

Process evaluation of complex interventions

UK Medical Research Council (MRC) guidance

Prepared on behalf of the MRC Population Health Science Research Network by:

Graham Moore^{1,2}, Suzanne Audrey^{1,3}, Mary Barker⁴, Lyndal Bond⁵, Chris Bonell⁶, Wendy Hardeman⁷, Laurence Moore⁸, Alicia O’Cathain⁹, Tannaze Tinati⁴, Danny Wight⁸, Janis Baird³

1 Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHer), 2 Cardiff School of Social Sciences, Cardiff University. 3 School of Social and Community Medicine, University of Bristol. 4 MRC Lifecourse Epidemiology Unit (LEU), University of Southampton. 5 Centre of Excellence in Intervention and Prevention Science, Melbourne. 6 Institute of Education, University of London. 7 Primary Care Unit, University of Cambridge. 8 MRC/CSO Social & Public Health Sciences Unit (SPHSU), University of Glasgow. 9 School of Health and Related Research (SchARR), University of Sheffield.

FOREWORD	6
ACKNOWLEDGEMENTS	7
KEY WORDS.....	8
EXECUTIVE SUMMARY.....	9
<i>AIMS AND SCOPE</i>	<i>9</i>
<i>WHY IS PROCESS EVALUATION NECESSARY?.....</i>	<i>9</i>
<i>WHAT IS PROCESS EVALUATION?</i>	<i>10</i>
<i>HOW TO PLAN, DESIGN, CONDUCT AND REPORT A PROCESS EVALUATION</i>	<i>11</i>
PLANNING A PROCESS EVALUATION	11
DESIGNING AND CONDUCTING A PROCESS EVALUATION	13
ANALYSIS	15
INTEGRATION OF PROCESS EVALUATION AND OUTCOMES DATA.....	16
REPORTING FINDINGS OF A PROCESS EVALUATION	16
SUMMARY	17
 1. INTRODUCTION: WHY DO WE NEED PROCESS EVALUATION OF COMPLEX INTERVENTIONS?	 18
<i>BACKGROUND AND AIMS OF THIS DOCUMENT</i>	<i>18</i>
<i>WHAT IS A COMPLEX INTERVENTION?.....</i>	<i>19</i>
<i>WHY IS PROCESS EVALUATION NECESSARY?.....</i>	<i>19</i>
<i>THE IMPORTANCE OF ‘THEORY’: ARTICULATING THE CAUSAL ASSUMPTIONS OF COMPLEX INTERVENTIONS</i>	<i>21</i>
<i>KEY FUNCTIONS FOR PROCESS EVALUATION OF COMPLEX INTERVENTIONS.....</i>	<i>22</i>
IMPLEMENTATION: HOW IS DELIVERY ACHIEVED, AND WHAT IS ACTUALLY DELIVERED?	22
MECHANISMS OF IMPACT: HOW DOES THE DELIVERED INTERVENTION PRODUCE CHANGE?.....	23
CONTEXT: HOW DOES CONTEXT AFFECT IMPLEMENTATION AND OUTCOMES?.....	23
A FRAMEWORK FOR LINKING PROCESS EVALUATION FUNCTIONS	24
FUNCTIONS OF PROCESS EVALUATION AT DIFFERENT STAGES OF THE DEVELOPMENT-EVALUATION-IMPLEMENTATION PROCESS	25
FEASIBILITY AND PILOTING	25
EFFECTIVENESS EVALUATION	26
POST-EVALUATION IMPLEMENTATION.....	27
PRAGMATIC POLICY TRIALS AND NATURAL EXPERIMENTS	27
SUMMARY OF KEY POINTS.....	28
 SECTION A - PROCESS EVALUATION THEORY	 30

2. FRAMEWORKS, THEORIES AND CURRENT DEBATES IN PROCESS EVALUATION 30

FRAMEWORKS WHICH USE THE TERM 'PROCESS EVALUATION'	30
INTERVENTION DESCRIPTION, THEORY AND LOGIC MODELLING	31
DESCRIBING COMPLEX INTERVENTIONS	31
PROGRAMME THEORY: CLARIFYING ASSUMPTIONS ABOUT HOW THE INTERVENTION WORKS	32
FROM THEORY TO IMPLEMENTATION	36
DEFINITIONS OF 'IMPLEMENTATION' AND 'FIDELITY'	36
DONABEDIAN'S STRUCTURE-PROCESS-OUTCOME MODEL	36
THE RE-AIM FRAMEWORK.....	37
CARROLL AND COLLEAGUES' CONCEPTUAL FRAMEWORK FOR FIDELITY.....	37
THE OXFORD IMPLEMENTATION INDEX	38
ORGANISATIONAL CHANGE AND THE IMPLEMENTATION PROCESS	38
THE FIDELITY AND ADAPTATION DEBATE	39
HOW DOES THE INTERVENTION WORK? THEORY-DRIVEN EVALUATION APPROACHES	41
THEORY-BASED EVALUATION AND REALIST(IC) EVALUATION	42
CAUSAL MODELLING: TESTING MEDIATION AND MODERATION.....	43
COMPLEXITY SCIENCE PERSPECTIVES	44
SUMMARY OF KEY POINTS	45

3. USING FRAMEWORKS AND THEORIES TO INFORM A PROCESS EVALUATION 46

IMPLEMENTATION	46
HOW IS DELIVERY ACHIEVED?	46
WHAT IS ACTUALLY DELIVERED?	46
MECHANISMS OF IMPACT: HOW DOES THE DELIVERED INTERVENTION WORK?	47
Understanding how participants respond to, and interact with, complex interventions	47
Testing mediators and generating intervention theory	47
Identifying unintended and/or unanticipated consequences.....	48
CONTEXTUAL FACTORS	48

SUMMARY OF KEY POINTS..... 49

SECTION B - PROCESS EVALUATION PRACTICE..... 50

4. HOW TO PLAN, DESIGN, CONDUCT AND ANALYSE A PROCESS EVALUATION 50

PLANNING A PROCESS EVALUATION	51
--	-----------

WORKING WITH PROGRAMME DEVELOPERS AND IMPLEMENTERS	51
COMMUNICATION OF EMERGING FINDINGS BETWEEN EVALUATORS AND IMPLEMENTERS	52
INTERVENTION STAFF AS DATA COLLECTORS: OVERLAPPING ROLES OF THE INTERVENTION AND EVALUATION	53
RELATIONSHIPS WITHIN EVALUATION TEAMS: PROCESS EVALUATION AND OTHER EVALUATION COMPONENTS.....	54
RESOURCES AND STAFFING	56
PATIENT AND PUBLIC INVOLVEMENT.....	57
DESIGNING AND CONDUCTING A PROCESS EVALUATION.....	57
DEFINING THE INTERVENTION AND CLARIFYING CAUSAL ASSUMPTIONS	57
LEARNING FROM PREVIOUS PROCESS EVALUATIONS. WHAT DO WE KNOW ALREADY? WHAT WILL THIS STUDY ADD?59	
DECIDING CORE AIMS AND RESEARCH QUESTIONS	61
Implementation: what is delivered, and how?	62
Mechanisms of impact: how does the delivered intervention work?	63
Contextual factors.....	63
Expecting the unexpected: building in flexibility to respond to emergent findings.....	64
SELECTING METHODS	66
Common quantitative methods in process evaluation	68
Common qualitative methods in process evaluation.....	70
Mixing methods	71
USING ROUTINE MONITORING DATA FOR PROCESS EVALUATION	73
SAMPLING	74
TIMING CONSIDERATIONS.....	74
ANALYSIS	75
ANALYSING QUANTITATIVE DATA.....	75
ANALYSING QUALITATIVE DATA.....	75
MIXING METHODS IN ANALYSIS.....	76
INTEGRATING PROCESS EVALUATION FINDINGS AND FINDINGS FROM OTHER EVALUATION COMPONENTS (E.G.	
OUTCOMES AND COST-EFFECTIVENESS EVALUATION)	77
ETHICAL CONSIDERATIONS.....	78
SUMMARY OF KEY POINTS.....	80

5. REPORTING AND DISSEMINATION OF PROCESS EVALUATION FINDINGS 81

HOW AND WHAT TO REPORT?	81
REPORTING TO WIDER AUDIENCES.....	82
PUBLISHING IN ACADEMIC JOURNALS.....	83
WHEN TO REPORT PROCESS DATA? BEFORE, AFTER OR ALONGSIDE OUTCOMES/COST-EFFECTIVENESS EVALUATION?84	
SUMMARY OF KEY POINTS.....	85

6. APPRAISING A PROCESS EVALUATION: A CHECKLIST FOR FUNDERS AND PEER REVIEWERS 86

WORKING WITH POLICY AND PRACTICE STAKEHOLDERS	86
--	-----------

<i>RELATIONSHIPS BETWEEN EVALUATION COMPONENTS.....</i>	<i>86</i>
<i>INTERVENTION DESCRIPTION AND THEORY</i>	<i>86</i>
<i>PROCESS EVALUATION AIMS AND RESEARCH QUESTIONS</i>	<i>87</i>
<i>SELECTION OF METHODS TO ADDRESS RESEARCH QUESTIONS</i>	<i>88</i>
<i>RESOURCE CONSIDERATIONS IN COLLECTING/ANALYSING PROCESS DATA</i>	<i>88</i>
<i>ANALYSIS AND REPORTING</i>	<i>89</i>
 SECTION C – CASE STUDIES.....	 90
 REFERENCES	 124
 APPENDIX 1 - ABOUT THE AUTHORS	 130
 APPENDIX 2 - DEVELOPING THE GUIDANCE: PROCESS AND MILESTONES	 133

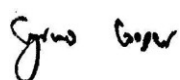
Foreword

The MRC Population Health Sciences Research Network (PHSRN) was established by the UK Medical Research Council in 2005. The work of the network focuses on methodological knowledge transfer in population health sciences. PHSRN has been responsible for producing guidance, on behalf of MRC, for the conduct of population health research including, in 2008, the updated guidance on the evaluation of complex interventions and, in 2011, guidance on the evaluation of natural experiments.

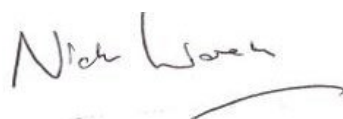
The updated MRC guidance on the evaluation of complex interventions called for definitive evaluation to combine evaluation of outcomes with that of process. This reflected recognition that in order for evaluations to inform policy and practice, emphasis was needed not only on whether interventions ‘worked’ but on how they were implemented, their causal mechanisms and how effects differed from one context to another. It did not, however, offer detail on *how* to conduct process evaluation. In November 2010 an MRC PHSRN-funded workshop discussed the need for guidance on process evaluation. Following the workshop, at which there had been strong support for the development of guidance, a multi-disciplinary group was formed to take on the task.

The guidance was developed through an iterative process of literature review, reflection on detailed case studies of process evaluation and extensive consultation with stakeholders. The report contains guidance on the planning, design, conduct, reporting and appraisal of process evaluations of complex interventions. It gives a comprehensive review of process evaluation theory, bringing together theories and frameworks which can inform process evaluation, before providing a practical guide on how to carry out a process evaluation. A model summarising the key features of process evaluation is used throughout to guide the reader through successive sections of the report.

This report will provide invaluable guidance in thinking through key decisions which need to be made in developing a process evaluation, or appraising its quality, and is intended for a wide audience including researchers, practitioners, funders, journal editors and policy-makers.



Cyrus Cooper, Chair, MRC Population Health Sciences Research Network



Nick Wareham, Chair Elect, MRC Population Health Sciences Research Network

Acknowledgements

The task of developing this guidance was initiated in large part through discussions at the PHSRN workshop on process evaluation within public health intervention studies held in the MRC Lifecourse Epidemiology Unit at the University of Southampton in November 2010. We gratefully acknowledge the input from participants at that workshop in identifying core aims for the guidance. We are grateful to the following stakeholders for reviewing the full draft of the guidance and providing vital comments, which helped to shape the final guidance: Claire Chandler, Peter Craig, Janet Darbyshire, Jenny Donovan, Aileen Grant, Sean Grant, Tom Kenny, Ruth Hunter, Jennifer Lloyd, Paul Montgomery, Heather Rothwell, Jane Smith and Katrina Wyatt. The structure of the guidance was influenced by feedback on outlines from most of the above people, plus Andrew Cook, Fiona Harris, Matt Kearney, Mike Kelly, Kelli Komro, Barrie Margetts, Lynsay Mathews, Sarah Morgan-Trimmer, Dorothy Newbury-Birch, Julie Parkes, Gerda Pot, David Richards, Jeremy Segrott, James Thomas, Thomas Willis and Erica Wimbush. We are also grateful to the workshop participants at the UKCRC Public Health Research Centres of Excellence conference in Cardiff, the QUALitative Research in Trials (QUART) symposium in Sheffield, the Society for Social Medicine conference in Brighton, two DECIPHer staff forums in Cardiff organised by Sarah Morgan-Trimmer and Hannah Littlecott, and British Psychological Society-funded seminars on ‘Using process evaluation to understand and improve the psychological underpinnings of health-related behaviour change interventions’ in Exeter and Norwich, led by Jane Smith. We regret that we are not able to list the individual attendees at these workshops and seminars who provided comments which shaped the draft. We gratefully acknowledge Catherine Turney for input into knowledge exchange activities and for assistance with editing the drafts, and to Hannah Littlecott for assistance with editing drafts of the guidance. In acknowledging valuable input from these stakeholders, we do not mean to imply that they endorse the final version of the guidance. We are grateful to the MRC Population Health Sciences Research Network for funding the work (PHSRN45). We are grateful for the assistance and endorsement of the MRC Population Health Sciences Group, the MRC Methodology Research Panel and the NIHR Evaluation, Trials and Studies Coordinating Centre.

Key words

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 outcomes evaluation.

Complex intervention – an intervention comprising multiple components which interact to produce change. Complexity may also relate to the difficulty of behaviours targeted by interventions, the number of organisational levels targeted, or the range of outcomes.

Public health intervention – an intervention focusing on primary or secondary prevention of disease and positive health promotion (rather than treatment of illness).

Logic model – a diagrammatic representation of an intervention, describing anticipated delivery mechanisms (e.g. how resources will be applied to ensure implementation), intervention components (what is to be implemented), mechanisms of impact (the mechanisms through which an intervention will work) and intended outcomes.

Implementation – the process through which interventions are delivered, and what is delivered in practice. Key dimensions of implementation include:

- Implementation process – the structures, resources and mechanisms through which delivery is achieved;
- Fidelity – the consistency of what is implemented with the planned intervention;
- Adaptations – alterations made to an intervention in order to achieve better contextual fit;
- Dose – how much intervention is delivered;
- Reach – the extent to which a target audience comes into contact with the intervention.

Mechanisms of impact – the intermediate mechanisms through which intervention activities produce intended (or unintended) effects. The study of mechanisms may include:

- Participant responses – how participants interact with a complex intervention; Mediators – intermediate processes which explain subsequent changes in outcomes; Unintended pathways and consequences.

Context – factors external to the intervention which may influence its implementation, or whether its mechanisms of impact act as intended. The study of context may include:

- Contextual moderators which shape, and may be shaped by, implementation, intervention mechanisms, and outcomes;

Executive summary

Aims and scope

This document provides researchers, practitioners, funders, journal editors and policy-makers with guidance in planning, designing, conducting and appraising process evaluations of complex interventions. The background, aims and scope are set out in more detail in Chapter 1, which provides an overview of core aims for process evaluation, and introduces the framework which guides the remainder of the document. The guidance is then divided into two core sections: **Process Evaluation Theory (Section A)** and **Process Evaluation Practice (Section B)**. Section A brings together a range of theories and frameworks which can inform process evaluation, and current debates. Section B provides a more practical ‘how to’ guide. Readers may find it useful to start with the section which directly addresses their needs, rather than reading the document cover to cover. The guidance is written from the perspectives of researchers with experience of process evaluations alongside trials of complex public health interventions (interventions focused upon primary or secondary prevention of disease, or positive health promotion, rather than treatment of illness). However, it is also relevant to stakeholders from other research domains, such as health services or education. This executive summary will provide a brief overview of **why** process evaluation is necessary, **what** it is, and **how** to plan, design and conduct a process evaluation. It signposts readers to chapters of the document in which they will find more detail on the issues discussed.

Why is process evaluation necessary?

High quality evaluation is crucial in allowing policy-makers, practitioners and researchers to identify interventions that are effective, and learn how to improve those that are not. As described in Chapter 2, outcome evaluations such as randomised trials and natural experiments are essential in achieving this. But, if conducted in isolation, outcomes evaluations leave many important questions unanswered. For example:

- If an intervention is effective in one context, what additional information does the **policy-maker** need to be confident that:
 - another organisation (or set of professionals) will deliver it in the same way;
 - if they do, it will produce the same outcomes in new contexts?

- If an intervention is ineffective overall in one context, what additional information does the **policy-maker** need to be confident that:
 - the failure is attributable to the intervention itself, rather than to poor implementation;
 - the intervention does not benefit any of the target population;
 - if it was delivered in a different context, it would be equally ineffective?
- What information do **systematic reviewers** need to:
 - be confident that they are comparing interventions which were delivered in the same way;
 - understand why the same intervention has different effects in different contexts?

Additionally, interventions with positive overall effects may reduce or increase inequalities. While simple sub-group analyses may allow us to identify whether inequalities are affected by the intervention, understanding how inequalities are affected requires a more detailed understanding of cause and effect than is provided by outcomes evaluation.

What is process evaluation?

Process evaluations aim to provide the more detailed understanding needed to inform policy and practice. As indicated in Figure 1, this is achieved through examining aspects such as:

- **Implementation:** the structures, resources and processes through which delivery is achieved, and the quantity and quality of what is delivered¹;
- **Mechanisms of impact:** how intervention activities, and participants' interactions with them, trigger change;
- **Context:** how external factors influence the delivery and functioning of interventions.

Process evaluations may be conducted within feasibility testing phases, alongside evaluations of effectiveness, or alongside post-evaluation scale-up.

¹ The term **implementation** is used within complex intervention literature to describe both post-evaluation scale-up (i.e. the 'development-evaluation-implementation' process) and intervention delivery during the evaluation period. Within this document, discussion of implementation relates primarily to the second of these definitions (i.e. the quality and quantity of what is actually delivered during the evaluation).

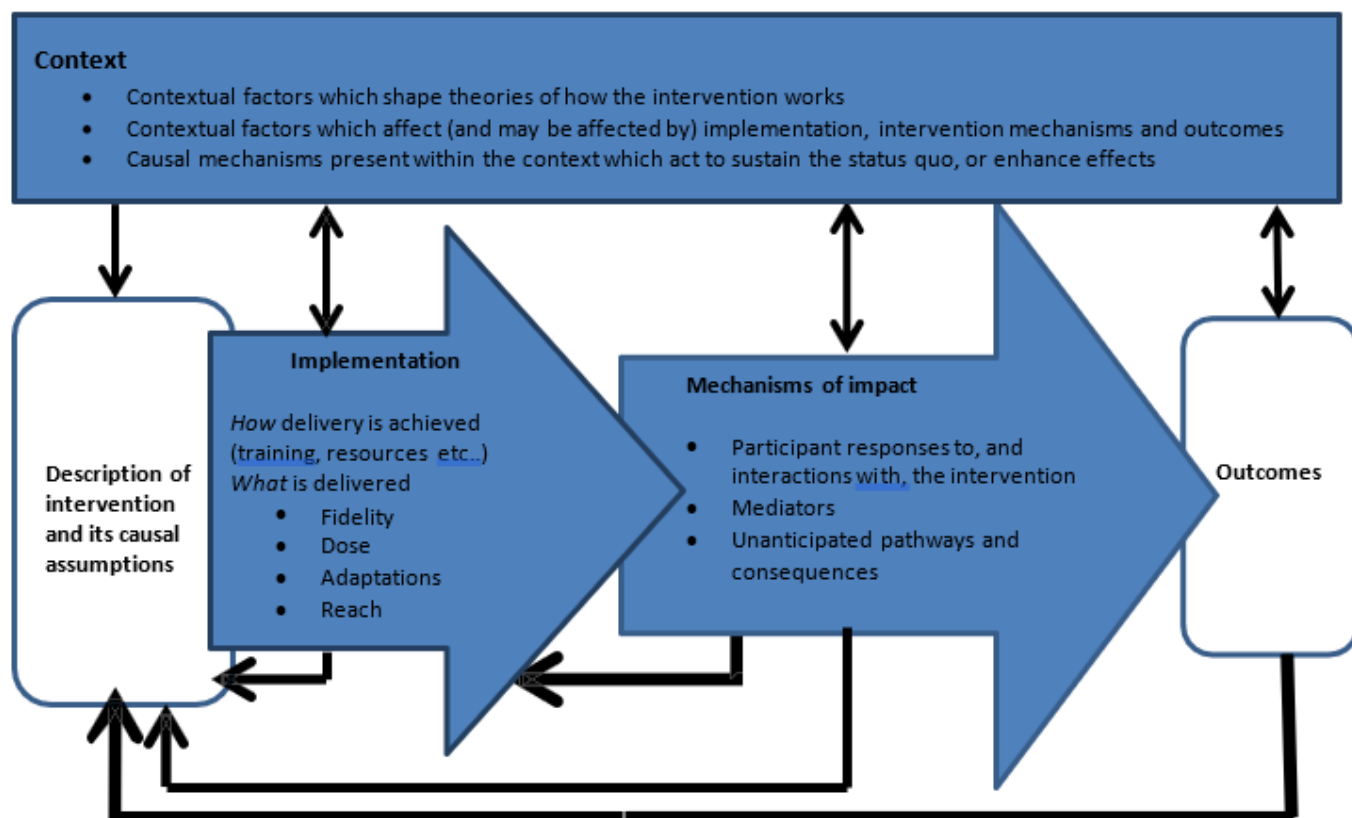


Figure 1. Key functions of process evaluation and relationships amongst them. Blue boxes represent components of process evaluation, which are informed by the causal assumptions of the intervention, and inform the interpretation of outcomes.

Extended Alt Text - Figure 1 above illustrates the relationship between context, intervention, implementation, mechanisms of impact and outcomes.

Context influences all parts of the intervention and is defined as:

- Contextual factors that shape theories of how the intervention works.
- Contextual factors that affect, and may be affected by, implementation, intervention mechanisms and outcomes.
- Causal mechanisms within the context that sustain the status quo or enhance intervention effects.

The intervention begins with a description of the intervention and its underlying causal assumptions.

Implementation describes how delivery is achieved, including training and resources, and what is delivered. Key implementation dimensions are:

- Fidelity
- Dose
- Adaptations

- Reach

Mechanisms of impact include:

- Participant responses to, and interactions with, the intervention.
- Mediators.
- Unanticipated pathways and consequences.

Outcomes arise from these mechanisms of impact.

Context interacts bidirectionally with implementation, mechanisms of impact and outcomes. Feedback loops also occur between outcomes, mechanisms of impact, implementation and intervention assumptions.

How to plan, design, conduct and report a process evaluation

Chapter 4 offers detailed guidance for planning and conducting process evaluations. The key recommendations of this guidance are presented in Box 1, and expanded upon below.

Planning a process evaluation

Relationships with intervention developers and implementers: Process evaluation will involve critically observing the work of intervention staff. Sustaining good working relationships, whilst remaining sufficiently independent for evaluation to remain credible, is a challenge which must be taken seriously. Reflecting on whether these relationships are leading evaluators to view the intervention too positively, or to be unduly critical, is vital. Planning for occasional critical peer review by a researcher with less investment in the project, who may be better placed to identify where the position of evaluators has started to affect independence, may be useful.

Key recommendations and issues to consider in planning, designing and conducting, analysing and reporting a process evaluation

When planning a process evaluation, evaluators should:

- Carefully define the parameters of relationships with intervention developers or implementers.
 - Balance the need for sufficiently good working relationships to allow close observation against the need to remain credible as an independent evaluator
 - Agree whether evaluators will play an active role in communicating findings as they emerge (and helping correct implementation challenges) or play a more passive role
- Ensure that the research team has the correct expertise, including
 - Expertise in qualitative and quantitative research methods
 - Appropriate inter-disciplinary theoretical expertise
- Decide the degree of separation or integration between process and outcome evaluation teams
 - Ensure effective oversight by a principal investigator who values all evaluation components
 - Develop good communication systems to minimise duplication and conflict between process and outcomes evaluations
 - Ensure that plans for integration of process and outcome data are agreed from the outset

When designing and conducting a process evaluation, evaluators should:

- Clearly describe the intervention and clarify its causal assumptions in relation to how it will be implemented, and the mechanisms through which it will produce change, in a specific context
- Identify key uncertainties and systematically select the most important questions to address.
 - Identify potential questions by considering the assumptions represented by the intervention
 - Agree scientific and policy priority questions by considering the evidence for intervention assumptions and consulting the evaluation team and policy/practice stakeholders
 - Identify previous process evaluations of similar interventions and consider whether it is appropriate to replicate aspects of them and build upon their findings
- Select a combination of quantitative and qualitative methods appropriate to the research questions
 - Use quantitative methods to quantify key process variables and allow testing of pre-hypothesised mechanisms of impact and contextual moderators
 - Use qualitative methods to capture emerging changes in implementation, experiences of the intervention and unanticipated or complex causal pathways, and to generate new theory
 - Balance collection of data on key process variables from all sites or participants where feasible, with detailed case studies of purposively selected samples
 - Consider data collection at multiple time points to capture changes to the intervention over time

When analysing process data, evaluators should:

- Provide descriptive quantitative information on fidelity, dose and reach
- Consider more detailed modelling of variations between participants or sites in terms of factors such as fidelity or reach (e.g. are there socioeconomic biases in who is reached?)
- Integrate quantitative process data into outcomes datasets to examine whether effects differ by implementation or pre-specified contextual moderators, and test hypothesised mediators
- Collect and analyse qualitative data iteratively so that themes that emerge in early interviews can be explored in later ones
- Ensure that quantitative and qualitative analyses build upon one another, with qualitative data used to explain quantitative findings, and quantitative data used to test hypotheses generated by qualitative data
- Where possible, initially analyse and report qualitative process data prior to knowing trial outcomes to avoid biased interpretation
- Transparently report whether process data are being used to generate hypotheses (analysis blind to trial outcomes), or for post-hoc explanation (analysis after trial outcomes are known)

When reporting process data, evaluators should:

- Identify existing reporting guidance specific to the methods adopted
- Report the logic model or intervention theory and clarify how it was used to guide selection of research questions
- Publish multiple journal articles from the same process evaluation where necessary
 - Ensure that each article makes clear its context within the evaluation as a whole
 - Publish a full report comprising all evaluation components or a protocol paper describing the whole evaluation, to which reference should be made in all articles
 - Emphasise contributions to intervention theory or methods development to enhance interest to a readership beyond the specific intervention in question
- Disseminate findings to policy and practice stakeholders

Deciding structures for communicating and addressing emerging issues: During a process evaluation, researchers may identify implementation problems which they want to share with policy-makers and practitioners. Process evaluators will need to consider whether they act as passive observers, or have a role in communicating or addressing implementation problems during the course of the evaluation. At the feasibility or piloting stage, the researcher should play an active role in communicating such issues. But when aiming to establish effectiveness under real world conditions, it may be appropriate to assume a more passive role. Overly intensive process evaluation may lead to distinctions between the evaluation and the intervention becoming blurred. Systems for communicating information and addressing emerging issues should be agreed at the outset.

Relationships within evaluation teams - process evaluations and other evaluation

components: Process evaluation will commonly be part of a larger evaluation which includes evaluation of outcomes and/or cost-effectiveness. The relationships between components of an evaluation must be defined at the planning stage. Oversight by a principal investigator who values all aspects of the evaluation is crucial. If outcomes evaluation and process evaluation are conducted by separate teams, effective communications must be maintained to prevent duplication or conflict. Where process and outcomes evaluation are conducted by the same individuals, openness about how this might influence data analysis is needed.

Resources and staffing: Process evaluations involve complex decisions about research questions, theoretical perspectives and research methods. Sufficient time must be committed by those with expertise and experience in the psychological and sociological theories underlying the intervention, and in the quantitative and qualitative methods required for the process evaluation.

Public and patient involvement: It is widely believed that increased attention to public involvement may enhance the quality and relevance of health and social science research. For example, including lay representatives in the project steering group might improve the quality and relevance of a process evaluation.

Designing and conducting a process evaluation

Defining the intervention and clarifying key assumptions: Ideally, by the time an evaluation begins, the intervention will have been fully described. A ‘logic model’ (a diagram describing the structures in place to deliver the intervention, the intended activities, and

intended short-, medium- and long-term outcomes; see Chapter 2) may have been developed by the intervention and/or evaluation team. In some cases, evaluators may choose not to describe the causal assumptions underpinning the intervention in diagrammatic form. However, it is crucial that a clear description of the intervention and its causal assumptions is provided, and that evaluators are able to identify how these informed research questions and methods.

What do we know already? What will this study add? Engaging with the literature to identify what is already known, and what advances might be offered by the proposed process evaluation, should always be a starting point. Evaluators should consider whether it is appropriate to replicate aspects of previous evaluations of similar interventions, building on these to explore new process issues, rather than starting from scratch. This could improve researchers' and systematic reviewers' ability to make comparisons across studies.

Core aims and research questions: It is better to identify and effectively address the most important questions than to try and answer every question. Being over-ambitious runs the risk of stretching resources too thinly. Selection of core research questions requires careful identification of the key uncertainties posed by the intervention, in terms of its implementation, mechanisms of impact and interaction with its context. Evaluators may start by listing assumptions about how the intervention will be delivered and how it will work, before reviewing the evidence for those assumptions, and seeking agreement within the evaluation team, and with policy and practice stakeholders, on the most important uncertainties for the process evaluation to investigate. While early and systematic identification of core questions will focus the process evaluation, it is often valuable to reserve some research capacity to investigate unforeseen issues that might arise in the course of the process evaluation. For example, if emerging implementation challenges lead to significant changes in delivery structures whose impacts need to be captured.

Selecting appropriate methods: Most process evaluations will use a combination of methods. The pros and cons of each method (discussed in more detail in Chapter 4) should be weighed up carefully to select the most appropriate methods for the research questions asked.

Common quantitative methods used by process evaluators include:

- structured observations;
- self-report questionnaires;

- secondary analysis of routine data.

Common qualitative methods include:

- one-to-one interviews;
- group interviews or focus groups;
- non-participant observation.

Sampling: While it is not always possible or appropriate to collect all process data from all of the participants in the outcomes evaluation, there are dangers in relying on a small number of cases to draw conclusions regarding the intervention as a whole. Hence, it is often useful to collect data on key aspects of process from all participants, in combination with in-depth data from smaller samples. ‘Purposive’ sampling according to socio-demographic or organisational factors expected to influence delivery or effectiveness is a useful approach.

Timing of data collection: The intervention, participants’ interactions with it, and the contexts in which these take place may change during the evaluation. Hence, attention should be paid to the time at which data are collected, and how this may influence the issues identified. For example, data collected early on may identify ‘teething problems’ which were rectified later. It may be useful to collect data at multiple different times to capture change in implementation or contextual factors.

Analysis

Mixing methods in analysis: While requiring different skills, and often addressing different questions, quantitative and qualitative data ought to be used in combination. Quantitative data may identify issues which require qualitative exploration, while qualitative data may generate theory to be tested quantitatively. Qualitative and quantitative components should assist interpretation of one another’s findings, and methods should be combined in a way which enables a gradual accumulation of knowledge of how the intervention is delivered and how it works.

Analysing quantitative data: Quantitative analysis typically begins with descriptive information (e.g. means, drop-out rates) on measures such as fidelity, dose and reach. Process evaluators may also conduct more detailed modelling to explore variation in factors such as implementation and reach. Such analysis may start to answer questions such as how

inequalities begin to widen/narrow at each stage. Integrating quantitative process measures into the modelling of outcomes may also help to identify links between delivery of specific components and outcomes, intermediate processes and contextual influences.

Analysing qualitative data: Qualitative analyses can provide in-depth understanding of mechanisms of action, how context affects implementation, or why those delivering or receiving the intervention do or do not engage as planned. Their flexibility and depth means qualitative approaches can be used to explore complex or unanticipated mechanisms and consequences. The length of time required for thorough qualitative analysis should not be underestimated. Ideally, collection and analysis of qualitative data should occur in parallel. This should ensure that emerging themes from earlier data can be investigated in later data collections, and that the researcher will not reach the end of data collection with an excessive amount of data and little time to analyse it.

Integration of process evaluation and outcomes findings

Those responsible for different aspects of the evaluation should ensure that plans are made for integration of data, and that this is reflected in evaluation design. If quantitative data are gathered on process components such as fidelity, dose, reach or intermediate causal mechanisms, these should ideally be collected in a way that allows their associations with outcomes and cost-effectiveness to be modelled in secondary analyses. Qualitative process analyses may help to predict or explain intervention outcomes. They may lead to the generation of causal hypotheses regarding variability in outcomes - for example, whether certain groups appear to have responded to an intervention better than others - which can be tested quantitatively.

Reporting findings of a process evaluation

The reporting of process evaluations is often challenging. Chapter 5 provides guidance on reporting process evaluations of complex interventions, given the large quantities of diverse data generated. Key issues are summarised below.

What to report: There is no 'one size fits all' method for process evaluation. Evaluators will want to draw upon a range of reporting guidelines which relate to specific methods (see Chapter 5 for some examples). A key consideration is clearly reporting relationships between quantitative and qualitative components, and the relationship of the process evaluation to other evaluation. The assumptions being made by intervention developers about how the

intervention will produce intended effects should be reported; logic models are recommended as a way of achieving this. Process evaluators should describe how these descriptions of the theory of the intervention were used to identify the questions addressed.

Reporting to wider audiences: Process evaluations often aim to directly inform the work of policy-makers and practitioners. Hence, reporting findings in lay formats to stakeholders involved in the delivery of the intervention, or decisions on its future, is vital. Evaluators will also want to reach policy and practice audiences elsewhere, whose work may be influenced by the findings. Presenting findings at policy-maker- and service provider-run conferences offers a means of promoting findings beyond academic circles.

Publishing in academic journals: Process evaluators will probably wish to publish multiple research articles in peer reviewed journals. Articles may address different aspects of the process evaluation and should be valuable and understandable as standalone pieces. However, all articles should refer to other articles from the study, or to a protocol paper or report which covers all aspects of the process evaluation, and make its context within the wider evaluation clear. It is common for process data not to be published in journals. Researchers should endeavour to publish all aspects of their process evaluations. Emphasising contributions to interpreting outcomes, intervention theory, or methodological debates regarding the evaluation of complex interventions, may increase their appeal to journal editors. Study websites which include links to manuals and all related papers are a useful way of ensuring that findings can be understood as part of a whole.

Summary

This document provides the reader with guidance in planning, designing and conducting a process evaluation, and reporting its findings. While accepting that process evaluations usually differ considerably, it is hoped that the document will provide useful guidance in thinking through the key decisions which need to be made in developing a process evaluation, or appraising its quality.

Introduction: why do we need process evaluation of complex interventions?

Background and aims of this document

In November 2010, a UK Medical Research Council (MRC) Population Health Science Research Network (PHSRN)-funded workshop met to discuss process evaluation of complex public health interventions, and whether guidance was needed. Workshop participants, predominantly researchers and policy-makers, strongly supported the development of a document to guide them in planning, designing, conducting, reporting and appraising process evaluations of complex interventions. There was consensus that funders and reviewers of grant applications would benefit from guidance to assist peer review. Subsequently, a group of researchers was assembled to lead the development of this guidance, with further support from the MRC PHSRN (see Appendix B for an overview of guidance development). The original aim was to provide guidance for process evaluations of complex public health interventions (interventions focused on primary or secondary prevention of disease or positive health improvement, rather than health care). However, this document is highly relevant to other domains, such as health services research and educational interventions, and therefore serves as guidance for anyone conducting or appraising a process evaluation of a complex intervention.

In consultations regarding the document's proposed content, it became clear that stakeholders were looking for guidance on different aspects of process evaluation. Some identified a need for an overview of theoretical debates, and synthesis of work in various fields providing guidance on process evaluation. Others emphasised the need for practical guidance on how to do process evaluation. This document addresses these dual concerns through two discrete but linked sections, '**Process Evaluation Theory**' (Section A) and '**Process Evaluation Practice**' (Section B). Section A reviews influential frameworks relevant to process evaluation, and current theoretical debates. We make no claims to exhaustiveness, but provide an overview of a number of core frameworks, including those with which we are familiar from our own work, and others identified by external stakeholders. Section B provides practical guidance on planning, designing, conducting, analysing and reporting a process evaluation. Readers looking primarily for a how-to guide may wish to start with

Section B, which signposts back to specific parts of Section A to consult for additional relevant information.

Before moving onto these two sections, this introductory chapter outlines what we mean by a complex intervention, and why process evaluation is necessary within complex intervention research, before introducing a framework for linking together process evaluation aims. This framework is revisited throughout Sections A and B.

What is a complex intervention?

While ‘complex interventions’ are most commonly thought of as those which contain several interacting components, ‘complexity’ can also relate to the implementation of the intervention and its interaction with its context. Interventions commonly attempt to alter the functioning of systems such as schools or other organisations, which may respond in unpredictable ways (Keshavarz et al., 2010). Key dimensions of complexity identified by the MRC framework (Craig et al., 2008a; Craig et al., 2008b) include:

- The number and difficulty (e.g. skill requirements) of behaviours required by those delivering the intervention;
- The number of groups or organisational levels targeted by the intervention;
- The number and variability of outcomes;
- The degree of flexibility or tailoring of the intervention permitted.

As will be elaborated in Chapter 2, additional distinctions have been made between ‘complex’ and ‘complicated’ interventions, with complex interventions characterised by unpredictability, emergence and non-linear outcomes.

Why is process evaluation necessary?

All interventions represent attempts to implement a course of action in order to address a perceived problem. Hence, evaluation is inescapably concerned with cause and effect. If we implement an obesity intervention, for example, we want to know to what extent obesity will decline in the target population. Randomised controlled trials (RCTs) are widely regarded as the ideal method for identifying causal relationships. Where an RCT is not feasible, effects may be captured through quasi-experimental methods (Bonell et al., 2009). In other cases, interventions are too poorly defined to allow meaningful evaluation (House of Commons Health Committee, 2009). However, where possible, RCTs represent the most internally valid means of establishing effectiveness.

Some critics argue that RCTs of complex interventions over-simplify cause and effect, ignoring the agency of implementers and participants, and the context in which the intervention is implemented and experienced (Berwick, 2008b; Clark et al., 2007; Pawson & Tilley, 1997). Such critics often argue that RCTs are driven by a ‘positivist’ set of assumptions, which are incompatible with understanding how complex interventions work in context (Marchal et al., 2013). However, these arguments typically misrepresent the assumptions made by RCTs, or more accurately, by the researchers conducting them (Bonell et al., 2013). Randomisation aims to ensure that there is no systematic difference between groups in terms of participant and contextual characteristics, reflecting acknowledgment that these factors influence intervention outcomes.

Nevertheless, it is important to recognise that there are limits to what outcomes evaluations can achieve in isolation. If evaluations of complex interventions are to inform future intervention development, additional research is needed to address questions such as:

- If an intervention is effective in one context, what additional information does the **policy-maker** need to be confident that:
 - the intervention **as it was actually delivered** can be sufficiently well described to allow replication of its core components;
 - another organisation (or set of professionals) will deliver it in the same way;
 - if they do, it will produce the same outcomes in these new contexts?
- If an intervention is ineffective overall in one context, what additional information does the **policy-maker** need to be confident that:
 - the failure is attributable to the intervention itself, rather than to poor implementation?
 - the intervention does not benefit any of the target population?
 - if it was delivered in a different context it would be equally ineffective?
- What information do **systematic reviewers** need to:
 - be confident that they are comparing interventions which were delivered in the same way?
 - understand why the same intervention has different effects in different contexts?

Recognition is growing that RCTs of complex interventions can be conducted within a more critical realist framework (Bonell et al., 2012), in which social realities are viewed as valid objects of scientific study, yet methods are applied and interpreted critically. An RCT can identify whether a course of action was effective in the time and place it was delivered, while concurrent process evaluation can allow us to interpret findings and understand how they might be applied elsewhere. Hence, combining process evaluations with RCTs (or other high quality outcomes evaluations) can enable evaluators to limit biases in estimating effects, while developing the detailed understandings of causality that can support a policymaker, practitioner or systematic reviewer in interpreting effectiveness data. The aforementioned MRC framework (Craig et al., 2008a; Craig et al., 2008b) rejects arguments against randomised trials, but recognises that ‘effect sizes’ alone are insufficient, and that process evaluation is necessary to understand **implementation, causal mechanisms** and the **contextual factors** which shape outcomes. The following section will discuss each of these functions of process evaluation in turn. First, the need to understand intervention theory in order to inform the development of a process evaluation is considered.

The importance of ‘theory’: articulating the causal assumptions of complex interventions

While not always based on academic theory, all interventions are ‘theories incarnate’ (Pawson and Tilley, 1997), in that they reflect assumptions regarding the causes of the problem and how actions will produce change. An intervention as simple as a health information leaflet, for example, may reflect the assumption that a lack of knowledge regarding health consequences is a key modifiable cause of behaviour. Complex interventions are likely to reflect many causal assumptions. Identifying and stating these assumptions, or ‘programme theories’, is vital if process evaluation is to focus on the most important uncertainties that need to be addressed, and hence advance understanding of the implementation and functioning of the intervention. It is useful if interventions, and their evaluations, draw explicitly on existing social science theories, so that findings can add to the development of theory. However, evaluators should avoid selecting ‘off-the-shelf’ theories without considering how they apply to the context in which the intervention is delivered. Additionally, there is a risk of focusing narrowly on inappropriate theories from a single discipline; for example, some critics have highlighted a tendency for over-reliance upon individual-level theorising when the aim is to achieve community, organisational or population-level change (Hawe et al., 2009).

In practice, interventions will typically reflect assumptions derived from a range of sources, including academic theory, experience and ‘common sense’ (Pawson and Tilley, 1997). As will be discussed in Chapters 2 and 4, understanding these assumptions is critical to assessing how the intervention works in practice, and the extent to which this is consistent with its theoretical assumptions. Intervention theory may have been developed and refined alongside intervention development. In many cases, however, causal assumptions may remain almost entirely implicit at the time an evaluation is commissioned. A useful starting point is therefore to collaborate with those responsible for intervention development or implementation, to elicit and document the causal assumptions underlying the intervention (Rogers et al., 2000). It is often useful to depict these in a logic model, a diagrammatic representation of the theory of the intervention (Kellogg Foundation, 2004) - see Chapter 2 for more discussion of using logic models in process evaluation.

Key functions for process evaluation of complex interventions

Implementation: how is delivery achieved, and what is actually delivered?

The term ‘implementation’ is used within the literature both to describe post-evaluation scale-up (i.e. the ‘development-evaluation-implementation’ process) and delivery of an intervention during a trial (e.g. ‘Process evaluation nested within a trial can also be used to assess fidelity and quality of implementation’ (Craig et al. 2008b; p12)). Throughout this document, the term refers primarily to the second of these definitions. The principal aim of an outcomes evaluation is to test the theory of the intervention, in terms of whether the selected course of action led to the desired change. Examining the quality (fidelity) and quantity (dose) of what was implemented in practice, and the extent to which the intervention reached its intended audiences, is vital in establishing the extent to which the outcomes evaluation represents a valid test of intervention theory (Steckler & Linnan, 2002). Current debates regarding what is meant by fidelity, and the extent to which complex interventions must be standardised or adapted across contexts, are described in detail in Chapter 2

In addition to *what* was delivered, there is a growing tendency for process evaluation frameworks to advocate examining *how* delivery was achieved (e.g. Carroll et al., 2007; Montgomery et al., 2013b). Complex interventions typically involve making changes to the behaviours of intervention providers, or the dynamics of the systems in which they operate,

which may be as difficult as the ultimate problems targeted by the intervention. To apply evaluation findings in practice, the policy-maker or practitioner will need information not only on what was delivered during the evaluation, but on how similar effects might be achieved in everyday practice. This may involve considering issues such as the training and support offered to intervention providers; communication and management structures; and, as discussed below, how these interact with their contexts to shape what is delivered.

Mechanisms of impact: how does the delivered intervention produce change?

MRC guidance for developing and evaluating complex interventions argues that only through close scrutiny of causal mechanisms is it possible to develop more effective interventions, and understand how findings might be transferred across settings and populations (Craig et al., 2008b). Rather than passively receiving interventions, participants interact with them, with outcomes produced by these interactions in context (Pawson and Tilley, 1997). Hence, understanding how participants interact with complex interventions is crucial to understanding how they work. Process evaluations may test and refine the causal assumptions made by intervention developers, through combining quantitative assessments of pre-specified mediating variables with qualitative investigation of participant responses. This can allow identification of unanticipated pathways, and in-depth exploration of pathways which are too complex to be captured quantitatively.

Context: how does context affect implementation and outcomes?

‘Context’ may include anything external to the intervention which impedes or strengthens its effects. Evaluators may, for example, need to understand how implementers’ readiness or ability to change is influenced by pre-existing circumstances, skills, organisational norms, resources and attitudes (Berwick, 2008a; Glasgow et al., 2003; Pawson & Tilley, 1997). Implementing a new intervention is likely to involve processes of mutual adaptation, as context may change in response to the intervention (Jansen et al., 2010). Pre-existing factors may also influence how the target population responds to an intervention. Smoke-free legislation, for example, had a greater impact on second-hand smoke exposure among children whose parents did not smoke (Akhtar et al., 2007). The causal pathways underlying problems targeted by interventions will differ from one context to another (Bonell et al., 2006), meaning that the same intervention may have different consequences if implemented in a different setting, or among different subgroups. Hence, the theme ‘context’ cuts across both of the previous themes, with contextual conditions shaping implementation and effects.

Even where an intervention itself is relatively simple, its interaction with its context may still be considered highly complex.

A framework for linking process evaluation functions

Figure 2 presents a framework for linking the core functions of process evaluation described above. Within this framework, developing and articulating a clear description of the causal assumptions of the intended intervention (most likely in a logic model) is conceived not as a part of process evaluation, but as vital in framing everything which follows.

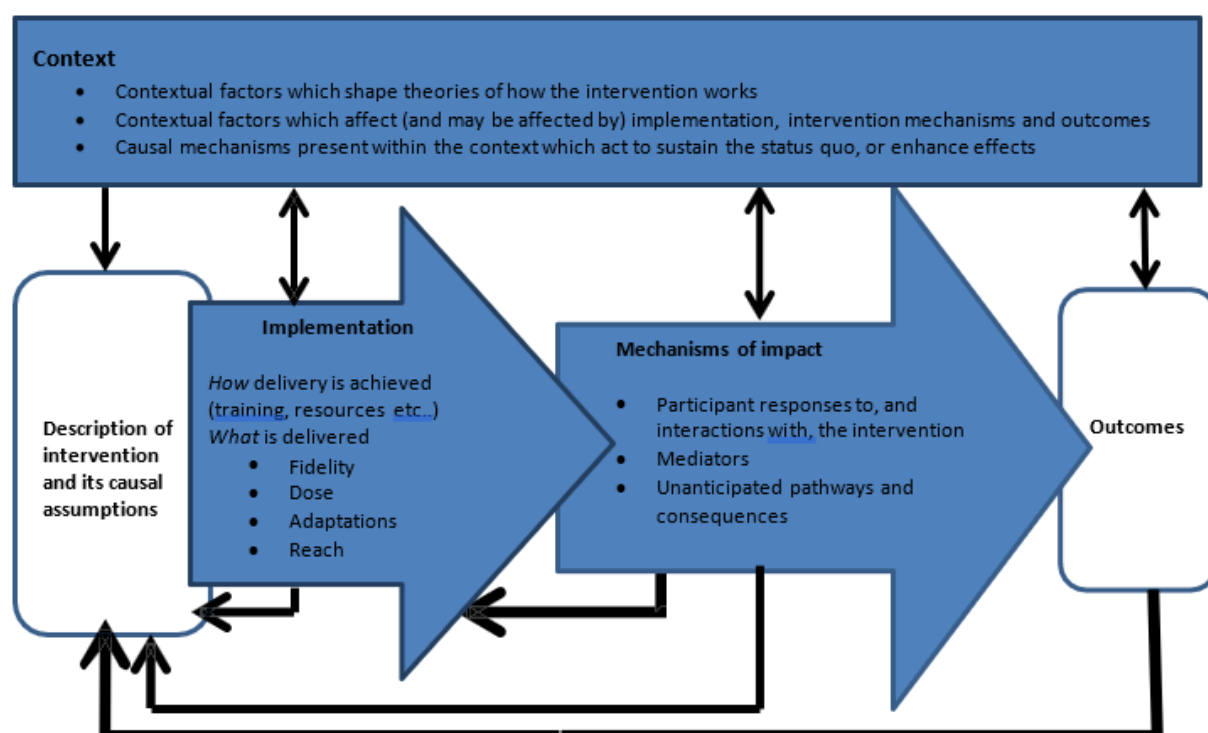


Figure 2. Key functions of process evaluation and relationships amongst them (blue boxes represent components of process evaluation, informed by the intervention description, which inform interpretation of outcomes).

Extended Alt Text - Figure 2 above illustrates the relationship between context, intervention, implementation, mechanisms of impact and outcomes.

Context influences all parts of the intervention and is defined as:

- Contextual factors that shape theories of how the intervention works.
- Contextual factors that affect, and may be affected by, implementation, intervention mechanisms and outcomes.
- Causal mechanisms within the context that sustain the status quo or enhance intervention effects.

The intervention begins with a description of the intervention and its underlying causal

assumptions.

Implementation describes how delivery is achieved, including training and resources, and what is delivered. Key implementation dimensions are:

- Fidelity
- Dose
- Adaptations
- Reach

Mechanisms of impact include:

- Participant responses to, and interactions with, the intervention.
- Mediators.
- Unanticipated pathways and consequences.

Outcomes arise from these mechanisms of impact.

Context interacts bidirectionally with implementation, mechanisms of impact and outcomes. Feedback loops also occur between outcomes, mechanisms of impact, implementation and intervention assumptions.

The ultimate goal of a process evaluation is to illuminate the pathways linking what starts as a hypothetical intervention, and its underlying causal assumptions, to the outcomes produced.

In order to achieve this, it is necessary to understand:

- **implementation**, both in terms of *how* the intervention was delivered (e.g. the training and resources necessary to achieve full implementation), and the quantity and quality of *what* was delivered;
 - the **mechanisms of impact** linking intervention activities to outcomes;
-

- how the **context** in which the intervention is delivered affects both what is implemented and how outcomes are achieved.

Although the diagram above presents a somewhat linear progression, feedback loops between components of the framework may occur at all stages, as indicated by the black arrows. As a clearer picture emerges of what was implemented in practice, intervention descriptions and causal assumptions may need to be revisited. Emerging insights into mechanisms triggered by the intervention may lead to changes in implementation. For example, in the National Exercise Referral Scheme in Wales (NERS, Case Study 5), professionals reported that many patients referred for weight loss became demotivated and dropped out, as two low intensity exercise sessions per week were unlikely to bring about substantial weight loss. Hence, many local coordinators added new components, training professionals to provide dietary advice.

Sections A (Process Evaluation Theory) and B (Process Evaluation Practice) will use this framework to shape discussion of process evaluation theory, frameworks and methods. First, the remainder of this chapter will discuss how aims of a process evaluation might vary according to the stage at which it is conducted.

Functions of process evaluation at different stages of the development-evaluation-implementation process

According to the MRC framework (Craig et al., 2008a; Craig et al., 2008b), feasibility testing should take place prior to evaluation of effectiveness, which should in turn precede scale-up of the intervention. The emphasis accorded to each of the functions of process evaluation described above, and the means of investigating them, may vary according to the stage at which process evaluation takes place.

Feasibility and piloting

Where insufficient feasibility testing has taken place, evaluation of effectiveness may fail to test the intended intervention because the structures to implement the intervention are not adequate (implementation failure), or the evaluation design proves infeasible (evaluation failure). Fully exploring key issues at the feasibility testing stage will ideally ensure that no major changes to intervention components or implementation structures will be necessary during subsequent effectiveness evaluation.

In addition to feasibility, process evaluations at this stage often focus on the acceptability of an intervention (and its evaluation). While it might seem that an intervention with limited acceptability can never be implemented properly, many effective innovations meet initial resistance. In SHARE (Sexual Health And RElationships, Case Study 3), teachers were highly resistant to the idea of providing condom demonstrations in classes, but in practice were happy to provide these when given a structure in which to do so. In NERS (Case Study 5), the move to national standardisation was resisted by many local implementers, but almost unanimously viewed positively one year later. Hence, there is a risk of not pursuing good ideas because of initial resistance if acceptability is regarded as fixed and unchanging. In some cases, process evaluation may involve developing strategies to counter resistance and improve acceptability. A recent trial of a premises-level alcohol harm reduction intervention (Moore et al., 2012b) provides an example of a process evaluation within an exploratory trial. The process evaluation explored the fidelity, acceptability and perceived sustainability of the intervention, and used these findings to refine the intervention's logic model.

Effectiveness evaluation

New challenges may be encountered at the stage of evaluating effectiveness. The increased scale of a fully powered evaluation is likely to mean greater variation in participant characteristics, contexts, and practitioners. Process evaluators will need to understand how this shapes the implementation and effectiveness of the intervention.

Emphasis, however, shifts from attempting to shape the intervention and its delivery structures, towards examining the internal validity of conclusions about effectiveness by examining the quantity and quality of what is delivered. Process evaluators may be increasingly conscious of minimising Hawthorne effects (where observation distorts what is delivered), only collecting the information needed to interpret outcomes (Audrey et al., 2006). Qualitative refinement of intervention theory may continue alongside evaluation of effectiveness, and it becomes possible to quantitatively test mediating mechanisms and contextual moderators. The evaluation of ASSIST (A Stop Smoking in Schools Study, Case Study 1) represents an example of a process evaluation within an evaluation of effectiveness, focusing on the views and experiences of participants and how variations in organisational contexts (schools) influenced implementation. Here, the process evaluation illuminated how the intervention theory (diffusion of innovations) was put into practice by young people.

Post-evaluation implementation

By this stage, there should be a clear and well-tested description of the intervention in place (probably in a logic model). This should set out what the intervention is, how to deliver it, the mechanisms through which the intervention works, and the contextual circumstances necessary for these mechanisms to be activated. Key remaining questions will centre on how to maintain fidelity in new settings (Bumbarger & Perkins, 2008). Reviews indicate that following evaluation, complex interventions are typically only sustained partially. How post-evaluation changes in implementation affect outcomes is usually unknown (Stirman et al., 2012). Understanding the diffusion of the intervention into new settings, the interaction of implementation processes with contextual circumstances, the transferability of evaluation findings into new contexts, and impacts of post-evaluation changes in implementation have on outcomes, become a key focus.

Pragmatic policy trials and natural experiments

‘Natural experiments’ are non-randomised evaluations of interventions delivered for purposes other than research (Craig et al., 2012). Examples include evaluations of smoke-free legislation (Haw et al., 2006). Pragmatic policy trials also aim to embed evaluation into real world interventions, with ‘nested’ randomisation incorporated when the policy is rolled out. The Primary School Breakfast Initiative (Murphy et al., 2011), and the National Exercise Referral Scheme (NERS, Case Study 5) in Wales (Murphy et al., 2012) are examples of pragmatic policy trials. The key strength of these methods is that they evaluate real world practice, and have high external validity. However, limited control over implementation poses significant challenges for process evaluation. Policy evaluations involve testing someone else’s ‘theory of change’, and substantial time may be needed to clarify what the intervention is and the assumptions being made. There may be greater likelihood of identifying flaws in implementation structures and intervention logic due to limited feasibility testing having taken place. In addition, they may involve rapid diffusion across multiple contexts. Hence, understanding how the intervention and its effects change shape as it moves from one setting to another, and how these changes affect the intervention, becomes critical.

When evaluating natural experiments, which involve non-randomised comparisons, particular attention should be paid to understanding how contextual factors differ between intervention and control settings (if a control setting is used). If, for example, we compare local authorities which have adopted a specific innovation to those that have not, which characteristics led to

the decision to adopt it? For instance, a greater organisational readiness may have led to increased enthusiasm in implementation, and greater effectiveness than where implemented more reluctantly in other settings. The NERS process evaluation (Case Study 5) served more formative functions than usual during an evaluation of effectiveness. Problems with implementation structures identified by process evaluation included underestimation of training and support requirements for implementing motivational interviewing. The process evaluation also paid substantial attention to how the intervention changed shape as it diffused into different local contexts.

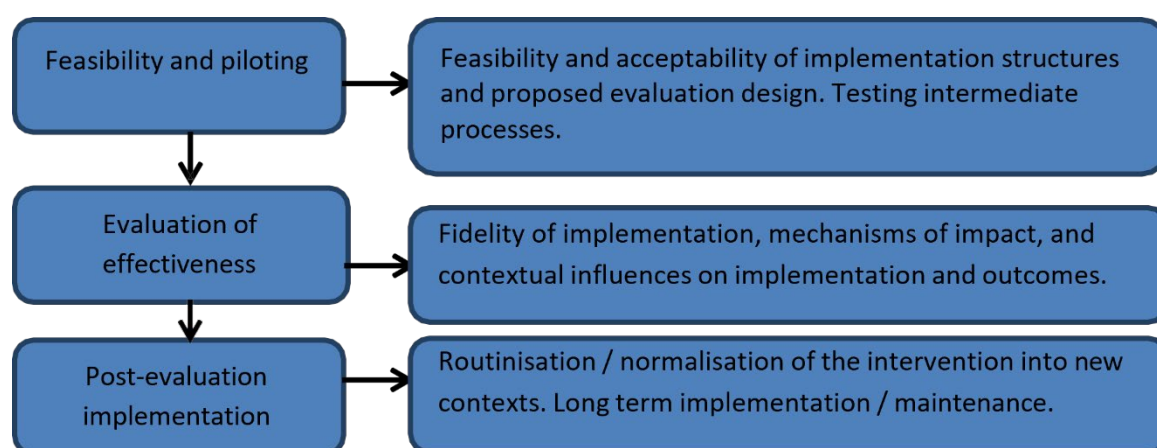


Figure 3. Functions of process evaluation at different evaluation stages.

Extended Alt Text - Figure 3 above shows three stages in intervention evaluation and implementation.

- The first stage, *Feasibility and Piloting*, focuses on assessing the feasibility and acceptability of implementation structures and the proposed evaluation design, including testing intermediate processes.
- The second stage, *Evaluation of Effectiveness*, focuses on assessing fidelity of implementation, mechanisms of impact, and contextual influences on implementation and outcomes.
- The third stage, *Post-evaluation Implementation*, focuses on routinisation or normalisation of the intervention in new contexts and on long-term implementation and maintenance.

The stages are sequential, with findings from each stage informing progression to the next.

Summary of key points

This chapter has described why we need process evaluation of complex interventions, and set out a framework to guide discussion throughout this document. It has argued that:

- An intervention may be complex in terms of the number of components it comprises, the nature of interactions between its components, challenges in its implementation, and how it interacts with its contexts.

- High quality outcomes evaluation is essential, but insufficient to provide the detailed understandings of how and why an intervention ‘worked’ (or did not), and for whom, which are necessary to inform policy and practice, and build an evidence base.
- A comprehensive and well-documented picture of what the intervention is, and the causal assumptions within it, is essential for the development of a high quality evaluation.
- Combining high quality outcomes evaluation with process evaluation allows evaluators to both capture overall effects, and understand **implementation**, the **mechanisms** through which intervention produces impacts, and how these are influenced by **context**.

Section A now draws together a number of key theories and frameworks which have informed process evaluation in recent years, and relates these back to the framework presented above. Readers looking to get to grips with process evaluation theory may find it most useful to start here. Readers looking primarily for practical guidance on how to do process evaluation may prefer to progress straight to Section B, which signposts back to relevant sections of Section A for more information on particular aspects of developing and conducting a process evaluation.

SECTION A - PROCESS EVALUATION THEORY

Frameworks, theories and current debates in process evaluation

At present, there is no unified definition of ‘process evaluation’. Studies using the term range from simple satisfaction questionnaires to complex mixed-method studies. As described in Chapter 1, the MRC argues that process evaluations ‘can be used to assess fidelity and quality of **implementation**, clarify causal **mechanisms**, and identify **contextual factors** associated with variation in outcomes’ (Craig et al., 2008a; our emphasis). Some influential frameworks that examine these core themes explicitly use the term ‘process evaluation’; others provide philosophical or methodological guidance in studying one or more of these themes but do not refer to themselves as process evaluation. While it is beyond the scope of this document to provide an exhaustive review, this section describes a number of influential frameworks and theoretical perspectives that researchers may draw upon in developing a process evaluation.

Frameworks which use the term ‘process evaluation’

A key aim of many early process evaluations was to monitor whether interventions were implemented as intended, in order to determine the extent to which outcomes evaluation represented a valid assessment of intervention theory (Finnegan et al., 1989; McGraw et al., 1989; Pirie et al., 1994). As recognition of the need for process evaluation increased, frameworks began to emerge, focusing attention on key priorities. Baranowski and Stables (2000) identified 11 priority areas for investigation: recruitment, maintenance, context, resources, implementation, reach, barriers, exposure, initial use, continued use and contamination. A similar framework, published soon after by Steckler and Linnan (2002), identified six priority areas: context (local factors that influence implementation), fidelity (the extent to which the intervention is delivered as conceived), dose delivered (the amount of intervention offered to participants), dose received (the extent of participants’ engagement in the intervention), reach and recruitment. More recently, a framework proposed by Grant and colleagues (2013a) emphasised areas for investigation when evaluating cluster randomised trials, but included some aims relevant to other methods. It went beyond many earlier frameworks in suggesting suitable methods for achieving these aims, and considering the timing of different aspects of a process evaluation. For example, intervention delivery (or

implementation) was considered to be suited to quantitative monitoring and qualitative exploration during the intervention. Examining responses to an intervention (in contrast to the quantitative and passive term ‘dose received’ within Steckler and Linnan’s framework, which in fact appears somewhat at odds with their own definition of ‘active engagement’) was considered best investigated qualitatively, during and following the intervention. The need to explore context qualitatively, both before and during intervention, was also emphasised, as was the quantitative and qualitative examination of unintended consequences. The importance of theorising and testing causal process was highlighted, with post-intervention quantitative analysis of causal processes seen as useful in testing intervention theory.

Intervention description, theory and logic modelling

Describing complex interventions

While not part of process evaluation, developing a clear definition of the intervention is central to planning a good quality process evaluation. It is common for evaluations to be undermined by limited description of the intervention under investigation (Michie et al., 2009). An investigation of the reporting of smoking cessation interventions found that fewer than half the components described in intervention manuals were described in the associated journal paper (Lorencatto et al., 2012). A recent review showed that most reports on RCTs of social and behavioural interventions do not provide links to intervention manuals (Grant et al., 2013b). Hence, the reader is left with data on whether or not an intervention works, but little insight into what the intervention is.

The need to fully describe complex interventions is highlighted in the Oxford Implementation Index (Montgomery et al., 2013b), which provides guidance for systematic reviewers on extracting information about interventions from evaluation articles prior to synthesis; without this, the reviewer cannot be sure which interventions are genuinely comparable. Michie and colleagues (2009) argue that making manuals publicly available, and greater uniformity in description of common behaviour change techniques, may help evaluators to achieve this. Their behaviour change technique taxonomy aims to improve homogeneity in reporting the ‘active ingredients’ of behavioural interventions (Michie et al., 2013), while the behaviour change wheel (Michie et al., 2011) attempts to categorise interventions according to the nature of the behaviour, intervention functions and policy categories. Work is also currently underway to extend CONSORT (Montgomery et al., 2013a; www.tinyurl.com/consort-study) reporting guidelines to incorporate reporting of social and psychological interventions.

Much of this work is best suited to traditional definitions of interventions as sets of activities delivered to individual participants. However, process evaluations are also necessary for interventions aiming to achieve change at community or organisational levels. For example, beginning from a hypothesis that greater use of restorative practices in schools may improve children's emotional wellbeing, one could identify a set of restorative practices and test whether delivering these activities improves children's emotional health. Alternatively, one might test this hypothesis through intervening at the level of the system, defining the intervention as a standard set of structures and processes to encourage the integration of restorative practices into the school (e.g. training for staff, changes to curriculum). Here, the aim is not to implement a discrete set of behaviours alongside what schools normally do, but to integrate these into everyday practice (Bonell et al. in press). Schools would not all deliver the exact same activities, but activities would aim to be consistent with the underlying theory and functions of restorative practice.

Programme theory: clarifying assumptions about how the intervention works

While not all interventions are based on formal theory, all are 'theories incarnate' (Pawson & Tilley, 1997), in that they assume that a course of action is a potential solution to a problem. Hence, in planning a process evaluation, it makes sense to start by carefully considering what these theories are. One means of making intervention assumptions clear is by constructing a logic model (a diagrammatic representation of the relationships between an intervention's resources, activities and intended outcomes (Kellogg Foundation, 2004)). There are a wide range of approaches to logic modelling, which vary in their language and the amount of emphasis they accord to the implementation process or intervention theory (e.g. Kirby, 2004; Renger & Titcomb, 2002). Logic models may depict the intended core components of the intervention, how they interact to produce change, anticipated short-, medium- and long-term outcomes, and resources and structures in place to ensure implementation. Developing a logic model may expose weak links or potential conflicts and contradictions in the hypothesised causal chain, or identify where stakeholders have differing understandings of the intended intervention. They may also enable evaluators to think critically about potential unintended consequences. An example of a logic model is presented below in **Error! Not a valid bookmark self-reference.**, which depicts the anticipated causal chain for the SPRING trial (an ongoing complex intervention study which aims to improve the diets of pregnant women, in which three of the author group are involved (MB, JB, TT)).

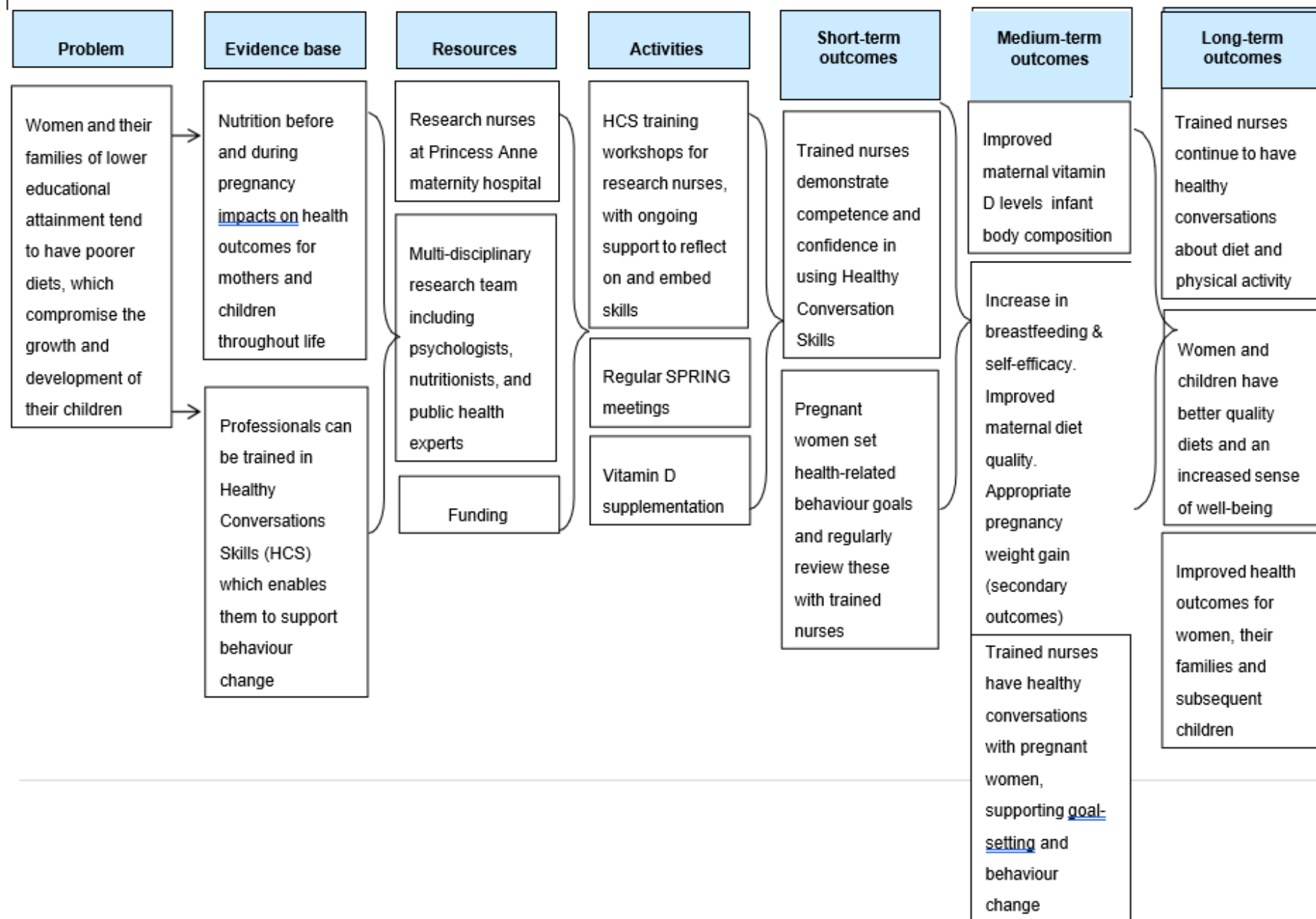


Figure 4 - Logic model for the Southampton Pregnancy Intervention for the Next Generation (SPRING)

Extended Alt Text for Figure 4 which shows the logic model for the Southampton Pregnancy Intervention for the Next Generation (SPRING)

Problem:

- Women and families with lower educational attainment often have poorer diets, which can adversely affect child growth and development.

Evidence base

- Nutrition before and during pregnancy influences health outcomes for mothers and children across the life course.
- Healthcare professionals can be trained in Healthy Conversation Skills (HCS) to support behaviour change.

Resources:

- Research nurses at Princess Anne Maternity Hospital.
- Multidisciplinary research team including psychologists, nutritionists and public health experts.
- Funding.

Activities:

- Healthy Conversation Skills training workshops for research nurses, with ongoing support to embed skills.
- Regular SPRING meetings.
- Vitamin D supplementation.

Short-term outcomes:

- Research nurses demonstrate competence and confidence in Healthy Conversation Skills.
- Pregnant women set health-related behavioural goals and review progress with trained nurses.

Medium-term outcomes:

- Improved maternal vitamin D levels and infant body composition.
- Increased breastfeeding and self-efficacy.
- Improved maternal diet quality.
- Appropriate pregnancy weight gain.

- Nurses routinely use Healthy Conversation Skills to support goal-setting and behaviour change.

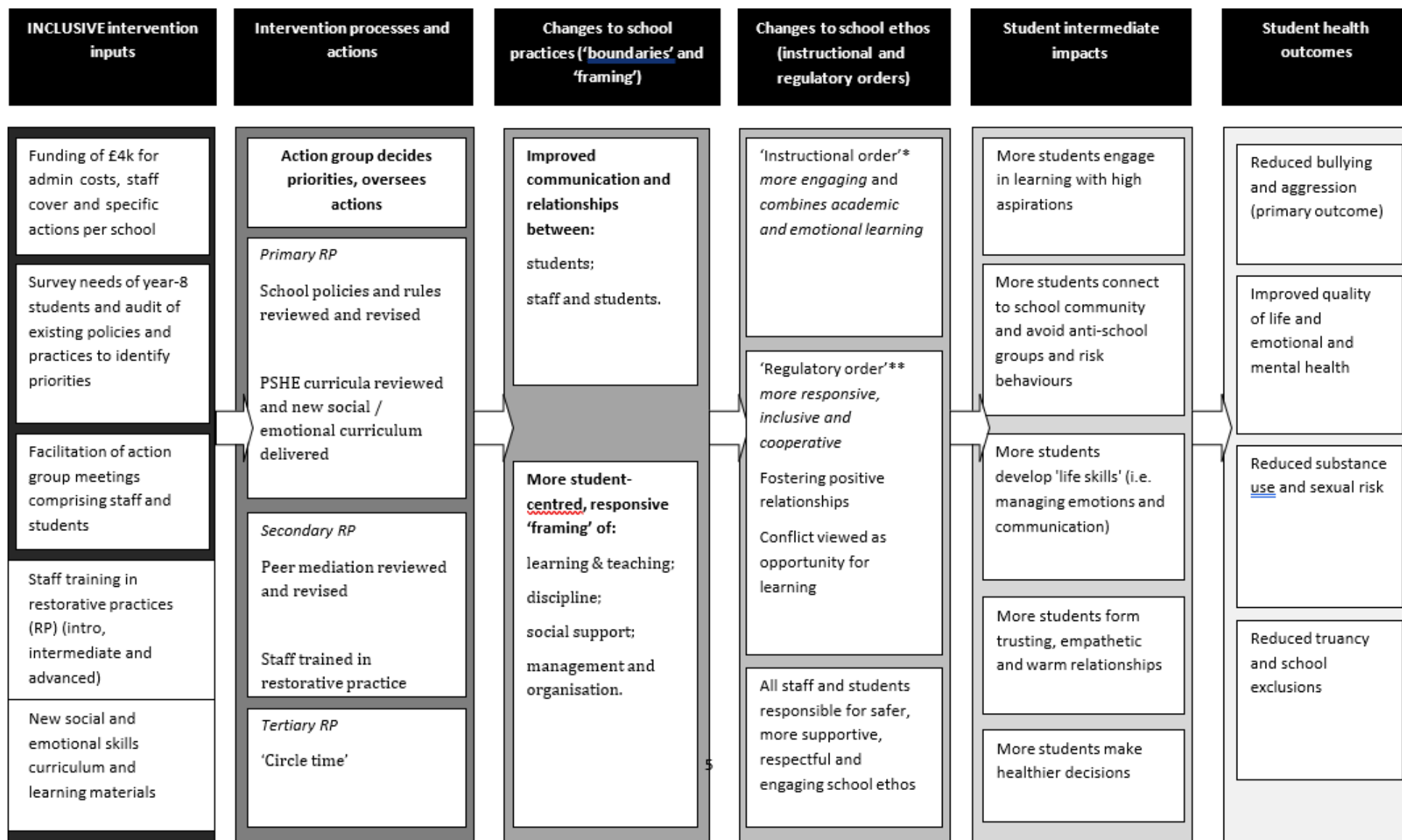
Long-term outcomes:

- Nurses continue to have healthy conversations about diet and physical activity.
- Women and children have improved diet quality and wellbeing.
- Improved health outcomes for women, families and subsequent children.

Hardeman and colleagues' (2005) causal modelling approach represents an extended form of logic model development. It has an explicit focus on statistical modelling, specifies methods to develop a causal model, and maps intervention content and measures onto causal pathways. A generic model links behavioural determinants causally to health outcomes and physiological and biochemical variables. This is developed into a model tailored to context, target population, behaviours and health outcomes. Causal pathways between intervention components, outcomes and process measures are mapped out and tested. This approach informed the development and evaluation of the ProActive intervention (Case Study 4).

The above examples draw predominantly upon psychological theory, which is useful for developing and evaluating interventions which work at the individual level. However, individual-level theorising becomes less useful for interventions which aim to improve health through intervening at other levels, such as the school or community (Hawe et al., 2009). The logic model for the pilot trial of the school-based INCLUSIVE intervention (Figure 5) drew predominantly upon sociological theory, focusing on system-level change (Bonell, in press). The intervention aimed to improve student health through promoting 'restorative practices' across the whole school, with the intervention comprising a set of standard structures and processes designed to trigger changes in organisational ethos and practice. Hence, full implementation related to the delivery of these key structures and processes, whereas the activities delivered to students as a result of organisational changes would vary between schools. The extent to which implementation of these structures and processes triggered changes in schools' practices and ethos are presented as key pathways linking intervention inputs to pupil health outcomes.

Figure 5. Logic model for INCLUSIVE intervention (Bonell et al. in press)



* i.e. learning and teaching in school ** i.e. discipline, social support and sense of community in school RP=Restorative Practice

Extended Alt Text for Figure 5 which presents a logic model describing the pathway through which an inclusive school intervention is expected to improve student health outcomes.

The intervention begins with inputs including funding, a needs assessment of Year 8 students, facilitation of staff-student action groups, restorative practices training, and social and emotional learning materials.

These inputs support intervention processes and actions. Action groups identify priorities and oversee activities. School policies are reviewed and revised, PSHE curricula are updated, peer mediation approaches are strengthened, staff are trained in restorative practices, and circle-time activities are delivered.

These actions lead to changes in school practices. Communication and relationships improve between students and between staff and students. Learning, discipline, social support and school organisation become more student-centred and responsive.

Changes in practices contribute to changes in school ethos. The school environment becomes more engaging, inclusive and cooperative. Positive relationships are fostered, conflict is regarded as an opportunity for learning, and responsibility for maintaining a safe and supportive environment is shared by staff and students.

Improved school ethos leads to intermediate impacts on students. More students engage with learning, connect positively with the school community, develop life skills, build trusting relationships and make healthier decisions.

These impacts contribute to improved student health outcomes, including reduced bullying and aggression, improved emotional wellbeing and quality of life, reduced substance use and sexual risk behaviours, and reduced truancy and school exclusions.

From theory to implementation

As described above, the principal concern of early process evaluation frameworks was capturing what was delivered in practice. Steckler and Linnan (2002) argue that capturing ‘implementation’ (conceived as a combination of fidelity, dose and reach) is central to avoiding dismissal of sound intervention theories due to a failure to implement them effectively (Type III error; Basch et al., 1985). This section briefly discusses variability in definitions of implementation and fidelity, before considering a number of frameworks which focus on implementation. The section also discusses the movement away from simply capturing what is delivered, towards understanding how implementation is achieved, and how interventions become part of the systems in which they are delivered. Finally, we provide an overview of key debates relating to the nature of standardisation and the extent to which complex interventions require adaptation in different contexts.

Definitions of ‘implementation’ and ‘fidelity’

Terms including ‘fidelity’, ‘adherence’, ‘integrity’ and ‘implementation’ are often used interchangeably to describe the extent to which an intervention is delivered as intended. However, multiple definitions appear in the literature. Moncher and Prinz (1991) define fidelity, in the context of outcomes research, as the extent to which the independent variable was manipulated as planned. Subsequent definitions also included participant engagement with the intervention, and the extent to which participants translated skills learned in a behavioural intervention into everyday life (Lichstein et al., 1994). Within their process evaluation framework, Steckler and Linnan (2002) define fidelity as the quality of delivery, emphasising the need to capture the qualitative nature (‘spirit’) of what was delivered, not just the technical aspects of delivery. Further definitions of fidelity offered by Bellg and colleagues (2004) include consideration of intervention design, training, delivery, receipt and enactment. A number of additional frameworks which focus on the quality of implementation are now discussed. Some of these emerged before, or in parallel to, early process evaluation frameworks, while others are more recent additions to this literature.

Donabedian’s structure-process-outcome model

While less commonly used in public health, Donabedian’s structure-process-outcome model has been widely used for assessing the quality of healthcare. Efforts to assess healthcare quality often focus on instrumental outcomes; for example, a low mortality rate in a hospital may be taken to indicate good quality care. However, Donabedian (1988) argues that

outcomes are a flawed measure of quality of care because they are a function of the power of a particular course of action to produce outcomes, and the extent to which that action was applied properly. Although healthcare professionals act with a degree of agency, whether a course of action can be applied properly will also depend on the extent to which structural aspects of the setting facilitate effective care. Hence, attention is focused on how relationships between context and actions of patients and care providers are associated with outcomes. The model also recognises that understanding implementation requires focus not only on what was delivered, but on the mechanisms through which it was delivered in context. As will be discussed, this is an increasing theme in more recent definitions of fidelity and implementation.

The RE-AIM framework

The RE-AIM framework (Glasgow et al., 2001; Glasgow et al., 1999) largely reflects frustrations with evaluators' tendency to focus on efficacy without considering how findings might translate into 'real world' impact. The RE-AIM framework presents public health impact as a function of an intervention's reach (proportion of the target population that participated in the intervention), effectiveness (success rate), adoption (proportion of eligible settings that adopt the intervention), implementation (extent to which the intervention is implemented as intended) and maintenance (extent to which the intervention is maintained over time). The language of the framework somewhat privileges quantification, with a 'public health impact score' conceived as the product of the five dimensions (Glasgow et al., 2006). It focuses attention on capturing the extent to which dimensions including implementation are achieved in the short or longer term, though less on the processes through which this occurs.

Carroll and colleagues' conceptual framework for fidelity

Writing from a health service research perspective, Carroll et al. (2007) propose a conceptual framework which defines fidelity (or adherence) as a combination of content; frequency and duration of delivery; and coverage. Though using different terminology, this definition is almost indistinguishable from what Steckler and Linnan defined as 'implementation' (a combination of fidelity, dose and reach). However, Carroll and colleagues' framework goes further in that it invites researchers to examine moderators of implementation, such as how effectively resources are applied. The authors highlight the need to consider issues such as intervention complexity, comprehensiveness of the intervention's description, effectiveness of strategies to facilitate implementation, and how participant response to intervention

moderates delivery. A modification by Hasson (2010) adds context and recruitment as moderators of implementation.

The Oxford Implementation Index

Another recent contribution to this field is the Oxford Implementation Index, which provides systematic reviewers with guidance in extracting data on implementation from primary studies (Montgomery et al., 2013b). The index provides a checklist focused on four domains: intervention design (e.g. whether core components are clearly specified); delivery by practitioners (e.g. staff qualifications, the quality and use of materials, dosage administered); uptake by participants; and contextual factors. Similar to Carroll's (2007) approach, and Bellg and colleagues' (2004) definitions of fidelity, the index includes consideration of intervention design, training, delivery, receipt and enactment. The delivery aspect of this index emphasises consideration of resources, structures and processes needed to achieve successful implementation, rather than just description of what is delivered.

Organisational change and the implementation process

As noted above, process evaluations and fidelity frameworks are increasingly concerned with not only whether an intervention is implemented correctly during the evaluation period, but the mechanisms through which implementation is achieved, and how this informs efforts to incorporate the intervention into routine practice after the evaluation. Translation of 'efficacious' behaviour change approaches into everyday practice has proved difficult (Moore et al., 2011). The MRC recommends that this integration, and any associated issues, is considered throughout the development and evaluation process (Craig et al., 2008a). Terms such as 'sustainability' (Steckler & Linnan, 2002; Baranowski & Stables, 2000) and 'maintenance' (Glasgow et al., 2001; Glasgow et al., 1999) have been used within frameworks to describe the potential for an intervention to become part of routine practice. Arguing from a systems perspective, Hawe and colleagues (2009) describe interventions as events within systems, which either leave a lasting footprint or wash out, depending how well system dynamics are harnessed. Theories from sociology and social psychology, such as diffusion of innovations theory (Greenhalgh et al., 2004; Rogers, 2003) and normalisation process theory (NPT; May & Finch, 2009) also emphasise the processes through which interventions become a fully integrated part of their setting, using the terms 'routinisation' or 'normalisation' respectively to describe these.

Building upon NPT, and drawing upon sociological and psychological theories including the theory of planned behaviour, social cognitive theory and diffusion of innovations theory, May (2013) proposes an integrated general theory of implementation. Successful implementation is seen as a result of the actions of agents, which in turn are shaped by capacity (social-structural resources available to agents), potential (social-cognitive resources available to agents) and capability (possibilities presented by the intervention). May argues that complex interventions are likely to become part of routine practice if: elements of the intervention can be made workable and integrated into everyday life; the social system provides the normative and relational capacity for implementers to cooperate and coordinate their actions; agents individually and collectively commit to the intervention; and agents' contributions to the intervention carry forward in time and space.

The fidelity and adaptation debate

All the above frameworks emphasise the need to understand what is implemented, and how, if evaluators are to understand how an intervention works. Frameworks which focus on the quality and quantity of implementation assume that the intervention must have certain standardised features in all settings. In support of this view, a review by Dane and Schneider (1998) found that while fidelity was rarely measured, primary and secondary prevention programmes which deviated furthest from protocols achieved the poorest outcomes. There are, however, significant unresolved debates regarding how best to conceive fidelity, and the extent to which adaptation across contexts is acceptable, or indeed necessary.

Drawing on evidence of links between fidelity and strength of effects, some argue that when we know little about which intervention components are 'active ingredients', allowing adaptations to take place might inhibit effectiveness (Mihalic, 2004). Hawe and colleagues (2004), however, argue that too much attention has been paid to rigidly standardising the form (e.g. content and modes of delivery) of interventions, and allowing this to change across contexts may ensure greater fidelity to its intended functions. An information campaign, for example, may achieve more consistent effects across settings if tailored to local literacy levels. While interventions are often defined in terms of a set of activities delivered to a target population, in some cases the core components of an intervention are a set of structures and processes intended to facilitate change at an organisational level. A process evaluation of such an intervention would need to look at the extent to which these structures and processes are standardised. The activities which follow would be expected to serve similar functions but

their form may differ across settings. Assessing whether changes triggered by the intervention were consistent with intended functions becomes central to understanding mechanisms of impact.

Durlak and DuPre (2008) argue for a compromise position which asserts that assessments of fidelity should focus on core intervention activities, while less central components can be altered to achieve ecological fit. This position is reflected within guidance from the US Centres for Disease Control (CDC) on the delivery of sexual health interventions. The CDC use a traffic light system to categorise adaptations as ‘red light’ (adaptations that compromise the functioning of the intervention), ‘yellow light’ (changes that should be made with caution, in consultation with experts on the theory of the intervention) and ‘green light’ (safe adaptations to allow better fit) (Firpo-Triplett, 2012). However, in practice, uncertainty regarding the active ingredients of an intervention, and how these interact to produce change, may be the reason for conducting an evaluation. Some might also question why one would include components which are not expected to affect outcomes. Mars and colleagues (2013), for example, conducted fidelity assessments of a course on self-management of musculoskeletal disorders, focusing on seven of 24 components considered most likely to affect change. If the remaining 17 were considered unlikely to contribute to the functioning of the intervention as a whole, this raises the question of why they were included. Some would also argue that attempting to separate out core components risks excessive atomisation; the effects of a complex intervention are not simply a sum of the effects of its individual components, but arise from the synergy between them.

Bumbarger and Perkins (2008) argue that rather than seeing fidelity and adaptation as opposites, evaluators need to distinguish between ‘innovation’ (skilled implementers actively attempting to make an intervention better fit their population or setting) and ‘drift’ (unintentional shortcomings, arising from barriers to full implementation). Consistent with this position, Mars and colleagues (2013) argue that within a self-management intervention for musculoskeletal disorders, precise adherence to intervention manuals often reflected a mechanistic, inflexible or unresponsive delivery style. Skilful implementers deviated from instructions in response to feedback from participants, while remaining consistent with the theoretical basis of the intervention. Some deviations from protocols, however, were adjudged to represent poor implementation. To Bumbarger’s innovation/drift dichotomy, one

could add a third category of ‘subversion’, where implementers actively choose not to adopt aspects which conflict with their values or theories of change. Stirman and colleagues (2013) have recently published a framework for capturing and categorising changes to interventions as they are adopted in new settings.

The MRC framework (Craig et al., 2008a; Craig et al., 2008b) recognises some of these challenges, identifying that one aspect of complexity is the degree of flexibility or tailoring permitted. Some have questioned the implicit suggestion that evaluators have control over whether implementers are ‘permitted’ to adapt a programme (Pawson, 2013). In reality, pragmatic evaluators must accept that interventions will be adapted as they move into new contexts. Evaluators need to monitor these adaptations, and attempt to understand why they occurred and how they may influence the functioning of the intervention. Discussions of how to distinguish between programme tailoring and poor fidelity are unlikely to be fully empirically resolved in the near future. It may never be possible to fully understand how variations in delivery affect outcomes, given that adaptations do not occur at random, and will be confounded by factors promoting or inhibiting intervention effects. A strong understanding of the theory of the intervention is a prerequisite for meaningful assessment of implementation, focused not just on the mechanics of delivery, but whether intervention remained consistent with its underlying theory.

How does the intervention work? Theory-driven evaluation approaches

As described, earlier frameworks such as Steckler and Linnan’s focused attention predominantly upon implementation, placing less emphasis on theory development. As described in Chapter 1, if only aggregate outcomes are presented, all we can know is whether an intervention package did more good than harm, in terms of pre-specified outcomes, in a specific context. Complementing this with implementation assessment provides a clearer picture of what caused these effects. However, this does not necessarily illuminate how, or for whom, the intervention worked. If ‘effective’, does it have to be reproduced in full, or are there structures, processes or activities which can be omitted? Were these effects achieved by the mechanisms hypothesised? If effects were limited, can we identify weak or absent causal pathways responsible for this? More recent efforts have been made to advocate use of process evaluation to test and develop intervention theories (Grant et al., 2013a). This section will discuss a number of influential theory-driven evaluation approaches.

Theory-based evaluation and realist(ic) evaluation

Two influential theory driven approaches to evaluation are theory-based evaluation (Weiss, 1997) and realistic evaluation (Pawson and Tilley, 1997). Theory-based evaluation (TBE) aims to examine how hypothesised causal chains play out in practice. Proponents of theory-based evaluation argue that this allows information to be gathered about the stages at which the causal chain might break down (Weiss, 1997). TBE may focus on ‘intervention theory’ (the mechanisms through which intervention activities produce change), ‘implementation theory’ (how successful implementation is achieved) or a combination of the two.

Realistic (or realist) evaluation emerged from criminology but is increasingly influential in other domains, including public health (Pawson and Tilley, 1997). Like TBE, it places change mechanisms at the heart of evaluation. However, it emphasises the contextually contingent nature of these mechanisms. Rooted in critical realist philosophy, it views interventions as ‘working’ by introducing mechanisms that are sufficiently suited to their context to produce change. Hence, evaluation aims to uncover context-mechanism-outcome configurations, in order to understand ‘what works, for whom, under what circumstances’. Such approaches may be useful in understanding how intended outcomes are achieved, and how unanticipated consequences emerge.

The adoption of realistic evaluation approaches has been limited within evaluative research by its tendency to be positioned in opposition to experimental methods. Given its explicit emphasis on broader theory development, it has been viewed by some as distinct from more inward-looking process evaluations, which are perceived as attempting to explain the outcomes of a specific intervention. Realistic evaluation is not so much a method as a philosophy for evaluation, and many evaluations that do not identify themselves as realist evaluations cite Pawson and Tilley as key influences (e.g. Moore et al., 2013). The aims of process evaluation as defined by the MRC framework include enabling evaluation to inform the development of effective interventions, through understanding their mechanisms and contextual contingencies. Hence, movement towards explaining mechanisms of impact and contextual contingencies appear to have permeated mainstream thinking.

In rejecting randomisation, and in common with other theory-driven methods, realistic evaluation is limited in its ability to disentangle events observed from what would have happened anyway. However, recent movement towards reconciliation (Blackwood et al., 2010) includes studies which combine realistic evaluation and RCT methodology (see for

example Byng et al., 2005). ‘Realist trials’ attempt to capture overall effects, and consider which intervention activities work, for whom, and under what circumstances, whilst also developing and validating theory (Bonell et al., 2012).

Causal modelling: testing mediation and moderation

In proposing the concept of realist trials, Bonell and colleagues (2012) argue that process evaluations often generate theory regarding intervention mechanisms, but do not quantitatively test it. The causal assumptions underpinning the intervention can be tested through mediation analysis (to examine mechanisms) or analysis of moderation (to examine contextual contingencies).

Incorporating mediation into an intervention’s theory of change means extending causal assumptions from ‘if intervention X is implemented, Y will occur’ to ‘if X is implemented, this will lead to change in the mediating variable, which will in turn lead to change in outcome Y’ (Baron & Kenny, 1986). For example, Gardner and colleagues (2010) examined mediation of the impacts of a parenting intervention, testing two primary causal pathways. It was hypothesised that impacts would be explained by improvements in positive parenting practices, reductions in negative parenting practices, or both. Analyses indicated that impacts on children’s behaviour were mediated by improvements in positive parenting practices, but not by reductions in harsh or negative parenting. While this involves breaking randomisation for the second phase of the causal chain (in this case, comparisons of how parenting practice predicted impacts were not themselves randomised), mediation analysis can offer valuable insights into how an intervention produces impacts.

Testing moderation involves examining which (if any) pre-existing characteristics predict who benefits from an intervention. For example, one may hypothesise that socioeconomic conditions moderate how participants interact with an intervention, and hence its effectiveness. Subgroup analysis and use of interaction terms within regression models could be used to test hypotheses of moderation. Gardner and colleagues (2010), for example, found that children from disadvantaged backgrounds often fared better following a parenting programme than those from more advantaged backgrounds. Provided that the moderator is evenly distributed between intervention and control conditions, and that dropout from the trial is not predicted by the moderator, such analyses can be conducted without compromising the balance between intervention and control groups achieved by randomisation.

Sub-group analyses are often seen by purists as statistically unsound (Petticrew et al., 2011), with trials powered on the basis of aggregate effects and subgroup analyses underpowered. However, reporting interaction effects can allow trends to be identified which, although not statistically significant in individual studies, point to meaningful differences if replicated across studies. For example, in Wales, Scotland and Northern Ireland, smokefree legislation was associated with slight increases in inequalities in second-hand smoke exposure. While non-significant in individual studies, odds ratios were virtually identical in all countries, reaching significance in pooled analysis (Moore et al., 2012a).

Complexity science perspectives

Complexity theorists argue that the characteristics distinguishing a ‘complex’ intervention from one that is merely ‘complicated’ include unpredictability, emergence (complex patterns of behaviour arising out of a combination of relatively simple interactions), and non-linearity of outcomes (Keshavarz et al., 2010). In distinguishing between a ‘complicated’ and a ‘complex’ problem, Glouberman and Zimmerman (2002) use the examples of sending a rocket to the moon (complicated) and raising a child (complex). Although the former requires substantial expertise, it can be achieved through expertly following formulae, with one successful attempt providing a high degree of certainty that it can be done again. In raising a child, however, formulae have little place, and raising one child provides limited assurance of future success.

Complex interventions attempt to change the dynamics of social systems, through influencing the behaviours of agents within those systems (Hawe et al., 2009). Outcomes may not occur in a linear manner, but may build or diminish over time as feedback loops occur. For example, legislation against using mobile phones in cars led to marked reductions in use in the short term, but limited enforcement led to a return to pre-legislation levels of use (Pawson et al., 2010). Complexity theorists also highlight the importance of ‘tipping points’ in understanding health behaviour. A person may have tried several times to quit smoking or lose weight, and then suddenly make a successful attempt as the result of a single event, or when ‘chunks of knowledge or attitude randomly coalesce to form a perfect motivational storm’ (Resnicow & Vaughan, 2006; p9). Interventions may therefore be delivered at the ‘right’ moment to trigger change, whereas delivery to the same person at a different time would have had no impact. Complexity theorists would also argue that complex interventions are often treated in a somewhat atomistic manner, being broken down by evaluators into their

constituent parts. As described in Chapter 1, complex interventions are by definition intended to be greater than the sum of their parts, with multiple components acting in synergy to produce change. Hence, attempts to understand parts of the intervention should always be considered in relation to the functioning of the intervention as a whole.

A criticism of the MRC framework (Craig et al., 2008a; Craig et al., 2008b) is that while using the term ‘complex’, it did not engage with complexity science. The authors argue that this is because the contributions of complexity science to evaluation remain on a theoretical level, and there are few empirical examples for it to inform guidance. Process evaluation may offer a means of providing some of these empirical examples by, for example, using qualitative data to capture feedback loops and investigate complex causal pathways.

Summary of key points

This chapter has drawn together a range of process evaluation frameworks, and related theories and frameworks whose aims overlap with those of process evaluation.

- A number of attempts have been made to provide frameworks for process evaluation, typically focusing on quantification of implementation, receipt and context.
- A broad range of additional theories and frameworks, which do not use the term process evaluation, provide insights into how we might understand implementation, mechanisms of impact, and context.
- Recent frameworks focusing on fidelity and implementation increasingly advocate understanding *how* implementation in context is achieved, as well as *what* is delivered.
- Theoretical tensions surrounding the nature of fidelity, and the extent to which allowing interventions to adapt to settings represents poor fidelity or beneficial local tailoring, remain unresolved.
- Theory-driven approaches to evaluation, such as realist evaluation, are of substantial interest to process evaluators aiming to understand how complex interventions work.

Chapter 3 will now relate the theories and frameworks described above to the three core themes for process evaluation described in Chapter 2.

Using frameworks and theories to inform a process evaluation

The frameworks and theories described in Chapter 2 place varying emphasis on the three components of process evaluation described in Chapter 1: implementation, mechanisms of impact, and context. A brief summary will now be given of how the frameworks described link to key aims of process evaluation.

Implementation

How is delivery achieved?

As described in Chapter 2, implementation frameworks increasingly highlight the need to understand how implementation is achieved, as well as what is delivered. The Oxford Implementation Index (Montgomery et al., 2013), and Carroll and colleagues' (2007) fidelity framework, focus on assessing whether the structures and resources in place are adequate to achieve successful implementation. More in-depth engagement with issues surrounding implementation process and how new interventions are integrated into their settings can be found in work on systems thinking (Hawe et al., 2009), diffusion of innovations (Rogers, 2003), normalisation process (May et al., 2009), or May's (2013) more recent general theory of implementation.

What is actually delivered?

Key dimensions of implementation within most of the frameworks discussed above include fidelity (the quality of what is delivered) and dose (the quantity of what is delivered). Some definitions of implementation incorporate 'reach' or 'coverage', (the extent to which the target audience come into contact with the intervention (Carroll et al., 2007; Steckler & Linnan, 2002)), while others see it as a separate dimension (Glasgow et al., 1999).

Distinguishing the actions of implementers from those of participants can be somewhat artificial; if an intervention is not delivered, it cannot reach participants; conversely, an intervention can rarely be implemented if no-one participates. The extent to which reach can be considered a necessary aspect of full implementation depends on how strongly it is linked to effectiveness. For example, an exercise referral scheme cannot be implemented if no one attends, though a national policy on minimum alcohol pricing may impact on people who are not aware that it even exists. Understanding emerging adaptations to the intended intervention involves exploring whether these improve its contextual fit or compromise its functioning (Hawe et al., 2004), or in Bumbarger and Perkins' (2008) terms, whether they represent innovation, or intervention drift. This may be best achieved through qualitative

methods, and requires a comprehensive understanding of intervention theory. As discussed in Chapter 2, there are substantial unresolved empirical tensions relating to the nature of fidelity and the need for adaptation; process evaluators should consider the theories and frameworks discussed above and justify the approach they take to understanding these issues.

Mechanisms of impact: how does the delivered intervention work?

Understanding how participants respond to, and interact with, complex interventions

Whether an intervention works or not will depend on how its intended audience responds to it. Within Steckler and Linnan's (2002) framework, participant responses to an intervention are largely discussed in terms of 'dose received'. Advocates of approaches such as realistic evaluation would reject the passive tone of this term; participants do not passively receive interventions, but exercise agency in interacting with them. The term 'dose' also implies a privileging of quantitative measurements, which is potentially problematic. Many process evaluations examine participant responses in terms of quantitative measures of 'acceptability' or 'satisfaction'. However, satisfaction should not be uncritically assessed, but examined with reference to its relationship to the mechanisms through which the intervention works. For example, in the case of alcohol or cigarette pricing policy, evaluators might anticipate consequences such as concentrations of black markets in cheap alcohol or cigarettes in areas where dissatisfaction is widespread. More probing qualitative methods may illuminate negative experiences, or provide in-depth understandings of how participants' interactions with interventions produce change. Grant and colleagues (2013a) emphasise the need for qualitative exploration of participant responses, during and after intervention, to understand how change is produced in the short and longer term.

Testing mediators and generating intervention theory

Key approaches that pay attention to understanding the mechanisms through which interventions work include theory-based evaluation and realistic evaluation. Neither is prescriptive about the use of quantitative or qualitative methods. Indeed, as acknowledged by recent work on 'realist trials' (Bonell et al., 2012), combining quantitative testing of hypothesised mediators and moderators with more inductive qualitative methods allows evaluators to test existing theory and generate new theory for future testing. Key causal assumptions may be quantitatively tested, while qualitative data may lead to refinement of the logic model. Additional mechanisms may be identified, which were too complex to be captured via quantitative methods or might not have been anticipated.

Identifying unintended and/or unanticipated consequences

In their recent framework for process evaluations in cluster RCTs, Grant and colleagues (2013a) acknowledge the need for process evaluation to capture unintended consequences, an emphasis missing from earlier frameworks. Within drug trials, it is common practice to measure potential side-effects. In evaluating complex interventions, some potential unintended consequences may be anticipated, and measures put in place to capture them quantitatively. Grant and colleagues (2013a) recommend that unintended consequences, including those that are unanticipated (which are likely to occur in a complex intervention) be captured via a range of methods, including qualitative analysis of observational and interview data, quantitative data collection during the trial, and the use of routine datasets.

Contextual factors

In frameworks such as Steckler and Linnan's (2002), context is considered primarily in terms of pre-existing conditions that may facilitate or impede implementation fidelity. An intervention may be delivered poorly in some contexts but well in others, because of better fit with some settings or target populations. In-depth engagement with the processes through which new interventions are integrated into their contexts can be achieved through engagement with literature on systems thinking, diffusion of innovations, or normalisation process. Realistic evaluation (Pawson & Tilley, 1997) moves beyond viewing contextual factors solely as moderators of implementation, towards also viewing them as moderating outcomes, meaning the same intervention may produce different outcomes in different contexts (Weiss et al., 2013). Participants are seen as agents, whose pre-existing circumstances, attitudes and beliefs will shape how they interact with the intervention. Hence, the aim of evaluation is to identify context-mechanism-outcome configurations, and to explain variability in intervention outcomes.

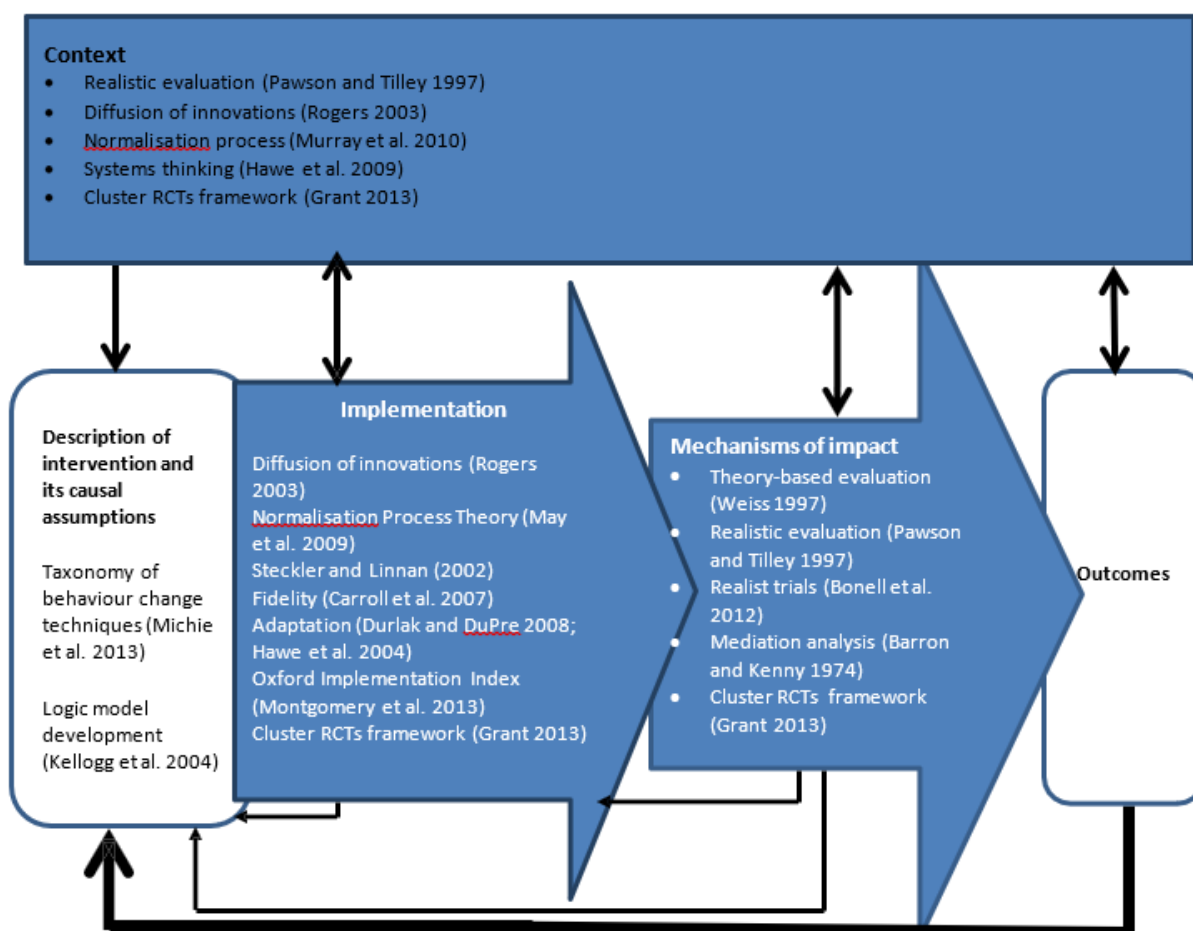


Figure 6. Examples of key frameworks for process evaluation and their relationship to each core function of process evaluation

Extended Alt Text for Figure 6 - this presents a conceptual framework for intervention development, implementation and evaluation.

Context sits above all components and influences intervention design, implementation, mechanisms of impact and outcomes. Context is informed by several theoretical perspectives including Realistic Evaluation, Diffusion of Innovations, Normalisation Process Theory, Systems Thinking and Cluster Randomised Controlled Trial frameworks.

The intervention begins with a description of the intervention and its causal assumptions. This stage draws on the Behaviour Change Techniques Taxonomy and logic model development methods.

The implementation stage focuses on how the intervention is delivered and evaluated. Relevant concepts include Diffusion of Innovations, Normalisation Process Theory, fidelity, adaptation, the Oxford Implementation Index and Cluster Randomised Controlled Trial frameworks.

Mechanisms of impact describe how and why the intervention produces outcomes. Approaches used to investigate mechanisms include theory-based evaluation, Realistic Evaluation, realist trials, mediation analysis and Cluster Randomised Controlled Trial frameworks.

Outcomes represent the intended and unintended effects of the intervention. The framework includes feedback loops, indicating that outcomes and mechanisms of impact can inform future implementation and intervention development. Context continues to influence all stages throughout the intervention lifecycle.

Summary of key points

As illustrated throughout this chapter, a broad range of frameworks and theories may be drawn upon in developing a process evaluation which serves the functions set out in MRC guidance. Examples of key frameworks relating to each aspect of process evaluation as defined in this document are presented in Figure 6 above. Section B will now provide practical guidance in how to design and conduct a process evaluation.

SECTION B - PROCESS EVALUATION PRACTICE

How to plan, design, conduct and analyse a process evaluation

This chapter provides guidance on how to plan, design and conduct a process evaluation. It does not provide a rigidly defined checklist; the diversity of the interventions evaluated by health researchers, and the uncertainties posed by them, means that not all process evaluations will look the same. However, it offers guidance in thinking through some of the common decisions that will need to be addressed when developing a process evaluation. The chapter begins by discussing issues to consider in planning a process evaluation, before considering questions of design and conduct. This should not be taken to indicate a linear process; given the unpredictability of the issues process evaluations will aim to investigate, flexible and iterative approaches to planning and execution are crucial.

Some potential pitfalls in planning and conducting a process evaluation are presented below. It would be a challenge to find an example of a process evaluation which has not fallen foul of at least some of these (all of our own case studies did). This chapter aims to provide the reader with insights into how to avoid or minimise them.

Planning and conducting a process evaluation: what can go wrong?

- Poor relationships with stakeholders limit the ability of the evaluator(s) to closely observe the intervention, or overly close relationships bias observations.
- Poor team working between quantitative and qualitative methodologists (or between outcomes and process evaluators) leads to parallel studies, which fail to sufficiently add value to one another.
- Employing an inexperienced member of staff to lead the process evaluation, with insufficient support from a team with expertise in quantitative and qualitative methods and social science theory, undermines the quality of the process evaluation.
- Absence of a clear description of the intervention and its underlying causal assumptions leads to a process evaluation which is not focused on the key uncertainties surrounding the intervention.
- Poor definition of research questions, and a lack of clarity over why certain data are being collected, leads to collection of too much data, some of which is not analysed.
- Over-reliance on a small number of case studies leads to a poor understanding of the intervention as a whole.
- Collection of more data than can be analysed, wastes effort and goodwill.
- Asking insufficiently probing questions about experiences of the intervention leads to superficial or false conclusions that everything is working as intended.

- An overly intensive process evaluation blurs the boundaries between the evaluation and the intervention, changing how it is implemented.
- Insufficient time is allowed for a thorough analysis of qualitative data.
- Qualitative data are used simply to illustrate quantitative data, leading to biased and superficial qualitative analysis.

Planning a process evaluation

Working with programme developers and implementers

Achieving a good quality evaluation is almost impossible without good working relationships with stakeholders involved in developing or delivering the intervention. While a wholly detached position is arguably untenable in any form of evaluation, this is particularly true of process evaluations, which aim to understand the inner workings of interventions.

Relationships between evaluators, and policy and practice stakeholders whose work the process evaluation aims to inform, are not always straightforward. Potential influences of these relationships on the research process, and indeed on the intervention, should be acknowledged.

Evaluation may involve becoming a critical observer of the work of those who developed or delivered the intervention. As reflected in SHARE (Sexual Health And Relationships, Case Study 3), evaluation is understandably often seen as threatening. Stakeholders may be invested in the intervention personally and professionally. For some, job security may depend on continuation of the intervention beyond the evaluation period. Researchers may have contributed significantly to intervention development, and may have an interest in showing it to work. They may in such circumstances be overly critical of practitioners who ‘fail’ to deliver the intervention than would a researcher less invested in the intervention. In some instances, stakeholders who developed the intervention may fund its evaluation, retaining some contractual control or other influence over aspects of the evaluation such as publication of findings.

Conflicts of interest may emerge if those with a vested interest in portraying an intervention positively exert too much influence on its evaluation. Sustaining good working relationships, while remaining sufficiently independent for evaluation to remain credible, is a challenge evaluators must take seriously. Ensuring process evaluation is understood as a means of allowing evaluation to inform efforts to improve interventions, rather than a pass or fail assessment, may alleviate some of these tensions. Agreeing the parameters of these

relationships early on may prevent problems later, and transparency about the relationship between the evaluation and the intervention is critical (Audrey et al., 2006). It is important to remain reflexive, and continuously question whether good or bad relationships between researchers and other stakeholders are leading to an overly positive or negative assessment of the intervention. It may be useful to seek occasional critical peer review by a more detached researcher with less investment in the project, who may be better placed to identify where researcher position has compromised the research.

Communication of emerging findings between evaluators and implementers

Another key aspect of the relationship between the evaluation and the intervention relates to structures for communication between stakeholders during the evaluation. Evaluators may learn of ‘incorrect’ implementation practices, or contextual challenges, which they feel should be immediately communicated to those responsible for implementing or overseeing the intervention. Here, evaluators are faced with a choice: to remain passive observers, or to play an active role in addressing ‘problems’. In a process evaluation at the stage of feasibility and piloting, which aims to test the feasibility of the intervention and its intended evaluation, the latter approach is appropriate. Arguably, in an evaluation which aims to establish effectiveness under real world conditions, it may be appropriate to assume a more passive role to avoid interfering with implementation and changing how the intervention is delivered.

There are notable exceptions; for example, where there are ethical implications in withholding information on harms. It might also be acceptable for evaluators to have a relatively high degree of influence on implementation if the structures and processes through which this is achieved can be captured and replicated should the intervention be scaled-up. For example, process evaluations may use monitoring and feedback systems which would form part of a fully scaled-up intervention. It may be that a specific role is created to enhance engagement between researchers and intervention stakeholders, and that the functions of this role in shaping implementation are carefully captured and replicated in the scaled-up intervention.

Whichever model is adopted, systems for communicating process information to key stakeholders should be agreed at the outset of the study, to avoid perceptions of undue interference or that vital information was withheld. Evaluators will need to consider carefully how, and to what extent, their engagement with implementers shapes how the intervention is delivered. Where feedback leads to changes in implementation, the impacts of these changes

on the intervention and its effectiveness should be considered. In the process evaluation of NERS, for example (Case Study 5), feedback on poor delivery of motivational interviewing triggered the inclusion of additional training. Impacts of training on practice became the focus of an emerging sub-study. In SIH (Case Study 2), impacts of training on staff practice was the main focus of the process evaluation. The logic model specified change in practice as a necessary step toward change in outcomes in women and children receiving support. Interim analysis assessed change in staff practice, with results fed back to the evaluation team.

Key considerations in working with policy and practice stakeholders to plan a process evaluation

When will process evaluation findings be communicated to policy / practice stakeholders (e.g. during the evaluation, or only at the end)?

Have structures for feedback been agreed among stakeholders?

Where feedback during a trial leads to changes in implementation, how will you capture these changes and their impact on effectiveness?

Are there structures in place to capture the influences of the evaluation on the intervention, and plans made for these processes (e.g. monitoring and feedback structures) to be included in the scaled-up intervention?

Will those involved in designing or implementing the intervention provide or collect data for the process evaluation?

Intervention staff as data collectors: overlapping roles of the intervention and evaluation

In some cases, the most efficient means of gathering data from an intervention across multiple settings may be to ask implementers to assist with data collections. This can bring substantial challenges. For example, ProActive (Case Study 4) and NERS (Case Study 5) both requested that implementers provide recordings of consultations. In both cases, substantial data were lost due to issues such as equipment failure or incomplete paperwork. In NERS, these difficulties were reduced in follow-up collections by clarifying data collection instructions, correcting any errors in paperwork at the earliest possible stage, ensuring that

data collection instructions were easy to follow, and minimising research burden on busy implementers.

Audrey and colleagues (2006) describe the challenge of overlap in the roles of evaluation and intervention in relation to ASSIST (A Stop Smoking in Schools Trial, Case Study 1). Health promotion trainers who designed and implemented ASSIST provided and collected process data, completing evaluations about the intervention and young people's responses. Attempts were made to minimise reporting bias by involving trainers in discussion about the aims of the research and the best ways to achieve these. This emphasised that performance of individual trainers was not being assessed, but that data were being sought about how the intervention might operate in the 'real world'. Post-intervention interviews with trainers revealed willingness to discuss shortcomings and suggest improvements; these suggested changes, in relation to the original 'training the trainers' event and schools-based follow-up visits, were incorporated into manuals for wider implementation of the intervention.

Relationships within evaluation teams: process evaluation and other evaluation components

Process evaluations most commonly form part of a package which includes outcomes and/or cost-effectiveness evaluation. This is likely to involve individuals from a diverse range of disciplinary and methodological backgrounds, and may be affected by status issues common to mixed-methods research. Conducting a process evaluation within a randomised trial, for example, may involve working with a clinical trials unit, where rigid policies and procedures conflict with process evaluators' desire to respond flexibly to emerging findings.

Within community randomised trials, tensions between outcomes evaluators and qualitative researchers may arise where, for example, qualitative data highlight poor implementation or potential harms, but are dismissed as insufficient grounds for changing course (Riley et al., 2005). O'Cathain and colleagues (2008a) characterise mixed-methods teams as multidisciplinary (parallel rather than fully integrated), interdisciplinary (different disciplines engaging in all aspects of the research and sharing viewpoints and interpretations) or dysfunctional (each methodological group fails to see the value of the others' work). The authors describe integration as more common where team members respect and see value in one another's work, and where the study is overseen by a principal investigator who values integration of methods.

Evaluation teams involve differing degrees of integration. Some separate process and outcomes evaluation teams; in others, process evaluators and outcomes evaluators are the same people. While the case can be made for either model, the relationships between the components of an evaluation, and the roles of the researchers, must be defined at the planning stage. Some key considerations in deciding the level of integration between outcomes and process evaluations are described below. Where allocated to separate teams, effective oversight of the evaluation as a whole, and communications between teams, must be maintained to prevent duplication or conflict. Where process and outcomes evaluation are conducted by the same individuals, there is a need for openness and reflexivity about how this might influence the conduct and interpretation of the evaluation.

Arguments for separation between outcomes / process evaluation teams include:

- Separation may reduce potential biases in analysis of outcomes data, which could arise from feedback on the functioning of the intervention.
- Where a controlled trial is taking place, process evaluators cannot be blinded to treatment condition. Those collecting or analysing outcomes data ought to be, where possible.
- Some (e.g. Oakley et al., 2006) argue that process data should be fully analysed without knowledge of trial outcomes, to prevent fishing for explanations and biasing interpretations. While it may not always be practical to delay outcomes analysis until process analyses are complete, if separate researchers are responsible for each, it may be possible for these to be conducted concurrently.
- Process evaluation may produce data which would be hard for those who have invested in the trial to analyse and report dispassionately.
- Where there are concerns about a trial among implementers or participants, it may be easier for process evaluators to build rapport with participants and understand their concerns if they have a degree of separation from the trial.

Arguments for integration of process and outcomes evaluation:

- Process evaluators and outcomes evaluators will want to work together to ensure that data on implementation can be integrated into analysis of outcomes.
- Data collection of intermediate outcomes and causal processes identified by process evaluators may be integrated into collection of outcomes data.
- Some relevant process measures may already be collected as part of the outcomes evaluation, such as data on participant characteristics and reach. It is important to avoid duplication of efforts and reduce measurement burden for participants.
- Integrating process and outcomes evaluation may limit the risk of one component of data collection compromising another. For example, if collection of process data is

causing a high measurement burden for participants, it may be possible to take measures to stop this leading to low response to outcomes assessments.

Resources and staffing

A common theme within the case studies presented in Section C is that process evaluations are often insufficiently resourced. This perhaps reflects a tendency to trim funding applications to competitively low cost through reducing the scope of the process evaluation, amidst (real or perceived) concerns that funders do not regard substantial process evaluation as providing good value for money. Perhaps for these reasons, responsibility for process evaluation is sometimes assigned to less experienced junior researchers.

Conducting a high-quality outcomes evaluation undeniably requires a wide range of skills. However, research questions are typically easily defined, and there is a much literature to turn to for guidance. Process evaluations, in contrast, involve deciding from a wide range of potentially important research questions, integrating complex theories that cross disciplinary boundaries, and combining quantitative and qualitative methods of data collection and analysis. Individual researchers are unlikely to be expert in all of the methodological skills and theoretical knowledge required for a high-quality process evaluation, particularly not in the early stages of their career. Hence, just as would be the case for a robust outcomes evaluation, sufficient funds, expertise and experience must be available to enable successful completion of the process evaluation. If a junior researcher is leading the process evaluation, they need to be supported by a team with expertise and experience in quantitative, qualitative and mixed methods, and relevant psychological and sociological theory. For the reasons above, evaluation needs to be overseen by a principal investigator who values all components of the research. In addition, consideration needs to be paid to whether sufficient resource has been costed for the collection and analysis of likely large quantities of data.

Resource and staffing considerations

Who is responsible for the process evaluation?

What experience of interdisciplinary and mixed-methods research is included within the evaluation team?

How will you ensure that researchers leading the process evaluation are sufficiently supported by experienced staff?

Will the study be led by a principal investigator who values the process evaluation?

Have sufficient hours and expenses been allocated for data collection and analysis (for example, travelling to and recording interviews⁵⁸ or focus groups, and transcription)?

Patient and public involvement

It is widely believed that patient and public involvement (PPI) may enhance the quality and relevance of health and social science research, and funders increasingly expect this to be included in research. Within evaluations of health interventions this may include, for example, lay advisors (e.g. school teachers, governors or students for a school-based intervention) who sit on project steering groups, or comment on research priorities, acceptability of research procedures, or readability of materials produced for the target population. There are substantial definitional and empirical uncertainties relating to PPI, and hence this document does not aim to provide guidance on its use within process evaluation. However, advice on public involvement in health related research can be sought from sources such as the NIHR-funded organisation INVOLVE (<http://www.invo.org.uk/>).

Designing and conducting a process evaluation

Defining the intervention and clarifying causal assumptions

A key prerequisite for designing a good quality process evaluation is a clear definition of the intended intervention. Where defined as a set of standardised activities delivered to a target audience (e.g. goal setting, monitoring and feedback), evaluators may be concerned with capturing the extent to which these are reproduced as per intervention manuals. Alternatively, an intervention may be defined as a set of structures and processes intended to improve health through facilitating changes in the dynamics of a system such as a school or workplace. Process evaluators would in such cases be interested in whether the structures and processes to facilitate these changes are followed with fidelity. Key steps in understanding the causal chain would then include identifying whether the activities resulting from these structures and processes remain consistent with intended functions, accepting that their exact form may vary according to local need.

Ideally, by the time an evaluation begins, formative research will have produced a thorough definition of the intervention. The intervention will have been fully described and, where appropriate, a protocol or manual drafted, using standardised terminology to describe intervention components. This manual may be made publicly available at the outset or once evaluation findings are published. The causal assumptions underpinning the intervention will have been clearly described, setting out the resources needed to implement the intervention, how they will be applied, how the intervention is intended to work, and the intended short-, medium- and long-term outcomes. Though evaluators may choose alternative ways of

describing the programme and its causal assumptions, the development of a logic model is highly recommended (for further discussion and examples of logic models, see Chapter 2). It is useful if the intervention and its evaluation draw explicitly on one or more sociological or psychological theories, so findings can add to the incremental development of theory. However, evaluators should avoid selecting one or more pre-existing theories without considering how they apply to the context in which the intervention is delivered. Additionally, there is a risk of focusing narrowly on inappropriate theories from a single discipline. For example, some evaluations have been criticised for drawing predominantly upon individual-level behaviour change theories, where the aim is to achieve community, organisational or population-level changes, for which the sociological literature may have offered a more appropriate starting point (Hawe et al. 2009).

If there is clarity over what the intervention is and how it is intended to work, designing a process evaluation should begin by reviewing descriptions of the intervention and its underlying theory, to decide what aspects of implementation, mechanisms of impact or context require investigation. If a comprehensive description is available, reviewing this with developers and implementers can help clarify whether understandings of the intervention are shared between implementers and evaluators, or indeed among different implementers. It is also beneficial to review relevant literature, and consider the plausibility of causal links proposed within the logic model or intervention description. This can help to identify whether evidence is particularly equivocal for any links, and the existence of any potential contradictions (e.g. components which may inhibit the effectiveness of other components).

In many cases, such as pragmatic trials of policy initiatives, there may not be a fully documented description of the intervention, and causal assumptions may not have been explicitly described at the time of commissioning an evaluation. In such instances, an important first step should be working with programme developers and implementers to develop a shared definition of the intervention, and to describe the causal assumptions underpinning it. This should not be the sole responsibility of process evaluators; outcomes evaluators may, for example, have developed a logic model to decide on primary and secondary outcomes, or to identify mediators for measurement. However, it is crucial that the evaluation team as a whole ensures that a clear description of the intervention and its causal assumptions is in place. Consulting stakeholders at multiple levels of implementation (e.g.

national policy representatives, local implementers) may reveal variation in understandings of what the intervention is, emphasis on each of its components, assumptions about underlying mechanisms, and perceptions of who is likely to benefit most. Where divergences in understandings of the intervention or causal assumptions become apparent, this will enable evaluators to anticipate where challenges may arise, and hence where the process evaluation should focus its attention.

Key considerations in defining the intervention and clarifying causal assumptions

- How well are the intended intervention and its components described? Are any available standardised definitions and taxonomies applied?
- Have you demonstrated how the intervention is conceptualised? Does it consist of a set of standard activities to be delivered, or a set of structures and processes to facilitate changes in practice throughout a system?
- Is there a logic model (or other clear method of representing the intervention's causal assumptions), or does one need to be developed?
- Have you drawn upon theory appropriate to the nature of the intervention (e.g. looking beyond individual-level theorising if system-level change is targeted)?
- How will you evaluate the plausibility of the causal assumptions within the logic model?
- What potentially weak links or contradictions can you identify in the implementation or assumptions about causal mechanisms?
- Are understandings of the content of the intervention, and assumptions about how the intervention works, shared between evaluators and programme developers at all levels of implementation?
- Where there appears to be variability in understandings of the intervention, how might this affect implementation?

Learning from previous process evaluations. What do we know already? What will this study add?

As with all research, a key starting point in developing a process evaluation should be to review the literature in order to identify what is known about the subject, and how the proposed study might advance this. It is an inefficient use of public money to focus inwardly on the specifics of an intervention while overlooking opportunities to advance the evidence base. At some point, the evaluation is likely to be included in systematic reviews, which attempt to synthesise evaluations of interventions that have similar components, or are

informed by similar theories of change. Waters and colleagues (2011) have argued that if such systematic reviews are to offer anything of value to decision-makers, implementation and contextual factors must be considered as part of the review process. These arguments are central to the recent Oxford Implementation Index, which provides guidance to systematic reviewers on extracting and synthesising information on implementation (Montgomery et al., 2013b). It is the responsibility of process evaluators to provide the information to enable reviewers to examine these issues closely.

Reviews may indicate that variation in the outcomes of similar interventions arises from subtle differences in implementation or context. However, if each study addresses different process questions, or uses non-comparable methods to address the same questions, ability to compare findings across studies will be compromised. Hence, a useful starting point in designing a process evaluation is to identify process evaluations of interventions which share similar components or related theories of change. As described in Chapter 2, not all relevant studies use the term ‘process evaluation’. Hence, evaluators should not overlook relevant work which does not call itself a process evaluation, such as qualitative studies examining implementation or participant experiences of similar interventions. It is likely that many such studies will have been identified during the process of developing the intervention, or articulating its theory of change. If previous process evaluations can be identified, evaluators should consider whether it is appropriate to replicate aspects of these evaluations, and build upon them to explore new questions or issues. Although there is no ‘one size fits all’ set of methods for process evaluation, one would expect a degree of overlap in the aims and methods of evaluations of similar interventions. There may be good reasons not to replicate a previous process evaluation; for example, critical examination may conclude that it did not address the most important process questions, or that the methods adopted were flawed. Nevertheless, building on previous evaluations, rather than starting from scratch, should be considered where possible.

Key considerations in locating a process evaluation within the evidence base

What is already known from other evaluations of this type of intervention, and what original contribution does your process evaluation aim to make?

How will your process evaluation add incrementally to understandings of intervention theory?

What information would a future systematic reviewer, or policymaker, need to make sense of the findings of your process evaluation and compare them to other evaluations of similar interventions?

Which process evaluations have been conducted of interventions sharing similar components or theories of change?

Can any aims and methods of these evaluations be replicated in your study?

How can you build on previous process evaluations, and identify important questions which further advance the evidence base?

Deciding core aims and research questions

Once a comprehensive description of the intervention and its causal assumptions has been agreed and clearly described (most likely in a logic model), attention turns to the identification of key uncertainties in relation to implementation, mechanisms of impact and contextual factors. This can give rise to an overwhelming array of potential research questions. It is important not to be overly optimistic and expect to leave no unanswered questions (Munro & Bloor, 2010). Instead, process evaluation should aim to offer important insights which advance understandings of intervention theory and practice, and raise questions for investigation, drawing on a clear understanding of the current evidence base. It is better to answer the most important questions well than to try to address too many questions, and do so unsatisfactorily. Early agreement of core research questions can reduce the tendency to collect more data than can realistically be analysed (described within several of the case studies in Section C), and also minimises the risk that excessively intensive data gathering will change the intervention.

Essentially, process evaluation questions should emerge from examining the assumptions behind the intervention, and considering the evidence for these. If the evaluation team has a strong working knowledge of the relevant theoretical and empirical literature, and a good range of expertise, discussions within the team should form a strong basis for identifying the key uncertainties to address. Hence, process evaluators should start by systematically listing the assumptions linking the proposed intervention to intended outcomes. Agreement should then be sought on the most important questions to investigate, through reviewing the literature, discussions within the evaluation team, and consulting policy and practice stakeholders and the target population. Some key considerations in deciding research

questions relating to the three core aims of process evaluation described in Chapter 2 are now addressed, before moving onto discussion of methods.

Implementation: what is delivered, and how?

Most process evaluations will aim to capture what is implemented in practice. In a feasibility or pilot study, evaluators will be particularly interested in identifying facilitators and barriers to implementation, so that strategies to ensure high quality implementation can be put in place in time for evaluation of effectiveness. Where evaluating effectiveness, implementation assessments will aim largely to provide assurances of internal validity, through capturing the quality (fidelity) and quantity (dose) of implementation, allowing outcomes to be understood in light of a clear picture of what was delivered. Process evaluations should aim to capture emerging adaptations to the intervention. Evaluators should consider how they will decide whether changes represent ‘innovations’ initiated deliberately to enhance effectiveness, unintentional implementation failures, or deliberate subversions due to limited acceptability (see Chapter 2 for discussion of debates surrounding the nature of fidelity and contextual adaptation).

Process evaluations will often include assessments of reach, in terms of, for example, proportions of the target audience who came into contact with the intervention. Evaluators should, by this stage, have a good understanding of how implementation is to be achieved. Understanding how the structures and resources, put in place to ensure successful implementation, work on a larger scale will offer key insights into how the intervention might be scaled-up after the trial. It may be that changes were made to the intervention in response to findings from feasibility testing, in which case evaluators will need to consider whether these changes had the desired effects.

Where resources do not allow for the implementation of all components to be monitored in detail, evaluators often choose to conduct more intensive assessment of ‘core’ intervention components. As described in Chapters 1 and 2, however, the evaluator should avoid losing sight of how components function within the intervention as a whole. If components are considered to contribute very little to the quality of implementation, one could question why they are present. Issues to consider in deciding how to allocate resources in evaluating implementation include:

- Which components represent the most complex changes to practice?
- Are there any components for which resource needs or challenges in delivery may have been underestimated?
- For which components do previous studies, or feasibility testing stages, indicate the greatest uncertainty regarding how to deliver them in routine practice?
- Are there any components for which there is relatively limited agreement among implementers on their roles in the overall functioning of the intervention, or any contradictory causal assumptions being made?
- Are there any components for which feasibility and acceptability appeared relatively low during feasibility testing? Have any measures been put in place to address these issues and, if so, do these need to be evaluated within the main evaluation?

Mechanisms of impact: how does the delivered intervention work?

Where interventions are assumed to produce change by means of participants' interactions with them, a key aspect of understanding how they work is examining these interactions. As described in Chapter 3, while there may be a role for quantitative measures of satisfaction or acceptability, evaluators are likely to want to ask more probing qualitative questions to understand how the audience interacted with the intervention. Inductive and exploratory questions will provide insights into unanticipated causal processes and consequences. Evaluators may also want to test key assumptions in the logic model through mediational analysis (e.g. whether a physical activity intervention produced behaviour change through mechanisms such as more positive beliefs about the behaviour(ProActive, Case Study 4) or whether changes in diets of women were contingent on exposure to staff trained in behaviour change techniques(SIH, Case Study 2)). They may also wish to link mechanisms to implementation data, such as by examining whether certain mechanisms were activated more effectively when intervention components were delivered with greater fidelity. It is likely that the intervention will include multiple anticipated mechanisms of impact. Hence, investigating all of them may not be feasible. As with implementation, greatest attention should be paid to links in the logic model for which the evidence is more equivocal, or on which there is relatively limited agreement.

Contextual factors

Contextual factors can influence the effectiveness of an intervention both indirectly, through shaping what is implemented, and directly, through shaping whether the delivered activities

trigger the anticipated mechanisms of impact. Some hypotheses may be informed by current evidence regarding likely moderators of implementation and effectiveness, or competing causal mechanisms which may weaken the effect of the intervention. For example, we might predict that legislation prohibiting smoking in public spaces will have the least impact on second-hand smoke exposure among children whose parents smoke at home. However, given the complex interactions of interventions with their contexts, many contextual factors, such as barriers and facilitators to implementation, and the circumstances under which ‘mechanisms’ were activated or suppressed, may be identified through engaging with implementers and participants.

Investigating the roles of context in shaping the implementation and effectiveness of complex interventions can be a bewildering task, and it is easy to get lost in trying to identify and evaluate every possible external factor with which the intervention might interact. It is helpful to draw upon an explicit theoretical framework to guide understandings of the interactions between implementation processes and the systems in which the intervention is implemented, and in turn contribute to the refinement of these theories. Examples of potentially relevant theories can be found in Chapter 3. Where investigating impacts of context on outcomes, it is helpful to relate contextual variations to *a priori* hypothesised causal mechanisms, or those emerging from qualitative analysis, in order to generate insights into context-mechanism-outcome patterns.

Expecting the unexpected: building in flexibility to respond to emergent findings

In many of the case studies in Section C, authors note that although in retrospect, research questions had not been sufficiently defined at the start of the process evaluation, important new questions emerged during the course of the evaluation. Figure 7 presents the research questions asked within the process evaluation of the National Exercise Referral Scheme in Wales (NERS, Case Study 5), and the methods used to address them. As indicated, some were specified in advance, others emerged as the study progressed. Early recognition that fidelity was limited led to additional research to understand the impacts of new training courses which attempted to improve implementation.

Similarly, in ProActive (Case Study 4), increasing recognition of the impact of fidelity on outcomes led to additional research to investigate delivery and participant responses. In both instances, additional funds were sought to pursue emerging issues which required more in-

depth analysis, or additional data collection. Within the evaluation of Sexual Health And Relationships (SHARE; Case Study 3), the research design allowed for additional qualitative data to be collected should issues emerge during the trial which needed to be explored. Furthermore, survey data included information on several important contextual variables, such as family life and school ethos, which could be analysed retrospectively to see if they helped explain the outcomes. Within SIH (Case Study 2), methods for evaluating change in staff practice were flexible, and adapted as the study progressed. Repeated observation of staff practice at one year evolved, during the course of the study, as the ideal method for assessing the effect of the training. Hence, while some evaluators described a need to focus process evaluation aims more explicitly from the outset, allowing the streamlining of data collection and analysis, a degree of flexibility in the research design (and, where possible, in funding arrangements), to allow evaluators to respond to emergent issues, appears to have been crucial.

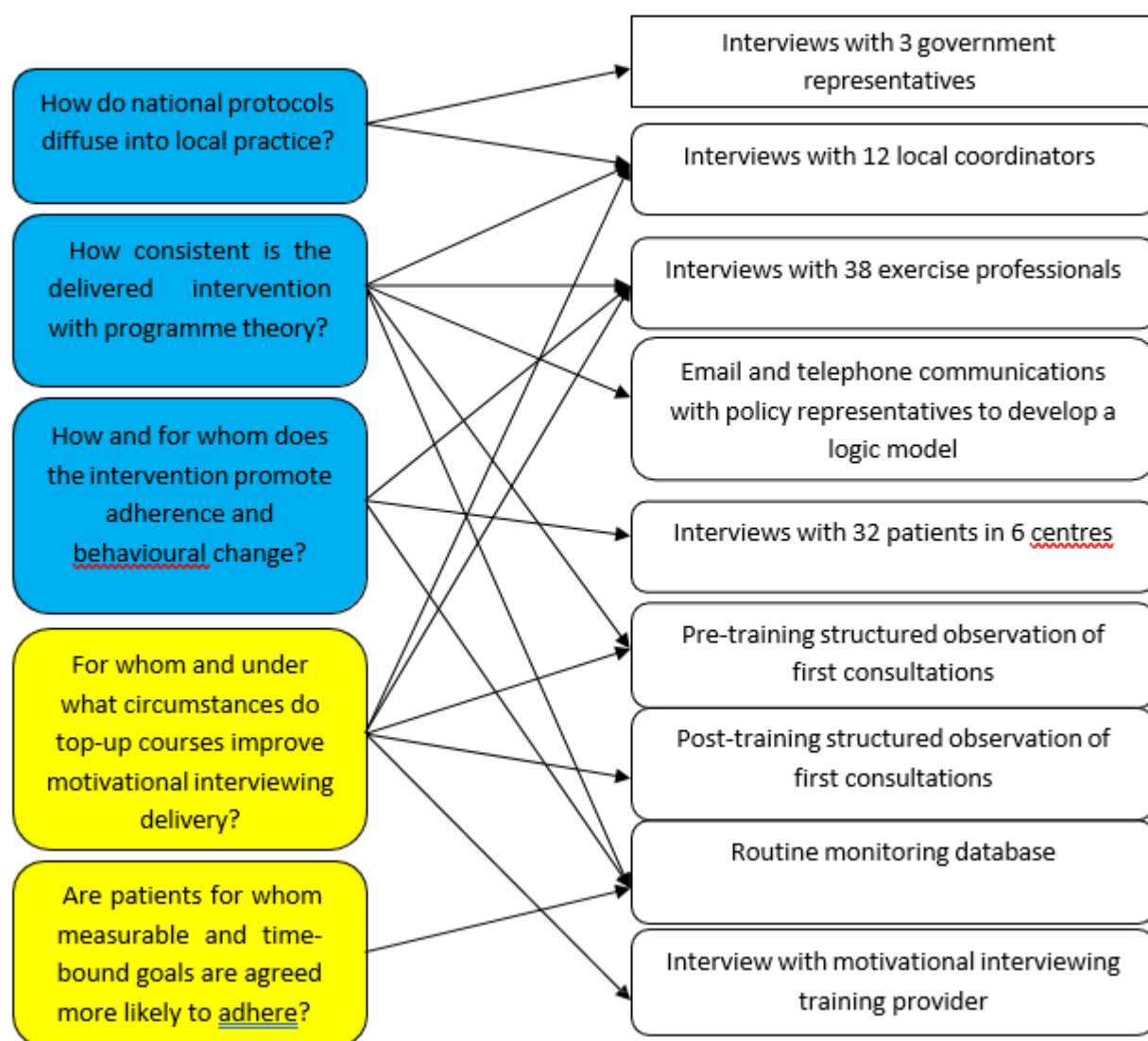


Figure 7. Research questions and methods adopted for the process evaluation of the National Exercise Referral Scheme in Wales. Pre-specified questions are in blue, questions which emerged during the course of the study are in yellow.

Extended Alt Text for Figure 7 which maps evaluation questions to data collection methods.

The first evaluation question asks how national protocols diffuse into local practice. Data sources include interviews with three government representatives and twelve local coordinators.

The second question asks how consistent the delivered intervention is with programme theory. Data sources include interviews with local coordinators and exercise professionals, communications with policy representatives to develop a logic

model, structured observations and routine monitoring data.

The third question asks how and for whom the intervention promotes adherence and behavioural change. Data sources include interviews with exercise professionals, interviews with thirty-two patients across six centres, structured observations and routine monitoring data.

The fourth question asks for whom and under what circumstances top-up courses improve motivational interviewing delivery. Data sources include pre-training and post-training structured observations, interviews and routine monitoring data, together with an interview with the motivational interviewing training provider.

The fifth question asks whether patients for whom measurable and time-bound goals are agreed are more likely to adhere. This question is addressed using the routine monitoring database.

Selecting methods

Quantitative and qualitative methods both have an important place, independently and in combination. There are numerous methods text books within the social sciences which provide detailed information on individual methods. Hence, this section does not provide comprehensive guidance on how to use and combine these. However, a brief overview of some common methods, and their pros and cons, will now be provided. Figure 8 links these methods to the aims of the process evaluation framework presented in Chapter 1, while Figure 9 presents methods and frameworks adopted by Case Study 5 (NERS, the National Exercise Referral Scheme in Wales).

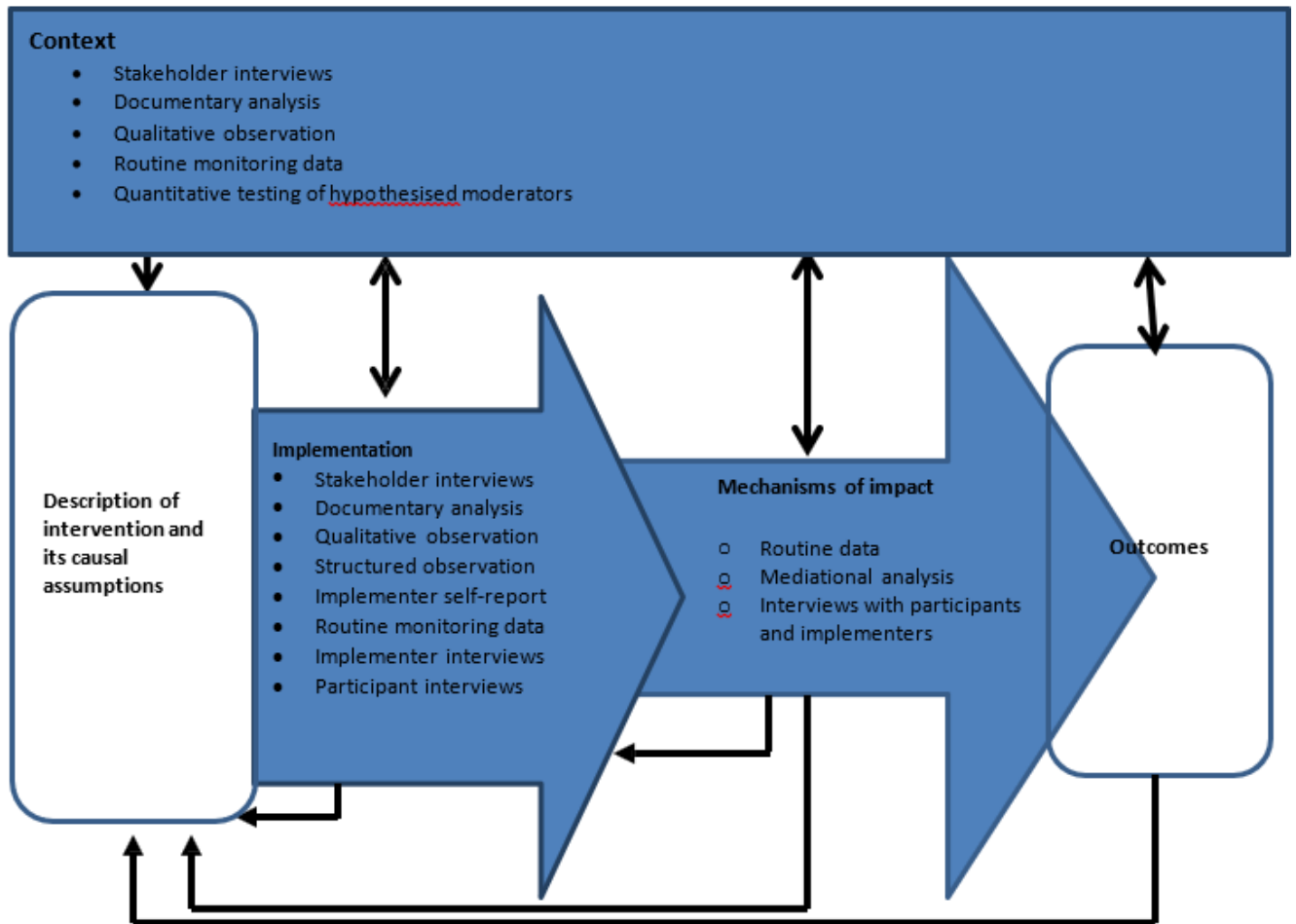


Figure 8. Examples of common methods for process evaluation and their relationship to each core function of process evaluation.

Extended Alt Text for Figure 8 which shows a process evaluation framework consisting of five components: context, intervention description and causal assumptions, implementation, mechanisms of impact, and outcomes.

Context influences all components and is studied through stakeholder interviews, documentary analysis, qualitative observation, routine monitoring data, and quantitative testing of hypothesised moderators.

The intervention description outlines the intervention and the causal assumptions that explain how change is expected to occur.

Implementation is examined using stakeholder interviews, documentary analysis, qualitative and structured observation, implementer self-report, routine monitoring data, implementer interviews, and participant interviews.

Mechanisms of impact are assessed using routine data, mediation analysis, and interviews with participants and implementers.

Outcomes represent the effects of the intervention. Feedback from outcomes and mechanisms of impact can be used to refine implementation and intervention design. Context continues to influence all stages throughout the evaluation process.

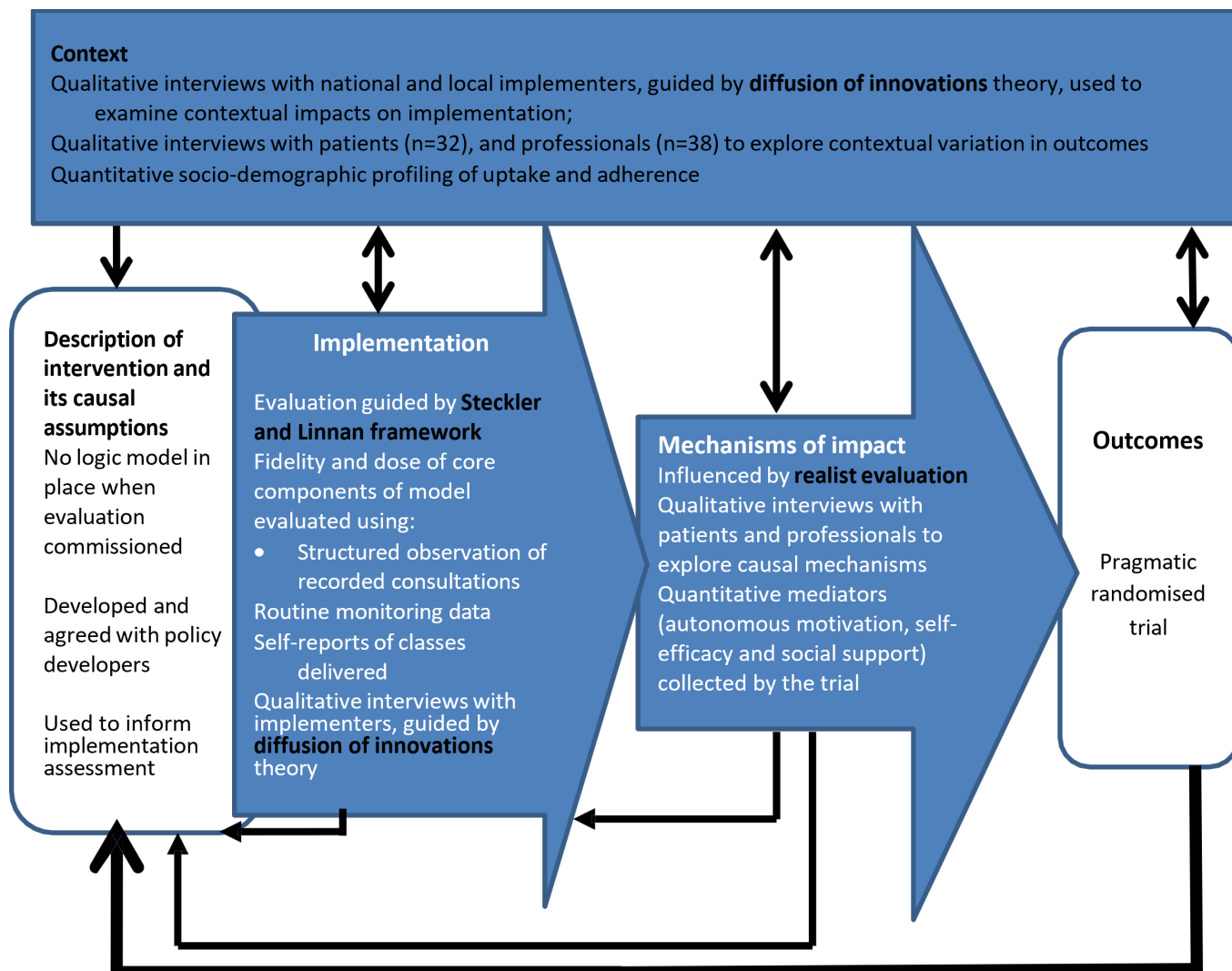


Figure 9. Frameworks and methods adopted for the NERS process evaluation

Extended Alt Text for Figure 9 which presents a process evaluation framework consisting of four main components: context, implementation, mechanisms of impact and outcomes.

Context is examined through qualitative interviews with national and local implementers informed by Diffusion of Innovations theory, qualitative interviews with patients and professionals exploring contextual variation in outcomes, and quantitative socio-demographic profiling of uptake and adherence.

Implementation is evaluated using the Steckler and Linnan framework. Fidelity and dose are assessed through structured observation of recorded consultations, routine monitoring data, self-reports of classes delivered, and qualitative interviews with implementers.

Mechanisms of impact are explored using principles of realist evaluation. Qualitative interviews with patients and professionals examine causal mechanisms, while quantitative measures assess mediators including autonomous motivation, self-efficacy and social support.

Outcomes represent the results of the intervention. Context influences implementation, mechanisms of impact and outcomes throughout the evaluation process.

Common quantitative methods in process evaluation

Commonly used quantitative methods in process evaluation include self-report questionnaires, structured observation (either direct observation or observation of recorded consultations) or secondary analyses of routine monitoring data. Process evaluators also increasingly use objective measures such as GPS trackers to understand context and intervention processes.

- **Self-report questionnaires** can be a simple, cheap and convenient way to gather information on key process variables. However, they may be subject to social desirability biases; an implementer may, for example, be reluctant to share information which indicates that they did not deliver something they were expected to. Furthermore, where intervention involves application of skilled techniques, implementers may not be well placed to rate their own competence. Self-report questionnaires may also be administered to participants to capture mediating mechanisms, or quantify participants' interactions with the intervention (e.g. reach and acceptability). Process evaluators should consider whether there are existing validated measures that serve the purposes of the study (e.g. standard measures of psychological mediating processes) and allow for comparison across studies. Where new bespoke measures are needed, efforts should be made to rigorously develop and validate these.
- **Structured observation** involves observing the delivery of intervention sessions, and coding the extent to which components are delivered, using a structured coding form. This provides a means of reducing the potential discrepancy between what implementers say they do, and what they actually do. However, knowing that one is being watched will almost inevitably lead to behaviour change (Hawthorne effects). Hence, such observation is best undertaken where it can be achieved relatively unobtrusively. Direct observation may be inappropriate for cases such as one-to-one consultations, where the presence of a researcher may adversely affect rapport. In such instances, examining video or audio recordings of consultations may be more appropriate, particularly as the quality of coding may then be checked by a second

researcher. Structured observation may be useful for evaluating implementers' acquisition of specific skills. Although behaviour is likely to be changed by observation, if implementers lack competence due to insufficient training or support, they will be unable to show competence, regardless of the presence of an observer. If validated measures for structured observation are available (e.g. for standard approaches such as motivational interviewing), these ought to be used.

- The benefits of **secondary analysis of routine monitoring data** are discussed below in sections on working with implementers to collect process data. These include avoidance of Hawthorne effects, and the potential of gaining data for the entire intervention period at low additional cost. However, their validity and reliability may be difficult to ascertain. Furthermore, recording may be affected by the intervention itself. For example, an anti-bullying intervention may lead to greater awareness of, and more detailed recording of bullying in schools, suggesting that bullying increased in intervention schools. Combining their use with smaller-scale observations to provide indications of their validity may be valuable.

Common qualitative methods in process evaluation

Common qualitative methods used in process evaluation include one-to-one interviews, focus groups and observations. Some pros and cons of these methods are discussed below:

- **Group interviews or focus groups** may produce interactions which provide deep insights into consensus and conflict in the views and experience of participants. The group setting also offers an opportunity to elicit a wider range of perspectives more quickly than individual interviews. However, group dynamics may lead participants to respond in a different manner than in a one-to-one interview, particularly when there is a hierarchy amongst participants. Where groups are formed of colleagues or other individuals who are in regular contact, this may enhance rapport and openness, but may also make participants more conscious of how they portray themselves to their peers. 'Lower status' participants may be less likely to contribute or express disagreement, leading to false consensus and overrepresentation of the views of 'higher status' participants. Group size may also compromise the depth in which a topic may be explored.

- **One-to-one interviews** may be useful where discussing more sensitive issues, or where there are concerns that a group dynamic may repress individuals rather than eliciting a wide range of views (due to, for example, unequal power relationships between group members). While individual interviews involve the collection of data from fewer individuals, they provide greater opportunity to explore individual experiences in depth. In some circumstances **paired interviews** may be appropriate. For example, if the views of young people are sought on a sensitive interview, they may feel more at ease if they can bring a trusted friend with them.
- **Non-participant observation** involves the researcher making detailed field notes about the implementation of an intervention and the responses of participants. This may be useful for capturing finer details of implementation, examining interactions between participants and intervention staff, and capturing aspects of the ‘spirit’ of implementation, rather than just the mechanics of its delivery. As with structured observation, the use of this method may be limited to situations where observation can be made relatively unobtrusively. Participants typically include informants such as implementers, intervention participants or key ‘gatekeepers’ (e.g. teachers or employers), allowing evaluators to explore experiences of the intervention from multiple perspectives. Intervention participants may be well positioned to provide insights into perceived strengths and weaknesses of the intervention, and how it helped or failed to help them achieve change. Those implementing the intervention may be able to provide insights into the emergence of social patterning in responses, how and why their implementation practices changed over time, and the features of their own context that affect the ease with which the intervention can be implemented. Those at higher levels of the implementation process (e.g. regional and national coordinators) may be in a position to identify a broader range of contextual barriers and facilitators.

Mixing methods

It is important to avoid conducting independent quantitative and qualitative studies, and to explicitly consider from the outset how they fit together to become a mixed-methods study (Creswell, 2005; Creswell & Clark, 2007). Bonell and colleagues (2012) advocate an iterative model in which early qualitative data identify causal processes and contextual factors, which may then be measured to test the hypotheses generated. This may not always be possible for

reasons of timing and resource, or due to delays in going back to ethics committees for approval of changes to methods. Nevertheless, qualitative and quantitative methods can be combined to increase understanding of outcomes and improve interventions. For example, if quantitative data indicate that disproportionately few members of minority ethnic groups are participating in an intervention, interviews and focus groups with key stakeholders and members of minority ethnic groups may be undertaken to tease out facilitators and barriers to participation. Measures may then be recommended to counteract identified barriers to participation, and subsequent quantitative data examined to assess whether there was an increase in uptake by members of minority ethnic groups. Table 1 illustrates the mixture of methods used within the evaluation of ASSIST (Case Study 1).

A key challenge in conducting process evaluations is that all data must be collected in a relatively short time. Quantitative data may identify challenges for which it is not possible to provide a qualitative explanation within the required timescale, whereas qualitative data may generate new hypotheses requiring further research which is not feasible given time constraints. A good quality process evaluation will therefore offer important partial insights and highlight priorities for future research.

Table 1. ASSIST process evaluation data collection: main sources and methods			
Source		Data collection tool	Stage of the trial
S T U D E N T S	Eligible students in all intervention and control schools	Self-complete behavioural questionnaires	Outcome data collection (Year 8 baseline) Outcome data collection (Year 8 post-intervention) Outcome data collection (Year 9) Outcome data collection (Year 10)
	Peer supporters in 30 intervention schools	Self-complete questionnaires	1 st and 4 th school-based PS follow-up sessions
	Peer supporters in four intervention schools selected for in-depth study	Semi-structured interviews Focus groups	Post intervention
	25% random sample of non-peer supporters in four intervention schools selected for in-depth study who indicated they had conversations about smoking with peer supporters	Semi-structured interviews	Post intervention
S C H O O L	Teachers supervising data collection in all intervention and control schools	Self-complete 'smoking policy' questionnaires	Outcome data collection (Year 8 baseline) Outcome data collection (Year 9) Outcome data collection (Year 10)
	Supervising teachers in intervention schools	Self-complete questionnaires	PS recruitment PS training

S T A F F	Contact teachers/key staff in four intervention schools selected for in-depth process evaluation	Semi-structured interviews	Year 8 baseline Year 8 post intervention
	Contact teachers in four control schools selected for in-depth process evaluation	Semi-structured interviews Self-complete questionnaires	Year 8 baseline Year 8 post intervention
A S S I S T E R S O N S	Health promotion trainers in all intervention schools	Self-complete questionnaires	Training the trainers PS recruitment PS training All school-based PS follow-up sessions Presentation of certificates/vouchers
	Health promotion trainers	Semi-structured interviews	Post intervention
	Researchers in four intervention schools selected for in-depth process evaluation	Non-participant observation	Training the trainers PS recruitment PS training All school-based PS follow-up sessions
Note: PS = peer supporter, a student nominated by their peers as influential, who was trained to diffuse the smoke-free message			

Using routine monitoring data for process evaluation

Aims of process evaluation will often overlap with management practices. For example, organisations responsible for delivering interventions will probably have already integrated some form of monitoring structures into their management practices in order to monitor the quality of implementation. Recent NICE guidance for behaviour change interventions recommends that all interventions should include structures for regular assessment of implementation (NICE, 2014). Where such data are available and can be shared with the evaluation team, their use may help to avoid issues such as Hawthorne effects (where the behaviour of the implementer is changed by awareness of being observed). This is not to suggest that monitoring does not change behaviour; however, if this monitoring is part of the structure of the intervention, any effect would be reproduced in a scaled-up version of the intervention. Use of routine monitoring data may reduce response biases, and prevent duplication of efforts, reducing the cost of the evaluation and burden on implementers. Furthermore, it may provide a cost-effective means of obtaining information on the full duration of the evaluation, allowing analyses of change over time, which may not be possible where observations are based on snapshots of implementation processes at one or two points in time.

While there are clear advantages to using routine data for process evaluation, the biggest risk in their use is that it is not always easy to ascertain their quality. Hence, it is often appropriate to conduct smaller-scale observations in order to validate the data collected from routine monitoring of the intervention. Additional challenges may arise from negotiating complex governance processes relating to the use of such data for research purposes. Where possible, it is useful to work with programme developers and implementers to develop high quality monitoring structures which provide routine data that can be analysed as part of a process evaluation. Where evaluators have limited input into the design of intervention monitoring structures, it is helpful to ascertain what monitoring data are available and whether there are any components whose delivery is not routinely monitored.

Considerations in using routine data for process evaluation

Can you use routine monitoring data to evaluate implementation?

Is there opportunity to influence the shape of monitoring structures to serve dual purposes of routine monitoring and providing good quality process data?

Can the validity and reliability of routine data be evaluated?

Is there sufficient time and resource to negotiate any necessary data governance structures to facilitate data sharing?

If asking implementers to collect data on your behalf, have you ensured that instructions can be followed with little effort, and are designed to minimise reporting bias?

Sampling

Sampling is an important consideration in conducting qualitative research in the context of large-scale evaluations (e.g. implementer interviews), or in conducting small-scale quantitative sub-studies (e.g. structured observations or validation sub-studies). It is often unnecessary or impractical to include all relevant stakeholders. In the NERS process evaluation (Case Study 5), all exercise professionals were invited to take part in interviews, largely because they had not previously been consulted on the scheme to the same degree as many other stakeholders. However, the large response led to an overwhelming volume of data, far more than was necessary for theoretical saturation. There are also risks in relying on a few small case studies to draw conclusions regarding the intervention as a whole (Munro & Bloor, 2010). Hence, it may be more appropriate to use random sampling, purposive sampling of sites or individual participants (according to core characteristics which are expected to impact the implementation or effects of the intervention) or a combination of the two. During ASSIST (Case Study 1), in-depth process evaluation was conducted in four purposively selected schools, out of the 30 schools that implemented the intervention. Within these schools, students were randomly sampled to take part in interviews and focus groups to avoid, for example, teachers nominating ‘well-behaved’ students for interview.

Timing considerations

Another key concern in designing and conducting any process evaluation is the timing of data collection. The intervention, participants’ interactions with it, and the contexts in which these are situated are not static entities, but continuously change shape during an evaluation. Hence, careful consideration needs to be given to how data are situated in the time at which they were collected. If the evaluator collects data only during the early stages of the evaluation, findings may largely reflect ‘teething problems’ that were addressed as the evaluation progressed. In the case studies in Section C, large-scale evaluations such as NERS,

SHARE and ASSIST combined brief measures of implementation throughout the evaluation with in-depth case studies, to overcome the tension between coverage and depth.

Considerations in deciding when to collect process data

How will implementers' perceptions of the intervention, and hence their practices, change over time as they begin to receive feedback from the target audience on what does and does not work?

Will the organisation change gradually over time to allow full integration of the intervention?

Are resources available to collect data at multiple time-points in order to capture changes over time, and can this be done without placing too much burden on respondents or changing how the intervention is delivered?

Analysis

Analysing quantitative data

Analysis of quantitative data within process evaluations ranges from descriptive to explanatory. Descriptive information is often provided on quantitative measures of process measures such as fidelity, dose and reach. Process evaluators may also conduct more detailed modelling to explore how delivery, reach or acceptability vary according to contexts or participant characteristics, offering insights into how inequalities are affected by the intervention.

Analysing qualitative data

Analysis of process evaluation is described in some of the case studies in Section C as being hampered by the collection of large volumes of qualitative data and insufficient resources to analyse it well. Hence, when designing studies and preparing funding applications, it is critical that appropriate staff, time and resources are allocated to the analysis of qualitative data. Evaluators should take advantage of the flexibility and depth of qualitative methods in order to explore complex mechanisms of delivery and impact, contextual factors and unanticipated consequences. Ideally, collection and analysis of qualitative data should be an iterative process, with both occurring in parallel. On a theoretical level, this means that emerging themes can be investigated in later interviews. On a practical level, this means that the researcher will not reach the end of data collection with a huge amount of data, just a few weeks before the study ends. There are numerous texts on the analysis of qualitative data

(Coffey & Atkinson, 1996), and hence it is not our intention to provide a detailed overview of the approach(es) one should choose. Nevertheless, the approach selected should be justified by the evaluator. It is often good practice to factor in time and resource for second coding in order to examine its validity, as well as considering quality assurance frameworks against which the analysis may be checked by reviewers (see Chapter 5).

Mixing methods in analysis

While requiring different technical skills and to a large extent addressing different process questions, efforts should be made to combine quantitative and qualitative analyses rather than presenting parallel mono-method studies. Quantitative data may identify issues which inform qualitative data collection and analysis, while qualitative data may generate hypotheses to be tested with quantitative data. Essentially, qualitative and quantitative components of a process evaluation should facilitate interpretation of one another's findings, and, where possible, inform how subsequent data are collected or analysed. For example:

- Qualitative data may identify strengths and weaknesses in the structures in place to implement the intervention.
- Quantitative data may then confirm whether or not the intervention was effectively implemented.
- Knowing what was delivered allows qualitative data on participant responses to be understood in light of a clear definition of the intervention with which participants interacted.
- Qualitative data on participant responses may generate hypotheses regarding causal mechanisms and how patterning in responses to the intervention emerged across contexts.
- Where data are available, quantitative analyses may test these emerging hypotheses.

Most case studies presented in Section C used a mixture of methods. In ASSIST (Case Study 1), quantitative data suggested that the smoking prevention programme was more effective with students who were 'experimenters' than regular smokers. Qualitative data identified the strategies used by the students tasked with diffusing the smoke-free message and revealed that they targeted friends and peers who were non-smokers and experimenters, rather than

students who belonged to smoking cliques. In the NERS process evaluation (Case Study 5), qualitative data identified a range of contextual and socio-demographic factors which exercise professionals or patients felt were linked to adherence to the scheme. Quantitative data also indicated that motivational interviewing was poorly delivered, with subsequent qualitative data collected to explore why this was the case.

Integrating process evaluation findings and findings from other evaluation components (e.g. outcomes and cost-effectiveness evaluation)

Integration of process and outcomes findings has often been limited. Although qualitative findings are sometimes used to illuminate trial outcomes, often this is not visible within peer-reviewed publications (Lewin et al., 2009; O'Cathain et al., 2013). Process evaluators should work with those responsible for other aspects of the evaluation to ensure that plans are made for integration from the outset, and that these are reflected in how the evaluation is conducted. This will include addressing key issues such as ensuring there is sufficient expertise in the team, a genuine interdisciplinary team environment, and a principal investigator who values and oversees all aspects of the evaluation.

Where quantitative process data are collected, these should be designed to enable associations with outcomes and cost-effectiveness to be modelled in secondary analyses. For example, if fidelity varied substantially between practitioners or areas, evaluators may examine whether better delivery produced better outcomes. Process data may facilitate 'on-treatment' analyses (comparing on the basis of intervention receipt rather than purely by intention-to-treat).

While flawed by the fact that it breaks randomisation, it may usefully be presented alongside traditional intention-to-treat analyses. In NERS (Case Study 5) for example, intervention effects were shown to be limited to those participants who completed the intervention. The RIPPLE evaluation by Strange and colleagues (2006), for example, examined differences in the impact of a sex education programme according to the quality of delivery of its key components. while the SHARE evaluation collected sufficiently comprehensive data on implementation to conduct an 'on-treatment' analysis of outcomes (Wight et al. 2002; Case Study 3).

Integration of quantitative process measures into analysis of outcomes or cost-effectiveness is challenging if assessments of implementation are based upon data gathered at only a few times or sites. For example, if fidelity data consist of case study observations in five or six

schools, there is likely to be insufficient power, variation or representativeness to move beyond description of fidelity. Hence, where possible, data for key measures should be obtained across all sites. However, as described above, improved coverage (for example, through reliance upon routinely collected data) must be balanced against variability in data quality.

Qualitative components should also be designed to relate outcomes data to understanding the intervention's implementation, and how its outcomes were produced. As described above, one means of achieving this is to seek perspectives of stakeholders purposively sampled according to characteristics which are anticipated to influence implementation and outcomes. Qualitative process analysis may serve predictive or post-hoc explanatory functions in relation to outcomes evaluation. Where conducted prior to outcomes analysis, this process analysis may provide insights into why we might expect to see positive or negative overall intervention effects. Qualitative data may also lead to the generation of hypotheses regarding reasons for variability in outcomes - for example, whether certain groups of participants appear to have responded to the intervention better than others. Hypotheses regarding such patterning may be tested quantitatively in secondary analysis. The wide range of work qualitative research has delivered when used with randomised controlled trials has been mapped (O'Cathain 2013).

Issues in relation to the timing of analysis of process and outcomes data are discussed within several case studies in Section C. Some chose to analyse qualitative process data independently from trial outcomes, in order to avoid biasing these analyses, as recommended by Oakley and colleagues (2006). However, others highlighted the value of post-trial analysis of causal pathways and implementation in allowing emerging issues to be explored (e.g. ProActive, Case Study 4). Given that many evaluators comment that process evaluations generate far more data than can be adequately analysed within the timescales of the main study, it would be wasteful not to further analyse data from the process evaluation once the outcomes of a trial are known. Hence, a more tenable position is that where possible, the core pre-planned process analyses should be answered without knowledge of outcomes, but analysis in relation to secondary or emerging questions performed later, with transparent reporting of how knowledge of outcomes shaped research questions and analysis.

Ethical considerations

A number of ethical considerations have been raised throughout this chapter. In particular, we have discussed challenges in negotiating the relationship with stakeholders who have a vested

interest in the success of the intervention, and ensuring that independence is maintained in evaluating complex interventions. The position of the evaluator, and influence on the research process of relationships with the research funder or intervention developers, should be transparently reported.

In addition, process evaluations typically involve collecting rich data from a limited pool of potential participants. This raises issues of confidentiality, as it is possible that someone with a good working knowledge of the intervention and the settings in which it was delivered may be able to identify individual participants from evaluation data. Data may involve criticisms of persons who hold a position of authority over the participant, and a failure to safeguard anonymity may jeopardise working relationships. Hence, close attention needs to be paid to ensuring that anonymity is maintained wherever possible (both for individual participants and for clusters such as schools or workplaces). If there is any doubt as to whether anonymity may be compromised, this should be discussed with the participant, and written confirmation obtained prior to publication that the participant is happy for their data to be used. Issues of anonymity should also be considered carefully where using routine data, with measures put in place to ensure that no identifiable data are shared between the intervention and evaluation team.

The role of the evaluator either as a passive observer of the intervention, or actively feeding back information to enable implementers to improve delivery, is discussed above. Related to this, another key ethical issue which evaluators should consider is what actions will be taken if process data show that an intervention is causing harm. Trials may have predefined stopping rules specifying that the trial is to be terminated if intermediate quantitative outcomes data indicate harms. However, it is impossible to anticipate all the possible outcomes of a complex intervention, and qualitative data may be the best way of capturing unanticipated and potentially undesirable outcomes. Evaluators should consider what weight should be given to such data, and whether any rules can be identified for deciding when evidence of harms is sufficient for the intervention, and its evaluation, to stop or significantly change course.

Summary of key points

This chapter has provided readers with practical guidance on key issues to consider in planning, designing and conducting a process evaluation. It has been argued that success in planning a process evaluation requires:

- effectively negotiating relationships with stakeholders such as policymakers and implementers;
- effective interdisciplinary working within the evaluation team;
- careful consideration of resource requirements and the mix of expertise within the evaluation team.

Designing and conducting a process evaluation requires:

- a clear definition of the intervention and its causal assumptions;
- consideration of what the process evaluation will add to the existing evidence base (including how study information might be used in future evidence synthesis);
- early definition of the most important research questions to address (drawing upon intervention theory, the current evidence base and consultations with wider stakeholders), while allowing the flexibility to address emerging questions;
- selection of an appropriate combination of quantitative and qualitative methods to address the questions identified.

Chapter 5 will now discuss key issues in reporting process evaluations.

Reporting and dissemination of process evaluation findings

A key challenge for process evaluation is reporting and disseminating large quantities of data to a wide range of audiences. Evaluators will typically need to share findings with the funders of the research, and with stakeholders from policy and practice who may be interested in the immediate implications of the evaluation for their work. When process evaluation is conducted by academic researchers, there will also be a strong desire and significant pressure to publish findings in high impact peer-reviewed journals. This chapter aims to:

- signpost the reader to relevant guidance on how and what to report in process evaluations;
- consider strategies for disseminating findings to wider audiences and publishing in academic journals;
- consider issues in the timing of reporting a process evaluation.

How and what to report?

Providing guidance on reporting standards for process evaluation is challenging as there is no ‘one size fits all’ method, or combination of methods, for process evaluation. Evaluators will therefore want to draw upon a range of existing reporting guidelines which relate to specific methods. A regularly updated database of reporting guidelines for health research is available on the website of the Enhancing the Quality and Transparency Of health Research network <http://www.equator-network.org/home/>). Reporting guidelines for qualitative research (Tong et al., 2007) will be relevant to almost all process evaluations. Where using implementation data to explain outcomes, or exploring mediators and moderators of effects, guidelines for reporting statistical methods (Lang & Secic, 2006), multi-level modelling (Jackson, 2010) or subgroup analysis within trials (Wang et al., 2007) may be relevant. As described in Chapter 4, when evaluators choose to check agreement between routinely collected data and researcher collected data from smaller sub-groups, guidelines on reliability (Kottner et al., 2011) may be relevant. There are currently no guidelines for reporting mixed-methods studies, although O’Cathain and colleagues’ (2008b) paper on the quality of mixed-methods studies in health comes closest. Armstrong and colleagues (2008) argue that existing reporting guidelines such as CONSORT typically do not meet the specific needs of complex intervention research, including limited emphasis on process issues, although as described in Chapter 2, work is currently underway to extend CONSORT reporting guidelines to social

and psychological interventions. This will include emphasis on reporting of process evaluation.

Key additional considerations in deciding what to report in a process evaluation include the relationships between quantitative and qualitative components, and the context of the process evaluation in the overall evaluation. The logic model, or other clear depiction of the causal assumptions of the intervention, should be reported in evaluation reports and at least one process evaluation article. Reporting this clearly enables external scrutiny of the assumptions made by the intervention developers, and the extent to which the process evaluation has identified and addressed the most important uncertainties.

Reporting to wider audiences

While process evaluation aims to inform the incremental development of intervention theory, and to contribute to a wider evidence base, it also aims to directly inform the actions of policy-makers and practitioners. Hence, it is vital to report findings in lay formats to stakeholders involved in the delivery of the intervention, or strategic decisions on its future beyond the evaluation. In the evaluation of the National Exercise Referral Scheme in Wales (NERS, Case Study 5), a summary report, including process data, cost-effectiveness data and outcomes, was provided to the Welsh Government and other intervention stakeholders. This took the form of a comprehensive executive summary, drawing together key findings of all aspects of the intervention, with a series of draft journal articles from the process evaluation, outcomes evaluation and cost-effectiveness evaluation included as appendices. National coordinators, and a designated member of the implementation team in each area involved in the trial (or which would subsequently be delivering the intervention), were then invited to attend a meeting, at which findings from the entire evaluation were presented, and implications for practice were discussed. Hence, findings were disseminated first to those who had a direct stake in them, allowing them to ask questions, to challenge the findings, and discuss the implications for practice.

In addition to those directly involved in the intervention, evaluators will probably want to reach policy and practice audiences whose future work may be influenced by the findings. Evaluators are increasingly using social media and blogging to reach wider audiences with their research. Presenting findings at conferences organised by policy and practice organisations also offers a means of promoting findings beyond academic circles.

Publishing in academic journals

While process evaluators will want to pay attention to reaching wider audiences, there are good reasons to publish findings in academic journals. Academic researchers are expected by their institutions to regularly produce articles in high impact, peer-reviewed journals. Demonstrating that findings have been through a critical peer review process may provide assurances of research quality, enhancing the credibility of conclusions. Findings in peer-reviewed journals will also be more easily retrieved by reviewers aiming to synthesise findings from evaluations.

It is unlikely that findings of the entire evaluation will be published in a single article. Outcomes evaluations will typically be submitted to scientific journals with word limits of 3000-4000 words. As highlighted by Armstrong and colleagues (2008), online journals increasingly offer opportunities to submit supporting material. Arguably, describing the methods and results of only the outcomes study in sufficient detail is a challenge within such a short article. It certainly does not allow for reporting substantial additional data from a process evaluation, which may generate too much data to be included within a single article. Some UK funders of health-related research (e.g. National Institute for Health Research Health Technology Assessment) now publish comprehensive reports of entire evaluation packages in a form which is closer to a book than a journal article (Isaacs et al., 2007).

Evaluators are likely to publish multiple research articles from the process evaluation in peer-reviewed journals. The key challenge is dividing the process evaluation into components which are valuable as standalone pieces and answer distinct research questions, while not losing sight of the broader picture, and ensuring effective cross-referencing between articles. Whilst Stange and colleagues (2006) recommend publishing sequential articles in the same journal, it will not always be possible to negotiate this with journal editors. Journals often prefer to emphasise one methodological tradition, or one theoretical aspect of the evaluation (Bryman, 2006). Hence, it is often necessary to divide findings across multiple journals; this was done for all of the case studies in Section C.

Ideally, even when submitting to different journals, the researcher would also want to be able to control the order of publication of process evaluation articles, so that these are published in a logical sequence which tells the story of the intervention. However, this is unlikely to be under the control of the research team. Two articles submitted to different journals at the same time may ultimately be published a year or more apart, even if both are accepted by

their first choice journal. Each article should hence make its context within the wider evaluation clear, and be explicit about its contribution to the overall process evaluation. All journal articles should refer to other articles published from the study, or at least to a protocol paper or report which covers all aspects of the evaluation. Study websites which include links to manuals and all related papers are a useful way of ensuring that findings can be understood as part of a whole.

In reality, it is not uncommon for process data never to be reported in academic articles, or for only some components to be reported (Lewin et al., 2009). This may reflect a tendency to place greater emphasis on the outcomes than the process evaluation, and to view process evaluation as unsuited to addressing important empirical or theoretical questions beyond the interpretation of the relevant outcomes study. Emphasising the extent to which process evaluations contribute to the theory development, or to methodological debates, may ensure that they have a broader appeal to journal editors. This could include, for example, making recommendations for the implementation of similar interventions, or reflecting on methodological lessons learned which may help other evaluators to improve their work.

When to report process data? Before, after or alongside outcomes/cost-effectiveness evaluation?

As described above, publishing process and outcomes data together can be a significant challenge. If they are to be separated, which should be reported first? Some authors recommend that all process data should be reported without knowledge of outcomes, in order to avoid biasing its interpretation (Oakley et al. 2006). For example, in the evaluation of an adolescent sexual health programme in Tanzania (Case Study 6), preliminary findings of an extensive process evaluation were documented in an internal report, and findings on the core components published (Plummer et al., 2007), prior to un-blinding of the trial outcomes data.

However, this is not always achievable. In some cases, policy and practice stakeholders and funders may be principally interested in outcomes, and hence push for these data to be analysed and reported first. Where trial follow-ups are relatively short, there may be insufficient time to produce a detailed analysis of qualitative data before follow-up data are fully collected, particularly if, as suggested by Grant and colleagues (2013a), qualitative data are collected from participants after the intervention.

Even if publications are submitted in the intended order, there is no guarantee that articles will be accepted in their first or second choice journal. This may result in peer reviewers recommending additional analysis of process data, which can no longer be conducted without knowledge of outcomes. Furthermore, if the outcomes paper has been accepted for publication and process papers have not, it would make little sense for process papers to ignore outcomes data, as these will now be in the public domain. Hence, some revision of the introduction or discussion sections of process papers may be needed in order to shed light on how process data explain outcomes. Given the large volumes of data generated by process evaluation, it is quite possible that secondary analyses will be conducted and reported for some years after outcomes are known. For example, ASSIST (Case study 1) published three key papers from the process evaluation before the main trial paper, and then published subsequent papers. Insisting that no further process data can be reported once outcomes are known is clearly untenable. However, evaluators should transparently report whether analyses were conducted prior to knowledge of outcomes, or the impact of knowledge of outcomes on the analysis and interpretation of process data.

Summary of key points

This chapter has signposted the reader to a range of guidelines they may wish to consult in considering how to report a process evaluation. It is challenging to provide general guidance for reporting process evaluation, given the diversity of methods adopted for evaluating different interventions. However, key issues to consider include explicitly reporting how methods were combined, and ensuring that the context of the process evaluation within the overall evaluation is clear. Process evaluators should maximise the likelihood of their research reaching multiple audiences, and any relevant findings influencing policy and practice. Academic publication is likely to involve breaking the process evaluation down into smaller parts, but careful cross-referencing between papers, and to full reports, should ensure that the bigger picture, and the contribution of each article to the whole, is not lost. Evaluators need to reflect on, and report, how analysis and interpretation were shaped by knowledge of outcomes.

Appraising a process evaluation: a checklist for funders and peer reviewers

This section draws upon the ‘how to’ chapter of this document and aims to provide funders and peer reviewers with a concise summary of considerations when deciding whether or not to recommend a process evaluation bid or article for funding or publication. This may also act as a useful summary of issues for researchers to review when developing a process evaluation. It is important to be mindful that funding application forms often have limited space to specify the proposed design, conduct and analysis of the process evaluation. Many of the issues below may have been considered by the research team but not included in a funding bid due to limited space. While it may be inappropriate to dismiss a process evaluation because not all of the issues below are described in full, reviewers may wish to query whether consideration has been given to any that are not mentioned.

Working with policy and practice stakeholders

- Are there any potential conflicts of interest arising from the relationship between evaluators and policy/practice stakeholders?
 - Have the authors described how they will address these and ensure that the evaluation remains independent?
 - Does the proposal set out a clear plan for communicating findings to policy and practice stakeholders during the evaluation?

Relationships between evaluation components

- Is the relationship between the process evaluation and other evaluation components clearly defined and justified?
 - Will process and outcomes evaluation be conducted by the same team or by separate teams?
 - If the former, how will researchers ensure that knowledge of outcomes or process does not bias analysis of the other?
 - If the latter, is there clear oversight of the two components?
- Is it clear that the principal investigator values all aspects of the evaluation, and will provide effective oversight of all aspects of the evaluation?

Intervention description and theory

- Is the intended intervention fully described?
 - Are standardised terminology and definitions of intervention components adopted where possible?

- Are the structures and processes involved in intervention delivery fully described?
- If appropriate, will a full intervention manual be made publicly available?
- Is a clear, plausible, set of causal assumptions specified and justified (for example, in a logic model)?
 - Does this draw upon appropriate theories?
 - If not, are there plans to develop a theory as part of the research?
 - Have the authors planned to review these assumptions with policy and practice stakeholders to explore agreement and divergence on what the intervention is, and how it will work?

Process evaluation aims and research questions

- Are the research questions clear, important and well justified with reference to the theory of the intervention and the status of the evidence base? What decisions will they inform?
- Have the authors considered whether previous process evaluations have been conducted of interventions involving similar components or theories of change?
 - Have they adopted comparable aims and methods, or justified not doing so?
- Has the theory of the intervention (or logic model) been used to identify key areas of uncertainty for investigation by process evaluation?
 - Have the authors considered which components may prove most challenging to implement (e.g. which represent more fundamental change, or for which there is least agreement on what they are and the purposes they serve)?
 - Have the authors considered for which causal assumptions evidence is most equivocal?
 - How will unanticipated consequences be captured?
- Is there linkage between research aims? Do they fit together to address the overall study aim?
- If conducted alongside an outcomes evaluation, is the added value of the process evaluation explained? Is it clear how the research will enhance the interpretation of outcomes?
 - Will process evaluation provide sufficient assurances regarding the internal validity of the outcomes evaluation?
 - Will it enable policymakers/practitioners to understand how the intervention might be applied in different contexts?

- Have the authors stated how and when they will combine process and outcomes data?

Selection of methods to address research questions

- Are the quantitative and qualitative methods selected appropriate to the research questions?
 - Will implementation be captured in sufficient detail to establish consistency with the theory of the intervention?
 - Are existing validated measures used where possible? Are plans to validate new measures included?
 - How will emerging changes, adaptations or additions to the intervention be captured?
 - Are the quantitative methods appropriate? (e.g. 'tick box' self-report by implementers of intervention delivery should be avoided if possible).
 - Are the qualitative methods appropriate?
 - Have the authors considered how change in practice as a result of being observed or measured will be minimised?
 - Have the authors considered the timing of data collection, and its impact on the data collected?
 - Have the authors investigated whether any routine programme monitoring data can be used? If so, are there plans to check their validity and reliability?
- Have the authors stated how quantitative and qualitative methods will be combined?
- Have the authors considered how they will respond if challenges emerge during the evaluation - for example, if serious implementation failures are identified which need deeper investigation?

Resource considerations in collecting/analysing process data

- Who will lead or conduct the process evaluation? Do they have, or have direct access to, appropriate expertise and experience?
- Does the research team have sufficient expertise in quantitative and qualitative methods, and relevant social science theory?
- Is sufficient time, funding and staff resource included for data collection, analysis (including sufficient time to conduct good quality analysis of qualitative data, with quality checks by a second coder where appropriate) and reporting?

Analysis and reporting

- Has consideration been given to the use of quantitative process measures for modelling variations in outcomes and/or cost-effectiveness?
- Is the relationship between qualitative data components and outcomes and/or cost-effectiveness analysis clear?
- Is there a coherent strategy for dissemination to an academic audience and wider stakeholders?

SECTION C – CASE STUDIES

This section provides case studies of process evaluations of complex interventions. Members of the author group completed a template on the context, methods, findings, and strengths and weaknesses of their particular process evaluation, and were asked to reflect on what they would do differently next time. Key themes were identified, which were then discussed by the group in order to shape recommendations. The case studies relate to the following interventions (initials indicate the member(s) of the author group involved):

1. A Stop Smoking in Schools Trial (ASSIST; SA, LM)
2. The Southampton Initiative for Health (JB, MEB, TT)
3. Sexual Health And RElationships (SHARE; DW)
4. ProActive: a physical activity intervention (WH)
5. Welsh National Exercise Referral Scheme (NERS; GM, LM)
6. MEMA kwa Vijana (Good things for young people; DW)

For readers wanting more detail, references to all published articles on the relevant interventions are provided. While we acknowledge a degree of similarity between some of these interventions, the case studies illustrate some different approaches taken to conducting process evaluations of somewhat similar interventions.

Case Study 1. A Stop Smoking in Schools Trial (ASSIST)

Study context

The intervention

A school-based, peer-led intervention to reduce adolescent smoking, based on diffusion of innovations theory. Influential students aged 12 to 13 years were identified by their peers using a nomination questionnaire. Nominated students were invited to train as ‘peer supporters’ and diffuse the smoke-free message to members of their year group. Further training and support were provided for the peer supporters at four school-based sessions who recorded brief details of their conversations in a diary. After ten weeks, they were thanked for their efforts, and received certificates of participation and gift vouchers.

Nature of evaluation (e.g. feasibility study, effectiveness evaluation, policy trial)

Pragmatic cluster RCT in 59 schools (10,730 students at baseline) incorporating process evaluation, cost-effectiveness evaluation and social network analysis.

Study overview

Theoretical framework used to guide process evaluation design

The design was influenced by Steckler and Linnan's process evaluation framework (2002) and Pawson and Tilley's work on realist evaluation (1997).

Methods used

Process data were collected systematically from a 'training the trainers' event; throughout the identification, training and support of the peer supporters throughout the trial. Data were collected from teachers, trainers, peer supporters and other students. Methods included participant and non-participant observation, behavioural questionnaires, evaluation forms, interviews and focus groups.

Key findings

The peer nomination process

The process evaluation highlighted the need to ensure school staff understood the importance of the peer nomination process and of allowing the nominated students, even those perceived as challenging or disaffected, to train as peer supporters.

School recruitment, retention and response rates

Aged 12 to 13 years was considered suitable for the intervention. Teachers welcomed the use of external trainers, suggesting that this relieved the burden on teaching staff and created additional interest amongst the students, and that young people may be reluctant to discuss smoking behaviour with teachers.

Fidelity

Each stage of the intervention was delivered in every intervention school, the overall desired peer supporter recruitment levels were reached across the study, and attrition was low. However, this general picture masked potentially important variation: targets for peer supporter recruitment were not met in all schools, and competition for available space, and the requirements of the school curriculum and timetable, impacted on fidelity and 'dose'. Recommendations in relation to these issues were made for wider implementation.

Challenges and solutions

Involving policy/practice stakeholders

The ASSIST health promotion trainers contributed their expertise to the design and implementation of the intervention, but there was little additional involvement of policy or practice stakeholders in intervention development. Lack of external stakeholder involvement may have resulted to some resistance by practitioners to take up the intervention during wider implementation.

Deciding on research questions

The research team tabulated the main stages of the intervention and outcome evaluation, and identified process questions for key stages. These related to: context (neighbourhood and school environment, smoking policies and procedures); practical arrangements (venues, timing, staff ratios); variations in content and style of delivery; interactions between participants (students, trainers, peer supporters, school staff); responses of participants to the intervention; activities of peer supporters (diffusion of smoke-free message); and students' views and experiences of the intervention.

Designing the process evaluation

A sub-group of the research team took responsibility for developing the process evaluation but all members of the multi-disciplinary team were encouraged to comment. Process evaluation methods and tools were tested during the pilot phase of the trial. In the main trial, two researchers worked part-time on the process evaluation. Because of limited time and resources, the health promotion trainers played an important role in collecting process data. A training session was organised before the main trial started, at which the role of the trainers in collecting and providing routine process data was agreed.

Collecting process data

The health promotion trainers administered self-complete questionnaires to peer supporters and teachers throughout the intervention, and also completed questionnaires about their own experiences. The two process evaluation researchers collated these data from the 30 schools that received the intervention and conducted in-depth research in four intervention schools, including observations, interviews and focus groups.

Potential Hawthorne effects

Observations, interviews and focus groups had the potential to influence the peer supporters' commitment and behaviour. Focusing in-depth process evaluation in four intervention schools, and conducting interviews and focus groups post-intervention, limited this effect.

During the trial, the need for a standardised intervention predominated and restricted opportunities for the trainers to modify the intervention to suit different groups, as they would usually. However, the process evaluation provided opportunities (evaluation forms, post-intervention interviews) for the trainers to contribute professional opinions for the wider implementation of ASSIST.

Overlapping roles

The health promotion trainers provided data about how they implemented the intervention and how participants responded to it. Emphasising that individual trainers were not being ‘judged’, but data were being sought to improve wider implementation, appeared to reduce the potential for unrealistically positive appraisal.

Distinguishing between intervention and evaluation

The team agreed that the peer supporter diaries were part of the intervention, and that other data were required to examine peer supporters’ activities. Qualitative analyses supported this decision: peer supporters indicated they had ‘made up’ some conversations and forgotten to include others.

Analysis of process data

A data analysis group, with qualitative and quantitative research skills, developed a data management and analysis plan. Descriptive statistics were compiled from attendance registers and questionnaires. Interviews and focus group recordings were fully transcribed and analysed using constant comparison from grounded theory.

Reporting process data

Numerous presentations have been made to academic and practitioner audiences, and peer-reviewed papers published. Training manuals have been produced to guide wider implementation: *Training the Trainers* and *The ASSIST Programme Manual*.

Integration of process and outcomes data

Key process data were analysed and published prior to the main outcome evaluation to reduce bias from knowledge of outcomes. Subsequent papers have referred to both process and outcome data.

Interpreting the primary outcome

Outcome data showed reductions in smoking uptake amongst occasional and experimental smokers, but little impact on students who were regular smokers at baseline. Process data

offer an explanation: in protecting themselves from potential derision or hostility, peer supporters concentrated their attention on peers who they felt could be persuaded not to take up regular smoking, rather than those they considered already ‘addicted’ or members of smoking cliques.

Strengths of this process evaluation

- The pilot phase allowed testing and refinement of the process evaluation before the main trial.
- The process evaluation was considered integral to the RCT and not an ‘add-on’;
- Data were collected systematically across the trial at each stage, with in-depth evaluation at selected sites to examine the process in more detail;
- Mixed methods complemented one another;
- Core programme components were identified, which formed the basis of manuals for wider implementation.

What we would do differently next time

- Greater involvement of policy/practice stakeholders throughout the design and implementation of the intervention;
- A larger budget allocated to the process evaluation to reduce the blurring of roles within the team;
- The development and use of a logic model may have better illustrated the underpinning theory.

Key references

- Audrey, S., Holliday, J. and Campbell, R. (2006) Teachers' perspectives on the implementation of an effective school-based, peer-led smoking intervention, *Health Education Journal*, 67, 74-90.
- Audrey, S., Holliday, J. and Campbell, R. (2006) It's good to talk: An adolescent perspective of talking to their friends about being smoke-free, *Social Science and Medicine*, 63, 2, 320-344.
- Audrey, S., Holliday, J., Parry-Langdon, N. and Campbell, R. (2006) Meeting the challenges of implementing process evaluation within randomised controlled trials: the example of ASSIST (A Stop Smoking In Schools Trial), *Health Education Research*, 21, 366-377.

- Audrey, S., Cordall, K., Moore, L., Cohen, D. and Campbell, R. (2004) The development and implementation of a peer-led intervention to prevent smoking among secondary school students using their established social networks, *Health Education Journal*, 63, 266-284.
- Holliday, J., Audrey, S., Moore, L., Parry-Langdon, N., Campbell, R. and Holliday, J. (2009) High fidelity: How should we consider variations in the delivery of school-based health promotion interventions?, *Health Education Journal*, 68, 44-62.
- Starkey, F., Audrey, S., Holliday, J., Moore, L. and Campbell, R. (2009) Identifying influential young people to undertake effective peer-led health promotion: the example of A Stop Smoking in Schools Trial (ASSIST), *Health Education Research*, 24, 977-988.
- Campbell, R., Starkey, F., Holliday, J., Audrey, S., Bloor, M., Parry- Langdon, N., Hughes, R. and Moore, L. (2008) An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial, *The Lancet*, 371, 1595-1602.

See also: <http://decipher.uk.net/en/content/cms/research/research-projects/assist/>

Case Study 2. Southampton Initiative for Health

Study context

The intervention

The Southampton Initiative for Health aimed to improve diets and physical activity levels of women from disadvantaged backgrounds, and thereby improve the growth and development of their children. The intervention trained Sure Start Children's Centre (SSCC) staff, who work with women and children from disadvantaged families, in behaviour change techniques, equipping staff with 'healthy conversation skills'. These help women address barriers to behaviour change, and set goals for themselves, increasing self-efficacy and sense of control.

The five core skills are:

1. To be able to identify and create opportunities to hold 'healthy conversations';
2. To use open discovery questions;
3. To reflect on practice;
4. To listen rather than provide information;

5. To support goal setting through SMARTER planning (Specific, Measurable, Action-oriented, Realistic, Timed, Evaluated, Reviewed goals).

Training was delivered by a team of researchers experienced in group work and behaviour change. Trainees then received ongoing support, including a phone call from one of the trainers to find out how skills were being implemented in practice, and attended a workshop approximately three months after training.

Nature of evaluation (e.g. feasibility study, effectiveness evaluation, policy trial)

Before and after non-randomised controlled trial (definitive effectiveness study).

Study overview

Theoretical framework used to guide design

Process and outcome evaluation design were guided by a logic model. The process evaluation drew on Kirkpatrick's evaluation model, which describes four levels specifically adapted for the evaluation of training programmes. These four levels, and how they translate to this intervention, are as follows: 1. reaction (the initial response from participating staff), 2. learning (knowledge of the new skills in which they are being trained), 3. behaviour (use of the new skills in practice), and 4. whether the new skills change behaviour in the women with whom the staff work at SSCCs. This last level concerns outcome rather than process so will not be discussed here.

Methods used

Training impact:

- Short term: Before and after training, trainees were asked to write down the first thing they would say in response to four written statements relating to diet and physical activity challenges from SSCC clients. Responses were scored on the extent to which they allowed a client to reflect on their issues and identify a solution.
- Medium term: Post-training, trainees were telephoned and asked to relate behaviour change conversations they had had since the training. Conversations were coded for evidence of competency in key skills.
- Long term: Researchers observed conversations between practitioners and clients at SSCC groups a year after training. Observations were made of trained staff in

Southampton, and were compared with observations of staff in SSCCs in control areas. A pro forma was used to capture date, time and location of each group and use of 'healthy conversation skills' by staff.

Implementation success:

- Reach: This was assessed by looking at the number of training courses held over the intervention period, and the proportion of eligible staff that completed the training.
- Fidelity: Observations were made during training sessions of the number of times the trainer(s) modelled key skills.
- Acceptability: Questionnaires and focus group discussions were used to assess the value the trainees placed on the training.

Key findings

Reach

Twenty-two courses of 'healthy conversation skills' training, each consisting of three sessions, were delivered to 148 practitioners working across 14 SSCCs in Southampton, a reach of 70%.

Fidelity

One complete training programme was observed in November 2010. The majority of each training session (76% overall) was spent with the trainees involved in activities or speaking. Thirty-one percent of the time trainers were speaking, they modelled the use of open discovery questions.

Acceptability

Trainees gave training a median value rating of eight out of ten (IQR 7-9), suggesting they had found it valuable. Content analysis of phone call transcripts showed that 84% of trainees gave positive feedback. Focus groups with trainees were carried out; the data are being analysed.

Training impact

Short term:

Numbers of 'open discovery questions' used by trainees increased significantly after training, from 16 to 321, and instances of giving information or making suggestions decreased from 428 to 130 ($p < 0.001$ for both). The number of trainees using open discovery questions

increased from 13 to 114, and 78% of the trainees who had used no open discovery questions before training used at least one afterwards.

Medium term:

The median overall competence rating on key skills for the whole group of 139 trainees was 55% (IQR 35-70). Trainees had moderate to high levels of skill in finding opportunities to have healthy conversations (median score 3, IQR 2-3) and in using open discovery questions (median score 2, IQR 0-4); however, these conversations did not include SMARTER goal-setting (median score 1, IQR 0-2).

Long term:

One year after training, 168 conversations with clients were observed, involving 70 trainees at 12 SSCCs in Southampton. At the same time, 89 conversations were observed involving 41 practitioners at ten SSCCs in control areas, where staff had not been trained. Skills were used significantly more in conversations conducted by practitioners in the intervention site than they were by those in the control site.

Challenges and solutions

Involving policy/practice stakeholders

The intervention was developed in consultation with local policy and practice stakeholders (City Council, PCT, Sure Start Children's Centres). Research findings were discussed with stakeholders at regular intervals throughout the study.

Deciding on research questions

The research questions for the process evaluation were generated by the logic model. We could not meaningfully assess the impact of the training programme on the health behaviour of women attending SSCCs without first establishing whether the training had been successful in changing SSCC staff practice.

Designing the process evaluation

We designed the process evaluation to answer the key questions, using appropriate methods. For example, to assess the impact of training on staff practice we used a questionnaire to assess immediate reaction to training, analysis of transcripts of phone calls to assess competency, and observation to assess long-term change in use of skills after training.

Collecting process data

We collected process data throughout the delivery of the training intervention to determine acceptability, fidelity and reach of the training.

Analysis of process data

Quantitative data relating to reach, fidelity, acceptability, and change in staff practice over time will be re-examined in the light of qualitative data from the focus groups.

Reporting process data

Findings of the process evaluation were reported to local practitioners and stakeholders throughout the study, and have been disseminated to academic audiences via national and international conferences. Two papers on the process evaluation have been published in peer-reviewed journals, and a further two are currently under review.

Integration of process and outcomes data

The analysis of process data was completed before the analysis of outcome data. Process data showed that the training intervention had been successful in changing staff practice, and enabled us to understand the role of exposure to the intervention in determining outcomes.

Strengths of this process evaluation

- We used mixed methods to assess process and change at various stages in the study, as set out by our logic model.
- Kirkpatrick's evaluation model allowed us to structure the process evaluation appropriately.
- Qualitative data gave us insight into the experiences of staff during and following the training.

What we would do differently next time

We would carry out additional formative work before introducing the intervention to understand better how women use Sure Start Children's Centres because our assumptions about Sure Start attendance and exposure to trained staff were inaccurate.

Key references

Barker, M., Baird, J., Lawrence, W., Jarman, M., Black, C., Barnard, K., Cradock, S., Davies, J., Margetts, B., Inskip, H. and Cooper, C. (2011) The Southampton Initiative for Health: a complex intervention to improve the diets and increase the physical activity levels of women from disadvantaged communities, *Journal of Health Psychology*, 16, 178-191.

- Tinati, T., Lawrence, W., Ntani, G., Black, C., Cradock, S., Jarman, M., Pease, A., Inskip, H., Cooper, C., Baird, J. and Barker, M. (2012) Implementation of new skills to support lifestyle changes – what helps and what hinders?, *Health and Social Care in the Community*, 20, 4, 430-7.
- Black, C., Lawrence, W., Cradock, S., Ntani, G., Tinati, T., Jarman, M., Inskip, H., Cooper, C., Barker, M. and Baird, J. (2012) Healthy Conversation Skills: increasing confidence and competence in front line staff, *Public Health Nutrition*, Sep 19, 1-8 (epub ahead of print).
- Lawrence W., Black C., Tinati T., Cradock S., Begum R., Jarman M., Pease A., Margetts B., Davies J., Inskip H., Cooper C., Baird J., Barker M. ‘Making every contact count’: longitudinal evaluation of the impact of training in behaviour change on the work of health and social care practitioners. *Journal of Health Psychology* (in press)

Case Study 3. Sexual Health and Relationships: Safe, Happy and REsponsible

(SHARE)

Study context

The intervention

SHARE (Sexual Health and Relationships: Safe, Happy and REsponsible) was a 20 session teacher-delivered sex education programme for 13-15-year-olds. Its theoretical bases were the theory of planned behaviour (TPB), interactionism, and sociological analysis of gender. It aimed to develop practical knowledge, change attitudes and, unlike conventional school sex education, develop sexual negotiation and condom use skills, primarily through interactive video. Teachers received five days intensive training. SHARE was intended to improve the quality of sexual relationships and reduce unsafe sex and unwanted pregnancies. It was developed over two years, involving two pilots each in four schools, and extensive consultation with practitioners, researchers and the education establishment.

Nature of evaluation (e.g. feasibility study, effectiveness evaluation, policy trial)

This process evaluation complemented a phase 3 RCT (effectiveness evaluation) involving 25 Scottish secondary schools and 7,630 pupils. Outcomes at six months post-intervention were questionnaire self-reports, and at four years post-intervention were NHS data on conceptions and terminations.

Study overview

Theoretical framework used to guide design

The key research questions came from the theoretical bases of the intervention, and issues emerging from the formative evaluation and from other sexual health interventions. During the trial *Realistic Evaluation* (see Chapter 2) was published. This provided a theoretical rationale for our pragmatic research questions, and confirmed our focus on context, and on pupils' differential responses to the intervention.

Methods used

Pupil and teacher data were collected at three levels – some data were collected in all schools, some in half the schools, and some in four case study schools. In all schools, a longitudinal pupil survey generated quantitative data on both outcomes and processes, including: the extent and quality of implementation; how the programme was received; indicators on the intended causal pathway (knowledge, attitudes, intentions, etc.); and contextual factors (socio-economic and family circumstances, peer groups, attitudes to school, etc.). Head teachers and Heads of Guidance were interviewed in all schools, and all third- and fourth-year sex education teachers completed one questionnaire on their confidence and experience, and another on the sex education lessons they taught.

In each of half the schools, three teachers were interviewed and for two teachers two successive lessons were observed.

In each of the four case study schools, additional research involved three further teacher interviews, and four group discussions and 16 interviews with pupils. These included 'high-' and 'low-achieving' pupils, both male and female.

In the intervention arm, at least one day of each teacher training course was observed. Numerous informal conversations with the main teacher trainer were noted. Teachers were asked to complete the questionnaire on confidence and perceived skills in delivering sex education before training, immediately after training, and then two years later.

Key findings

Compared with controls, SHARE recipients had greater practical sexual health knowledge and less regret of sexual relationships, but no change in any reported behaviours or in

conception or termination rates. Process evaluation findings suggested several factors contributing to the lack of behavioural impact:

Context

Faithful delivery was facilitated by SHARE's compatibility with existing Personal and Social Education (PSE) provision and senior management support, but hindered by scarce curriculum time, brevity of lessons, and low priority accorded to PSE. Pupils' engagement was undermined by the low importance they attached to PSE, and girls in particular had more influential sources of knowledge, values and advice.

Mechanism

The SHARE teacher training was weakest in convincing teachers of the theoretical basis of the intervention, meaning some hardly used role-play and hence gave weak behavioural messages. Some were still not confident in delivery, and 20 sessions was probably too few to overcome long-term, pervasive influences on pupils (e.g. family, friends, local culture and media). Overall there was very little change in cognitions predicting behaviour change, .

Differential response

Given the motivational requirement to learn new skills, SHARE would probably be most effective just at the start of sexual relationships, but highly heterogeneous relationship experience within school classes prevented optimum delivery for most. Additionally, the most alienated pupils, who were least influenced by teachers, had the most risky sexual lifestyles. Sufficient data in each arm of the trial were collected on the extent and quality of sex education for an on-treatment analysis; this had the same results as the intention-to-treat analysis.

Challenges and solutions

Involving policy /practice stakeholders

Some health promotion officials were antagonistic to rigorous outcome evaluations and tried to discourage schools from participating. This was overcome by liaising directly with head teachers, and offering teacher training costs to intervention schools and the equivalent (£3,000) to control schools.

The Scottish educational and health promotion establishments were reluctant to acknowledge that, although probably the best sex education available, SHARE did not influence behaviour.

Deciding on research questions

Four factors were considered critical to interpreting trial outcomes: the extent and quality of delivery (collecting sufficient data for an on-treatment analysis); how sex education was received by different kinds of pupils; the intended causal mechanism; and the school and wider context.

Designing the process evaluation

The main objectives and methods of the process evaluation were outlined in the funding proposal to the MRC. Subsequently, a subgroup of co-investigators and researchers developed the design, taking into account schools' levels of tolerance for data collection. This was endorsed by the trial advisory committee. The tension between breadth/representativeness and depth was resolved by collecting data at three levels: in all schools, half the schools and four case study schools. The design allowed for extra data collection in response to emerging issues.

Collecting process data

Although schools initially committed to the planned data collection, in practice it was difficult getting teachers to systematically detail their sex education delivery. Some teachers were resistant to lesson observations, with some refusing altogether. This was addressed by ensuring total anonymity, and this process data was probably the most useful. Some pupils, particularly younger boys, were very awkward discussing sex, and the need to fit each interview into one school period hindered rapport. Where possible we organised interviews and group discussions in double periods.

Analysis of process data

Researchers' involvement in developing an intervention is likely to bias them towards demonstrating effectiveness. Conversely, it might lead them to have higher standards of programme fidelity than disinterested researchers. Two process evaluation researchers had been involved in SHARE's development, one leading it. This conflict of interest was made explicit to the other researchers, to alert them to possible bias. The large volume of data was a problem in itself, given the limited number of evaluation staff, and the considerable time required to analyse qualitative data. More strategic prioritisation of topics for analysis and writing might have helped.

Reporting process data

Process data were analysed, presented at conferences and written up for publication from the start of the trial. Many findings were disseminated before the analysis of outcomes.

Integration of process and outcomes data

Unless the process or outcome data are privileged in epistemological terms, each is likely to influence the analysis of the other. Since process data were largely qualitative the authors felt that there was a greater danger of bias in interpreting them, so ideally they would be analysed first. Several process issues were written up prior to analysing the outcomes, but many were not. Subsequently, the survey data were analysed to examine how contextual factors shaped the outcomes, such as family influences and school effects.

Strengths of this process evaluation

- Quantitative and qualitative data on the extent and quality of sex education delivery in both arms of the trial allowed an on-treatment analysis of the outcomes;
- A combination of self-reported and observational data provided rich, highly valid data to evaluate the teacher training component of the intervention;
- In-depth interviews and a wide range of questionnaire measures provided information on numerous contextual factors affecting pupils' sexual behaviour and engagement with sex education;
- Fairly good data were collected on all the main stages along the intended causal pathway.

What we would do differently next time

We would attempt to focus data collection more precisely on key issues that help interpret the outcomes, be more realistic about the staff time available, and develop an analysis plan at the outset. Prior to analysing outcomes by arm of the trial we would publish a process evaluation report on how far each stage of the causal mechanism operated as intended, how the intervention was received by different subgroups, and the contextual factors facilitating or hindering it. However, this would require more process evaluation researcher time to avoid delaying analysis of outcomes.

Key references

Process evaluation

- Buston, K. and Wight, D. (2006) The salience and utility of school sex education to young men, *Sex Education*, 6, 2, 135–150.
- Wight, D. and Dixon, H. (2004) SHARE: The rationale, principles and content of a research-based teacher-led sex education programme, *Education and Health*, 22, 1, 3-7.
- Buston, K. and Wight, D. (2004) Pupils' participation in sex education lessons, *Sex Education*, 4, 3, 285-301.
- Wight, D. and Obasi, A. (2003) 'Unpacking the Black Box: the importance of process data to explain outcomes' in *Effective Sexual Health Intervention*, Eds. Stephenson, J., Imrie, J. and Bonell, C. pp.151-66, Oxford University Press.
- Wight, D. and Buston, K. (2003) Meeting needs but not changing goals: evaluation of inservice teacher training for sex education, *Oxford Review of Education*, 29, 4, 521-43.
- Buston, K. and Wight, D. (2002) The salience and utility of school sex education to young women, *Sex Education*, 2, 3, 233-50.
- Buston, K., Wight, D. and Hart, G. (2002) Inside the sex education classroom: the importance of class context in engaging pupils, *Culture, Health and Sexuality*, 4, 3, 317-36.
- Buston, K., Wight, D., Hart, G. and Scott, S. (2002) Implementation of a teacher-delivered sex education programme: obstacles and facilitating factors, *Health Education Research*, 17, 1, 59-72.

Outcomes

- Henderson, M., Wight, D., Raab, G., Abraham, C., Parkes, A., Scott, S. and Hart, G. (2007) Impact of a theoretically based sex education programme (SHARE) delivered by teachers on NHS registered conceptions and terminations, *British Medical Journal*, 334, 133-136.
- Wight, D., Raab, G., Henderson, M., Abraham, C., Buston, K., Hart, G. and Scott, S. (2002) The limits of teacher-delivered sex education: interim behavioural outcomes from a randomised trial, *British Medical Journal*, 324, 1430-33.

See also: <http://share.sphsu.mrc.ac.uk/>

Case Study 4. ProActive: a physical activity intervention

Study context

The intervention

ProActive was a theory-based intervention to increase everyday physical activity among sedentary adults at risk of type 2 diabetes, delivered by trained facilitators (Williams et al. 2004 ; Kinmonth et al. 2008). The intervention was developed and piloted over three years. The hypothesised mechanism of effect was based on the theory of planned behaviour (TPB). Behaviour change techniques (BCTs) were mapped onto the TPB, self-regulation theory, relapse prevention theory, and operant theory.

Nature of evaluation (e.g. feasibility study, effectiveness evaluation, policy trial)

Randomised controlled trial with individual randomisation. We recruited 365 adults (30-50 years) through primary care registers. All participants received a theory-based leaflet encouraging increases in physical activity (comparison arm). Two-thirds of participants also received a one-year intervention delivered either at home and by phone, or by phone and post. The primary outcome was physical activity, assessed objectively. The process evaluation was an integral component of the trial.

Study overview

Theoretical framework used to guide design

We developed a causal modelling approach which informed choice of process measures and hypothesised causal model: from behavioural determinants based on the TPB, through behaviour, to physiological/biochemical variables and health and disease outcomes. BCTs, outcome and process measures were mapped onto this model [3]. Epidemiological studies by our group informed the choice of target behaviour (everyday physical activity) and target group (an easily identifiable group in primary care who were at risk of the consequences of being inactive and/or becoming obese). Fidelity assessment was informed by the treatment fidelity framework by Bellg et al. (2004) .

Methods used

The process evaluation focused on explaining any intervention effects, and illuminating how the intervention was implemented and received in practice.

- Theory-based mediators of intervention effects (trial analysis) and predictors of physical activity (cohort analysis) were assessed by questionnaire, including a TPB questionnaire to assess beliefs about becoming more physically active.
- Intervention delivery: facilitators recorded all contacts on standardised forms. We assessed BCT use by facilitators from 108 transcribed sessions with 27 out of the 244 intervention participants (Hardeman et al. 2005).
- Intervention receipt: participants' acceptability and satisfaction were assessed by questionnaire. Participant talk in 108 sessions was coded under 17 theoretical components (Hardeman et al. 2008).
- Participants' use of skills taught in the intervention (enactment) was assessed by questionnaire.

Key findings

The intervention was no more effective at increasing objectively measured physical activity than the leaflet alone. One explanation was the substantial increase in activity in the entire trial cohort, by an average of 20 minutes of brisk walking per day. This may have been due in part to repeated intensive measurement and trial participation combined with the leaflet. The process evaluation provided three additional explanations:

1. Small effects on mediators: intervention effects on beliefs were limited in size, and mostly found only at six months and not sustained at 12 months (Michie et al. 2008).
2. Insufficient intervention delivery: the fidelity study showed that on average 44% of the intervention package was delivered (Hardeman et al. 2005).
3. Unimportant intervention targets: in this cohort, participants' beliefs, targeted by the intervention, failed to predict physical activity levels and change (Hardeman et al. 2011).

Challenges and solutions

Involving policy / practice stakeholders

Policy and practice stakeholders were involved in all development phases. Their involvement during trial evaluation was limited.

Deciding on research questions

The hypothesised causal model informed *a priori* research questions for outcome and process evaluation. The research team realised the importance of assessing intervention delivery and

receipt (fidelity) during trial evaluation, and obtained additional funding for an independent assessment.

Designing the process evaluation

Design was informed by the hypothesised causal model. Questionnaire burden was reduced by prioritising measures along the hypothesised causal pathway. Questionnaire-based measures were collected at baseline, and at six and 12 months. Suitable coding frames for fidelity assessment were unavailable; we developed coding frames tailored to our intervention. Sensitivities to being tape-recorded were addressed by frequent discussions with facilitators. An independent expert conducted the fidelity assessment to reduce bias. A small sample was selected for tape-recording due to feasibility (time, resources) and acceptability.

Collecting process data

Baseline and 12-month measurement were conducted at the measurement centre; six-month questionnaires were completed at home. Collecting fidelity data was challenging as four tape-recorded sessions were required for each participant. Twenty-five out of the 52 participants had incomplete sets of tapes due to forgetting to tape-record, equipment failure, no consent, dropouts, or transfer to another facilitator.

Analysis of process data

Analysis of process measures was specified in a written plan. No guidance was available for analysis of large amounts of fidelity data (208 facilitator behaviours across four sessions) collected from a small sample of participants assigned to selected facilitator. We resolved this by mapping behaviours onto 14 BCTs, followed by independent validation. A similar approach was followed for analysis of participant talk. Frequent consultation with trial statisticians was vital.

Reporting process data

The main trial paper reported intervention effects on questionnaire-based process measures. Due to insufficient journal space and a range of research questions, most process data were reported in subsequent publications. Findings from the process evaluation were also presented at national and international conferences, and disseminated in newsletters for practices and participants.

Integration of process and outcomes data

Intervention effects on process measures were analysed as part of trial analysis. The trial results informed research questions for further analysis of process data, and additional

funding was sought. Integration of fidelity data with outcome data was limited by the small sample of participants, owing to limited time and resources for fidelity assessment.

Strengths of this process evaluation

Process evaluation was based on an explicit, *a priori* hypothesised causal model; this helped to illuminate various explanations for trial results. Process evaluation generated methodological developments which can be generalised to other complex interventions: a causal modelling approach for intervention development and evaluation, and methods for theory-based fidelity assessment. The fidelity study highlighted the complexity of intervention content and associated challenges in delivery, and assessed delivery of BCTs as the active content of interventions.

What we would do differently next time

Qualitative research in parallel with the trial might have been beneficial. For instance, interviews with participants immediately after the 12-month follow-up may have identified reasons for the substantial increase in physical activity in the cohort, and lack of intervention effect. Interviews with facilitators may have provided insight into factors influencing fidelity and potential solutions. Intervention delivery and fidelity assessment could benefit from an *a priori* definition of core, optional and proscribed intervention components. Fidelity assessment would ideally be conducted across the whole period of intervention delivery to facilitate consistent delivery over time and across facilitators. Finally, although we conducted a large pilot study which identified salient beliefs and predictors of intentions (Sutton et al. 2003), very few variables predicted physical activity change (Simmons et al. 2010). Consideration of a broader range of predictors and follow-up measure of behaviour might have identified other key intervention targets.

Key references

- Williams K, Prevost T, Griffin S, Hardeman W, Hollingworth W, Spiegelhalter D, Sutton S, Ekelund U, Wareham NJ, and Kinmonth AL (2004). Randomised controlled trial of the efficacy of a family-based, domiciliary intervention programme to increase physical activity among individuals at high risk of diabetes [ISRCTN 61323766]. *BMC Public Health*, 4, 48-84.
- Kinmonth AL, Wareham NJ, Hardeman W, Sutton S, Prevost AT, Fanshawe T, Williams K, Ekelund U, and Griffin SG (2008). Efficacy of a theory-based behavioural intervention

- to increase physical activity in an at-risk group in primary care *ProActive (UK)*: A randomised trial. *The Lancet*, 371, 41-48.
- Hardeman W, Sutton S, Griffin S, Johnston M, White AJ, Wareham NJ, and Kinmonth AL (2005). A causal modelling approach to the development of theory-based behaviour change programmes for trial evaluation. *Health Education Research*, 20, 6, 676-687.
- 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 Behavior Change Consortium. *Health Psychology* 2004, 23: 443-451.
- Hardeman W, Michie S, Fanshawe T, Prevost AT, McLoughlin K, and Kinmonth AL. Fidelity of delivery of a physical activity intervention: Predictors and consequences (2008). *Psychology & Health*, 23, 1, 11-24.
- Michie S, Hardeman W, Fanshawe, T Prevost T, Taylor L, and Kinmonth AL. Investigating theoretical explanations for behaviour change: The case study of *Proactive* (2008). *Psychology & Health*, 23, 1, 25-39
- Hardeman W, Kinmonth AL, Michie S, Sutton S on behalf of the *ProActive* Project team (2009). Impact of a physical activity intervention program on cognitive predictors of behaviour among adults at risk of Type 2 diabetes (ProActive randomised controlled trial). *The International Journal of Behavioral Nutrition and Physical Activity*, 6: 16.
- Hardeman W, Kinmonth AL, Michie S, Sutton S on behalf of the *ProActive* project team (2011). Theory of Planned Behaviour cognitions do not predict self-reported or objective physical activity levels or change in the *ProActive* Trial. *British Journal of Health Psychology* 16, 135–150.
- Sutton S, French D, Hennings SJ, Wareham NJ, Griffin S, Hardeman W, Kinmonth AL (2003). Eliciting salient beliefs in research on the Theory of Planned Behaviour: The effect of question wording. *Current Psychology*, 22, 3, 234-251.
- Simmons RK, Van Sluijs E, Hardeman W, Sutton S, Griffin S on behalf of the *ProActive* project team (2010). Who will increase their physical activity? Predictors of change in objectively measured physical activity over 12 months in the *ProActive* cohort. *BMC Public Health* 10:226.

Case Study 5. The Welsh National Exercise Referral Scheme (NERS)

Study context

The intervention

Exercise referral scheme, trialled in 12 Welsh local health boards, comprising:

- i) Health professional advice/referral;
- ii) Consultations including motivational interviewing and goal setting;
- iii) A 16-week exercise programme, under the supervision of a qualified professional;
- iv) Reviews of progress at four weeks;
- v) Signposting to exit routes at 16 weeks;
- vi) Reviews of progress at eight and 12 months.

Nature of evaluation (e.g. feasibility study, effectiveness evaluation, policy trial)

Pragmatic randomised controlled trial.

Study overview

Theoretical framework used to guide design

The study was influenced by Steckler and Linnan's process evaluation framework, diffusion of innovations theory and realistic evaluation. The final framework is described below:

Component	Methodological and analytical focus
Programme theory	Clear elicitation of programme theory, in terms of key planned programme components, causal pathways and intended outcomes.
Diffusion	Qualitative exploration of how diffusion activities, contextual factors and actions of implementers shape delivery across local contexts.
Implementation	Quantitative measurement of the consistency of delivery with programme theory, and quantity of intervention delivered.
Participant experience	Qualitative exploration of how the intervention is experienced by patients, and the causal processes through which change is promoted in context.
Reach	Quantitative measurement of patterning in programme reach by patient characteristics, and variability in programme delivery.

Methods

- **Theory** – agreement of logic model with national programme developers.
- **Diffusion** – qualitative interviews with national and local scheme coordinators.

- **Implementation** – fidelity and dose quantified using routine monitoring data, implementer self-reports of whether exercise classes were offered, and structured observation of motivational interviewing.
- **Patient experiences and mechanisms of change** – interviews with professionals delivering the scheme throughout Wales, and patients attending the scheme in six case study centres.
- **Reach** – routine monitoring data and patient demographic information from baseline trial data.

Key findings

- Limitations in communication, support and training had an impact on the fidelity of some components.
- Fidelity checks indicated a common core of discounted, supervised, group exercise, but motivational interviewing and goal setting were delivered poorly.
- Nevertheless, NERS was successful in improving physical activity among patients referred for CHD risk factors, and reducing depression and anxiety.
 - Key mechanisms of change linked to professional support (building confidence in using unfamiliar machinery, and preventing overexertion), and patient-only group exercise environments (providing empathy and realistic role models).
- Qualitative data generated hypotheses regarding patterning of programme reach, which were quantitatively modelled.

Key challenges

Involving policy/practice stakeholders

Policy and practice stakeholders were involved at all stages of evaluation. The evaluation was designed in consultation with the Welsh Government, to whom findings were fed back throughout the evaluation.

Deciding on research questions

Initial research questions initially focused on fidelity and patient experiences of NERS. However, new questions emerged during the course of the evaluation. Notably, early fidelity checks indicated that motivational interviewing was not delivered, leading to additional training courses. Additional data was collected to examine if practice changed. The study

benefited to a certain extent from a degree of flexibility; however, its primary research questions were not sufficiently well defined from the outset.

Designing the process evaluation

The evaluation's design was hampered by the fact that the lead process evaluator (a PhD student) began only after implementation had begun, leading to rushed decisions on aims and methods.

Additionally, it proved challenging to define fidelity. Agreement of a logic model with policy representatives clarified what the intervention was, guiding fidelity checks.

Collecting process data

Quantitative fidelity checks drew upon routine monitoring structures. In order to obtain additional information on fidelity, implementers were requested to provide recordings of consultations. Limited opportunity for piloting meant this took longer than anticipated, and incomplete paperwork sometimes rendered recordings unusable. For the second round of recordings, more time was allowed, and instructions clarified.

The order in which qualitative process evaluation data was collected was determined by pragmatic considerations, and concerns regarding the likelihood of blurring distinction between evaluation and intervention. The value of longitudinal qualitative data in exploring implementation processes over time was weighed up against the risk of overburdening implementers, and Hawthorne effects. Only one interview was conducted with each implementer, after the scheme had been running for 6-12 months, allowing retrospective reflection on diffusion and implementation.

Analysis of process data

The framework developed for this study was sequential, rotating between qualitative exploration of causal processes and quantification of intermediate outcomes. Qualitative data on the diffusion helped predict and explain variation in fidelity, and qualitative data on mechanisms of change were interpreted in light of a quantified programme. Qualitative data, in turn, generated hypotheses regarding socio-demographic patterning in uptake and adherence, which were tested quantitatively using routine monitoring data and baseline trial data

Reporting process data

Findings were disseminated in lay form to implementers via presentations and a government report. These were subsequently divided into multiple journal articles, which cross-referenced one another and the outcomes evaluation. It proved challenging to select the order in which these articles were published, meaning this did not necessarily reflect the sequential analyses described above. Process data were analysed prior to knowledge of outcomes, but the outcomes paper was approved for publication by funders several months before process papers, and was hence published first. This led to some redrafting of the discussion, to include the implications of process data in light of knowledge of outcomes, although analyses were not revisited at this stage. Findings from the study were also disseminated via national and international conferences.

Integration of process and outcomes data

Process data were analysed fully prior to analysis of outcomes data, but not published until afterwards. Hence, while the analysis of process data was not biased by knowledge of outcomes, qualitative findings and trial outcomes were discussed in relation to one another. The potential for integration of fidelity measures into outcomes analysis was limited by the low quality of routine data, and the fact that researcher collected data only collected from small sub-samples.

Strengths of this process evaluation

- Collection of qualitative data from all levels of implementation facilitated in-depth insights.
- The evaluation extended and tailored an existing evaluation framework to fit the context of the policy trial.
- Careful consideration was given to how mixed methods complemented one another, and how elements of the process evaluation framework fitted together to tell the overall story of the intervention.

What we would do differently next time

The main challenges with the design and conduct of this process evaluation emerged from the fact that process evaluation did not begin until the programme was already underway in some areas. This raised the following issues:

- The logic model should have been developed earlier. It was constructed in response to an identified need for focus, rather than being used from the outset as a planning tool.

- While it focused on links between intervention activities and mechanisms of impact, the logic model should have focused more on delivery mechanisms.
- Although the use of routine monitoring data to examine implementation reduced the amount of researcher intrusion into the running of the intervention, better planning would have allowed measures to be put in place to validate these data.
- Pre-planning integration of process measures into outcomes analysis would also have focused attention more on obtaining high quality measures which covered a sufficient sample of cases.
- Earlier definition of core questions would have helped streamline the collection of data.

References

Process evaluation

- Moore, G. F., Raisanen, L., Moore, L., Din, N. U. and Murphy, S. (2013) Mixed-method process evaluation of the Welsh National Exercise Referral Scheme, *Health Education*, 113, 6, 476-501.
- Moore, G. F., Moore, L. and Murphy, S. (2012) Integration of motivational interviewing into the National Exercise Referral Scheme in Wales: a mixed-method study, *Behavioural and Cognitive Psychotherapy*, 40, 3, 313-330.
- Moore, G.F., Moore, L and Murphy, S. (2011) Facilitating adherence to physical activity: exercise professionals' experiences of the National Exercise Referral Scheme in Wales. A qualitative study. *BMC Public Health*, 11, 935.

Trial and cost-effectiveness evaluation

- Murphy, S., Tudor Edwards, R., Williams, N., Raisanen, L., Moore, G. F., Linck, P., Hounscome, N., Ud Din, N. and Moore, L. (2012) An evaluation of the effectiveness and cost effectiveness of the National Exercise Referral Scheme in Wales, UK: a randomised controlled trial of a public health policy initiative, *Journal of Epidemiology & Community Health*, 66, 11, 1082.

See also: <http://www.decipher.uk.net/en/content/cms/research/research-projects/ners-evaluation/>

Case Study 6. MEMA kwa Vijana (Good things for young people)

Study context

The intervention

An adolescent sexual health programme in rural Mwanza, northern Tanzania, consisting of:

- community-based activities;
- a teacher-led, peer-assisted curriculum of 10-15 40-minute sessions per year for the three upper years of primary school;
- training and supervision of health facility clinicians to encourage youth friendliness;
- promotion and sale of subsidised condoms by out-of-school youth.

It was loosely based on social learning/cognitive theory and the theory of reasoned action.

Nature of evaluation (e.g. feasibility study, effectiveness evaluation, policy trial)

Cluster randomised trial from 1998-2002, including 20 clusters, 63 schools and 9,645 young people. The main outcomes were: incidence of HIV, other sexually transmitted infections (STIs) and pregnancy, and self-reported sexual behaviour, attitudes and knowledge.

A follow-up survey was conducted in 2007-8 with 13,814 15-30-year-olds.

Study overview

Theoretical framework used to guide design

An MRC- and DfID-funded process evaluation complemented an RCT of the intervention. Its primary research objectives were to:

- conduct an ethnographic study of young people's sexual relationships,
- establish the risk factors for STIs;
- develop and evaluate quantitative and qualitative methods to study sexual behaviour in rural Africa.

The routine monitoring incorporated into the trial included collecting data on implementation, participant responses, and contextual factors. However, investigating the intended mechanisms was limited by the difficulty of developing valid quantitative measures of cognition.

Methods used

Participant observation took place for 158 person-weeks in nine villages. Four were visited for approximately seven weeks per year from 2000 to 2002 by pairs of researchers. They lived with the family of a same-sex trial participant and accompanied him/her and peers in daily activities. Ninety-two trial participants were also interviewed in-depth: 76 selected randomly, 13 because HIV-positive and three because pregnant while in school. Seventy-one were interviewed twice, two years apart. Six single-sex groups of young villagers each participated in three or four discussions addressing particularly sensitive topics, such as why girls have sex and the range of sexual activities practised. Other data were collected through simulated patient visits to health clinics, group discussions with intervention pupils, observation of training sessions, internal monitoring and evaluation, and annual surveys of implementers. Seven to nine years after being exposed to *MEMA kwa Vijana*, 23 students were followed up to investigate their sexual histories and how they had been influenced by the programme. In addition, extensive quantitative, monitoring data on each intervention component were collected by the implementation team, and process evaluation surveys with medical and education practitioners were conducted by the trial research team.

Key findings

The trial showed improvements in sexual health knowledge, social norms and self-reported behaviour, but no reduction in pregnancy or STIs. The survey in 2007/8 had broadly similar findings.

The process evaluation provided rich quantitative and qualitative data on the extent and quality of programme implementation and young people's engagement with it. The ethnographic study highlighted the importance of contextual factors in the intervention's limited impact. Limited data were collected on the intervention's impact on the theoretical determinants of behaviour change (mechanisms). Key findings in relation to each intervention component are as follows:

Community mobilisation aimed to reduce opposition to the programme. It comprised a one-week visit to each ward for introductory meetings and the creation of a *MEMA kwa Vijana* advisory committee. Few refused to participate, but participant observation suggested that many adults knew very little about the purpose and content of the intervention, and some themes, in particular condom use, were controversial.

Training courses for teachers, class peer educators, health workers and condom promoter-distributors were well implemented. Almost all teachers delivered most of the sessions, some adopted interactive teaching styles, and corporal punishment decreased, especially during intervention sessions. However, some continued non-participatory teaching, corporal punishment or, occasionally, sexual abuse. One of the most important contextual factors shaping the intervention was the conservatism of Tanzanian education – for example, condoms could not be shown or depicted in primary school. Contextual factors also meant girls were less engaged in the intervention than boys, through the paucity of female teachers, girls' inhibition from participating in mixed-sex activities, and concern with 'respectability'. Class peer educators performed the drama serial well, and generally understood the intervention content much better than their classmates. However, they could not answer complex questions, and other pupils sometimes ignored or rejected their opinions. In some villages severe poverty meant the incentives peer-educators received, such as meals or T-shirts, undermined their status as 'peers' and their validity as role models or educators. Most classes visited a health facility as recommended, where condoms were demonstrated and made available. However, some health workers examined all the girls for pregnancy, a practice widespread before the intervention but discouraged by it. Adolescents remained very reluctant to use health services, fearing poor confidentiality and stigma. The simulated patient visits showed that, for those who did, provision was variable and limited, but of higher quality than in control communities.

Few pupils or villagers were aware of *MEMA kwa Vijana* condom promoter-distributors, predominantly male volunteers elected by young villagers. They reported difficulty selling condoms and were rumoured not to use condoms themselves. Due to lack of demand, unsustainability and cost, this component was discontinued. Overall the main barriers to sexual behaviour change seemed to be structural: women's lower social status; their economic dependence on men; sex as an economic resource for women; poor educational and health infrastructure; low salience of HIV; limited agency; low status of youth; masculine esteem of sexual activity; and negative condom beliefs. At an individual level, social norms were extremely important, but these were elements of a broader culture.

Challenges and solutions

Funding

A process evaluation was included in the original trial proposal but removed due to restricted funding and the prioritising of the outcome evaluation. Subsequently, separate funding was sought from DFID and the MRC. Delays in securing this funding meant the qualitative process evaluation could not start until after the intervention had started.

Recruiting skilled research team

As in many low income countries, it was very difficult to recruit suitably experienced researchers, especially since we were far from the commercial capital. The original proposal was for qualitative data to be collected by two Sukuma-speaking graduates (the local language). Despite advertising locally and nationally, and offering high salaries, such graduates could not be recruited, largely due to poor local education. The two most promising Swahili-speaking applicants (the national language) were hired instead, neither of whom had degrees in sociology or anthropology. As they were unable to follow informal Sukuma conversations, or to formally interview villagers who spoke little Swahili, we also recruited two Sukuma speakers. Although the research team were highly committed and intelligent, they did not have the considerable research skills required. They found it challenging to analyse data while collecting them, in order to direct further investigation, or to critically reflect on their own cultural perspectives in order to try and understand their informants' assumptions and world views.

Achieving depth and breadth

The objectives and methods of the process evaluation were outlined in the funding proposals but detailed plans were only developed during the trial by the co-investigators and senior researchers. Quantitative monitoring data were collected systematically from all schools and health facilities. For qualitative data, the tension between breadth/representativeness and depth was resolved by focusing the participant observation research on three pairs of villages. One control and one intervention village were selected with each of the following characteristics: large roadside village near a gold mine, large roadside fishing village near Lake Victoria, remote dispersed village.

Social desirability bias

Social desirability bias is of particular concern in evaluations where respondents know which activities or attitudes are being promoted. It is even more acute in low income countries

where there may be little awareness of impartial social research, and a culture of showing respect by presenting oneself as different social contexts require. This was evident from triangulation on young people's reported sexual behaviour, which found very poor reliability. It was also likely to have affected the monitoring data collected by the implementation team, since practitioners' relationships with their trainers and supervisors were very likely to have affected how they reported their activities. To address this we maximised the observation of intervention activities, and prioritised these data over self-reported data.

Evaluation as an intervention

There was concern that the evaluative group discussions with teachers might have motivated teachers to improve their delivery. There was no similar professional development in the control group that could have been used to balance the evaluation impact. How much the intervention impact was due to the process of its evaluation could probably only be answered definitively by having four groups, with each allocated to either intervention or non-intervention arms, and to either minimum or maximum evaluation arms.

Analysis of process data

The volume of data collected made it impossible to include all data in every analysis. For some topics, detailed testing of hypotheses was restricted to pre-selected portions of the data. These were selected either as a representative sample of the complete data set or on a theoretical basis, such as exclusively using participant observation data where these were likely to have greater validity.

Sharing findings for ongoing modifications to the intervention

The proposal for the trial only allowed one year for developing and testing the intervention. This involved intensive work to review theoretical bases for behavioural sexual health interventions and existing local and international sexual health interventions, adapt and test existing materials, and develop and test new ones. However, it later became clear that the intervention would have benefited from a better understanding of the social factors shaping sexual health, some of which were identified through the process evaluation. Given the severity of the HIV epidemic it was decided to prioritise intervention improvement over maintaining its consistency, so the intervention was modified as limitations became apparent. For example, the teachers' guide was adapted into separate guides for different age groups, and another intervention component, condom promotion and distribution, was developed.

This was implemented in the second year of the trial.

Integration of process and outcomes data

When the outcome data were analysed, only preliminary analyses of the process evaluation data had been conducted, posing the risk that their interpretation would be biased by knowledge of the trial outcomes. To counter this, prior to un-blinding, the process evaluation research team wrote a report summarising their views of the intervention's processes and impact. Six researchers each independently drafted a brief report based on their impressions and preliminary analyses, before working together to merge these into one document.

Strengths of this process evaluation

A combination of self-reports, observations and direct measurement provided rich, highly valid data to evaluate the different components of the intervention. Carefully selecting villages for participant observation and group discussions, theoretical sampling for in-depth interviews, and systematic time sampling maximised data representativeness while providing sufficient detail to gauge the intervention's salience to people's everyday lives. The field research team of mixed sex and ethnicity reduced the risk of idiosyncratic findings. Writing a process evaluation report prior to analysing outcomes by arm of the trial (discussed above) minimised the risk that the qualitative process data would be interpreted in the light of trial outcomes.

What we would do differently next time

Ideally the ethnographic study of young people's sexual relationships, and thorough analysis of the data, would have preceded and informed the development of the intervention. We could then focus the process evaluation more precisely on key issues that help interpret the outcomes.

We would start by clarifying with the intervention team their theory of change and use this to plan the process evaluation, focusing on key stages in the causal pathway. We would develop an analysis plan at the outset with realistic timescales for analysis of qualitative data.

References

Process evaluation

Obasi, A.I., et al., Rationale and design of the MEMA kwa Vijana adolescent sexual and reproductive health intervention in Mwanza Region, Tanzania. *AIDS Care*, 2006. **18**(4): p.

311-322.

Plummer, M.L., *Promoting abstinence, being faithful and condom use with young rural Africans: findings from a large qualitative study within an intervention trial in Tanzania*. 2012. Lanham, Maryland: Lexington.

Plummer, M.L. and D. Wight, *Young People's Lives and Sexual Relationships in Rural Africa: Findings from a large qualitative study in Tanzania*. 2011. Lanham, Maryland: Lexington.

Plummer, M., et al., A process evaluation of a school-based adolescent sexual health intervention in rural Tanzania: The MEMA kwa Vijana programme. *Health Education Research* 2007. **22** p. 500-512.

Terris-Prestholt, F., et al., Costs of an Adolescent Sexual Health Program in Mwanza, Tanzania. . *Sexually Transmitted Diseases*, 2006. **33**(10): p. S133-39.

Wamoyi, J, Mshana, G, Doyle, A.M, Ross, D.A.. Recall, relevance and application of an in-school sexual and reproductive health intervention 7–9 years later: perspectives of rural Tanzanian young people. *Health Promotion International*, 2012. 28 (3): 311-21.

Wight, D, Plummer, M.L., Ross, D.A.. The need to promote behaviour change at the cultural level: one factor explaining the limited impact of the *MEMA kwa Vijana* adolescent sexual health intervention in rural Tanzania. A process evaluation. *BMC Public Health* 2012. 12:788.

Wight, D. and Obasi, A. ‘Unpacking the Black Box: the importance of process data to explain outcomes.’ In *Effective Sexual Health Interventions*. 2003. Eds J. Stephenson, J. Imrie and C. Bonell, pp.151-66. Oxford University Press.

Outcomes

Doyle, A.M., et al., Long-term biological and behavioural impact of an adolescent sexual health intervention in Tanzania. *PLoS Medicine* 2010. **7**(6).

Larke, N., et al., Impact of MEMA kwa Vijana ...on Use of Health Services by Young People. *Journal Adolescent Health*, 2010. **47**: p. 512-22.

Ross, D.A., et al., Biological and behavioral impact of an adolescent sexual health intervention in Tanzania: a community-randomised trial. *AIDS*, 2007. **21**(14): p. 1943-1955.

See also: www.memakwavijana.org

References

- Akhtar, P.C., Currie, D.B., Currie, C.E. and Haw, S.J. (2007) Changes in child exposure to environmental tobacco smoke (CHETS) study after implementation of smoke-free legislation in Scotland: national cross sectional survey, *BMJ*, 335, 7619, 545.
- Armstrong R, Waters E, Moore L, Riggs E, Cuervo L G, Lumbiganon P, and Hawe P (2008) Improving the reporting of public health intervention research: advancing TREND and CONSORT. *Journal of Public Health*: fdm082.
- Audrey, S., Holliday, J., Parry-Langdon, N. and Campbell, R. (2006) Meeting the challenges of implementing process evaluation within randomized controlled trials: the example of ASSIST (A Stop Smoking in Schools Trial), *Health education research*, 21, 3, 366-377.
- Baranowski, T. and Stables, G. (2000) Process Evaluations of the 5-a-Day Projects, *Health Educ Behav*, 27, 2, 157-166.
- Baron, R.M. and Kenny, D.A. (1986) The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations, *Journal of personality and social psychology*, 51, 6, 1173.
- Basch, C.E., Sliepecevic, E.M., Gold, R.S., Duncan, D.F. and Kolbe, L.J. (1985) Avoiding Type III Errors in Health Education Program Evaluations: A Case Study, *Health Educ Behav*, 12, 3, 315-331.
- Bellg, A.J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D.S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D. and Czajkowski, S. (2004) Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium, *Health Psychology*, 23, 5, 443.
- Berwick, D.M. (2008a) The science of improvement, *Jama-Journal of the American Medical Association*, 299, 10, 1182-1184.
- Berwick, D.M. (2008b) The Science of Improvement, *JAMA*, 299, 10, 1182-1184.
- Blackwood, B., O'Halloran, P. and Porter, S. (2010) On the problems of mixing RCTs with qualitative research: the case of the MRC framework for the evaluation of complex healthcare interventions, *Journal of Research in Nursing*, 15, 6, 511-521.
- Bonell, C., Fletcher, A., Morton, M., Lorenc, T. and Moore, L. (2012) Realist randomised controlled trials: A new approach to evaluating complex public health interventions, *Social Science & Medicine*.
- Bonell, C., Fletcher, A., Morton, M., Lorenc, T. and Moore, L. (2013) Methods don't make assumptions, researchers do: a response to Marchal et al, *Social Science and Medicine*.
- Bonell, C., Fletcher, A., Fitzgerald-Yau, N, Hale, D, Allen, E, Elbourne, D, Jones, R, Bond, L, Wiggins, M, Miners, A, Legood, R, Scott, S, Christie, D, Viner, R, (in press) Initiating change locally in bullying and aggression through the school environment (INCLUSIVE): pilot randomised controlled trial. Health Technology Assessment, in press., *Health Technology Assessment*.
- Bonell, C., Oakley, A., Hargreaves, J., Strange, V. and Rees, R. (2006) Research methodology: Assessment of generalisability in trials of health interventions: suggested framework and systematic review, *BMJ: British Medical Journal*, 333, 7563, 346.
- Bonell, C.P., Hargreaves, J.R., Cousens, S.N., Ross, D.A., Hayes, R.J., Petticrew, M. and Kirkwood, B. (2009) Alternatives to randomisation in the evaluation of public-health interventions: design challenges and solutions, *Journal of Epidemiology and Community Health*.

- Bryman, A. (2006) Integrating quantitative and qualitative research: how is it done?, *Qualitative Research*, 6, 1, 97-113.
- Bumbarger, B. and Perkins, D. (2008) After randomised trials: issues related to dissemination of evidence-based interventions, *Journal of Children's Services*, 3, 2, 55-64.
- Byng, R., Norman, I. and Redfern, S. (2005) Using realistic evaluation to evaluate a practice level intervention to improve primary health care for patients with long-term mental illness, *Evaluation*, 11, 1, 69 - 93.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. and Balain, S. (2007) A conceptual framework for implementation fidelity, *Implementation Science*, 2, 1, 40.
- Clark, A.M., MacIntyre, P.D. and Cruickshank, J. (2007) A critical realist approach to understanding and evaluating heart health programmes, *Health (London)*, 11, 4, 513-539.
- Coffey, A.A. and Atkinson, P. (1996) *Making sense of qualitative data: Complementary research strategies*: Sage.
- Craig, P., Cooper, C., Gunnell, D., Haw, S., Lawson, K., Macintyre, S., Ogilvie, D., Petticrew, M., Reeves, B. and Sutton, M. (2012) Using natural experiments to evaluate population health interventions: new Medical Research Council guidance, *Journal of epidemiology and community health*, 66, 12, 1182-1186.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. and Petticrew, M. (2008a) Developing and evaluating complex interventions: new guidance: Medical Research Council.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. and Petticrew, M. (2008b) Developing and evaluating complex interventions: the new Medical Research Council guidance, *BMJ*, 337, a1655.
- Cresswell, J.W. (2005) Mixed methods research and applications in intervention studies.: ; Cambridge, July 11-12.
- Creswell, J.W. and Clark, P.L.K. (2007) *Designing and conducting mixed methods research*, London: Sage.
- Dane, A.V. and Schneider, B.H. (1998) Program integrity in primary and early secondary prevention: are implementation effects out of control?, *Clinical psychology review*, 18, 1, 23-45.
- Donabedian, A. (1988) The quality of care: How can it be assessed?, *JAMA*, 260, 12, 1743-1748.
- Durlak, J.A. and DuPre, E.P. (2008) Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation, *American journal of community psychology*, 41, 3-4, 327-350.
- Finnegan, J.R.J.R., Murray, D.M., Kurth, C. and McCarthy, P. (1989) Measuring and Tracking Education Program Implementation: The Minnesota Heart Health Program Experience, *Health Educ Behav*, 16, 1, 77-90.
- Firpo-Triplett, R., Fuller, T.R., (2012) A Guide to Adapting Evidence-Based Sexual Health Curricula. General Adaptation Guidance, Scotts Valley, CA: ETR Associates/CDC.
- Fitzgerald-Yau, N., Fletcher, A., Hale, D., Bonell, C. and Viner, R. Creating a restorative culture in schools.
- Gardner, F., Hutchings, J., Bywater, T. and Whitaker, C. (2010) Who benefits and how does it work? Moderators and mediators of outcome in an effectiveness trial of a parenting intervention, *Journal of Clinical Child & Adolescent Psychology*, 39, 4, 568-580.
- Glasgow, R.E., Klesges, L.M., Dzewaltowski, D.A., Estabrooks, P.A. and Vogt, T.M. (2006) Evaluating the impact of health promotion programs: using the RE-AIM framework

- to form summary measures for decision making involving complex issues, *Health Education Research*, 21, 5, 688-694.
- Glasgow, R.E., Lichtenstein, E. and Marcus, A.C. (2003) Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition, *Am J Public Health*, 93, 8, 1261-1267.
- Glasgow, R.E., McKay, H.G., Piette, J.D. and Reynolds, K.D. (2001) The RE-AIM framework for evaluating interventions: what can it tell us about approaches to chronic illness management?, *Patient Education and Counseling*, 44, 2, 119-127.
- Glasgow, R.E., Vogt, T.M. and Boles, S.M. (1999) Evaluating the public health impact of health promotion interventions: the RE-AIM framework, *Am J Public Health*, 89, 9, 1322-1327.
- Glouberman, S. and B. Zimmerman (2002) *Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like? Commission on the Future of Health Care in Canada*, Discussion Paper 8.
- Grant, A., Treweek, S., Dreischulte, T., Foy, R. and Guthrie, B. (2013a) Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting, *Trials*, 14, 1, 15.
- Grant, S.P., Mayo-Wilson, E., Melendez-Torres, G.J. and Montgomery, P. (2013b) Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications, *PloS one*, 8, 5, e65442.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P. and Kyriakidou, O. (2004) Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations, *The Milbank Quarterly*, 82, 4, 581-629.
- Hardeman, W., Sutton, S., Griffin, S., Johnston, M., White, A., Wareham, N.J. and Kinmonth, A.L. (2005) A causal modelling approach to the development of theory-based behaviour change programmes for trial evaluation, *Health education research*, 20, 6, 676-687.
- Hasson, H. (2010) Systematic evaluation of implementation fidelity of complex interventions in health and social care, *Implement Sci*, 5, 1, 67.
- Haw, S.J., Gruer, L., Amos, A., Currie, C., Fischbacher, C., Fong, G.T., Hastings, G., Malam, S., Pell, J. and Scott, C. (2006) Legislation on smoking in enclosed public places in Scotland: how will we evaluate the impact?, *Journal of Public Health*, 28, 1, 24-30.
- Hawe, P., Shiell, A. and Riley, T. (2004) Complex interventions: how "out of control" can a randomised controlled trial be?, *British Medical Journal*, 328, 7455, 1561-1563.
- Hawe, P., Shiell, A. and Riley, T. (2009) Theorising Interventions as Events in Systems, *American Journal of Community Psychology*, 43, 3, 267-276.
- House of Commons Health Committee (2009) Health inequalities: third report of session 2008-9, London: Stationery Office.
- Isaacs, A.J., Critchley, J.A., Tai, S.S., Buckingham, K., Westley, D., Harridge, S.D.R., Smith, C. and Gottlieb, J.M. (2007) Exercise Evaluation Randomised Trial (EXERT): a randomised trial comparing GP referral for leisure centre-based exercise, community-based walking and advice only, *Health Technology Assessment*, 11, 10, 1-+.
- Jackson, D.L. (2010) Reporting results of latent growth modeling and multilevel modeling analyses: Some recommendations for rehabilitation psychology, *Rehabilitation Psychology*, 55, 3, 272.
- Jansen, Y.J.F.M., Foets, M.M.E. and de Bont, A.A. (2010) The contribution of qualitative research to the development of tailor-made community-based interventions in primary care: a review, *The European Journal of Public Health*, 20, 2, 220-226.

- Kellogg Foundation, W.K. (2004) *Logic model development guide*, Battle Creek, MI: W.K Kellogg Foundation.
- Keshavarz, N., Nutbeam, D., Rowling, L. and Khavarpour, F. (2010) Schools as social complex adaptive systems: a new way to understand the challenges of introducing the health promoting schools concept, *Social science & medicine* (1982), 70, 10, 1467.
- Kirby, D. (2004) BDI logic models: A useful tool for designing, strengthening and evaluating programs to reduce adolescent sexual risk-taking, pregnancy, HIV, and other STDs, *ETR Associates*. Retrieved July, 10, 2008.
- Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., Roberts, C., Shoukri, M. and Streiner, D.L. (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed, *International Journal of Nursing Studies*, 48, 6, 661-671.
- Lang, T.A. and Secic, M. (2006) *How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers*: ACP Press.
- Lewin, S., Glenton, C. and Oxman, A.D. (2009) Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study, *BMJ: British Medical Journal*, 339.
- Lichstein, K.L., Riedel, B.W. and Grieve, R. (1994) Fair tests of clinical trials: A treatment implementation model, *Advances in Behaviour Research and Therapy*, 16, 1, 1-29.
- Lorencatto, F., West, R., Stavri, Z. and Michie, S. (2012) How Well Is Intervention Content Described in Published Reports of Smoking Cessation Interventions?, *Nicotine & Tobacco Research*.
- Marchal, B., Westhorp, G., Wong, G., Van Belle, S., Greenhalgh, T., Kegels, G. and Pawson, R. (2013) Realist RCTs of complex interventions—An oxymoron, *Social Science & Medicine*, 94, 124-128.
- Mars, T., Ellard, D., Carnes, D., Homer, K., Underwood, M. and Taylor, S.J.C. (2013) Fidelity in complex behaviour change interventions: a standardised approach to evaluate intervention integrity, *BMJ Open*, 3, 11.
- May, C. (2013) Towards a general theory of implementation, *Implementation Science*, 8, 1, 18.
- May, C. and Finch, T. (2009) Implementing, embedding, and integrating practices: an outline of normalization process theory, *Sociology*, 43, 3, 535-554.
- McGraw, S.A., McKinlay, S.M., McClements, L., Lasater, T.M., Assaf, A. and Carleton, R.A. (1989) Methods in Program Evaluation: The Process Evaluation System of the Pawtucket Heart Health Program, *Eval Rev*, 13, 5, 459-483.
- Michie, S., Fixsen, D., Grimshaw, J. and Eccles, M. (2009) Specifying and reporting complex behaviour change interventions: the need for a scientific method, *Implementation Science*, 4, 1, 40.
- Michie, S., Johnston, M., Abraham, C., Francis, J. and Eccles, M.P. (2013) The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions, *Annals of Behavioral Medicine*, 1-15.
- Michie, S., van Stralen, M.M. and West, R. (2011) The behaviour change wheel: A new method for characterising and designing behaviour change interventions, *Implementation Science*, 6, 1, 42.
- Mihalic, S. (2004) The importance of implementation fidelity, *Emotional & Behavioral Disorders in Youth*, 4, 83 - 86.
- Moncher, F.J. and Prinz, R.J. (1991) Treatment fidelity in outcome studies, *Clinical psychology review*, 11, 3, 247-266.

- Montgomery, P., Grant, S., Hopewell, S., Macdonald, G., Moher, D., Michie, S. and Mayo-Wilson, E. (2013a) Protocol for CONSORT-SPI: an extension for social and psychological interventions, *Implementation Science*, 8, 1, 99.
- Montgomery, P., Underhill, K., Gardner, F., Operario, D. and Mayo-Wilson, E. (2013b) The Oxford Implementation Index: a new tool for incorporating implementation data into systematic reviews and meta-analyses, *Journal of Clinical Epidemiology*, 66, 8, 874-882.
- Moore, G.F., Currie, D., Gilmore, G., Holliday, J.C. and Moore, L. (2012a) Socioeconomic inequalities in childhood exposure to secondhand smoke before and after smoke-free legislation in three UK countries, *Journal of Public Health*.
- Moore, G.F., Moore, L. and Murphy, S. (2011) Facilitating adherence to physical activity: exercise professionals' experiences of the National Exercise Referral Scheme in Wales, *BMC Public Health*.
- Moore, G.F., Raisanen, L., Moore, L., Din, N.U. and Murphy, S. (2013) Mixed-method process evaluation of the Welsh National Exercise Referral Scheme, *Health Education*, 113, 6, 2-2.
- Moore, S., Murphy, S., Moore, S., Brennan, I., Byrne, E., Shepherd, J. and Moore, L. (2012b) An exploratory randomised controlled trial of a premises-level intervention to reduce alcohol-related harm including violence in the United Kingdom, *BMC Public Health*, 12, 1, 1-17.
- Munro, A. and Bloor, M. (2010) Process evaluation: the new miracle ingredient in public health research?, *Qualitative Research*, 10, 6, 699-713.
- Murphy, S., Moore, G.F., Tapper, K., Lynch, R., Clark, R., Raisanen, L., Desousa, C. and Moore, L. (2011) Free healthy breakfasts in primary schools: a cluster randomised controlled trial of a policy intervention in Wales *Public Health Nutrition*.
- Murphy, S.M., Edwards, R.T., Williams, N., Raisanen, L., Moore, G., Linck, P., Hounscome, N., Din, N.U. and Moore, L. (2012) An evaluation of the effectiveness and cost effectiveness of the National Exercise Referral Scheme in Wales, UK: a randomised controlled trial of a public health policy initiative, *Journal of Epidemiology and Community Health*.
- National Institute for Health and Care Excellence (2014) Behaviour change: individual approaches - guidance. NICE: London
- O'Cathain, A., Murphy, E. and Nicholl, J. (2008a) Multidisciplinary, Interdisciplinary, or Dysfunctional? Team Working in Mixed-Methods Research, *Qualitative Health Research*, 18, 11, 1574-1585.
- O'Cathain, A., Murphy, E. and Nicholl, J. (2008b) The quality of mixed methods studies in health services research, *Journal of Health Services Research & Policy*, 13, 2, 92-98.
- O'Cathain, A., Thomas, K.J., Drabble, S.J., Rudolph, A. and Hewison, J. (2013) What can qualitative research do for randomised controlled trials? A systematic mapping review, *BMJ Open*, 3, 6.
- Pawson, R., Owen, L. and Wong, G. (2010) Legislating for health: locating the evidence, *Journal of public health policy*, 31, 2, 164-177.
- Pawson, R. and Tilley, N. (1997) *Realistic evaluation*, London: Sage.
- Petticrew, M., Tugwell, P., Kristjansson, E., Oliver, S., Ueffing, E. and Welch, V. (2011) Damned if you do, damned if you don't: subgroup analysis and equity, *Journal of Epidemiology and Community Health*.
- Pirie, P.L., Stone, E.J., Assaf, A.R., Flora, J.A. and Maschewskyschneider, U. (1994) Program-Evaluation Strategies for Community-Based Health Promotion Programs -

- Perspectives from the Cardiovascular-Disease Community Research and Demonstration Studies, *Health Education Research*, 9, 1, 23-36.
- Plummer, M.L., Wight, D., Obasi, A.I.N., Wamoyi, J., Mshana, G., Todd, J., Mazige, B.C., Makokha, M., Hayes, R.J. and Ross, D.A. (2007) A process evaluation of a school-based adolescent sexual health intervention in rural Tanzania: the MEMA kwa Vijana programme, *Health education research*, 22, 4, 500-512.
- Renger, R. and Titcomb, A. (2002) A three-step approach to teaching logic models, *American Journal of Evaluation*, 23, 4, 493-503.
- Resnicow, K. and Vaughan, R. (2006) A chaotic view of behavior change: a quantum leap for health promotion, *International Journal of Behavioral Nutrition and Physical Activity*, 3, 1, 25.
- Riley, T., Hawe, P. and Shiell, A. (2005) Contested ground: how should qualitative evidence inform the conduct of a community intervention trial?, *Journal of Health Services Research & Policy*, 10, 2, 103-110.
- Rogers, E.M. (2003) *Diffusion of Innovations*, New York: Free Press.
- Rogers, P.J., Petrosino, A., Huebner, T.A. and Hacs, T.A. (2000) Program theory evaluation: Practice, promise, and problems, *New Directions for Evaluation*, 2000, 87, 5-13.
- Stange, K.C., Crabtree, B.F. and Miller, W.L. (2006) Publishing Multimethod Research, *Ann Fam Med*, 4, 4, 292-294.
- Steckler, A. and Linnan, L. (2002) *Process evaluation for public health interventions and research*, San Francisco: Jossey-Bass.
- Stirman, S.W., Kimberly, J., Cook, N., Calloway, A., Castro, F. and Charns, M. (2012) The sustainability of new programs and innovations: a review of the empirical literature and recommendations for future research, *Implementation Science*, 7, 17, 1-19.
- Stirman, S.W., Miller, C.J., Toder, K. and Calloway, A. (2013) Development of a framework and coding system for modifications and adaptations of evidence-based interventions, *Implementation Science*, 8, 1, 65.
- Strange, V., Allen, E., Oakley, A., Bonell, C., Johnson, A., Stephenson, J. and The Ripple Study, T. (2006) Integrating Process with Outcome Data in a Randomized Controlled Trial of Sex Education, *Evaluation*, 12, 3, 330-352.
- Tong, A., Sainsbury, P. and Craig, J. (2007) Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups, *International Journal for Quality in Health Care*, 19, 6, 349-357.
- Wang, R., Lagakos, S.W., Ware, J.H., Hunter, D.J. and Drazen, J.M. (2007) Statistics in medicine—reporting of subgroup analyses in clinical trials, *New England Journal of Medicine*, 357, 21, 2189-2194.
- Waters, E., Hall, B.J., Armstrong, R., Doyle, J., Pettman, T.L. and de Silva-Sanigorski, A. (2011) Essential components of public health evidence reviews: capturing intervention complexity, implementation, economics and equity, *Journal of Public Health*, 33, 3, 462-465.
- Weiss, C. (1997) Theory-based evaluation: Past, present, and future, *New Directions for Program Evaluation*, 1997, 76, 41-55.
- Wight, D., Raab, G., Henderson, M., Abraham, C., Buston, K., Hart, G. and Scott, S. (2002) The limits of teacher-delivered sex education: interim behavioural outcomes from a randomised trial, *British Medical Journal*, 324, 1430-33.

Appendix 1 - About the authors

Guidance development lead

Graham Moore (Research Fellow, DECIPHer UKCRC Public Health Research Centre of Excellence, School of Social Sciences, Cardiff University) is an interdisciplinary, mixed-method researcher, with a background in public health, social science research methods, and psychology. Most of his work has involved evaluating Welsh Government health improvement policies, using RCTs, natural experiments and mixed method process evaluations, and drawing predominantly upon psychological and sociological theory. He is currently supported by an MRC Population Health Scientist Fellowship.

Chair

Janis Baird (Senior Lecturer in Public Health, MRC Lifecourse Epidemiology Unit, University of Southampton) is a public health physician whose research focuses on the translation of evidence of the developmental origins of disease into public health policy and practice. She is part of a team undertaking a series of complex intervention studies which aim to improve the health behaviours of women of childbearing age.

Other guideline development group members

Suzanne Audrey (Research Fellow, DECIPHer UKCRC Public Health Research Centre of Excellence, School of Social and Community Medicine, University of Bristol) is a qualitative researcher with expertise in developing and implementing process evaluation within randomised controlled trials, and advises on research methods with the MRC ConDuCT (COllaboration and iNnovation in DifficUlt and complex randomised Controlled Trials) hub.

Mary Barker (Senior Lecturer in Psychology, MRC Lifecourse Epidemiology Unit, University of Southampton) is a psychologist whose main interest is in intervening to change health behaviour. She has recently been part of a team evaluating a trial of a complex public health intervention to improve the diets and lifestyle of young women and their children. Her current work involves exploring the role of psychological variables in the process of change.

Lyndal Bond (Principal Research Officer, Centre of Excellence in Intervention and Prevention Science, Melbourne, VIC Australia) is a public health researcher whose work has focused on the development and evaluation of complex interventions in schools and

communities. Her research has included evaluating the health effects of social and policy interventions, and researching the implementation and sustainability of complex interventions. In her current role, her work involves taking a systems approach to understand and evaluate such interventions.

Chris Bonell (Professor of Sociology and Social Policy, Social Science Research Unit, Institute of Education, University of London) is a sociologist whose research interests include interventions to promote young people's health and wellbeing, especially via modifying school environments. Methodological interests include evaluating intervention transferability, and qualitative research within trials.

Wendy Hardeman (Senior Research Associate in Behavioural Science, The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge) is a behavioural scientist whose work focusses on translating evidence-based guidelines into clinical practice and policy. She has developed and evaluated theory- and evidence-based behaviour change interventions aimed at improving health and well-being at the individual and population level. Her key interests include intervention fidelity and specifying intervention content, using the Behaviour Change Technique Taxonomy (V1).

Laurence Moore (Director, MRC/CSO Social and Public Health Sciences Unit, University of Glasgow) is a social scientist and statistician with a particular interest in the development and evaluation of complex interventions to improve public health. He has completed mixed methods evaluations of diverse interventions and programmes, including RCTs of interventions in schools and in the primary health care and social care sectors.

Alicia O'Cathain (Professor of Health Services Research, School of Health and Related Research, University of Sheffield) is a health services researcher with expertise in the evaluation of new health services and mixed methods research. She recently completed an MRC-funded study of the use of qualitative research with randomised controlled trials (the QUART study) and is currently leading a process evaluation of telehealth for long term conditions.

Tannaze Tinati (Research Fellow, MRC Lifecourse Epidemiology Unit, University of Southampton) is a Chartered Psychologist and the programme manager for the MRC Population Health Sciences Research Network (PHSRN). She has a background in mixed methods research, cognition and health. She works in a multidisciplinary team, and recently

evaluated a behaviour change intervention. She is currently leading on the process evaluation of SPRING (Southampton PRegnancy Intervention for the Next Generation).

Daniel Wight (Children, Young People, Families and Health Programme Leader, MRC/CSO Social and Public Health Sciences Unit, University of Glasgow) trained in social anthropology, has worked in public health research for two decades. He has been involved in many forms of evaluation in the UK and Africa, mainly of behavioural sexual health interventions, including two large process evaluations complementing RCTs.

Appendix 2 - Developing the guidance: process and milestones

The guidance was developed through reviewing literature on key theories and frameworks informing process evaluation; reflecting upon case studies (from the authors' own experience) at Guideline Development Group meetings; stakeholder consultation on the draft structure of the guidance; and further consultation on a full draft of the guidance. The following timeline sets out the key milestones in this process:

- November 2010 – Workshop held in Southampton, funded by MRC PHSRN, to bring together public health researchers, policy-makers and other stakeholders, to discuss the need for guidance for process evaluation of complex public health interventions.
- January 2012 – Guideline Development Group (GDG) assembled, including ten individuals who attended the November workshop.
- July 2012 – Funding obtained from MRC PHSRN for a part-time cross-unit post to lead development of guidance (to which GDG member Graham Moore was appointed).
- October 2012 – Guideline Development Group meets to discuss plans for the guidance and develop a draft structure.
- November 2012 – Plans for the guidance form on the basis of workshop at QUALitative Research in Trials (QUART) symposium. Alicia O'Cathain joins the author group to provide additional expertise on qualitative research within trials, and a Health Services Research perspective.
- Oct 2012 - April 2013 – Draft guidance structure developed. Members of the GDG provide case studies from their own work, reflections upon which inform the development of the guidance. Initial review of theories and frameworks for process evaluation conducted.
- April 2013 – Stakeholder group consulted via email on proposed structure of the guidance, and asked to identify any perceived gaps in review of theories and frameworks.
- May 2013 - July 2013 – Full draft of guidance developed, and discussed at GDG meeting. Framework and draft case studies discussed at UKCRC Public Health Research Centres of Excellence workshop and BPS funded seminar on process evaluation in behaviour change interventions.
- July 2013 - Sep 2013 – Second draft produced for stakeholder consultation.

- September 2013 – workshop held at the Society for Social Medicine Annual Scientific Conference focussing on chapter ‘how to design, plan and conduct process evaluation’.
- October 2013 – Guidance draft circulated to individuals who commented on the draft structure, and to others identified at conference and workshops. Sent directly to 20 individuals, but snowballed beyond this. Reviews of the full document received from 13 individuals.
- December 2013 – Final meeting of GDG in London to agree changes to document in response to reviewer feedback. The majority of feedback related to minor issues of clarification, most of which were made, but there were certain issues on which there was consensus. These, and the GDG’s responses to them, are as follows:
 - The document was too long and contained too much dense theoretical material. → The main text of the document (excluding the executive summary, case studies, appendices and references) was shortened by 6000 words, and divided into sections on theory and practice, with more effective signposting so that readers can go straight to the parts which interest them.
 - A working definition of public health intervention was needed, and consideration of whether the document was applicable outside public health. → Definitions are included in the executive summary. While the position of the majority of the group as public health researchers is made clear, we have removed direct reference to public health where the arguments being made are applicable beyond this.
 - Some issues of terminology required clarification, specifically the use of the word ‘implementation’. → We now clarify early in the document that implementation is used more than one way in the literature, but that we are using it to relate to the delivery of an intervention during the evaluation period.