbury B, Leeuw FL. Unpacking black boxes: mechanisms and theory building in evaluation. *Am J Eval* 2010;**31**:363–81. <https://doi.org/10.1177/1098214010371972>
43. Borek AJ, Abraham C, Greaves CJ, Gillison F, Tarrant M, Morgan-Trimmer S, *et al*. Identifying change processes in group-based health behaviour-change interventions: development of the mechanisms of action in group-based interventions (MAGI) framework. *Health Psychol Rev* 2019;**13**:227–247. <https://doi.org/10.1080/17437199.2019.1625282>
44. Blamey A, Mackenzie M. Theories of change and realistic evaluation: peas in a pod or apples and oranges? *Evaluation* 2007;**13**:439–55. <https://doi.org/10.1177/1356389007082129>
45. Guba EG, Lincoln YS. *Fourth Generation Evaluation*. London: SAGE Publications Ltd; 1989.
46. Connell J, Kubisch A. Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects and Problems. In Fulbright-Anderson K, Kubisch AC, Connell JP, editors. *New Approaches to Evaluating Community Initiatives, Vol. 2, Theory, Measurement, and Analysis*. Washington, DC: Aspen Institute; 1998.

REFERENCES

47. Connell JP, Kubisch AC, Schorr LB, Weiss CH. *New Approaches to Evaluating Community Initiatives*. Washington, DC: Aspen Institute; 1995.
48. Pawson R, Tilley N. *Realistic Evaluation*. London: SAGE Publications Ltd; 1997.
49. Gallagher R, Appenzeller T. Beyond reductionism. *Science* 1999;**284**:79. <https://doi.org/10.1126/science.284.5411.79>
50. Williams B. Prosaic or profound? The adoption of systems ideas by impact evaluation. *IDS Bulletin* 2015;**46**:7–16. <https://doi.org/10.1111/1759-5436.12117>
51. Sutherland G, Stapleton JA, Russell MA, Jarvis MJ, Hajek P, Belcher M, Feyerabend C. Randomised controlled trial of nasal nicotine spray in smoking cessation. *Lancet* 1992;**340**:324–9. [https://doi.org/10.1016/0140-6736\(92\)91403-U](https://doi.org/10.1016/0140-6736(92)91403-U)
52. Nutbeam D, Macaskill P, Smith C, Simpson JM, Catford J. Evaluation of two school smoking education programmes under normal classroom conditions. *BMJ* 1993;**306**:102–7. <https://doi.org/10.1136/bmj.306.6870.102>
53. Douglas FC, Gray DA, van Teijlingen ER. Using a realist approach to evaluate smoking cessation interventions targeting pregnant women and young people. *BMC Health Serv Res* 2010;**10**:49. <https://doi.org/10.1186/1472-6963-10-49>
54. Graham H. Smoking in pregnancy: the attitudes of expectant mothers. *Soc Sci Med* 1976;**10**:399–405. [https://doi.org/10.1016/0037-7856\(76\)90097-4](https://doi.org/10.1016/0037-7856(76)90097-4)
55. Haw SJ, Gruer L, Amos A, Currie C, Fischbacher C, Fong GT, *et al*. Legislation on smoking in enclosed public places in Scotland: how will we evaluate the impact? *J Public Health* 2006;**28**:24–30. <https://doi.org/10.1093/pubmed/fdi080>
56. Haw SJ, Gruer L. Changes in exposure of adult non-smokers to secondhand smoke after implementation of smoke-free legislation in Scotland: national cross sectional survey. *BMJ* 2007;**335**:549. <https://doi.org/10.1136/bmj.39315.670208.47>
57. Akhtar PC, Currie DB, Currie CE, Haw SJ. Changes in child exposure to environmental tobacco smoke (CHETS) study after implementation of smoke-free legislation in Scotland: national cross sectional survey. *BMJ* 2007;**335**:545. <https://doi.org/10.1136/bmj.39311.550197.AE>
58. Pell JP, Haw S, Cobbe S, Newby DE, Pell AC, Fischbacher C, *et al*. Smoke-free legislation and hospitalizations for acute coronary syndrome. *N Engl J Med* 2008;**359**:482–91. <https://doi.org/10.1056/NEJMsa0706740>
59. Mackay DF, Haw S, Pell JP. Impact of Scottish smoke-free legislation on smoking quit attempts and prevalence. *PLOS ONE* 2011;**6**:e26188. <https://doi.org/10.1371/journal.pone.0026188>
60. Eadie DR, MacAskill SG, Heim D, Hastings GB. Responding to change: how did bar workers adapt to the smoke-free legislation in Scotland? *Int J Environ Health Res* 2010;**20**:13–26. <https://doi.org/10.1080/09603120903255279>
61. Adda J, Berlinski S, Machin S. Short-run economic effects of the Scottish smoking ban. *Int J Epidemiol* 2007;**36**:149–54. <https://doi.org/10.1093/ije/dyl258>
62. Moore GF, Evans RE, Hawkins J, Littlecott H, Melendez-Torres GJ, Bonell C, Murphy S. From complex social interventions to interventions in complex social systems: future directions and unresolved questions for intervention development and evaluation. *Evaluation* 2019;**25**:23–45. <https://doi.org/10.1177/1356389018803219>

63. Torjesen I. Drug development: the journey of a medicine from lab to shelf. *Pharma J* May 2015. <https://doi.org/10.1211/PJ.2015.20068196> URL: <https://pharmaceutical-journal.com/article/feature/drug-development-the-journey-of-a-medicine-from-lab-to-shelf> (accessed 2 March 2021).
64. Pfadenhauer LM, Mozygemba K, Gerhardus A, Hofmann B, Booth A, Lysdahl KB, *et al.* Context and implementation: a concept analysis towards conceptual maturity. *Z Evid Fortbild Qual Gesundheitswes* 2015;**109**:103–14. <https://doi.org/10.1016/j.zefq.2015.01.004>
65. Olds DL, Kitzman HJ, Cole RE, Hanks CA, Arcolego KJ, Anson EA, *et al.* Enduring effects of prenatal and infancy home visiting by nurses on maternal life course and government spending: follow-up of a randomized trial among children at age 12 years. *Arch Pediatr Adolesc Med* 2010;**164**:419–24. <https://doi.org/10.1001/archpediatrics.2010.49>
66. Eckenrode J, Campa M, Luckey DW, Henderson CR, Cole R, Kitzman H, *et al.* Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized trial. *Arch Pediatr Adolesc Med* 2010;**164**:9–15. <https://doi.org/10.1001/archpediatrics.2009.240>
67. Mejdoubi J, van den Heijkant SC, van Leerdam FJ, Heymans MW, Crijnen A, Hirasings RA. The effect of VoorZorg, the Dutch nurse-family partnership, on child maltreatment and development: a randomized controlled trial. *PLOS ONE* 2015;**10**:e0120182. <https://doi.org/10.1371/journal.pone.0120182>
68. Robling M, Bekkers MJ, Bell K, Butler CC, Cannings-John R, Channon S, *et al.* Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. *Lancet* 2016;**387**:146–55. [https://doi.org/10.1016/S0140-6736\(15\)00392-X](https://doi.org/10.1016/S0140-6736(15)00392-X)
69. Olds D. Building evidence to improve maternal and child health. *Lancet* 2016;**387**:105–7. [https://doi.org/10.1016/S0140-6736\(15\)00476-6](https://doi.org/10.1016/S0140-6736(15)00476-6)
70. Pfadenhauer LM, Gerhardus A, Mozygemba K, Lysdahl KB, Booth A, Hofmann B, *et al.* Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. *Implement Sci* 2017;**12**:21. <https://doi.org/10.1186/s13012-017-0552-5>
71. Greenhalgh T, Pawson R, Wong G, Westhorp G, Greenhalgh J, Manzano A, *et al.* *What Realists Mean by Context; or Why Nothing Works Everywhere or for Everyone*. Southampton: National Institute for Health Research; 2017.
72. Funnell SC, Rogers PJ. *Purposeful Program Theory. Effective Use of Theories of Change and Logic Models*. San Francisco, CA: Jossey-Bass; 2011.
73. Pluye P, Potvin L, Denis JL, Pelletier J, Mannoni C. Program sustainability begins with the first events. *Eval Progr Plan* 2005;**28**:123–37. <https://doi.org/10.1016/j.evalprogplan.2004.10.003>
74. Lawless A, Baum F, Delany-Crowe T, MacDougall C, Williams C, McDermott D, Eyk HV. Developing a framework for a program theory-based approach to evaluating policy processes and outcomes: health in all policies in South Australia. *Int J Health Policy Manag* 2018;**7**:510–21. <https://doi.org/10.15171/ijhpm.2017.121>
75. Craig P, Campbell M. *Evaluability Assessment: A Systematic Approach to Deciding Whether and How to Evaluate Programmes and Policies. Working Paper*. Edinburgh: What Works Scotland; 2015.
76. Bonell C, Jamal F, Melendez-Torres GJ, Cummins S. 'Dark logic': theorising the harmful consequences of public health interventions. *J Epidemiol Community Health* 2015;**69**:95–8. <https://doi.org/10.1136/jech-2014-204671>

77. Shearn K, Allmark P, Piercy H, Hirst J. Building realist program theory for large complex and messy interventions. *Int J Qual Methods* 2017;**16**:1609406917741796. <https://doi.org/10.1177/1609406917741796>
78. Andersen DF, Vennix JAM, Richardson GP, Rouwette EAJA. Group model building: problem structuring, policy simulation and decision support. *J Operation Res Soc* 2007;**58**:691–4. <https://doi.org/10.1057/palgrave.jors.2602339>
79. Checkland P. Soft systems methodology: a thirty year retrospective. *Syst Res Behav Sci* 2000;**17**:S11–58. [https://doi.org/10.1002/1099-1743\(200011\)17:1+<::AID-SRES374>3.0.CO;2-O](https://doi.org/10.1002/1099-1743(200011)17:1+<::AID-SRES374>3.0.CO;2-O)
80. Siokou C, Morgan R, Shiell A. Group model building: a participatory approach to understanding and acting on systems. *Public Health Res Pract* 2014;**25**:e2511404. <https://doi.org/10.17061/phrp2511404>
81. McGlashan J, Johnstone M, Creighton D, de la Haye K, Allender S. Quantifying a systems map: network analysis of a childhood obesity causal loop diagram. *PLOS ONE* 2016;**11**:e0165459. <https://doi.org/10.1371/journal.pone.0165459>
82. Dalkin S, Lhussier M, Williams L, Burton CR, Rycroft-Malone J. Exploring the use of Soft Systems Methodology with realist approaches: a novel way to map programme complexity and develop and refine programme theory. *Evaluation* 2018;**24**:84–97. <https://doi.org/10.1177/1356389017749036>
83. Davidoff F, Dixon-Woods M, Leviton L, Michie S. Demystifying theory and its use in improvement. *BMJ Qual Saf* 2015;**24**:228–38. <https://doi.org/10.1136/bmjqs-2014-003627>
84. Bartholomew Eldredge L, Markham C, Ruiter R, Fernandez M, Kok G, Parcel G. *Planning Health Promotion Programs: An Intervention Mapping Approach*. 4th Edition. San Francisco, CA: Jossey-Bass; 2016.
85. Mills T, Lawton R, Sheard L. Advancing complexity science in healthcare research: the logic of logic models. *BMC Med Res Methodol* 2019;**19**:55. <https://doi.org/10.1186/s12874-019-0701-4>
86. Moore GF, Evans RE. What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation of complex population health interventions. *SSM Popul Health* 2017;**3**:132–5. <https://doi.org/10.1016/j.ssmph.2016.12.005>
87. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. *J Adv Nurs* 2013;**69**:1005–22. <https://doi.org/10.1111/jan.12095>
88. Blazeby JM, Byrne J, Welbourn R. What is the most effective operation for adults with severe and complex obesity? *BMJ* 2014;**348**:g1763. <https://doi.org/10.1136/bmj.g1763>
89. Petticrew M, Knai C, Thomas J, Rehfuss EA, Noyes J, Gerhardus A, et al. Implications of a complexity perspective for systematic reviews and guideline development in health decision making. *BMJ Global Health* 2019;**4**:e000899. <https://doi.org/10.1136/bmjgh-2018-000899>
90. Gottfredson DC, Cook TD, Gardner FE, Gorman-Smith D, Howe GW, Sandler IN, Zafft KM. Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: next generation. *Prev Sci* 2015;**16**:893–926. <https://doi.org/10.1007/s11121-015-0555-x>
91. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;**326**:1167–70. <https://doi.org/10.1136/bmj.326.7400.1167>
92. Moodie R, Stuckler D, Monteiro C, Sheron N, Neal B, Thamarangsi T, et al. Profits and pandemics: prevention of harmful effects of tobacco, alcohol, and ultra-processed food and drink industries. *Lancet* 2013;**381**:670–9. [https://doi.org/10.1016/S0140-6736\(12\)62089-3](https://doi.org/10.1016/S0140-6736(12)62089-3)

93. Petticrew M, Katikireddi SV, Knai C, Cassidy R, Maani Hessari N, Thomas J, Weishaar H. 'Nothing can be done until everything is done': the use of complexity arguments by food, beverage, alcohol and gambling industries. *J Epidemiol Community Health* 2017;**71**:1078–83. <https://doi.org/10.1136/jech-2017-209710>
94. Greenhalgh T, Ozbilgin MF, Prainsack B, Shaw S. Moral entrepreneurship, the power-knowledge nexus, and the Cochrane 'crisis'. *J Eval Clin Pract* 2019;**25**:717–25. <https://doi.org/10.1111/jep.13124>
95. INVOLVE. *What is Public Involvement in Research?* 2015. URL: www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/ (accessed 15 October 2019).
96. Crocker JC, Ricci-Cabello I, Parker A, Hirst JA, Chant A, Petit-Zeman S, et al. Impact of patient and public involvement on enrolment and retention in clinical trials: systematic review and meta-analysis. *BMJ* 2018;**363**:k4738. <https://doi.org/10.1136/bmj.k4738>
97. Karazivan P, Dumez V, Flora L, Pomey MP, Del Grande C, Ghadiri DP, et al. The patient-as-partner approach in health care: a conceptual framework for a necessary transition. *Acad Med* 2015;**90**:437–41. <https://doi.org/10.1097/ACM.0000000000000603>
98. Wicks P, Richards T, Denegri S, Godlee F. Patients' roles and rights in research. *BMJ* 2018;**362**:k3193. <https://doi.org/10.1136/bmj.k3193>
99. Meyrick J, Gray D. Evidence-based patient/public voice: a patient and public involvement audit in the field of sexual health. *BMJ Sex Reprod Health* 2018;**44**:267–71. <https://doi.org/10.1136/bmj.srh-2018-200082>
100. Boivin A, Richards T, Forsythe L, Grégoire A, L'Espérance A, Abelson J, Carman KL. Evaluating patient and public involvement in research. *BMJ* 2018;**363**:k5147. <https://doi.org/10.1136/bmj.k5147>
101. Liabo K, Boddy K, Bortoli S, Irvine J, Boulton H, Fredlund M, et al. Public involvement in health research: what does 'good' look like in practice? *Res Involv Engagem* 2020;**6**:11. <https://doi.org/10.1186/s40900-020-0183-x>
102. INVOLVE. *Public Involvement in Research: Values and Principles Framework*. Eastleigh: INVOLVE; 2015.
103. National Institute for Health Research (NIHR). *UK Standards for Public Involvement*. Southampton: NIHR; 2019. URL: <https://sites.google.com/nihr.ac.uk/pi-standards/home> (accessed 22 May 2019).
104. National Institute for Health Research (NIHR). *Patient and Public Involvement in Health and Social Care Research: A Handbook for Researchers*. Southampton: NIHR; 2014.
105. Greenhalgh T, Hinton L, Finlay T, Macfarlane A, Fahy N, Clyde B, et al. Frameworks for supporting patient and public involvement in research: systematic review and co-design pilot. *Health Expect* 2019;**22**:785–801. <https://doi.org/10.1111/hex.12888>
106. Guise JM, O'Haire C, McPheeters M, Most C, Labrant L, Lee K, et al. A practice-based tool for engaging stakeholders in future research: a synthesis of current practices. *J Clin Epidemiol* 2013;**66**:666–74. <https://doi.org/10.1016/j.jclinepi.2012.12.010>
107. Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. *BMJ* 2017;**358**:j3453. <https://doi.org/10.1136/bmj.j3453>
108. Pollock A, Campbell P, Struthers C, Synnot A, Nunn J, Hill S, et al. Development of the ACTIVE framework to describe stakeholder involvement in systematic reviews. *J Health Serv Res Policy* 2019;**24**:245–55. <https://doi.org/10.1177/1355819619841647>

REFERENCES

109. Medical Research Council. *Industry Charter*. 2017. URL: <https://mrc.ukri.org/documents/pdf/mrc-industry-charter/> (accessed 29 August 2019).
110. National Institute for Health Research. *Model Industry Collaborative Research Agreement (mICRA Guidance)*. 2011. URL: www.nihr.ac.uk/documents/micra-guidance-2011/12161 (accessed 29 August 2019).
111. Medical Research Council. *MRC Industry Collaboration Agreement (MICA)*. 2015. URL: <https://mrc.ukri.org/innovation/mrc-industry-collaboration-agreement-mica/> (accessed 2 March 2021).
112. UK Prevention Research Partnership (UKPRP). *Researchers Collaborating with Industry - Guidance on Policy and Procedures*. Version 1.0 (June 2018). URL: <https://ukprp.org/wp-content/uploads/2019/08/UKPRP-Industry-Guidance-Report-v1.pdf> (accessed 2 March 2021).
113. Intellectual Property Office. *University and Business Collaboration Agreements: Lambert Toolkit*. 2016. URL: www.gov.uk/guidance/university-and-business-collaboration-agreements-lambert-toolkit (accessed 29 August 2019).
114. Williams B, Hummelbrunner R. *Systems Concepts in Action: A Practitioner's Toolkit*. Redwood City, CA: Stanford Business Books; 2010. <https://doi.org/10.1515/9780804776554>
115. Checkland P, Scholes J. *Soft Systems Methodology in Action*. Hoboken, NJ: John Wiley & Sons Inc; 1999.
116. Natural Sciences. *Methods and Tools for Co-Producing Knowledge*. Bern: Swiss Academy of Sciences. URL: https://naturalsciences.ch/topics/co-producing_knowledge/methods (accessed 29 August 2019).
117. Network for Transdisciplinary Research. *Methods for Coproducing Knowledge*. URL: <http://transdisciplinarity.ch/en/td-net/Methoden.html> (accessed 1 March 2021).
118. Claxton K, Cohen JT, Neumann PJ. When is evidence sufficient? *Health Affairs* 2005;**24**:93–101. <https://doi.org/10.1377/hlthaff.24.1.93>
119. Kessler R, Glasgow RE. A proposal to speed translation of healthcare research into practice: dramatic change is needed. *Am J Prev Med* 2011;**40**:637–44. <https://doi.org/10.1016/j.amepre.2011.02.023>
120. Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press; 2006.
121. Claxton K. Exploring uncertainty in cost-effectiveness analysis. *PharmacoEconomics* 2008;**26**:781–98. <https://doi.org/10.2165/00019053-200826090-00008>
122. Silverman E, Gostoli U, Picascia S, Almagor J, McCann M, Shaw R, et al. *Situating Agent-based Modelling in Population Health Research*. *airXiv*. 2020. <https://arxiv.org/abs/2002.02345>
123. Petticrew M, McKee M, Lock K, Green J, Phillips G. In search of social equipoise. *BMJ* 2013;**347**:f4016. <https://doi.org/10.1136/bmj.f4016>
124. Fischer AJ, Threlfall A, Meah S, Cookson R, Rutter H, Kelly MP. The appraisal of public health interventions: an overview. *J Public Health* 2013;**35**:488–94. <https://doi.org/10.1093/pubmed/ftd076>
125. Rutter H, Wolpert M, Greenhalgh T. Managing uncertainty in the covid-19 era. *BMJ* 2020;**370**:m3349. <https://doi.org/10.1136/bmj.m3349>
126. Martin GP, Hanna E, McCartney M, Dingwall R. Science, society, and policy in the face of uncertainty: reflections on the debate around face coverings for the public during COVID-19. *Critical Public Health* 2020;**5**:501–8. <https://doi.org/10.1080/09581596.2020.1797997>

127. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *J Med Internet Res* 2017;**19**:e232. <https://doi.org/10.2196/jmir.7126>
128. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, *et al.* Evaluating digital health interventions: key questions and approaches. *Am J Prev Med* 2016;**51**:843–51. <https://doi.org/10.1016/j.amepre.2016.06.008>
129. Mohr DC, Cheung K, Schueller SM, Hendricks Brown C, Duan N. Continuous evaluation of evolving behavioral intervention technologies. *Am J Prev Med* 2013;**45**:517–23. <https://doi.org/10.1016/j.amepre.2013.06.006>
130. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, *et al.* Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. *BMC Med* 2010;**8**:63. <https://doi.org/10.1186/1741-7015-8-63>
131. Collins LM, Murphy SA, Bierman KL. A conceptual framework for adaptive preventive interventions. *Prev Sci* 2004;**5**:185–96. <https://doi.org/10.1023/B:PREV.0000037641.26017.00>
132. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. 4th edition. Oxford: Oxford University Press; 2015.
133. Barnett ML, Dopp AR, Klein C, Ettner SL, Powell BJ, Saldana L. Collaborating with health economists to advance implementation science: a qualitative study. *Implement Sci Commun* 2020;**1**:82. <https://doi.org/10.1186/s43058-020-00074-w>
134. National Institute for Health and Care Excellence (NICE). *Developing NICE Guidelines: The Manual Chapter 7: Incorporating Economic Evaluation*. (Updated 15 October 2020.) London: NICE; 2014. URL: www.nice.org.uk/process/pmg20/resources/developing-nice-guidelines-the-manual-pdf-72286708700869 (accessed 1 March 2021).
135. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, *et al.* Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA* 2016;**316**:1093–103. <https://doi.org/10.1001/jama.2016.12195>
136. Jabeen S. Unintended outcomes evaluation approach: a plausible way to evaluate unintended outcomes of social development programmes. *Eval Program Plan* 2018;**68**:262–74. <https://doi.org/10.1016/j.evalprogplan.2017.09.005>
137. Anderson R, Hardwick R, Pearson M, Byng R. Using Realist Approaches to Explain the Costs and Cost-Effectiveness of Programmes. In Emme N, Greenhalgh J, Manzano A, Monaghan M, Dalkin S, editors. *Doing Realist Research*. London: SAGE Publications Ltd; 2018. <https://doi.org/10.4135/9781526451729.n8>
138. Cookson R, Mirelman AJ, Griffin S, Asaria M, Dawkins B, Norheim OF, *et al.* Using cost-effectiveness analysis to address health equity concerns. *Value Health* 2017;**20**:206–12. <https://doi.org/10.1016/j.jval.2016.11.027>
139. Cookson R, Drummond M, Weatherly H. Explicit incorporation of equity considerations into economic evaluation of public health interventions. *Health Econ, Policy Law* 2009;**4**:231–45. <https://doi.org/10.1017/S1744133109004903>
140. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices Task Force-2. *Med Decis* 2012;**32**:678–89. <https://doi.org/10.1177/0272989X12454941>
141. Tudor Edwards R, McIntosh E. *Applied Health Economics for Public Health Practice and Research*. Oxford: Oxford University Press; 2019. <https://doi.org/10.1093/med/9780198737483.001.0001>

REFERENCES

142. Deidda M, Geue C, Kreif N, Dundas R, McIntosh E. A framework for conducting economic evaluations alongside natural experiments. *Soc Sci Med* 2019;**220**:353–61. <https://doi.org/10.1016/j.socscimed.2018.11.032>
143. Ogilvie D, Adams J, Bauman A, Gregg EW, Panter J, Siegel KR, et al. Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. *J Epidemiol Community Health* 2020;**74**:203–8. <https://doi.org/10.1136/jech-2019-213085>
144. Escoffery C, Lebow-Skelley E, Haardoefer R, Boing E, Udelson H, Wood R, et al. A systematic review of adaptations of evidence-based public health interventions globally. *Implement Sci* 2018;**13**:125. <https://doi.org/10.1186/s13012-018-0815-9>
145. Stirman SW, Miller CJ, Toder K, Calloway A. Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implement Sci* 2013;**8**:65. <https://doi.org/10.1186/1748-5908-8-65>
146. Campbell R, Starkey F, Holliday J, Audrey S, Bloor M, Parry-Langdon N, et al. An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial. *Lancet* 2008;**371**:1595–602. [https://doi.org/10.1016/S0140-6736\(08\)60692-3](https://doi.org/10.1016/S0140-6736(08)60692-3)
147. Forsyth R, Purcell C, Barry S, Simpson S, Hunter R, McDaid L, et al. Peer-led intervention to prevent and reduce STI transmission and improve sexual health in secondary schools (STASH): protocol for a feasibility study. *Pilot Feasibility Stud* 2018;**4**:180. <https://doi.org/10.1186/s40814-018-0354-9>
148. White J, Hawkins J, Madden K, Grant A, Er V, Angel L, et al. Adapting the ASSIST model of informal peer-led intervention delivery to the Talk to FRANK drug prevention programme in UK secondary schools (ASSIST+ FRANK): intervention development, refinement and a pilot cluster randomised controlled trial. *Public Health Res* 2017;**5**(7).
149. Roberto CA, Kawachi I. *Behavioural Economics and Public Health*. Oxford: Oxford University Press; 2016. <https://doi.org/10.1093/med/9780199398331.001.0001>
150. Levy DT, Yuan Z, Li Y, St Claire AW, Schillo BA. The Minnesota SimSmoke tobacco control policy model of smokeless tobacco and cigarette use. *Am J Prev Med* 2019;**57**:e103–e115. <https://doi.org/10.1016/j.amepre.2019.06.003>
151. Li Y, Zhang D, Pagán JA. Social norms and the consumption of fruits and vegetables across New York City neighborhoods. *J Urban Health* 2016;**93**:244–55. <https://doi.org/10.1007/s11524-016-0028-y>
152. Tappenden P. *Conceptual Modelling For Health Economic Model Development*. *Health Economics and Decision Science Discussion Paper 12.05*. Sheffield: University of Sheffield; 2012.
153. Squires H, Chilcott J, Akehurst R, Burr J, Kelly MP. A framework for developing the structure of public health economic models. *Value Health* 2016;**19**:588–601. <https://doi.org/10.1016/j.jval.2016.02.011>
154. Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K, Sutton A. An iterative Bayesian approach to health technology assessment: application to a policy of preoperative optimization for patients undergoing major elective surgery. *Med Decis Making* 2006;**26**:480–96. <https://doi.org/10.1177/0272989X06290493>
155. Johnson R, Woodcock J, de Nazelle A, de Sa T, Goel R, Tainio M, et al. *A Guide to Value of Information Methods for Prioritising Research in Health Impact Modelling*. Ithaca, NY: Cornell University; 2019.

156. O’Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, *et al.* Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. *Pilot Feasibility Stud* 2019;5:41. <https://doi.org/10.1186/s40814-019-0425-6>
157. Duncan E, O’Cathain A, Rousseau N, Croot L, Sworn K, Turner KM, *et al.* Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. *BMJ Open* 2020;10:e033516. <https://doi.org/10.1136/bmjopen-2019-033516>
158. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, *et al.* RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. *Front Public Health* 2019;7:64. <https://doi.org/10.3389/fpubh.2019.00064>
159. Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annu Rev Public Health* 2017;38:39–56. <https://doi.org/10.1146/annurev-publhealth-031816-044327>
160. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). *Health Technol Assess* 2015;19(99). <https://doi.org/10.3310/hta19990>
161. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, *et al.* Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687. <https://doi.org/10.1136/bmj.g1687>
162. Campbell M, Katikireddi SV, Hoffmann T, Armstrong R, Waters E, Craig P. TIDieR-PHP: a reporting guideline for population health and policy interventions. *BMJ* 2018;361:k1079. <https://doi.org/10.1136/bmj.k1079>
163. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, Lancaster GA, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ* 2016;355:i5239. <https://doi.org/10.1136/bmj.i5239>
164. Simpson SA, Matthews L, Pugmire J, McConnachie A, McIntosh E, Coulman E, *et al.* An app-, web- and social support-based weight loss intervention for adults with obesity: the HelpMeDolt! feasibility RCT. *Public Health Res* 2020;8(3). <https://doi.org/10.3310/phr08030>
165. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, *et al.* Process evaluation of complex interventions: Medical Research Council guidance. *BMJ* 2015;350:h1258. <https://doi.org/10.1136/bmj.h1258>
166. Ogilvie D, Cummins S, Petticrew M, White M, Jones A, Wheeler K. Assessing the evaluability of complex public health interventions: five questions for researchers, funders, and policymakers. *Milbank Q* 2011;89:206–25. <https://doi.org/10.1111/j.1468-0009.2011.00626.x>
167. Hallingberg B, Turley R, Segrott J, Wight D, Craig P, Moore L, *et al.* Exploratory studies to decide whether and how to proceed with full-scale evaluations of public health interventions: a systematic review of guidance. *Pilot Feasibility Stud* 2018;4:104. <https://doi.org/10.1186/s40814-018-0290-8>
168. Moore L, Hallingberg B, Wight D, Turley R, Segrott J, Craig P, *et al.* Exploratory studies to inform full-scale evaluations of complex public health interventions: the need for guidance. *J Epidemiol Community Health* 2018;72:865–6. <https://doi.org/10.1136/jech-2017-210414>
169. McIntosh E, Clarke P, Frew E, Louviere JJ. Applied Methods of Cost-Benefit Analysis in Health Care. In Gray A, Briggs A, editors. *Handbooks in Health Economic Evaluation*. Oxford: Oxford University Press; 2010.

REFERENCES

170. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, *et al.* Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMJ* 2013;**346**:f1049. <https://doi.org/10.1136/bmj.f1049>
171. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, *et al.* Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health* 2005;**8**:521–33. <https://doi.org/10.1111/j.1524-4733.2005.00045.x>
172. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 2018;**210**:2–21. <https://doi.org/10.1016/j.socscimed.2017.12.005>
173. UK Research and Innovation (UKRI). *Pathways to Impact*. URL: <https://webarchive.nationalarchives.gov.uk/20200203113953/https://www.ukri.org/innovation/excellence-with-impact/pathways-to-impact> (accessed 1 March 2021).
174. Barnett HJM, Taylor DW, Haynes RB, Sackett DL, Peerless SJ, Ferguson GG, *et al.* Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N Engl J Med* 1991;**325**:445–53. <https://doi.org/10.1056/NEJM199108153250701>
175. Ashburn A, Pickering R, McIntosh E, Hulbert S, Rochester L, Roberts HC, *et al.* Exercise- and strategy-based physiotherapy-delivered intervention for preventing repeat falls in people with Parkinson's: the PDSAFE RCT. *Health Technol Assess* 2019;**23**(36). <https://doi.org/10.3310/hta23360>
176. Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J* 2003;**20**:54–60. <https://doi.org/10.1136/emj.20.1.54>
177. Salisbury C, Man MS, Bower P, Guthrie B, Chaplin K, Gaunt DM, *et al.* Management of multimorbidity using a patient-centred care model: a pragmatic cluster-randomised trial of the 3D approach. *Lancet* 2018;**392**:41–50. [https://doi.org/10.1016/S0140-6736\(18\)31308-4](https://doi.org/10.1016/S0140-6736(18)31308-4)
178. Haugen AS, Søfteland E, Almeland SK, Sevdalis N, Vonen B, Eide GE, *et al.* Effect of the World Health Organization checklist on patient outcomes: a stepped wedge cluster randomized controlled trial. *Ann Surg* 2015;**261**:821–8. <https://doi.org/10.1097/SLA.0000000000000716>
179. Kreif N, Grieve R, Hangartner D, Turner AJ, Nikolova S, Sutton M. Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Econ* 2016;**25**:1514–28. <https://doi.org/10.1002/hec.3258>
180. Allender S, Millar L, Hovmand P, Bell C, Moodie M, Carter R, *et al.* Whole of Systems Trial of Prevention Strategies for childhood obesity: WHO STOPS childhood obesity. *Int J Environ Res Public Health* 2016;**13**:E1143. <https://doi.org/10.3390/ijerph13111143>
181. Randell R, Greenhalgh J, Hindmarsh J, Dowding D, Jayne D, Pearman A, *et al.* Integration of robotic surgery into routine practice and impacts on communication, collaboration, and decision making: a realist process evaluation protocol. *Implement Sci* 2014;**9**:52. <https://doi.org/10.1186/1748-5908-9-52>
182. Randell R, Honey S, Hindmarsh J, Alvarado N, Greenhalgh J, Pearman A, *et al.* A realist process evaluation of robot-assisted surgery: integration into routine practice and impacts on communication, collaboration and decision-making. *Health Serv Deliv Res* 2017;**5**(20). <https://doi.org/10.3310/hsdr05200>
183. Mukumbang FC, Marchal B, Van Belle S, van Wyk B. 'Patients are not following the [adherence] club rules anymore': a realist case study of the antiretroviral treatment adherence club, South Africa. *Qual Health Res* 2018;**28**:1839–57. <https://doi.org/10.1177/1049732318784883>

184. Popay J, Whitehead M, Carr-Hill R, Dibben C, Dixon P, Halliday E, *et al.* The impact on health inequalities of approaches to community engagement in the New Deal for Communities regeneration initiative: a mixed-methods evaluation. *Public Health Res* 2015;**3**(12). <https://doi.org/10.3310/phr03120>
185. Rosas S, Knight E. Evaluating a complex health promotion intervention: case application of three systems methods. *Crit Public Health* 2019;**29**:337–52. <https://doi.org/10.1080/09581596.2018.1455966>
186. McGill E, Marks D, Sumpter C, Egan M. Consequences of removing cheap, super-strength beer and cider: a qualitative study of a UK local alcohol availability intervention. *BMJ Open* 2016;**6**:e010759. <https://doi.org/10.1136/bmjopen-2015-010759>
187. Sumpter C, McGill E, Dickie E, Champo E, Romeri E, Egan M. Reducing the Strength: a mixed methods evaluation of alcohol retailers' willingness to voluntarily reduce the availability of low cost, high strength beers and ciders in two UK local authorities. *BMC Public Health* 2016;**16**:448. <https://doi.org/10.1186/s12889-016-3117-7>
188. Blackman T, Wistow J, Byrne D. A qualitative comparative analysis of factors associated with trends in narrowing health inequalities in England. *Soc Sci Med* 2011;**72**:1965–74. <https://doi.org/10.1016/j.socscimed.2011.04.003>
189. Singal AG, Higgins PD, Waljee AK. A primer on effectiveness and efficacy trials. *Clin Transl Gastroenterol* 2014;**5**:e45. <https://doi.org/10.1038/ctg.2013.13>
190. Mann C, Shaw ARG, Guthrie B, Wye L, Man MS, Chaplin K, Salisbury C. Can implementation failure or intervention failure explain the result of the 3D multimorbidity trial in general practice: mixed-methods process evaluation. *BMJ Open* 2019;**9**:e031438. <https://doi.org/10.1136/bmjopen-2019-031438>
191. Penney T, Adams J, Briggs A, Cummins S, Harrington R, Monsivais P, *et al.* *Evaluation of the Impacts on Health of the Proposed UK Industry Levy on Sugar Sweetened Beverages: Developing a Systems Map and Data Platform, and Collection of Baseline and Early Impact Data.* Southampton: National Institute for Health Research; 2018.
192. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Tugwell P. Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;**13**:132. <https://doi.org/10.1186/1745-6215-13-132>
193. The COMET Initiative. *Core Outcome Measures in Effectiveness Trials*. URL: www.comet-initiative.org (accessed 5 November 2020).
194. Cook JA, Julious SA, Sones W, Hampson LV, Hewitt C, Berlin JA, *et al.* DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *BMJ* 2018;**363**:k3750. <https://doi.org/10.1136/bmj.k3750>
195. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999;**319**:312–15. <https://doi.org/10.1136/bmj.319.7205.312>
196. Creswell JW, Clark VLP. *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: SAGE Publications Ltd; 2017.
197. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. *BMJ* 2010;**341**:c4587. <https://doi.org/10.1136/bmj.c4587>
198. Guetterman TC, Feters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *Ann Fam Med* 2015;**13**:554–61. <https://doi.org/10.1370/afm.1865>

REFERENCES

199. Richards DA, Bazeley P, Borglin G, Craig P, Emsley R, Frost J, *et al.* Integrating quantitative and qualitative data and findings when undertaking randomised controlled trials. *BMJ Open* 2019;**9**:e032081. <https://doi.org/10.1136/bmjopen-2019-032081>
200. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, *et al.* Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. *J Epidemiol Community Health* 2012;**66**:1182–6. <https://doi.org/10.1136/jech-2011-200375>
201. Brand SL, Musgrove A, Jeffcoate WJ, Lincoln NB. Evaluation of the effect of nurse education on patient-reported foot checks and foot care behaviour of people with diabetes receiving haemodialysis. *Diabet Med* 2016;**33**:204–7. <https://doi.org/10.1111/dme.12831>
202. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ* 2015;**350**:h391. <https://doi.org/10.1136/bmj.h391>
203. Thorlund K, Haggstrom J, Park JJ, Mills EJ. Key design considerations for adaptive clinical trials: a primer for clinicians. *BMJ* 2018;**360**:k698. <https://doi.org/10.1136/bmj.k698>
204. Kairalla JA, Coffey CS, Thomann MA, Muller KE. Adaptive trial designs: a review of barriers and opportunities. *Trials* 2012;**13**:145. <https://doi.org/10.1186/1745-6215-13-145>
205. Almirall D, Nahum-Shani I, Sherwood NE, Murphy SA. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Transl Behav Med* 2014;**4**:260–74. <https://doi.org/10.1007/s13142-014-0265-0>
206. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A ‘SMART’ design for building individualized treatment sequences. *Annu Rev Clin Psychol* 2012;**8**:21–48. <https://doi.org/10.1146/annurev-clinpsy-032511-143152>
207. Bärnighausen T, Røttingen JA, Rockers P, Shemilt I, Tugwell P. Quasi-experimental study designs series – paper 1: introduction: two historical lineages. *J Clin Epidemiol* 2017;**89**:4–11. <https://doi.org/10.1016/j.jclinepi.2017.02.020>
208. Basu S, Meghani A, Siddiqi A. Evaluating the health impact of large-scale public policy changes: classical and novel approaches. *Annu Rev Public Health* 2017;**38**:351–70. <https://doi.org/10.1146/annurev-publhealth-031816-044208>
209. Chattopadhyay R, Duflo E. Women as policy makers: evidence from a randomized policy experiment in India. *Econometrica* 2004;**72**:1409–43. <https://doi.org/10.1111/j.1468-0262.2004.00539.x>
210. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol* 2017;**46**:348–55. <https://doi.org/10.1093/ije/dyw098>
211. Bor J, Moscoe E, Mutevedzi P, Newell ML, Bärnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology* 2014;**25**:729–37. <https://doi.org/10.1097/EDE.000000000000138>
212. Wing C, Simon K, Bello-Gomez RA. Designing difference in difference studies: best practices for public health policy research. *Annu Rev Public Health* 2018;**39**:453–69. <https://doi.org/10.1146/annurev-publhealth-040617-013507>
213. Miller G, Pinto DM, Vera-Hernández M. *Risk Protection, Service Use, and Health Outcomes Under Colombia’s Health Insurance Program for the Poor*. No. 0898-2937: Cambridge, MA: National Bureau of Economic Research; 2009. <https://doi.org/10.3386/w15456>

214. Hilton Boon M, Craig P, Thomson H, Campbell M, Moore L. Regression discontinuity designs in health: a systematic review. *Epidemiology* 2021;**32**:87–93. <https://doi.org/10.1097/EDE.0000000000001274>
215. Bouttell J, Craig P, Lewsey J, Robinson M, Popham F. Synthetic control methodology as a tool for evaluating population-level health interventions. *J Epidemiol Community Health* 2018;**72**:673–8. <https://doi.org/10.1136/jech-2017-210106>
216. Burges Watson D, Adams J, Azevedo LB, Haighton C. Promoting physical activity with a school-based dance mat exergaming intervention: qualitative findings from a natural experiment. *BMC Public Health* 2016;**16**:609. <https://doi.org/10.1186/s12889-016-3308-2>
217. McCoy C. Do drug courts work? For what, compared to what? Qualitative results from a natural experiment. *Vict Offender* 2009;**5**:64–75. <https://doi.org/10.1080/15564880903423102>
218. Bärnighausen T, Tugwell P, Røttingen JA, Shemilt I, Rockers P, Geldsetzer P, et al. Quasi-experimental study designs series – paper 4: uses and value. *J Clin Epidemiol* 2017;**89**:21–9. <https://doi.org/10.1016/j.jclinepi.2017.03.012>
219. Tugwell P, Knottnerus JA, McGowan J, Tricco A. Big-5 quasi-experimental designs. *J Clin Epidemiol* 2017;**89**:1–3. <https://doi.org/10.1016/j.jclinepi.2017.09.010>
220. Sedgwick P. What is an ‘n-of-1’ trial? *BMJ* 2014;**348**:g2674. <https://doi.org/10.1136/bmj.g2674>
221. McDonald S, Quinn F, Vieira R, O’Brien N, White M, Johnston DW, Sniehotta FF. The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: a systematic literature overview. *Health Psychol Rev* 2017;**11**:307–23. <https://doi.org/10.1080/17437199.2017.1316672>
222. Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. *Med Care* 2011;**49**:761–8. <https://doi.org/10.1097/MLR.0b013e318215d90d>
223. Starkey F, Audrey S, Holliday J, Moore L, Campbell R. Identifying influential young people to undertake effective peer-led health promotion: the example of A Stop Smoking In Schools Trial (ASSIST). *Health Educ Res* 2009;**24**:977–88. <https://doi.org/10.1093/her/cyp045>
224. White M, Cummins S, Rayner M, Smith R, Rutter H, Adams J, et al. *Evaluation of the Health Impacts of the UK Treasury Soft Drinks Industry Levy (SDIL). Project Protocol*. Southampton: National Institute of Health Research; 2018.
225. Bruch E, Atwell J. Agent-based models in empirical social research. *Sociol Methods Res* 2013;**44**:186–221. <https://doi.org/10.1177/0049124113506405>
226. Tracy M, Cerdá M, Keyes KM. Agent-based modeling in public health: current applications and future directions. *Annu Rev Public Health* 2018;**39**:77–94. <https://doi.org/10.1146/annurev-publhealth-040617-014317>
227. Kypridemos C, Collins B, McHale P, Bromley H, Parvulescu P, Capewell S, O’Flaherty M. Future cost-effectiveness and equity of the NHS health check cardiovascular disease prevention programme: microsimulation modelling using data from Liverpool, UK. *PLOS Med* 2018;**15**:e1002573. <https://doi.org/10.1371/journal.pmed.1002573>
228. Kypridemos C, Guzman-Castillo M, Hyseni L, Hickey GL, Bandosz P, Buchan I, et al. Estimated reductions in cardiovascular and gastric cancer disease burden through salt policies in England: an IMPACTNCD microsimulation study. *BMJ Open* 2017;**7**:e013791. <https://doi.org/10.1136/bmjopen-2016-013791>

229. Allen K, Pearson-Stuttard J, Hooton W, Diggle P, Capewell S, O'Flaherty M. Potential of trans fats policies to reduce socioeconomic inequalities in mortality from coronary heart disease in England: cost effectiveness modelling study. *BMJ* 2015;**351**:h4583. <https://doi.org/10.1136/bmj.h4583>
230. Smith RD, Keogh-Brown MR, Barnett T, Tait J. The economy-wide impact of pandemic influenza on the UK: a computable general equilibrium modelling experiment. *BMJ* 2009;**339**:b4571. <https://doi.org/10.1136/bmj.b4571>
231. Holmes J, Meng Y, Meier PS, Brennan A, Angus C, Campbell-Burton A, et al. Effects of minimum unit pricing for alcohol on different income and socioeconomic groups: a modelling study. *Lancet* 2014;**383**:1655–64. [https://doi.org/10.1016/S0140-6736\(13\)62417-4](https://doi.org/10.1016/S0140-6736(13)62417-4)
232. Blackman T, Wistow J, Byrne D. Using qualitative comparative analysis to understand complex policy problems. *Evaluation* 2013;**19**:126–40. <https://doi.org/10.1177/1356389013484203>
233. Byrne D. *Qualitative Comparative Analysis: A Pragmatic Method for Evaluating Intervention*. A CECAN Evaluation and Policy Practice Note for Policy Analysts and Evaluators. Guildford: Centre for the Evaluation of Complexity Across the Nexus; 2016.
234. Rihoux B, Ragin CC. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, CA: SAGE Publications Ltd; 2008. <https://doi.org/10.4135/9781452226569>
235. Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Realist randomised controlled trials: a new approach to evaluating complex public health interventions. *Soc Sci Med* 2012;**75**:2299–306. <https://doi.org/10.1016/j.socscimed.2012.08.032>
236. McGill E, Marks D, Er V, Penney T, Petticrew M, Egan M. Qualitative process evaluation from a complex systems perspective: a systematic review and framework for public health evaluators. *PLOS Med* 2020;**17**:e1003368. <https://doi.org/10.1371/journal.pmed.1003368>
237. Charles J, Jones A, Lloyd-Williams H. Return on Investment, Social Return on Investment and the Business Case for Prevention. In Edwards R, McIntosh E, editors. *An Introduction to the Principles of Health Economics and Economic Evaluation Relevant to the Specific Challenges of Public Health and its Aims of Improving Population Health*. Oxford: Oxford University Press; 2019. <https://doi.org/10.1093/med/9780198737483.003.0012>
238. Marsh K, IJzerman M, Thokala P, Baltussen R, Boysen M, Kaló Z, et al. Multiple criteria decision analysis for health care decision making – emerging good practices: report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health* 2016;**19**:125–37. <https://doi.org/10.1016/j.jval.2015.12.016>
239. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;**8**(36). <https://doi.org/10.3310/hta8360>
240. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J, ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force – 4. *Value Health* 2012;**15**:821–7. <https://doi.org/10.1016/j.jval.2012.04.013>
241. Mathes T, Antoine SL, Prengel P, Bühn S, Polus S, Pieper D. Health technology assessment of public health interventions: a synthesis of methodological guidance. *Int J Technol Assess Health Care* 2017;**33**:135–46. <https://doi.org/10.1017/S0266462317000228>

242. Caro JJ, Briggs AH, Siebert U, Kuntz KM, ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling good research practices – overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. *Med Decis Making* 2012;**32**:667–77. <https://doi.org/10.1177/0272989X12454577>
243. Squires H, Boyd KA. The Use of Modelling Approaches for the Economic Evaluation of Public Health Interventions. In McIntosh E, editor. *An Introduction to the Principles of Health Economics and Economic Evaluation Relevant to the Specific Challenges of Public Health and its Aims of Improving Population Health*. Oxford: Oxford University Press; 2019. <https://doi.org/10.1093/med/9780198737483.003.0011>
244. Montgomery P, Grant S, Mayo-Wilson E, Macdonald G, Michie S, Hopewell S, Moher D, CONSORT-SPI Group. Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 extension. *Trials* 2018;**19**:407. <https://doi.org/10.1186/s13063-018-2733-1>
245. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;**345**:e5661. <https://doi.org/10.1136/bmj.e5661>
246. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, *et al*. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;**337**:a2390. <https://doi.org/10.1136/bmj.a2390>
247. Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, *et al*. CONSORT extension for reporting N-of-1 trials (CENT) 2015 statement. *BMJ* 2015;**350**:h1738. <https://doi.org/10.1136/bmj.h1738>
248. Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, *et al*. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ* 2018;**363**:k1614. <https://doi.org/10.1136/bmj.k1614>
249. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Ann Intern Med* 2017;**167**:40–7. <https://doi.org/10.7326/M17-0046>
250. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA* 2013;**309**:814–22. <https://doi.org/10.1001/jama.2013.879>
251. Dimairo M, Pallmann P, Wason J, Todd S, Jaki T, Julious SA, *et al*. The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ* 2020;**369**:m115. <https://doi.org/10.1136/bmj.m115>
252. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT, *et al*. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *JAMA* 2018;**319**:483–94. <https://doi.org/10.1001/jama.2017.21903>
253. Gamble C, Krishan A, Stocken D, Lewis S, Juszcak E, Doré C, *et al*. Guidelines for the content of statistical analysis plans in clinical trials. *JAMA* 2017;**318**:2337–43. <https://doi.org/10.1001/jama.2017.18556>
254. Grant A, Treweek S, Dreischulte T, Foy R, Guthrie B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. *Trials* 2013;**14**:15. <https://doi.org/10.1186/1745-6215-14-15>

REFERENCES

255. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J, Greenhalgh T. RAMESES II reporting standards for realist evaluations. *BMC Med* 2016;**14**:96. <https://doi.org/10.1186/s12916-016-0643-1>
256. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, *et al.* SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013;**158**:200–7. <https://doi.org/10.7326/0003-4819-158-3-201302050-00583>
257. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med* 2014;**89**:1245–51. <https://doi.org/10.1097/ACM.0000000000000388>
258. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;**147**:573–7. <https://doi.org/10.7326/0003-4819-147-8-200710160-00010>
259. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, *et al.* A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol* 2009;**62**:464–75. <https://doi.org/10.1016/j.jclinepi.2008.12.011>
260. BMC. *Sample Size Calculations and Specifying the Target Difference for Randomized Trials. An Article Collection Published in Trials.* 2018. URL: www.biomedcentral.com/collections/randomizedtrialsamplesize (accessed 2 March 2021).
261. Remme M, Martinez-Alvarez M, Vassall A. Cost-effectiveness thresholds in global health: taking a multisectoral perspective. *Value Health* 2017;**20**:699–704. <https://doi.org/10.1016/j.jval.2016.11.009>
262. Terris-Prestholt F, Neke N, Grund JM, Plotkin M, Kuringe E, Osaki H, *et al.* Using discrete choice experiments to inform the design of complex interventions. *Trials* 2019;**20**:157. <https://doi.org/10.1186/s13063-019-3186-x>
263. University of Birmingham. *ICECAP Capability Measures.* URL: www.birmingham.ac.uk/research/activity/mds/projects/haps/he/icecap/index.aspx (accessed 2 March 2021).
264. Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Qual Life Res* 2012;**21**:167–76. <https://doi.org/10.1007/s11136-011-9927-2>
265. Deidda M, Boyd KA, Minnis H, Donaldson J, Brown K, Boyer NRS, McIntosh E, BeST study team. Protocol for the economic evaluation of a complex intervention to improve the mental health of maltreated infants and children in foster care in the UK (the BeST? services trial). *BMJ Open* 2018;**8**:e020066. <https://doi.org/10.1136/bmjopen-2017-020066>
266. Kreif N, Grieve R, Sadique MZ. Statistical methods for cost-effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. *Health Econ* 2013;**22**:486–500. <https://doi.org/10.1002/hec.2806>
267. World Health Organization. *Scaling Up Projects and Initiatives for Better Health: From Concepts to Practice.* Geneva: World Health Organization; 2016.
268. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, *et al.* Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;**8**(6). <https://doi.org/10.3310/hta8060>
269. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci* 2009;**4**:50. <https://doi.org/10.1186/1748-5908-4-50>

270. Eccles MP, Mittman BS. Welcome to implementation science. *Implement Sci* 2006;**1**:1. <https://doi.org/10.1186/1748-5908-1-1>
271. Theobald S, Brandes N, Gyapong M, El-Saharty S, Proctor E, Diaz T, *et al.* Implementation research: new imperatives and opportunities in global health. *Lancet* 2018;**392**:2214–28. [https://doi.org/10.1016/S0140-6736\(18\)32205-0](https://doi.org/10.1016/S0140-6736(18)32205-0)
272. Huis A, Holleman G, van Achterberg T, Grol R, Schoonhoven L, Hulscher M. Explaining the effects of two different strategies for promoting hand hygiene in hospital nurses: a process evaluation alongside a cluster randomised controlled trial. *Implement Sci* 2013;**8**:41. <https://doi.org/10.1186/1748-5908-8-41>
273. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ, *et al.* Standards for Reporting Implementation studies (StaRI) statement. *BMJ* 2017;**356**:i6795. <https://doi.org/10.1136/bmj.i6795>
274. Aveyard P, Lewis A, Tearne S, Hood K, Christian-Brown A, Adab P, *et al.* Screening and brief intervention for obesity in primary care: a parallel, two-arm, randomised trial. *Lancet* 2016;**388**:2492–500. [https://doi.org/10.1016/S0140-6736\(16\)31893-1](https://doi.org/10.1016/S0140-6736(16)31893-1)
275. Diepeveen S, Ling T, Suhrcke M, Roland M, Marteau TM. Public acceptability of government intervention to change health-related behaviours: a systematic review and narrative synthesis. *BMC Public Health* 2013;**13**:756. <https://doi.org/10.1186/1471-2458-13-756>
276. Presseau J, Sniehotta FF, Francis JJ, Campbell NC. Multiple goals and time constraints: perceived impact on physicians' performance of evidence-based behaviours. *Implement Sci* 2009;**4**:77. <https://doi.org/10.1186/1748-5908-4-77>
277. Petticrew M, Whitehead M, Macintyre SJ, Graham H, Egan M. Evidence for public health policy on inequalities: 1: the reality according to policymakers. *J Epidemiol Community Health* 2004;**58**:811–16. <https://doi.org/10.1136/jech.2003.015289>
278. May C. Mobilising modern facts: health technology assessment and the politics of evidence. *Social Health Illn* 2006;**28**:513–32. <https://doi.org/10.1111/j.1467-9566.2006.00505.x>
279. Pick W. Lack of evidence hampers human-resources policy making. *Lancet* 2008;**371**:629–30. [https://doi.org/10.1016/S0140-6736\(08\)60280-9](https://doi.org/10.1016/S0140-6736(08)60280-9)
280. Kelly MP, Speller V, Meyrick J. *Getting Evidence into Practice in Public Health*. London: NHS Health Development Agency; 2004.
281. Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. *BMJ* 2004;**328**:343–5. <https://doi.org/10.1136/bmj.328.7435.343>
282. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care* 2012;**50**:217–26. <https://doi.org/10.1097/MLR.0b013e3182408812>
283. Landes SJ, McBain SA, Curran GM. An introduction to effectiveness-implementation hybrid designs. *Psychiatry Res* 2019;**280**:112513. <https://doi.org/10.1016/j.psychres.2019.112513>
284. Green BB, Coronado GD, Schwartz M, Coury J, Baldwin LM. Using a continuum of hybrid effectiveness-implementation studies to put research-tested colorectal screening interventions into practice. *Implement Sci* 2019;**14**:53. <https://doi.org/10.1186/s13012-019-0903-5>
285. Villeval M, Bidault E, Shoveller J, Alias F, Basson JC, Frasse C, *et al.* Enabling the transferability of complex interventions: exploring the combination of an intervention's key functions and implementation. *Int J Public Health* 2016;**61**:1031–8. <https://doi.org/10.1007/s00038-016-0809-9>

286. Imison C, Curry N, Holder H, Castle-Clarke S, Nimmons D, Appleby J, *et al.* *Shifting the Balance of Care: Great Expectations*. London: Nuffield Trust; 2017.
287. Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, *et al.* Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychol* 2004;**23**:443–51. <https://doi.org/10.1037/0278-6133.23.5.443>
288. Fagg J, Cole TJ, Cummins S, Goldstein H, Morris S, Radley D, *et al.* After the RCT: who comes to a family-based intervention for childhood overweight or obesity when it is implemented at scale in the community? *J Epidemiol Community Health* 2015;**69**:142–8. <https://doi.org/10.1136/jech-2014-204155>
289. Hull L, Goulding L, Khadjesari Z, Davis R, Healey A, Bakolis I, Sevdalis N. Designing high-quality implementation research: development, application, feasibility and preliminary evaluation of the implementation science research development (ImpRes) tool and guide. *Implement Sci* 2019;**14**:80. <https://doi.org/10.1186/s13012-019-0897-z>
290. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. *BMC Psychol* 2015;**3**:32. <https://doi.org/10.1186/s40359-015-0089-9>
291. Nilsen P. Making sense of implementation theories, models and frameworks. *Implement Sci* 2015;**10**:53. <https://doi.org/10.1186/s13012-015-0242-0>
292. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q* 2004;**82**:581–629. <https://doi.org/10.1111/j.0887-378X.2004.00325.x>
293. Meyers DC, Durlak JA, Wandersman A. The quality implementation framework: a synthesis of critical steps in the implementation process. *Am J Community Psychol* 2012;**50**:462–80. <https://doi.org/10.1007/s10464-012-9522-x>
294. Gray CM, Hunt K, Mutrie N, Anderson AS, Leishman J, Dalgarno L, Wyke S. Football Fans in Training: the development and optimization of an intervention delivered through professional sports clubs to help men lose weight, become more active and adopt healthier eating habits. *BMC Public Health* 2013;**13**:232. <https://doi.org/10.1186/1471-2458-13-232>
295. Gray CM, Hunt K, Mutrie N, Anderson AS, Treweek S, Wyke S. Weight management for overweight and obese men delivered through professional football clubs: a pilot randomized trial. *Int J Behav Nutr Phys Act* 2013;**10**:121. <https://doi.org/10.1186/1479-5868-10-121>
296. Hunt K, Gray CM, Maclean A, Smillie S, Bunn C, Wyke S. Do weight management programmes delivered at professional football clubs attract and engage high risk men? A mixed-methods study. *BMC Public Health* 2014;**14**:50. <https://doi.org/10.1186/1471-2458-14-50>
297. Hunt K, Wyke S, Gray CM, Anderson AS, Brady A, Bunn C, *et al.* A gender-sensitized weight loss and healthy living programme for overweight and obese men delivered by Scottish Premier League football clubs (FFIT): a pragmatic randomised controlled trial. *Lancet* 2014;**383**:1211–21. [https://doi.org/10.1016/S0140-6736\(13\)62420-4](https://doi.org/10.1016/S0140-6736(13)62420-4)
298. Hunt K, Wyke S, Bunn C, Donnachie C, Reid N, Gray CM. Scale-up and scale-out of a gender-sensitized weight management and healthy living program delivered to overweight men via professional sports clubs: the wider implementation of Football Fans in Training (FFIT). *Int J Environ Res Public Health* 2020;**17**:E584. <https://doi.org/10.3390/ijerph17020584>
299. Hoddinott P, Britten J, Pill R. Why do interventions work in some places and not others: a breastfeeding support group trial. *Soc Sci Med* 2010;**70**:769–78. <https://doi.org/10.1016/j.socscimed.2009.10.067>

300. Hoddinott P, Britten J, Prescott GJ, Tappin D, Ludbrook A, Godden DJ. Effectiveness of policy to provide breastfeeding groups (BIG) for pregnant and breastfeeding mothers in primary care: cluster randomised controlled trial. *BMJ* 2009;**338**:a3026. <https://doi.org/10.1136/bmj.a3026>
301. Stephens TJ, Peden CJ, Pearse RM, Shaw SE, Abbott TEF, Jones EL, *et al.* Improving care at scale: process evaluation of a multi-component quality improvement intervention to reduce mortality after emergency abdominal surgery (EPOCH trial). *Implement Sci* 2018;**13**:142. <https://doi.org/10.1186/s13012-018-0823-9>
302. Peden CJ, Stephens T, Martin G, Kahan BC, Thomson A, Rivett K, *et al.* Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery (EPOCH): a stepped-wedge cluster-randomised trial. *Lancet* 2019;**393**:2213–21. [https://doi.org/10.1016/S0140-6736\(18\)32521-2](https://doi.org/10.1016/S0140-6736(18)32521-2)
303. Maini R, Mounier-Jack S, Borghi J. How to and how not to develop a theory of change to evaluate a complex intervention: reflections on an experience in the Democratic Republic of Congo. *BMJ Global Health* 2018;**3**:e000617. <https://doi.org/10.1136/bmjgh-2017-000617>
304. Spero JC, McQuide PA, Matte R. Tracking and monitoring the health workforce: a new human resources information system (HRIS) in Uganda. *Hum Resour Health* 2011;**9**:6. <https://doi.org/10.1186/1478-4491-9-6>
305. Franco LM, Bennett S, Kanfer R. Health sector reform and public sector health worker motivation: a conceptual framework. *Soc Sci Med* 2002;**54**:1255–66. [https://doi.org/10.1016/S0277-9536\(01\)00094-6](https://doi.org/10.1016/S0277-9536(01)00094-6)
306. World Health Organization. *Working Together for Health – The World Health Report 2006*. Geneva: World Health Organization; 2006. URL: www.who.int/whr/2006/en/ (accessed 1 March 2021).
307. Cook PA, Hargreaves SC, Burns EJ, de Vocht F, Parrott S, Coffey M, *et al.* Communities in charge of alcohol (CICA): a protocol for a stepped-wedge randomised control trial of an alcohol health champions programme. *BMC Public Health* 2018;**18**:522. <https://doi.org/10.1186/s12889-018-5410-0>
308. Ebenso B, Manzano A, Uzochukwu B, Etiaba E, Huss R, Ensor T, *et al.* Dealing with context in logic model development: reflections from a realist evaluation of a community health worker programme in Nigeria. *Eval Program Plann* 2019;**73**:97–110. <https://doi.org/10.1016/j.evalprogplan.2018.12.002>
309. Jones S, Riste L, Barrowclough C, Bartlett P, Clements C, Davies L, *et al.* Reducing relapse and suicide in bipolar disorder: practical clinical approaches to identifying risk, reducing harm and engaging service users in planning and delivery of care – the PARADES (Psychoeducation, Anxiety, Relapse, Advance Directive Evaluation and Suicidality) programme. *Programme Grants Appl Res* 2018;**6**(6). <https://doi.org/10.3310/pgfar06060>
310. Great Britain. *Mental Capacity Act 2005*. London: The Stationery Office; 2005.
311. Yardley L, Ainsworth B, Arden-Close E, Muller I. The person-based approach to enhancing the acceptability and feasibility of interventions. *Pilot Feasibility Stud* 2015;**1**:37. <https://doi.org/10.1186/s40814-015-0033-z>
312. Minnis H, Bryce G, Phin L, Wilson P. The ‘Spirit of New Orleans’: translating a model of intervention with maltreated children and their families for the Glasgow context. *Clin Child Psychol Psychiatry* 2010;**15**:497–509. <https://doi.org/10.1177/1359104510376124>
313. Boyd KA, Balogun MO, Minnis H. Development of a radical foster care intervention in Glasgow, Scotland. *Health Promot Int* 2016;**31**:665–73. <https://doi.org/10.1093/heapro/dav041>

314. Pritchett R, Fitzpatrick B, Watson N, Cotmore R, Wilson P, Bryce G, *et al.* A feasibility randomised controlled trial of the New Orleans intervention for infant mental health: a study protocol. *Sci World J* 2013;**2013**:838042. <https://doi.org/10.1155/2013/838042>
315. Turner-Halliday F, Kainth G, Young-Southward G, Cotmore R, Watson N, McMahon L, Minnis H. Clout or doubt? Perspectives on an infant mental health service for young children placed in foster care due to abuse and neglect. *Child Abuse Negl* 2017;**72**:184–95. <https://doi.org/10.1016/j.chiabu.2017.07.012>
316. Turner-Halliday F, Watson N, Minnis H. *Process Evaluation of the New Orleans Intervention Model for Infant Mental Health in Glasgow*. London: National Society for the Prevention of Cruelty to Children; 2016.
317. Minnis H, Wilson P, Boyd K, Messow C, Bryce G, Donaldson J, *et al.* *The BeST? Services Trial: Evaluation of the New Orleans Intervention Model for Infants Mental Health in Glasgow*. Chief Scientist Office, CZH/4/629; 2016. URL: www.cso.scot.nhs.uk/outputs/focus-on-research-summaries/focus-on-research-generic-health-relevance/ (accessed 4 November 2020).
318. Minnis H, Boyd KA, Fitzpatrick B, *et al.* *Protocol 15PRT/6090: The Best Services Trial (BeST?): Effectiveness and Cost-effectiveness of the New Orleans Intervention Model for Infant Mental Health – NCT02653716*. 2016. URL: www.thelancet.com/doi/story/10.1016/html.2016.11.15.4368 (accessed 4 November 2020).
319. Signal LN, Walton MD, Ni Mhurchu C, Maddison R, Bowers SG, Carter KN, *et al.* Tackling ‘wicked’ health promotion problems: a New Zealand case study. *Health Promot Int* 2013;**28**:84–94. <https://doi.org/10.1093/heapro/das006>
320. Blackman T. *Placing Health: Neighbourhood Renewal, Health Improvement and Complexity*. Bristol: Policy Press; 2006. <https://doi.org/10.2307/j.ctt9qgp55>
321. Byrne D. What is complexity science? Thinking as a realist about measurement and cities and arguing for natural history. *Emergence* 2001;**3**:61–76. https://doi.org/10.1207/S15327000EM0301_05
322. Byrne DS. *Complexity Theory and the Social Sciences: An Introduction*. Hove: Psychology Press; 1998.
323. Coffin PO, Santos GM, Behar E, Hern J, Walker J, Matheson T, *et al.* Randomized feasibility trial of directly observed versus unobserved hepatitis C treatment with ledipasvir-sofosbuvir among people who inject drugs. *PLOS ONE* 2019;**14**:e0217471. <https://doi.org/10.1371/journal.pone.0217471>
324. Honeycutt S, Hermstad A, Carvalho ML, Arriola KRJ, Ballard D, Escoffery C, *et al.* Practice to evidence: using evaluability assessment to generate practice-based evidence in rural South Georgia. *Health Educ Behav* 2017;**44**:454–62. <https://doi.org/10.1177/1090198116673360>
325. Mann C, Shaw A, Guthrie B, Wye L, Man MS, Hollinghurst S, *et al.* Protocol for a process evaluation of a cluster randomised controlled trial to improve management of multimorbidity in general practice: the 3D study. *BMJ Open* 2016;**6**:e011260. <https://doi.org/10.1136/bmjopen-2016-011260>
326. Man MS, Chaplin K, Mann C, Bower P, Brookes S, Fitzpatrick B, *et al.* Improving the management of multimorbidity in general practice: protocol of a cluster randomised controlled trial (the 3D study). *BMJ Open* 2016;**6**:e011261. <https://doi.org/10.1136/bmjopen-2016-011261>
327. Thorn J, Man MS, Chaplin K, Bower P, Brookes S, Gaunt D, *et al.* Cost-effectiveness of a patient-centred approach to managing multimorbidity in primary care: a pragmatic cluster randomised controlled trial. *BMJ Open* 2020;**10**:e030110. <https://doi.org/10.1136/bmjopen-2019-030110>

328. Sutton M, Nikolova S, Boaden R, Lester H, McDonald R, Roland M. Reduced mortality with hospital pay for performance in England. *N Engl J Med* 2012;**367**:1821–8. <https://doi.org/10.1056/NEJMsa1114951>
329. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J Am Stat Assoc* 2010;**105**:493–505. <https://doi.org/10.1198/jasa.2009.ap08746>
330. Allender S, Owen B, Kuhlberg J, Lowe J, Nagorcka-Smith P, Whelan J, Bell C. A community based systems diagram of obesity causes. *PLOS ONE* 2015;**10**:e0129683. <https://doi.org/10.1371/journal.pone.0129683>
331. Rose J, Riolo R, Hovmand P, Cherng S, Ferrer R, Katerndahl DA, et al. Modeling the Paradox of Primary Care. In Sturmberg JP, Martin CM, editors. *Handbook of Systems and Complexity in Health*. New York, NY: Springer; 2013. pp. 815–25. https://doi.org/10.1007/978-1-4614-4998-0_46
332. Jayne D, Pigazzi A, Marshall H, Croft J, Corrigan N, Copeland J, et al. Effect of robotic-assisted vs conventional laparoscopic surgery on risk of conversion to open laparotomy among patients undergoing resection for rectal cancer: the ROLARR randomized clinical trial. *JAMA* 2017;**318**:1569–80. <https://doi.org/10.1001/jama.2017.7219>
333. Collinson FJ, Jayne DG, Pigazzi A, Tsang C, Barrie JM, Edlin R, et al. An international, multicentre, prospective, randomised, controlled, unblinded, parallel-group trial of robotic-assisted versus standard laparoscopic surgery for the curative treatment of rectal cancer. *Int J Colorectal Dis* 2012;**27**:233–41. <https://doi.org/10.1007/s00384-011-1313-6>
334. Mukumbang FC, Van Belle S, Marchal B, Van Wyk B. Realist evaluation of the antiretroviral treatment adherence club programme in selected primary healthcare facilities in the metropolitan area of Western Cape Province, South Africa: a study protocol. *BMJ Open* 2016;**6**:e009977. <https://doi.org/10.1136/bmjopen-2015-009977>
335. Bloor M, Frankland J, Langdon NP, Robinson M, Allerston S, Catherine A, et al. A controlled evaluation of an intensive, peer-led, schools-based, anti-smoking programme. *Health Educ J* 1999;**58**:17–25. <https://doi.org/10.1177/001789699905800103>
336. Holliday J, Audrey S, Moore L, Parry-Langdon N, Campbell R. High fidelity? How should we consider variations in the delivery of school-based health promotion interventions? *Health Educ J* 2009;**68**:44–62. <https://doi.org/10.1177/0017896908100448>
337. Hollingworth W, Cohen D, Hawkins J, Hughes RA, Moore LA, Holliday JC, et al. Reducing smoking in adolescents: cost-effectiveness results from the cluster randomized ASSIST (A Stop Smoking In Schools Trial). *Nicotine Tob Res* 2012;**14**:161–8. <https://doi.org/10.1093/ntr/ntr155>
338. Butler M, Epstein RA, Totten A, Whitlock EP, Ansari MT, Damschroder LJ, et al. AHRQ series on complex intervention systematic reviews – paper 3: adapting frameworks to develop protocols. *J Clin Epidemiol* 2017;**90**:19–27. <https://doi.org/10.1016/j.jclinepi.2017.06.013>
339. Noyes J, Gough D, Lewin S, Mayhew A, Michie S, Pantoja T, et al. A research and development agenda for systematic reviews that ask complex questions about complex interventions. *J Clin Epidemiol* 2013;**66**:1262–70. <https://doi.org/10.1016/j.jclinepi.2013.07.003>
340. Shepperd S, Lewin S, Straus S, Clarke M, Eccles MP, Fitzpatrick R, et al. Can we systematically review studies that evaluate complex interventions? *PLOS Med* 2009;**6**:e1000086. <https://doi.org/10.1371/journal.pmed.1000086>
341. Kelly MP, Noyes J, Kane RL, Chang C, Uhl S, Robinson KA, et al. AHRQ series on complex intervention systematic reviews – paper 2: defining complexity, formulating scope, and questions. *J Clin Epidemiol* 2017;**90**:11–18. <https://doi.org/10.1016/j.jclinepi.2017.06.012>

342. Squires JE, Valentine JC, Grimshaw JM. Systematic reviews of complex interventions: framing the review question. *J Clin Epidemiol* 2013;**66**:1215–22. <https://doi.org/10.1016/j.jclinepi.2013.05.013>
343. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review – a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;**10**:21–34. <https://doi.org/10.1258/1355819054308530>
344. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.1. 2020. URL: www.training.cochrane.org/handbook (accessed 4 November 2020). <https://doi.org/10.1002/9781119536604>
345. Norris SL, Rehfuss EA, Smith H, Tunçalp Ö, Grimshaw JM, Ford NP, et al. Complex health interventions in complex systems: improving the process and methods for evidence-informed health decisions. *BMJ Glob Health* 2019;**4**:e000963. <https://doi.org/10.1136/bmjgh-2018-000963>
346. Booth A, Noyes J, Flemming K, Moore G, Tunçalp Ö, Shakibazadeh E. Formulating questions to explore complex interventions within qualitative evidence synthesis. *BMJ Glob Health* 2019;**4**:e001107. <https://doi.org/10.1136/bmjgh-2018-001107>
347. Flemming K, Booth A, Hannes K, Cargo M, Noyes J. Cochrane Qualitative and Implementation Methods Group guidance series – paper 6: reporting guidelines for qualitative, implementation, and process evaluation evidence syntheses. *J Clin Epidemiol* 2018;**97**:79–85. <https://doi.org/10.1016/j.jclinepi.2017.10.022>
348. Noyes J, Booth A, Moore G, Flemming K, Tunçalp Ö, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. *BMJ Glob Health* 2019;**4**:e000893. <https://doi.org/10.1136/bmjgh-2018-000893>
349. Booth A, Moore G, Flemming K, Garside R, Rollins N, Tunçalp Ö, et al. Taking account of context in systematic reviews and guidelines considering a complexity perspective. *BMJ Glob Health* 2019;**4**:e000840. <https://doi.org/10.1136/bmjgh-2018-000840>
350. Rehfuss EA, Stratil JM, Scheel IB, Portela A, Norris SL, Baltussen R. The WHO-INTEGRATE evidence to decision framework version 1.0: integrating WHO norms and values and a complexity perspective. *BMJ Glob Health* 2019;**4**:e000844. <https://doi.org/10.1136/bmjgh-2018-000844>
351. Tugwell P, Knottnerus JA, Idzerda L. Complex interventions – how should systematic reviews of their impact differ from reviews of simple or complicated interventions? *J Clin Epidemiol* 2013;**66**:1195–6. <https://doi.org/10.1016/j.jclinepi.2013.09.003>
352. Guise JM, Chang C, Butler M, Viswanathan M, Tugwell P. AHRQ series on complex intervention systematic reviews – paper 1: an introduction to a series of articles that provide guidance and tools for reviews of complex interventions. *J Clin Epidemiol* 2017;**90**:6–10. <https://doi.org/10.1016/j.jclinepi.2017.06.011>
353. Lewin S, Hendry M, Chandler J, Oxman AD, Michie S, Shepperd S, et al. Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT_SR). *BMC Med Res Methodol* 2017;**17**:76. <https://doi.org/10.1186/s12874-017-0349-x>
354. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;**151**:264–9, W64. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>

355. Guise JM, Butler M, Chang C, Viswanathan M, Pigott T, Tugwell P, Complex Interventions Workgroup. AHRQ series on complex intervention systematic reviews – paper 7: PRISMA-CI elaboration and explanation. *J Clin Epidemiol* 2017;**90**:51–8. <https://doi.org/10.1016/j.jclinepi.2017.06.017>
356. Guise JM, Butler ME, Chang C, Viswanathan M, Pigott T, Tugwell P, Complex Interventions Workgroup. AHRQ series on complex intervention systematic reviews – paper 6: PRISMA-CI extension statement and checklist. *J Clin Epidemiol* 2017;**90**:43–50. <https://doi.org/10.1016/j.jclinepi.2017.06.016>
357. France EF, Cunningham M, Ring N, Uny I, Duncan EAS, Jepson RG, *et al.* Improving reporting of meta-ethnography: the eMERGe reporting guidance. *BMC Med Res Methodol* 2019;**19**:25. <https://doi.org/10.1186/s12874-018-0600-0>
358. Montgomery P, Movsisyan A, Grant SP, Macdonald G, Rehfuess EA. Considerations of complexity in rating certainty of evidence in systematic reviews: a primer on using the GRADE approach in global health. *BMJ Glob Health* 2019;**4**:e000848. <https://doi.org/10.1136/bmjgh-2018-000848>

Appendix 1 Oversight of the project

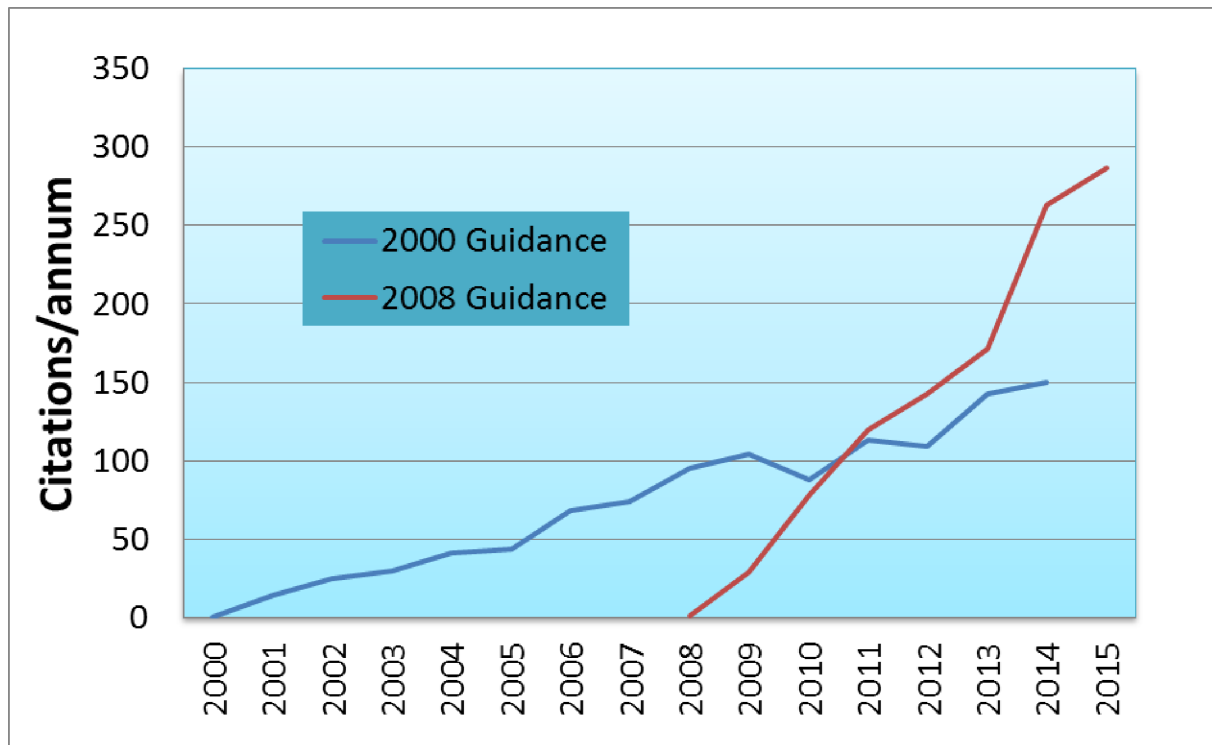
Who	Role
Project team	To undertake all aspects of the project including leading the data gathering, data analysis and writing the framework document
SAG (see <i>Appendix 3</i> for members)	To act as the oversight body for the work, specifically to: <ul style="list-style-type: none"> • provide expert advice to the project team • monitor progression of the project in relation to overall timelines • approve the proposed methodology • approve the membership of the writing group • approve/comment on the publication policy • approve/comment on the main project outputs • contribute to the promotion and dissemination of the updated framework • follow the terms of reference as agreed
Writing group (authors)	To be part of the author group for the work, adhering to specified authorship criteria. This included: <ul style="list-style-type: none"> • providing written and critical commentary related to area of expertise • delivering a significant contribution to the main project outputs
MRC PHSG and MRP Boards	Reviewed and approved the framework

Appendix 2 Updating the Medical Research Council guidance on complex interventions: an outline proposal for developing an integrated body of guidance

1. Background: the need for updating and integrating the MRC guidance

The MRC guidance on evaluating complex interventions (CIs) was first published in 2000, and updated in 2008. Over this period, interest in complex interventions has grown rapidly, and more specialised guidance has been published or is being developed in several related areas, including intervention development, intervention description, process evaluation, natural experimental approaches, dealing with intervention context, group-delivered interventions, realist trials, surgical trials, exploratory trials, trial management, disability research and knowledge translation [see **References** for examples]. A number of other areas that were identified as gaps in the 2008 guidance, such as research priority-setting and the application of complex systems science to health interventions, have also attracted interest, but are not yet covered by accessible guidance for producers or users of evidence. Both the 2000 and 2008 core guidance documents continue to be highly cited (figure), but given the pace and extent of methodological development, there is a strong case for updating the core guidance, linking it with related developments, and also addressing some of the remaining weaknesses and gaps in the existing guidance.

Aim & objective: To identify and summarise aspects of the 2006 MRC complex intervention guidance that require updating, with the aim of using this gap analysis as a starting point for discussion (through workshops and authorship group) to achieve consensus on focus points and resulting updates for the new guidance.



Source: Web of Science. The '2000 Guidance' refers to Campbell *et al.* (2000) and the '2008 Guidance' refers to Craig *et al.* (2008).

2. Proposed outputs

The 2000 and 2008 guidance each comprised a long version, published online by the MRC, and a short version published by the BMJ. The MRC's process evaluation and natural experiments guidance followed a similar model. Citations of the short versions greatly outnumber citations of the long versions, although the longer versions have important additional detail and have a very high number of downloads. For the updated guidance we recommend two products to disseminate the updated core guidance:

- a) A **downloadable pdf** version of the core guidance.
- b) A **journal article** (ideally published simultaneously in a number of journals, as the CONSORT statements are) which would describe the importance and need for the guidance as well as outlining its content, also referring potential users to the detailed online pdf version and signposting resource.

In addition, we will continue discussions about creating an online resource comprising the updated core guidance with links to signpost other related source documents (e.g. the MRC's process or natural experiments guidance) or to brief summary statements prepared specifically, with links to other useful published resources already available online.

3. Proposed stages of development

- *Gap analysis*: based on scoping reviews to find publications that identify gaps and weaknesses in the existing guidance, or that provide more detailed guidance on specific topics. This will take the form of *horizon scanning and further literature review*: a brief scoping review will be followed by discussion at the Scientific Advisory Group (SAG) meeting (24/11/17). Our initial search (google scholar, forward/backward citations) will focus on: (a) New approaches/progress since previous guidance; (b) Criticisms of existing guidance; (c) Other gaps. Prior to the SAG meeting we will develop a brief list of topics to address in the updated guidance. We will present these topics to the SAG (by sending a summary in advance and by presentation of an overview at the meeting). Following discussion at the SAG meeting we will finalise a list of topics that we will explore in more depth.
- *Expert workshop*: The findings from the gap analysis will inform the agenda for an expert workshop to be held in early 2018. Each of the 'topics/themes' identified for update should be represented (by an expert) at the workshop i.e. we will invite experts based on these themes. The aim of the expert workshop is to achieve consensus on topics that should be newly covered or updated by the new guidance and as a basis for the project team, along with the rest of the authorship group, to produce the updates and additions. Follow-up consultation (email and a consensus meeting) will be used to achieve consensus on the details of the updated guidance.
- *Identify relevant case studies*: Worked examples of the development, implementation, or evaluation of a complex intervention. This will be started at the workshop and finalised through consultation afterwards.
- Convening a steering group to oversee the work, review and approve drafts, and ensure stakeholder commitment; and a writing group to draft the update and linked summaries
- Drafting the update and summaries
- Drafting the journal article and managing the publication process

4. Representation

- (a) Steering group:** The steering group will meet on an infrequent basis (e.g. two or three times over timeline of activity). The steering group members are: (chair) Martin White (NIHR Public Health Research Programme), David French (MRC–NIHR Methodology Research Programme), Jo Rycroft-Malone (NIHR Health Services Delivery Research Programme), Mark Petticrew (Co-author of previous guidance), Martin Ashton-Key (NIHR HTA Programme & Consultant Adviser, NETSCC), Janis Baird (Co-author of previous process evaluation guidance), Jane Blazeby (MRC Hubs for Trial Methodology Research), Samuel Rowley (Observer), Gavin Malloch (Observer). All will be offered authorship (alongside formal authorship criteria).
- (b) Writing group:** The writing group will meet on a more regular basis (tbc), be engaged in the progress of the guidance and contribute to the writing process. Members of the writing group will be acknowledged as authors on the final guidance. The project team will be members of the writing group, with additional authors identified as appropriate, e.g. to fill gaps in expertise.
- (c) Project team:** Employed researchers Kathryn Skivington & Lynsay Matthews; senior staff from the host institution Sharon Simpson, Peter Craig, Laurence Moore.

References

Craig P., C. Cooper, et al. (2012). "Using natural experiments to evaluate population health interventions: new Medical Research Council guidance." *Journal of Epidemiology and Community Health* 66: 1182-1186.

Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;337:a1655.

Hoddinott, P., Allan, K., Avenell, A., Jane Britten, J., 2010. Group interventions to improve health outcomes: a framework for their design and delivery. *BMC Public Health* 10, 800-9.

Hoffmann, T. C., P. P. Glasziou, et al. (2014). "Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide." *BMJ* 348.

Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for the design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694-6.

McCulloch, P., Altman, D.G., Campbell, W.B., Flum, D.R., Glasziou, P., Marshall, J.C., Nicholl, J., for the Balliol Collaboration, 2009. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 374, 1105–12.

Moore G et al 2015. Process evaluation of complex interventions. Medical Research Council Guidance. *BMJ* 350:h1258. doi:10.1136/bmj.h1258.

Treweek, S Altman, D, Bower, P et al. (2015) Making randomised trials more efficient: report of the first meeting to discuss the Trial Forge platform. *Trials* 16:261

Appendix 3 Advisors/consultees

Members of the Scientific Advisory Group

Martin White (chairperson), Martin Ashton-Key, Janis Baird, Jane M Blazeby, David P French, Mark Petticrew and Jo Rycroft-Malone.

Participants of the expert workshop

Jean Adams, Rob Anderson, Till Baernighausen, Janis Baird, Jane Blazeby, Chris Bonell, Kathleen Boyd, Rona Campbell, Neil Craig, Steven Cummins, Frank DeVocht, Matt Egan, Matt Field, David P French, Michael Grayling, Judith Green, Marie Johnston, Frank Kee, Mike Kelly, Natasha Kriznik, James Lewsey, Alastair Leyland, Ronan Lyons, Petra Meier, Susan Michie, Graham Moore, Jane Noyes, Chris Owen, Mark Petticrew, Harry Rutter, Jo Rycroft-Malone, Falko Sniehotta, David Taylor-Robinson, Rebecca Walwyn, Katrina Wyatt and Lucy Yardley.

In addition, input was received from (although invited, could not attend the workshop) Marion Campbell, Ruth Dundas, Pat Hoddinot, Russ Jago, Alicia O'Cathain and Eva Annette Rehfuess.

Systems meeting

Steve Cummins, Matt Egan, Vanessa Er, Mike Kelly, Karen Lock and Harry Rutter (follow-up meeting only).

Intervention development meeting

Graham Moore, Alicia O'Cathain, Lucy Yardley and Ed Duncan.

Provided final feedback

Matt Egan, Matti Heino, Graham Moore, Dave Richards, Mark Robinson, Mike Robling and Jeremy Segrott.

Appendix 4 Online consultation questionnaire

Q1. Please indicate which stakeholder category you belong to (tick one or more boxes)

Researcher
Funder
Journal editor
Practitioner
Policy-maker
Service user (Patient or Public)
Other (please specify)

Q2. What is your main field of expertise? (tick one or more boxes)

Public health
Clinical medicine
Health services research
Health technology assessment
Policy-making
Systems-based research
Patient or public involvement
Other (please specify)

Q3. Figure 2: This diagram presents the key elements for developing and evaluating complex interventions. Is this diagram a useful way of capturing the research process (please elaborate)? [TEXT BOX HERE]

Q4. Definition of complexity and research perspectives: The updated framework presents a new definition of complexity, and suggests different perspectives of the research process. How useful is it to have the framework for addressing complexity presented in this way? [TEXT BOX HERE]

Q5. Figure 3. This figure relates to the concept of using different research perspectives (as noted in Q3). Does *Figure 3* make sense (please elaborate)? [TEXT BOX HERE]

Q6. Missing elements: If you think there are key elements missing from the guidance, can you briefly outline what they are? [TEXT BOX HERE]

Q7. Signposting to other guidance: We are keen to signpost to further guidance where appropriate. If we have missed any specific guidance that may be helpful to readers can you please note it here, thank you. [TEXT BOX HERE]

Q8. Clarity: Was anything in the guidance particularly unclear (please elaborate)? [TEXT BOX HERE]

Q9. Case studies: We are collating case studies to include in the guidance. We would welcome suggestions in the following areas:

- Modifying interventions to improve their intervention design and/or evaluation design
- The use of programme theory throughout the research process
- Involvement of stakeholders throughout the research process
- Economic considerations throughout the research process
- The exploration and use of context throughout the research process
- Addressing uncertainties throughout the research process
- Complexity informed approaches to intervention research

Q10. Other feedback: If you have other comments that may be useful as we finalise the guidance, please provide brief details [TEXT BOX HERE]

Appendix 5 Approaching complexity in evidence synthesis

The synthesis of complex intervention studies can be particularly challenging and there are numerous important considerations applicable to conducting evidence synthesis of complex interventions, including (but certainly not limited to: see *Signposts to further reading 12* for detailed discussion and guidance):

- Context – a similar intervention could be implemented in different contexts with very different results; therefore, in synthesising these results it may be important to consider the context in which they were achieved. There are different ways of taking account of context within evidence syntheses: see *Signposts to further reading 12* for further reading. In terms of developing the protocol for a complex intervention evidence synthesis, it may be necessary to adapt traditional frameworks, for example to extract data to consider contextual complexity, such as on timing and setting, dependent on each review's aims.³³⁸
- Uncertainty – similar to undertaking primary intervention research, whether or not it makes sense to take a systems perspective is dependent on what the users of the review actually want to know, that is where their uncertainty lies and their related priorities.^{89,339–341}
- Stakeholder engagement – taking account of multiple perspectives is key to understanding complexity; stakeholders can be involved in each stage, for example determining the scope of the review, framing and defining the review questions, understanding context and interpreting results. It is also important to ensure that the review is relevant and useful to the needs of those with an interest in using the findings, preventing research waste.^{341,342}
- Theory-based review – a different approach to the standard evidence synthesis is a 'theory-led' approach, for example a realist review.³⁴³ Realist synthesis takes a different approach to systematic review in that the focus is on coming to an understanding about the theory behind the interventions, exploring the circumstances in which interventions work or do not work, and for whom. The Cochrane Handbook also provides guidance to support the synthesis of studies aiming to increase understanding of intervention complexity, for example through synthesis of qualitative evidence.³⁴⁴

Useful discussions and guides about approaching complexity in evidence syntheses have been published; see the box below for further details.

SIGNPOSTS TO FURTHER READING 12 Complexity in systematic reviews

- A special issue in *BMJ Global Health* on concepts and methods for evidence-informed health decisions,³⁴⁵ including:
 - Determining whether or not a complex systems perspective would be useful for your evidence synthesis: consider the questions laid out by Petticrew *et al.*⁸⁹ in box 2 of their paper.
 - Considering complexity within qualitative evidence synthesis^{346,347} and in mixed-methods research.³⁴⁸
 - Tools and frameworks for considering context within evidence synthesis.³⁴⁹
 - WHO-INTEGRATE: a new evidence-to-decision framework tool, particularly developed to be used with complex interventions, to support the transparent consideration of all decision criteria.³⁵⁰
 - Special issues of the *Journal of Epidemiology and Community Health*: one largely devoted to considering complexity in systematic reviews of interventions³⁵¹ and one to complex interventions in systematic reviews developed by the Agency for Healthcare Research and Quality.³⁵²
 - A tool for assessing the complexity of interventions within systematic reviews: intervention complexity assessment tool for systematic reviews.³⁵³

SIGNPOSTS TO FURTHER READING 12 Complexity in systematic reviews (*continued*)

- Version 6.1 of the Cochrane Handbook includes new chapters on intervention complexity, qualitative evidence, network meta-analysis and synthesising findings using alternative methods to meta-analysis.³⁴⁴
- PRISMA-CI is an extension of PRISMA, the accepted standard for encouraging consistency and transparency in systematic reviews.³⁵⁴ PRISMA-CI incorporates issues specific to complex interventions, alongside an 'explanation and elaboration' guideline to support its use.^{355,356}
- Meta-ethnography can inform the development, evaluation and implementation of complex interventions. The eMERGe reporting guidance aims to improve the reporting of meta-ethnography.³⁵⁷
- GRADE is being adapted to address non-randomised studies of complex interventions, alongside guidance on how sources of complexity can be considered when using GRADE to rate certainty of evidence.³⁵⁸

GRADE, Grading of Recommendations Assessment, Development and Evaluation; PRISMA, Grading of Recommendations Assessment, Development and Evaluation; PRISMA-CI, Grading of Recommendations Assessment, Development and Evaluation – Complex Interventions.

Appendix 6 Checklist for developing and evaluating complex interventions

This checklist is intended as a tool to help researchers prepare funding applications, research protocols and journal publications. It may also help reviewers to assess whether or not the recommendations have been followed.

Item	If NO, please justify. If YES, briefly describe how this has been addressed	Reported on page number(s)
Addressing uncertainties		
<ol style="list-style-type: none"> 1. Have you determined the aim(s)/purpose(s) of the intervention? 2. Have you identified the key uncertainties given existing evidence about the intervention and the context in which it will be tested or implemented? 3. Do the research questions and methods address the key uncertainties? 4. Does the choice of research perspective (efficacy, effectiveness, theory-based, systems) reflect the key uncertainties that have been identified? 		
Engaging stakeholders		
<ol style="list-style-type: none"> 1. Have you engaged stakeholders in the design/identification of the intervention and the development of the research protocol? 2. Have you engaged stakeholders in the conduct of the research and the dissemination of findings? 3. Have all stakeholders declared any potential conflicts of interest? 		
Considering context		
<ol style="list-style-type: none"> 1. Have you identified all the dimensions of context that may influence how the intervention achieves its effects? 2. Have you considered how context may affect the scaling up or scaling out of the intervention? 		
Developing and refining programme theory		
<ol style="list-style-type: none"> 1. Have you developed a programme theory for your intervention that describes the key components and mechanisms of the intervention and how it interacts with the context in which it will be implemented? 2. Have you updated the programme theory to incorporate the new evidence gathered by the study? 		
Refining the intervention		
<ol style="list-style-type: none"> 1. Have you refined the intervention so that it is optimised for the context in which it will be implemented? 2. Have you specified how far and in what ways the intervention can be refined during implementation without undermining the programme theory? 		
Economic considerations		
<ol style="list-style-type: none"> 1. Have you considered whether or not the value of the evidence, in terms of informing future decision-making, justifies the cost of the research? 2. Have you identified an economic evaluation framework that is appropriate to the expected outcomes of the intervention? 		

Item	If NO, please justify. If YES, briefly describe how this has been addressed	Reported on page number(s)
Phase-specific considerations		
Developing interventions – have you used a formal framework (such as INDEX) to guide development of the intervention?		
Identifying interventions – for policy and practice interventions, have you performed an evaluability assessment to determine whether or not and how an evaluation should be undertaken?		
Feasibility – have you defined and used clear progression criteria to guide decisions about whether to proceed to an evaluation study?		
Evaluation – have you chosen an appropriate study design to answer the research questions and provide robust evidence to inform decision-making about further intervention refinement, evaluation or implementation?		
Implementation – have constraints and enablers of implementation been considered at all phases, from intervention development, through feasibility and effectiveness testing, to large-scale roll-out?		

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

*This report presents independent research funded by the National Institute for Health Research (NIHR).
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the
Department of Health and Social Care*

Published by the NIHR Journals Library