Developing and Evaluating Complex Interventions

Draft of updated guidance

**Funding:** This project is jointly funded by the Medical Research Council (MRC) and National Institute for Health Research (NIHR).

[NOTE FULL AUTHORSHIP CONTRIBUTION STATEMENT TO BE PROVIDED]

**Project team (in alphabetical order):**
- Peter Craig, MRC/CSO Social and Public Health Sciences Unit
- Lynsay Matthews, MRC/CSO Social and Public Health Sciences Unit
- Laurence Moore, MRC/CSO Social and Public Health Sciences Unit
- Sharon Simpson, MRC/CSO Social and Public Health Sciences Unit
- Kathryn Skivington, MRC/CSO Social and Public Health Sciences Unit

**Other contributing authors (in alphabetical order):**
- Janis Baird, MRC Lifecourse Epidemiology Unit
- Jane Blazeby, MRC Hub for Trials Methodology
- Kathleen Boyd, Health Economics and Health Technology Assessment Unit, University of Glasgow
- Neil Craig, NHS Health Scotland
- David French, MRC Methodology Research Programme
- Jo Rycroft-Malone, NIHR Health Services and Delivery Research Programme
- Emma McIntosh, Health Economics and Health Technology Assessment Unit, University of Glasgow
- Mark Petticrew, London School of Hygiene and Tropical Medicine
- Martin White, MRC Epidemiology Unit

**Scientific Advisory Group (in alphabetical order)**
- Martin Ashton-Key, NIHR Evaluation, Trials and Studies Coordinating Centre
- Janis Baird, MRC Lifecourse Epidemiology Unit
- Jane Blazeby, MRC Hub for Trials Methodology
- David French, MRC Methodology Research Programme
- Mark Petticrew, London School of Hygiene and Tropical Medicine
- Jo Rycroft-Malone, NIHR Health Services and Delivery Research Programme
- Martin White (Chair), MRC Epidemiology Unit

**Observers:** Gavin Malloch, MRC Programme Manager; Samuel Rowley, MRC-NIHR Methodology Research Programme Advisory Group
Keywords: Development; feasibility; evaluation; implementation; complexity; programme theory; context; systems; stakeholders; intervention modification, complex intervention
# TABLE OF CONTENTS

1. **Introduction** ................................................................. 5
   1.1. Overview ....................................................................... 5
   1.2. Scope of the guidance .................................................. 6
   1.3. Structure of the guidance .............................................. 7

2. **Framework for developing and evaluating complex interventions** ............. 8
   2.1. Overview ....................................................................... 8
   2.2. What is a ‘complex intervention’? ................................. 10
   2.3. Efficacy perspective ..................................................... 14
   2.4. Effectiveness perspective .............................................. 15
   2.5. Realist perspective ....................................................... 15
   2.6. Systems perspective ..................................................... 16

3. **Overarching considerations** .................................................. 19
   3.1. Stakeholders ............................................................... 19
   3.2. Programme theory ....................................................... 22
   3.3. Context .......................................................................... 26
   3.4. Economic considerations ............................................. 28
   3.5. Intervention modification ............................................ 33
   3.6. Uncertainties ............................................................... 35

4. **Phases of research** ................................................................... 37
   4.1. DEVELOPING OR IDENTIFYING THE INTERVENTION .......... 37
      4.1.1. Developing an intervention ........................................ 37
      4.1.2. Identifying an intervention ........................................ 42
   4.2. FEASIBILITY .................................................................. 47
      4.2.1. Why is a feasibility stage needed? ............................. 48
      4.2.2. Feasibility of the intervention design ......................... 49
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Start Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.3.</td>
<td>Feasibility of the evaluation design</td>
<td>49</td>
</tr>
<tr>
<td>4.2.4.</td>
<td>Economic considerations for feasibility phase</td>
<td>50</td>
</tr>
<tr>
<td>4.2.5.</td>
<td>Progression criteria</td>
<td>51</td>
</tr>
<tr>
<td>4.3.</td>
<td>EVALUATION</td>
<td>56</td>
</tr>
<tr>
<td>4.3.1.</td>
<td>Effectiveness and usefulness</td>
<td>56</td>
</tr>
<tr>
<td>4.3.2.</td>
<td>Research perspectives &amp; evaluation</td>
<td>57</td>
</tr>
<tr>
<td>4.3.3.</td>
<td>Types of evaluation design</td>
<td>59</td>
</tr>
<tr>
<td>4.3.4.</td>
<td>Choice of evaluation outcomes</td>
<td>63</td>
</tr>
<tr>
<td>4.3.5.</td>
<td>Economic Evaluations / Analyses</td>
<td>65</td>
</tr>
<tr>
<td>4.3.6.</td>
<td>Approaching complexity in systematic reviews</td>
<td>69</td>
</tr>
<tr>
<td>4.3.7.</td>
<td>Reporting guidelines</td>
<td>69</td>
</tr>
<tr>
<td>4.4.</td>
<td>IMPLEMENTATION</td>
<td>72</td>
</tr>
<tr>
<td>4.4.1.</td>
<td>Designing research with implementation in mind</td>
<td>72</td>
</tr>
<tr>
<td>4.4.2.</td>
<td>Implementation as an intervention</td>
<td>73</td>
</tr>
<tr>
<td>4.4.3.</td>
<td>Understanding implementation</td>
<td>74</td>
</tr>
<tr>
<td>4.4.4.</td>
<td>Economic considerations</td>
<td>75</td>
</tr>
<tr>
<td>5.</td>
<td>Case studies</td>
<td>80</td>
</tr>
<tr>
<td>6.</td>
<td>Appendix 1. Glossary of terms</td>
<td>81</td>
</tr>
<tr>
<td>7.</td>
<td>Appendix 2. Checklist for developing and evaluating complex interventions</td>
<td>88</td>
</tr>
<tr>
<td>8.</td>
<td>References</td>
<td>90</td>
</tr>
</tbody>
</table>
1. Introduction

1.1. Overview

In 2006, the MRC published guidance for developing and evaluating complex interventions (Craig and et al, 2006), building on the framework that had been published in 2000 (Craig and et al, 2000). The aim was to help researchers and research funders recognise and adopt appropriate methods to improve the quality of research to develop and evaluate complex interventions and thereby maximise their impact. The guidance documents have been highly influential, and the accompanying papers published in the BMJ are widely cited (Campbell et al., 2000, Craig et al., 2008). This guidance has also been complemented by MRC guidance on ‘process evaluation’ (Moore et al., 2015) and ‘natural experiments’ (Craig et al., 2012a).

A review of the existing guidance was undertaken in 2018. This involved a scoping review of the literature (‘gap analysis’) and a facilitated workshop that was attended by experts from disciplines relevant to complex intervention development and evaluation [details on methods to be added to Appendix]. Two key aspects of the 2006 framework were identified as in need of rethinking:

- The definition of a complex intervention is now somewhat outdated. Revision of the definition and the scope of the guidance is required, in particular drawing upon a wider understanding of the complex contexts (systems) in which interventions are delivered and evaluated.
- The predominant focus was on designing evaluations to minimise bias; a helpful advance would be, in addition, to consider how to maximise the usefulness of evidence for decision-making.
- Additional elements identified as needing further guidance, included:
  - The use of theory
  - How to research complex interventions where interventions are not developed by researchers
  - Stakeholder engagement
  - Modification of interventions
  - Economic considerations
- How to approach complexity, or take a systems perspective, in systematic reviews

Some of these points were not covered in previous guidance because the ideas were new and examples of practice were lacking. The workshop and scoping review concluded that there had been considerable advances since 2006. Although much of the previous guidance remained valuable, it was felt that the accumulation of knowledge and experience on these issues should inform an updated version.

1.2. Scope of the guidance

This guidance aims to support the planning, development, feasibility testing, evaluation and implementation of complex interventions. It aims to reach a broad audience, including:

(i) Health researchers, broadly defined to include those working in public health, health policy, social care, health care, and health services research. This includes those who focus on researcher-led and non-researcher-led interventions (e.g. policy evaluations). The aim is to guide their overall approach to intervention development, evaluation, and implementation, and providing information to support the choice of appropriate study designs and methods.

(ii) Research funders of health and social care, and public health interventions: to highlight the importance of funding a broader range of research questions, study designs, methods, and outcomes.

(iii) Practitioners e.g. public health and primary and secondary care professionals, clinicians, and non-clinical practitioners such as those working in allied health services, social care, or the community sector: providing guidance on identifying and understanding complexity, and support in the implementation of interventions.

(iv) Decision makers e.g. within management or policy making: supporting them to use research evidence to inform policy decisions and to build evaluation into the process of policy and service development.

Rather than replicate other relevant guidance we signpost to useful and appropriate sources where available. While some aspects of good practice are clear, methods for developing, evaluating, and implementing complex interventions are evolving, and on many important issues there is not yet consensus on best practice. We therefore anticipate this version of the
Updated Guidance: Developing and Evaluating Complex Interventions [DRAFT FOR CONSULTATION]

guidance will stimulate constructive debate, leading to further developments. This guidance is intended to be of practical relevance; therefore we include case study examples where possible.

1.3. Structure of the guidance

The guidance is presented as follows:

- In section 2 we set out the revised framework for developing and evaluating complex interventions, including the updated definition of ‘complex intervention’.
- In section 3 we provide an overview of the key overarching considerations that have been identified for developing and evaluating complex interventions.
- In section 4 we go into further detail on the phases of developing, evaluating, and implementing a complex intervention. Within each phase, we suggest further questions that researchers considering research on a complex intervention should ask themselves.
- In section 5 we present a range of case studies, using a variety of study designs, from a range of disciplines, carried out in a range of settings.
- Appendices: we present a glossary of terms used throughout the document and the definitions that we have used, and a checklist to support the development and evaluation of complex interventions.
2. Framework for developing and evaluating complex interventions

2.1. Overview

The following section presents an overview of the main aspects on which we provide guidance i.e. the key phases and overarching considerations when developing and evaluating complex interventions. We recognise that users of this guidance may have constraints in terms of time, resource, and other factors. We highlight that even where it is not possible to address all the relevant issues, it is valuable to consider all aspects that are feasible within such constraints.

The main challenges identified for revising this guidance were to consider how context dependence and complex systems should be taken into account; and how to ensure evidence generated is most useful for decision makers. The most recent iteration of the guidance (2006) indicated the importance of context, but did not feature it centrally in the considerations for developing and evaluating interventions (see Figure 1 for diagram published in 2006 guidance document).

Figure 1: Key elements of the development and evaluation process from 2006 guidance.

In the light of the significant scientific and theoretical developments in thinking about complex interventions and complexity since 2006; we have reworked this diagram to better highlight the centrality of considerations of context, programme theory, stakeholders, modification, economics, and uncertainties. The new diagram (shown in Figure 2) retains the main phases—development, feasibility, evaluation, and implementation—though adds ‘intervention identification’ to explicitly ensure the guidance is relevant to interventions (e.g.
public policies) that are not developed by the researcher or where the intervention design is not within their control. In this revised version, we have not included arrows directly between each of these research phases. Instead each phase is connected to a set of overarching considerations—uncertainty, programme theory, stakeholders, economics, and modification—that should be visited throughout the research, particularly before moving between phases (e.g. between feasibility to evaluation). As with the 2006 diagram, there is no defined starting point for the research process, and movement between phases is not necessarily expected to be linear or cyclical. The diagram also reflects the importance of intervention context, situating the development, evaluation and implementation of interventions within a broader context, which should be considered in the research.

**Context:** any feature of the circumstances in which an intervention is conceived, developed, implemented and evaluated. For example, this may be a social, political, economic or geographical context (Craig, Di Ruggiero et al. 2018).

Figure 2: Key elements for developing and evaluating complex interventions.
2.2. **What is a ‘complex intervention’?**

The previous guidance defined complexity in terms of features of the intervention, for example the number of components. We now propose that these intervention characteristics (see Box 1) should be thought of as dimensions of complicatedness rather than of complexity. However, complexity is not solely a property of the intervention itself, but also of the context in which it is located, and the research perspective taken (Greenhalgh and Papoutsi, 2018, Hawe et al., 2009, Petticrew, 2011). There are several explanatory research perspectives that can be employed in the development and evaluation of interventions (see Box 2), which we summarise as lying on a continuum between ‘efficacy’ and ‘systems’ (i.e. these are not necessarily completely distinct perspectives). The research perspective represents the type of question that the research aims to address, and needs to be considered at the start of the research process.

Box 1: Characteristics of interventions: simple to complicated

<table>
<thead>
<tr>
<th>Characteristics of interventions may include:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of and interactions between phases within the experimental and control interventions</td>
</tr>
<tr>
<td>Number and difficulty of behaviours, expertise and skills (including technical e.g. particular surgical techniques, and non-technical e.g. communication) required by those delivering or receiving the intervention</td>
</tr>
<tr>
<td>Number of groups or organisational levels targeted by the intervention</td>
</tr>
<tr>
<td>Number and variability of outcomes</td>
</tr>
<tr>
<td>Degree of flexibility or tailoring of the intervention or its components that is permitted</td>
</tr>
</tbody>
</table>
Box 2: Continuum of research perspectives

Research perspectives used in the development and evaluation of interventions sit on a continuum from efficacy to systems and should not be thought of as mutually exclusive perspectives. These perspectives differ in terms of external validity, and include:

- **Efficacy perspective**: To what extent does the intervention produce the intended outcome(s) in experimental settings?
- **Effectiveness perspective**: To what extent does the intervention produce the intended outcome in real world settings?
- **Realist perspective**: What works, for whom, under which circumstances, and why?
- **Systems perspective**: How does the intervention interact with the system to produce change?

This guidance aims to support more ‘complexity-informed’ research (Greenhalgh and Papoutsi, 2018), i.e. which gives sufficient consideration to complexity rather than attempts to control for it. The choice of research perspective taken should be informed by an awareness of where real-world uncertainties related to the issue or challenge, and thus the intervention, lie. Here it is useful to draw upon systems thinking.

**System**: a set of things that are interconnected in such a way that they produce their own pattern of behaviour over time (Meadows, 2008). For example a school, hospital, body, government.

All interventions take place within systems. Relationships between agents is a key feature of systems, i.e. people in social systems, cells within the body etc. Some systems are complex and adaptive, in that they are defined by system-level properties such as feedback, emergence, adaption, and self-organisation (examples of these are given in Table 1). These properties mean that change is not always predictable, nor necessarily understood by linear causal pathways. In a social system, agents, i.e. people, interact with each other in non-linear ways, and their actions are interconnected so that the actions of one agent alter the context for other agents (Plsek and Greenhalgh, 2001).
Table 1: Properties of complex adaptive systems, with examples.

<table>
<thead>
<tr>
<th>Properties of complex adaptive systems</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Emergence</strong></td>
<td></td>
</tr>
<tr>
<td>Complex systems have emergent properties that are a feature of the system as a whole, which cannot be predicted from individual parts or the sum of individual parts.</td>
<td>Emergence of herd immunity from a vaccination programme. This is when vaccine coverage reaches a level that stops infections from spreading. Even those who are not vaccinated can be protected by herd immunity. Enhanced recovery pathways (ERP) are introduced to optimise early discharge and improve patient outcomes. This leads to changes across pre-operative care where patient expectations are managed, in surgical wards, and in physiotherapy – both for the patients considered for ERP and all patients – thus whole culture of surgical practice changes towards ‘enhanced recovery’.</td>
</tr>
<tr>
<td><strong>Feedback</strong></td>
<td></td>
</tr>
<tr>
<td>Where one change reinforces, promotes, balances or diminishes another.</td>
<td>A smoking ban in public places reduces the visibility and convenience of smoking. Fewer young people start smoking due to its reduced appeal, thus further reducing its visibility and so on in a reinforcing loop (Rutter, Savona et al.).</td>
</tr>
<tr>
<td><strong>Adaptation</strong></td>
<td></td>
</tr>
<tr>
<td>A change of system behaviour in response to an intervention</td>
<td>Retailers adapted to the ban on multi-buy discounts for alcohol by placing discounts on individual alcohol products, offering them at the same price individually as they would have been if part of a multi-buy offer (Robinson et al., 2014).</td>
</tr>
<tr>
<td><strong>Self-organisation</strong></td>
<td></td>
</tr>
<tr>
<td>Global organisation and order achieved as the product of spontaneous local interaction rather than of a preconceived plan or external control.</td>
<td>An isolated group of individuals will typically self-organise leaders to co-ordinate activity and direction even without external instruction e.g. parkrun (<a href="http://www.parkrun.org.uk/">http://www.parkrun.org.uk/</a>) Alcoholics Anonymous: Individually focussed treatment did not address some social aspects of alcohol dependency, recovery groups were self-organised in a collective effort.</td>
</tr>
</tbody>
</table>
Complexity-informed research can encourage awareness of real-world uncertainties. Researchers are encouraged not only to consider how the intervention is expected to work (i.e. the internal intervention logic), but to consider the overall system, what parts of the system could influence the intervention, and how the intervention could lead to wider system change (i.e. consideration of the properties of systems shown in Table 1).

Figure 3 shows the ‘complicatedness’ of the intervention on the x axis (which refers to the characteristics in Box 1), and the research perspective (detailed in Box 2) on the y axis, with increasing consideration of complexity from bottom to top. The shading on Figure 3 represents whether the research perspective can appropriately assess all aspects of the intervention. For example, it is unlikely that a complicated intervention e.g. a programme of anti-bullying initiatives in schools, with flexibility in delivery, or a multi-component intervention involving a multidisciplinary team to improve health-related quality of life for people with multimorbidity, can be adequately evaluated with an efficacy or effectiveness perspective alone. Although this perspective would provide an estimate of effect, it would not capture the interactions with context that are important to evaluate how the intervention achieves its effects. However, there is no clear distinction for ‘simple’ interventions, as here it is dependent on the questions posed (which is dependent on where the uncertainty lies). On the one hand, where the research is responding to questions about proof of concept within a restricted context e.g., the research questions are intended to resolve uncertainty surrounding whether the intervention produces intended outcomes in ideal conditions, little is likely to be gained from evaluating with a realist or systems perspective e.g. ‘does drug A work in controlled lab conditions?’ could sufficiently be answered by adopting an efficacy perspective. On the other hand, some questions that could appropriately be asked of an intervention that would be considered simple in terms of its components, would most appropriately be answered with systems perspectives e.g. does a change in food policy have an impact on obesity? This highlights the possibility that the complexity does not solely lie within the intervention, nor indeed the system, but in the interaction between the two.
Figure 3: Framework for addressing complexity within evaluation. [NOTE ‘EVALUATION PERSPECTIVE TO CHANGE TO RESEARCH PERSPECTIVE & REALIST TO ‘What works, for whom, under which circumstances, and why’]

2.3. Efficacy perspective

Efficacy is a useful perspective for identifying whether an intervention can work, though should not be confused with research that explores whether there will be any real-world effectiveness or change associated with the intervention. An efficacy perspective is adopted in order to maximise internal validity and provide causal explanation. An efficacy perspective is useful for proof of concept, for example to determine whether or not interventions are promising candidates for testing in large scale effectiveness studies. An example is testing whether a vaccine prevents infection under optimal conditions. This would be done prior to embedding the vaccine in a delivery programme and testing the effectiveness of the programme in a real-world setting. Even if taking an efficacy perspective, it is important to have an awareness of the wider system in which the vaccine will be delivered as there is little point in knowing about its efficacy if there is no context in which it could actually be implemented.
2.4. **Effectiveness perspective**

An effectiveness perspective answers questions about whether the intervention produces the intended effects, but is broader in that it aims to do so in the real world rather than in a controlled environment. Effectiveness studies (in clinical research referred to as pragmatic trials) therefore aim to be more generalizable than efficacy studies and as such are likely to recruit a broader population and setting, i.e. they aim to provide the intervention to the individuals who would receive it in practice rather than excluding on the basis of age or existing health conditions (within the bounds of safety). In addition, the intervention itself is delivered with some flexibility and not strictly controlled. An effectiveness perspective alone

An example is research by Salisbury et al (2018), which reports on the effectiveness of the 3D study, a randomised controlled trial of an intervention to improve care for people with multimorbidity. This trial explored the intervention in a range of contexts i.e. 33 GP practices in England and Scotland, and as well as the primary outcome (quality of life), it assessed three domains of secondary outcomes as well as measurement of care processes (continuity of care and number of consultations), and some disease-specific measures. The study followed pre-specified analyses and reported on all outcomes. The effectiveness study alone provides a robust estimate of effect, but cannot provide understanding of how the effects were produced. However, the 3D study provides an example of an effectiveness study within a broader programme of research, which included evaluations of cost-effectiveness, carer experience, and a qualitative process assessment (further papers forthcoming). [CASE STUDY TO BE ADDED]

2.5. **Realist perspective**

In a realist evaluation researchers develop theories, which they then evidence, to take account of how context and mechanisms interact to produce outcomes i.e. causal mechanisms of an intervention are context specific. As in efficacy and effectiveness perspectives, the realist perspective is still concerned with answering ‘what works’ questions, but crucially takes into account the context in which the mechanisms operate and aims to explore why interventions work, and for whom. ‘Realist evaluation’ was originally coined by Ray Pawson and Nick Tilley in their seminal work ‘Realistic Evaluation’ (1997). There has been considerable debate over what a realist perspective to intervention research entails, with
some drawing on the original Realistic Evaluation and asserting that experimental methods are applicable to this approach and others disagreeing with this stance (which interested readers can explore further e.g. in Bonell et al., 2013, Bonell et al., 2012, Bonell et al., 2018, Jamal et al., 2015, Marchal et al., 2013, Van Belle et al., 2016).

We recommend that evaluations that seek to answer questions about what works, for whom, under which circumstances, and why, can be answered with use of experimental trial designs if they are appropriately combined with methods that seek to go beyond questions of effectiveness. Experimental methods have developed to incorporate this thinking within evaluation, i.e. to clarify causal mechanisms by understanding process (Moore et al., 2014) and to better understand the context i.e. the circumstances or environment in which the intervention is situated (Craig et al., 2018). An example of this in practice, is the aforementioned Salisbury 3D study, which answered questions of effectiveness, but also broader questions related to how the intervention worked (or otherwise) (Salisbury et al., 2018). Examples of research that takes this further, incorporating realist questions about the mechanisms and contextual factors related to the intervention, alongside experimental trials, are underway (Chandler et al., 2013, Rycroft-Malone et al., 2018). Alternative approaches are not situated alongside trials, but use case studies and largely qualitative data to explore the dynamic interplay between the context, mechanisms, and outcomes of an intervention (Greenhalgh et al., 2009). [CASE STUDY TO BE ADDED]

2.6. Systems perspective

A systems perspective suggests that interventions can only be fully understood with an examination of the system(s) in which they are embedded. A systems perspective treats interventions as events within, or disruptions to, systems (Hawe et al., 2009). The properties of a system cannot be fully explained by an understanding of each of the system’s individual parts (Gallagher and Appenzeller, 1999), and therefore is concerned with an awareness and understanding of the whole system; “The essential point is that the theory driving the intervention is about the dynamics of the context or system, not the psyche or attributes of the individuals within it.”. Key to taking a systems perspective is consideration of relationships between intervention and context, engaging with multiple perspectives, and having an awareness of system boundaries. The research does not need to consider the
whole system, unless this is something that is desired and is relevant to understanding the intervention. An example can be seen with the systems perspective planned in the evaluation of the Soft Drinks Industry Levy (SDIL) (Penney et al., 2018). The SDIL was implemented by the UK Government in 2018—manufacturers are taxed on sugar-sweetened beverages they produce depending on the amount of sugar they contain. Taking the default effectiveness evaluation perspective i.e. looking solely at effectiveness in relation to the primary aim (to reduce population-level sugar consumption), would miss the broader purpose of the intervention and its position in the wider system. It would risk missing many important interactions, such as the response by industry and other government actions, which will shape longer-term outcomes. [CASE STUDY TO BE ADDED]

Although there has been debate about whether realist and systems perspectives are compatible, with one argument being that realist approaches are too reductionist and linear (see Westhorp 2012 for discussion), we suggest that they should not be considered incompatible, and that both perspectives and associated toolkits of methods can be drawn upon for evaluation (as argued by Westhorp 2012 and others). The starting point for a systems perspective is the system rather than the intervention, and taking a systems perspective could usefully add to the scope of an evaluation, concentrating on effort on obtaining multiple perspectives and uncovering unintended outcomes (Jabeen, 2017). As shown later (see section 4.3.5), updated guidance on conducting economic evaluations of complex interventions also aligns with this systems thinking by recommending broader frameworks such as cost-consequences and cost-benefit analyses.
**Key points**

- Consider the research perspective with regards to the intervention, its place in the wider system in which it is implemented, and the interaction between the two.
- The choice of research perspective is dependent on the uncertainties surrounding the intervention, (i.e. there is no right perspective for all circumstances).
- Complexity-informed research, with an awareness of the system(s) can encourage: (i) researchers to develop research questions that take into account the wider contextual factors that influence an intervention; and (ii) researchers, funders, practitioners and policy makers to develop, evaluate, and implement interventions using the most appropriate tools and methods.
3. Overarching considerations

We have identified six overarching factors as central to the effective development and evaluation of complex interventions. These are: (i) involvement of relevant stakeholders at appropriate stages; (ii) the development and ongoing iteration of programme theory; (iii) consideration of the context; (iv) appropriate health economic considerations; (v) ongoing intervention modification; and (vi) consideration of the uncertainty surrounding the intervention. This section is structured to provide an overview of these considerations, with following section linking these overarching considerations to different phases of research.

3.1. Stakeholders

Whether research findings lead to changes in practice or policy is heavily dependent on stakeholder engagement throughout the research process.

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

Effort is required to involve stakeholders at the beginning and to keep them engaged throughout. Reciprocity is vital and the goal should be the development of collaborative relationships that continue through the phases of development, evaluation and implementation (Armstrong, 2006).

Many complex health and social care interventions are delivered outside the health sector – for example in prisons or schools, in the transport system, through legal regulation or through the tax system. Likewise, policymakers may not be directly involved in the delivery or implementation of the intervention but may be closely involved in its development and have a keen interest in the outcomes. Engaging with a multidisciplinary group of stakeholders is therefore important for understanding the issue or challenge, the intervention, the wider context, the system properties, and the potential areas for change, including the development of a logic model (Moore et al., 2015). The process of theorising the intervention and its place within a system, will help to identify the relevant stakeholders.
The risk of not engaging relevant stakeholders in complex intervention research is that the intervention will not be fully theorised or considered from relevant perspectives. For example, outcomes of surgical interventions for severe and complex obesity depend upon patient engagement with new lifestyle and diet, and are therefore optimised by multidisciplinary care, e.g. psychologists and dieticians in addition to the traditional surgical teams (Blazeby et al., 2014). The development and evaluation of a surgical intervention like this example therefore requires multidisciplinary stakeholder engagement.

The purpose of stakeholder engagement will differ depending on the context and phase of the research, but could include: drawing on stakeholders' practical experience; obtaining their views on the problem being addressed by the intervention and potential solutions, development of programme theory, facilitating understanding of the context and the questions that they would find useful, as well as ensuring all 'voices' are heard. It can also help with prioritising questions, building consensus, building relationships to make evaluation more straightforward e.g. by accessing gatekeepers, supporting intervention refinement, and finally considering issues around transferability and implementation. Taken further, co-production of interventions with effective collaboration between a group of stakeholders supports the move from interventions being driven, in a closed manner, by researchers, to a model whereby interventions are developed in partnership with a range of stakeholders, and thus a move to being more ‘socially as well as scientifically robust’ (Greenhalgh et al., 2019). Practical tools from soft systems methodology can be drawn upon to support the engagement of stakeholders (see Box 3 for more information on soft systems methodology). These, as well as guidance on co-production, are described further in section 4.1 ‘Developing or Identifying the Intervention’.
Box 3: Soft systems methodology

Soft Systems Methodology (Checkland and Scholes, 1999, Williams and Hummelbrunner, 2010): methods for bringing together multiple different views of situations, in order to address problem situations. It addresses the following questions:

• What are the different ways in which a situation can be framed?

• How does each of these ways, on its own, provide a means of comprehending how a situation behaves?

• What are the implications for any changes to the situation?

Key points

• ‘Stakeholder’ is a wide-reaching term for anyone who has a stake in the intervention and/or its evaluation or implementation; therefore, careful consideration at the outset of the relevant stakeholders (for different phases of the research) is important.

• Stakeholder engagement needs to be genuine rather than to satisfy criteria for funding, publication etc.

• Involvement of appropriate stakeholders is crucial to each phase of research, and ultimately to the prospects of the research leading to changes in policy or practice.

• Soft systems methodologies can be drawn upon for working with multidisciplinary stakeholders.

Signposts to further reading/guidance

• INVOLVE is a national advisory group funded by NIHR to support active public and patient involvement (PPI) in NHS, public health and social care research. There is an online resource to support researchers to involve the public in how research is identified, prioritised, designed, conducted, and disseminated.

- There is also a NIHR handbook for researchers on PPI in health and social care research: https://www.nihr.ac.uk/about-us/CCF/funding/how-we-can-help-you/RDS-PPI-Handbook-2014-v8-FINAL.pdf.
- Researchers can draw upon soft systems methodologies to engage stakeholders, further information:
  - Systems concepts in action (Williams and Hummelbrunner, 2010)
  - Cynefin Framework: https://cognitive-edge.com/
  - Cognitive mapping tools: https://cmap.ihmc.us/
  - Soft systems methodology in action (Checkland and Scholes, 1999).

### 3.2. Programme theory

#### Terminology

Researchers and other stakeholders use theory-related terms differently and interchangeably, and there is a lack of clarity around the definition of terms in common use including: theory; theory of change; intervention theory; programme theory; mechanisms of action; logic model, amongst others. The definitions as used in this guidance are detailed in the glossary (Appendix 1). Whilst we acknowledge that it is unlikely there will be agreement across disciplines, for the purpose of clarity, we will focus on using the following two terms throughout the guidance: **programme theory** and **logic model**.

**Programme theory:** Describes how an intervention is expected to lead to its effects and under what conditions. This can illustrate what mechanisms are expected to generate the outcomes and what features of the context are expected to affect whether or not those mechanisms operate (Rogers, 2008).

**Logic model:** A visual representation of the programme theory, typically presenting mechanisms by which an intervention, and the context in which it operates, is thought to influence outcomes of interest (Moore et al., 2014).

#### Developing programme theory
A programme theory describes how an intervention is expected to lead to its effects and under what conditions. It articulates: (i) the mechanisms of the intervention; (ii) the interaction of components; and (iii) features of the context that are expected to influence those mechanisms (Funnell and Rogers, 2011). A programme theory ideally considers system-level factors; this helps those involved think about the above issues as well as the implementation and transferability of the intervention. The programme theory informs the evaluation, including the choice of outcomes, and collection of relevant data in order to explore the theorised causal pathways. Programme theory should be ‘adaptive’, continually updated, developed, and refined as understanding of the intervention and its interaction with the system changes over time. Best practice is to develop programme theory, co-produced with stakeholders, at the beginning of the research project, and to refine it during the different stages to produce an updated programme theory at project completion. A refined programme theory should be considered a meaningful evaluation outcome. Improved theory will help inform transferability of interventions across settings and help produce evidence and understanding that is more applicable to decision makers.

Although researchers are not always involved in developing initial programme theory (if there is one), for example in the case of policy development, they will still need to theorise a policy intervention before attempting to evaluate it. This usually involves defining the problem definition, and supporting the articulation and refinement of a programme theory. This should be done in the same way as for researcher-developed interventions, but here involvement of stakeholders is even more critical. For example, Lawless et al (2017) developed a programme theory-based framework for evaluating policy led processes and outcomes. Documents can also be drawn upon (e.g. policy documents or existing funding applications, to refine programme theory). This ties in with evaluability assessment, one of the tools that drives the theorising of an existing or proposed policy intervention (Craig and Campbell, 2015). Evaluability assessment is discussed further in section 4.1.2.

This guidance does not describe how to develop programme theory. For further guidance on this, please see the signpost box below. Some key things to consider, however, include:

- Identifying the problem and processes through which this could be changed.
- Developing the programme theory alongside the intervention development.
- Involving a multidisciplinary team and relevant stakeholders.
- Choosing appropriate evaluation outcomes, ensuring measurement and testing of theory.
- Refining the programme theory on an ongoing basis using process evaluation data, outcome data and further stakeholder involvement.
- Overall, a programme theory should be ‘adaptive’, continually refined throughout the different phases of development and evaluation. Development of programme theory can therefore be considered cumulative for a particular intervention. A benefit of programme theory is that theories developed for similar interventions have the potential to inform higher level theories (e.g. psychological or sociological theory).

**Representing and articulating programme theory**

A logic model is a diagrammatic representation of aspects of the programme theory. It is not the programme theory itself; the programme theory should always be articulated in text without the need for a logic model. However, logic models can be useful in the initial stages of developing interventions and programme theory, and for involving policymakers and the public.

Logic models can be useful, but are often criticised for being overly simplistic and showing only linear pathways of causation, and for failing to capture features of systems such as adaptiveness and feedback loops. Examples of a simple versus more complex logic model(s) are provided in the appendices [*NB to be added post-consultation*]. It is also important to consider and theorise unintended consequences of an intervention, which may include the development of ‘dark logic’ within the logic model (Bonell et al., 2015). ‘Dark logic’ describes the potential harmful consequences or adverse effects that may arise from an intervention, and their underlying mechanisms. Helpful overviews of logic modelling in relation to complexity have been published (Mills et al., 2019, Hawe, 2015).

In addition to logic models, other approaches exist to develop and represent programme theory. For example, soft systems methodologies (Box 3) may complement realist thinking to map the complexity of an intervention, and refine the programme theory with stakeholder engagement (Dalkin et al., 2018).
Key points

- Develop programme theory at the outset, as a collaborative process involving researchers and stakeholders.
- The development of programme theory is an iterative process. Produce an updated programme theory at the end of each stage of development and evaluation.
- Visual representation of the programme theory, in the form of logic model(s), can be useful for summarising programme theory and communicating with stakeholders. However, programme theory should always be articulated well in text, and not just in a logic model(s).

Signposts to further reading/guidance

- The use of programme theory within intervention development: ‘Taxonomy of approaches to developing interventions to improve health: a systematic methods overview’ (part of the INDEX study) (O’Cathain et al., 2019).
- The 6SQuID model for intervention development provides detail on house of theory: ‘Six steps in quality intervention development (6SQuID)’ (Wight et al., 2016).
- A paper exploring the use of logic models in interventions with a complexity perspective: ‘Advancing complexity science in healthcare research: the logic of logic models’ (Mills et al., 2019).
- An editorial reflecting on the use of theory within complex health interventions: ‘What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation of complex population health interventions’ (Moore and Evans, 2017).
3.3. Context

**Context:** “any feature of the circumstances in which an intervention is conceived, developed, implemented and evaluated” (Craig et al., 2018).

Features of the context that may be relevant could relate to, for example, the immediate setting of the intervention, ethical standards and rules, political factors, cultural practices. Taking context into account is crucial throughout: in theorising the interventions effects, intended and unintended; in the development phase to make it more likely that interventions are more suitable, successful, and implementable; at the evaluation phase to understand how interventions ‘work’, in terms of how they interact with their context; and at the reporting phase by providing contextual details surrounding the intervention, to make reports more useful to decision makers (Craig et al., 2018). Circumstances surrounding the intervention may change after the study has begun, for example through the introduction of a new policy, or, in the case of trials with long-term follow-up, more fundamental changes in the economic or political context. The context may also react to the intervention (e.g. a supermarket may make changes in placement, price or promotion of foods in response to regulation of in store promotions of less healthy foods). It is important therefore to develop a good understanding of the context in which the study is being carried out, and to monitor and document any changes throughout.

What works in one setting may not be as effective, or may even be harmful, elsewhere. The impact of a new intervention will depend on what provision already exists and interventions may have to be explicitly modified to fit different contexts (see example of the Family Nurse Partnership in Box 4).

Box 4: Risk of underestimating the importance of context: example

Underestimation of the importance of context, or a failure to take account of all of the relevant contextual dimensions, can lead to difficulties with implementation or a lack of effectiveness, even if an intervention can be implemented successfully in a new context. The Nurse Family Partnership had been shown in a number of US trials (Olds et al., 2010, Eckenrode et al., 2010) and one European trial (Mejdoubi et al., 2015) to be effective in improving a range of pregnancy, parenting, maternal health and social outcomes before it
was implemented in the UK as the Family Nurse Partnership (FNP). A large well-conducted trial of FNP found that, despite successful implementation, there was no improvement across the four primary outcomes. The report concluded that benefits could not be assumed in ‘different health-care settings and service populations, even when objective programme fidelity from US trials was maintained’ (Robling et al., 2016).

Example taken from ‘Taking account of context in population health intervention research: guidance for producers, users and funders of research’ (Craig et al., 2018).

It is important to note that the context of both the intervention and the research should be considered. This approach encourages researchers to move away from a narrow view of the intervention, which could result in interventions that are effective in the specific circumstances of the study but ineffective when implemented in real-world settings. Considering these wider contextual factors at the earliest stage can improve our understanding of how an intervention works, and how it might work in different circumstances. It allows interventions to be developed and evaluated with flexible and adaptive methods (see section 3.5 ‘intervention modification’), resulting in outcomes that are more useful and meaningful.

**Key points**

- The circumstances surrounding an intervention have a role in its success, or otherwise.
- An understanding of the context of the intervention should be developed at the outset, and continuously revisited and documented throughout the research process.
- Good understanding, description and explanation of context enables judgements to be made about transferability.
3.4. **Economic considerations**

Consideration should be given to the economic aspects of developing and evaluating complex interventions from the outset. It is important to identify early in the process the potential resource consequences of the interventions being evaluated and the potential outcomes for those people, organisations and systems directly and indirectly affected by the intervention. Early consideration of economic aspects—alongside the processes of stakeholder engagement, exploration of programme theory, development of a logic model (also known as a conceptual model (Roberts, et al. 2012), awareness of the systems and public sector perspective (see updated NICE recommendations)—can help shape the development of appropriate economic research questions in synchrony with complex intervention development. This, in turn, enables design of an appropriate economic evaluation of the intervention. It is important to identify the form of economic evaluation best suited to the intervention being evaluated and ensure that the collection of data required for that form of economic evaluation is planned into the study design from the outset. Early engagement, ideally throughout the entire process of development, will help identify the most appropriate form(s) of economic evaluation and to plan appropriate economic data collection to answer the decision question(s) of relevance to the central stakeholders and decision makers.

Economic evaluation is the systematic comparison of alternative interventions in terms of their costs and consequences. There is ample guidance on carrying out economic evaluation and this will not be repeated here (Drummond, 2015). Instead, this guidance focuses on additional considerations in the context of economic evaluation of complex interventions, updating it in light of the most recent methods guidance on the economic evaluation of public health (Tudor Edwards and McIntosh, 2019), economic evaluation of social care.
interventions as well as new and re-emerging methodological developments for economic evaluations relevant to complex interventions more generally.

Central to all forms of economic evaluation is the identification, measurement and valuation of the costs and outcomes of the interventions being evaluated. Identification involves knowing and listing all the potentially important costs and impacts to inform decisions about which should be measured given their potential importance and which can be measured given the methodological requirements of doing so in a valid and reliable way. Measurement requires quantifying the amount of resource and outcomes identified e.g. days in hospital or quality of life. Valuation relates to the final stage of placing values on the identified and measured costs and outcomes (e.g. unit costs and preference-weighted utilities). Stakeholder engagement at the identification stage is particularly useful in the context of complexity. The resources required to commission or deliver complex interventions and the indirect resource consequences of them are likely to be many and varied. Likewise, for complex interventions, the range of intended and unintended consequences for beneficiaries and those indirectly affected is likely to be broad and not necessarily obvious to evaluators (Jabeen, 2017). Engagement of provider organisations, those likely to experience additional costs (or savings) and those who are likely to experience the benefits or negative impacts of an intervention, is useful to ensure important costs and consequences aren’t missed. It is also essential to clarify the appropriate alternative intervention against which the intervention of interest might most usefully be compared.

Programme theories, typically expressed in one or more logic models, can be used to identify the cost and outcome variables that should and can be taken forward into the measurement and valuation stages of the economic evaluation. Indeed, in economic terminology this is referred to as conceptual modelling which essentially develops a logic model to (i) conceptualise the decision problem, and (ii) conceptualise and design the economic analysis itself, identifying what variables are measurable and identifying of key data sources, before later development into a decision model for economic analysis (Roberts, 2012). The use of programme theories as part of a conceptual modelling framework is particularly important for economic evaluation of public or complex interventions (Tappenden, 2012; Squires 2016). Economic evidence from existing studies, both economic
evaluations of comparable interventions, or insights from behavioural economics, can be used to inform the programme theories.

Decisions about what to measure and how also depend crucially on the type of economic evaluation carried out (see Box 5). Dopp et al (2019) provide a useful taxonomy of mixed-methods studies relevant to economic evaluation. Key to this is including qualitative methods both early in the process and latterly in the design and use of economic evaluation.

Box 5: Types of economic evaluation

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

For complex interventions, the broad range of potential outcomes and the populations and organisations affected means that narrower forms of economic evaluation, such as CEA or CUA, are unlikely to fully reflect the full range of costs and consequences of the intervention of interest. CBA and CCA are the ‘broader’ forms of economic evaluation suited to complex intervention evaluations, although they have limitations of their own. Ascribing monetary values to all the dimensions of benefit in a CBA is challenging and costly. CCA is the most straightforward methodologically and has the intuitive appeal of presenting outcome (consequence) data across the range of factors decision makers are likely to be concerned with in making a policy decision. However, without valuation of these outcomes (e.g. preference weighting as in QALYs or monetary valuation in CBA) decision making in the absence of known thresholds requires judgement. Further, if some outcomes are better for
some interventions and other outcomes are better for others, the policy implications of a CCA can be ambiguous.

Drawing on experience gained from producing guidance on the cost-effectiveness of public health and social care measures, the National Institute of Health and Care Excellence (NICE) has placed more emphasis on using cost–consequences and cost–benefit analyses in their recent guidance update (NICE, 2017). These types of economic evaluation are not mutually exclusive: CBA (or CUA and CEA) can be carried out and embedded in a wider CCA providing data on a more comprehensive inventory of outcomes.

The remainder of this guidance explores practical considerations for conducting economic evaluation in the context of complex interventions. Ways in which economic evaluation of complex interventions might differ from standard guidance on economic evaluation are highlighted throughout the guidance. New methods for carrying out economic evaluation of complex interventions are highlighted, including the important role of systems modelling work, given the ethical and practical constraints often faced in setting up experimental or quasi-experimental studies of complex health interventions. Further new methods include guidance for conducting economic evaluations alongside natural experiments (Deidda et al., 2019) and a new handbook on the economic evaluation of public health interventions (Tudor Edwards and McIntosh, 2019).
**Key points**

- From the outset of developing and evaluating complex interventions identify economic aspects, and design considerations for the future development of a formal economic evaluation.
- Include economists in the early stages of development and design of complex interventions.
- Development of an economic logic (conceptual) model alongside the program theory and logic model will help with understanding of the decision problem and shape the design of future feasibility and evaluation studies.
- Economic considerations are relevant at each of the stages set out in this guidance.
- Broader economic evaluation frameworks, such as cost-benefit and cost-consequence analyses, are likely to be the forms of economic evaluation best suited to the evaluation of complex interventions. Embedded cost-utility and cost-effectiveness analyses are also informative.

**Signposts to further reading/guidance**

- An introduction to the principles of health economics and economic evaluation relevant to the specific challenges of public health and its aims of improving population health (Tudor-Edwards and McIntosh, 2019).
- Guidance for conducting economic evaluations alongside natural experiments (Deidda et al., 2019).
3.5. Intervention modification

The effectiveness and success of an intervention is related to the system and context in which it is implemented. Therefore, it does not make sense to transfer an intervention that was successful in one context directly to another without considering how aspects of that context may impact on the intervention and vice versa (Evans et al., 2019).

Ongoing modification is encouraged so that an intervention being evaluated and/or implemented is the most optimal for a desired population and context. Ongoing modification can mitigate the risk of missed opportunities for understanding research, and reduce the delay before findings influence policy and practice. This is particularly important, for example, in the current climate of technological advances, where digital interventions offer potential for reach and impact. Such interventions require a rapid route to implementation to avoid them becoming ‘out-of-date’ (Michie et al., 2017). These types of studies, although needed, come with challenges such as complicated design and analysis, and the need to gather data quickly.

This guidance encourages those involved with complex interventions to use an approach of ongoing modification. This allows interventions to be adapted and optimised at the development, feasibility, evaluation and implementation stages. Members of the study team need to agree what level of modification is acceptable. This may differ between studies, however, may include agreement that any modification retains the key principles of the original intervention, such as theorised mechanisms of action, or use of theory (Pluye et al., 2005). As Hawe and colleagues have argued, the form of the intervention can be flexible as long as the key functions of the programme theory are delivered (Hawe et al., 2004).

Depending on the nature of the research question(s), and the interests and requirements of stakeholders, intervention modification can be undertaken via either: (i) a reactive adaptive approach; or (ii) an empirical optimisation phase (see Table 2).

[DETAIL TO BE ADDED – considerations when an ‘identified’ intervention undergoes modification].
Table 2: Reactive modification versus empirical optimisation approach

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Reactive modification</th>
<th>Empirical optimisation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A ‘living protocol’ that adapts with feedback in real time.</td>
<td>Identification of active intervention components delivered at the optimal dose.</td>
</tr>
<tr>
<td></td>
<td>A final product that has taken on board reflections throughout the process.</td>
<td>Rapid testing of theory.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Considerations</th>
<th>Decision needed about how much an intervention can be modified before it is no longer the same intervention.</th>
<th>Can be used in any setting but particularly useful for digital interventions.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Qualitative research can be used to refine and augment hypotheses, protocols and economic decision models prior to quantitative evaluation.</td>
<td>Choice of single or combination of multiple methods e.g. A-B testing, Multiphase Optimisation Strategy (MOST), adaptive interventions.</td>
</tr>
<tr>
<td></td>
<td>Ongoing iteration can take place throughout the process with theory testing and mini evaluations.</td>
<td>- A-B testing: allows rapid comparison of current intervention with previous versions (ref.).</td>
</tr>
<tr>
<td></td>
<td>Normalisation Process Theory can be one useful framework to support the identification of factors to improve at different stages of an intervention (Murray et al., 2010a).</td>
<td>- MOST: builds and evaluates interventions via identifying the optimal components and dose delivery leading to the best outcomes (Collins et al., 2007).</td>
</tr>
<tr>
<td></td>
<td>Involve stakeholders in decision-making processes related to intervention refinement.</td>
<td>- Adaptive interventions: assesses outcomes of components delivered in different sequences and/or at different doses (e.g. the Sequential Multiple Assignment Randomised Trial (SMART) (Collins et al., 2007).</td>
</tr>
</tbody>
</table>

| Case study | [add example] | [add example] |
Key points

- Modification is permitted, may be helpful, and should be done systematically and in a theory-informed way.
- On-going modification can speed up the implementation of research findings.
- Agree acceptable boundaries of modification beforehand.
- It benefits all stakeholders for an optimised version of an intervention to be evaluated and/or implemented.
- Intervention modification can be achieved via either an adaptive approach, or an empirical optimisation phase.

Signposts to further reading/guidance

- Normalisation process theory: a framework for developing, evaluating and implementing complex interventions (Murray et al., 2010b) (see also www.normalizationprocess.org/)
- A conceptual framework for adaptive preventive interventions (Collins et al., 2004).
- The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART): New Methods for More Potent eHealth Interventions (Collins et al., 2007).
- Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop (Michie et al., 2017).
- Continuous Evaluation of Evolving Behavioral Intervention Technologies (Mohr et al., 2013).
- [Add A-B testing ]

3.6. Uncertainties

There are various types of uncertainty, related to each research phase. Intervention development research is about reducing the uncertainty concerning the components of the
intervention. Feasibility research is about reducing uncertainty concerning the evaluation methods. The aim of intervention evaluation is to reduce uncertainty about whether the intervention leads to benefit rather than harm, to such an extent that decisions about the intervention can be made (Threlfall et al., 2014). Evaluation is not solely concerned with determining whether an intervention ‘works’, but in the overall effect i.e. in which circumstances it works, and for whom, addressing questions around reach and uptake of interventions. Implementation research is about reducing uncertainties or barriers to sustainable delivery of an intervention in different contexts.

The uncertainty regarding the intervention supports the choice of research perspective. Research questions should be prioritised based on where the key uncertainties lie, and these should be uncovered through existing theory and evidence, and experiential evidence through the engagement of multiple perspectives (i.e. relevant stakeholder groups). Uncertainties should be revisited throughout the research, not solely at the outset, as it is likely that the uncertainties will change over time.

In determining whether it is worthwhile to embark on further research activities, Value of Information methods are potentially useful for identifying areas of particular uncertainty, and identifying a monetary value limit for the costs of any further research activities (see section 3.4). Qualitative methods here can inform early economic decision models, which in turn can lead to Value of Information methods.

**Key points**

- Identify key uncertainties that the research aims to address and formulate research questions accordingly.
- Uncertainties may change throughout, therefore it is important to revisit and document them and any resulting changes in the research study.
- Engage multiple perspectives to develop a detailed understanding of where uncertainty lies for different stakeholder groups.
4. Phases of research

The following section outlines four phases of research: (i) developing or identifying the intervention; (ii) feasibility testing; (iii) evaluation; and (iv) implementation. As illustrated in Figure 2 these phases are not linear – they can progress in either direction, and are often overlapping. Each phase below is accompanied by two tables: one addressing the overarching considerations identified in section 3; and the other signposting to further guidance and resources.

4.1. DEVELOPING OR IDENTIFYING THE INTERVENTION

This phase of the research process was previously referred to as intervention development (Craig et al., 2008). However, this did not address evaluation of studies where researchers were not involved in their development. This is often the case in policy led interventions and natural experiments; we have therefore added intervention identification alongside intervention development.

4.1.1. Developing an intervention

Prior to conducting an evaluation an intervention should first be developed to a stage where it is expected to have an effect. There are different resources available to guide this process (see signpost box below), in particular the taxonomy for intervention development approaches (part of the MRC’s INDEX study) (O’Cathain et al., 2019). In summary, a robust process of intervention development should involve the following:

(i) *Develop theory of the problem:* It is important to understand the problem in order to propose informed solutions. Problems range in their ease of definition and how they could potentially be addressed (Wight et al., 2016). Input from a variety of sources may therefore be necessary to develop a strong theory of the problem. Sources include: the scientific evidence, in particular qualitative approaches; stakeholders (e.g. policymakers, clinicians, patients and the public); and your own professional work experience. Initial ideas about the intervention may change as understanding of the problem increases. For example, exploring the social distribution of the problem,
or defining the demands an intervention places on recipients could identify that certain types of intervention may widen inequalities.

(ii) **Explore the evidence base**: Identify the relevant, existing evidence base. The key aim of this is to: explore what interventions have already attempted to address the problem; which have been evaluated; what evidence exists to support key intervention components; and what evidence of effectiveness and cost-effectiveness exists, and from which contexts. Ideally this will involve carrying out a systematic review, if one does not already exist. You may be lucky and find a recent high-quality review that is relevant to your intervention, but you may have to conduct one yourself, or update an existing one, and maintain and update it as the evaluation proceeds.

(iii) **Develop theory-led research questions**: Theory can also inform the relevant research questions that need to be answered to enhance the evidence base related to the problem and increase the likelihood of future effectiveness and implementation. There is some evidence that interventions based on theory are more effective, although the evidence is somewhat mixed, possibly due to poor implementation of theory (Davis et al., 2015). Start by identifying one or more potential theories. These may include theories related to behaviour change, implementation, psychology, sociology etc. (O’Cathain et al., 2018 (forthcoming), Croot et al., 2018 (forthcoming)). Your understanding of how the chosen theories represent the intervention may change as this process progresses. Theories are a helpful starting point, but may be used flexibly. Economic considerations can also be integrated into this process in helping to shape an appropriate research question. For example in incentive studies (ref) the role of qualitative work can help inform on the amount of incentives most likely to influence a change in behaviour (ref), as well as choice architecture, understanding preferences and the trade-offs associated with choices (Discrete choice experiments) (ref) and the role of time preference in designing interventions [add example]. Establishing the key areas of uncertainty and research questions at an early stage helps inform the most appropriate intervention and evaluation methods.

(iv) **Consider the context and the system**: Adopting a systems lens guides the intervention development team to consider the wider context within which the intervention operates (see Table 3 below). Meta-syntheses of qualitative studies can be useful for identifying contextual factors (Taylor et al., 2011). There is also guidance available to
help identify what elements of context may influence an intervention (Craig et al., 2018). In brief, these may include factors associated with: individuals; wider population; intervention environment; geography; social and cultural influences; political or economic influences; and broader implementation issues, such as, policy. It is important to understand these factors to prevent the development of interventions which are ineffective or which fail to be implemented. The system or parts of the system could be mapped to help facilitate this process. Relevant information can be gathered via qualitative engagement with stakeholders, for example using soft systems methodology (see Box 3, page 19), exploring the evidence base and/or primary data collection. Some contexts change rapidly, such as digital technology, where new platforms and software are being developed and there are regular shifts in what is popular, and this should be considered (Michie et al., 2017).

(v) **Develop programme theory:** All interventions should have a programme theory – which describes how change will be achieved and under what conditions. In the early stages of the research this may be underdeveloped, but should become more detailed as understanding increases via exploring the evidence base, mapping the system, and receiving input from stakeholders (Hawe, 2015, Rogers, 2008). There may be many competing or partly overlapping theories and finding the most appropriate ones will require expertise in the relevant disciplines within the team. The programme theory should be considered dynamic and updated at all stages of the development and evaluation process as new information becomes available. Stakeholder input, as well as primary and secondary data collection, can be used to test and refine the programme theory.

(vi) **Economic considerations:** Early consideration of economic aspects alongside the processes of stakeholder engagement, adoption of a systems lens and exploration of programme theory and development of a logic model can help shape the development of appropriate research question in parallel with the intervention development. This in turn enables design of an appropriate economic evaluation of the intervention. Behavioural economic insights (such as prospect theory) and the understanding that human decisions are strongly influenced by context, including the way in which choices are presented may help evolve the programme theory, and
assist with the development of appropriate theory led research questions. While the evidence base is explored (see above), there should ideally be a review of the economic evidence (nationally and internationally) during the development stage. This can help establish the economic evidence base, and is important in terms of establishing what is appropriate to the current context, identifying gaps, uncertainties and unknowns, identifying cost drivers and suitable outcome measures. Prior to undertaking an economic evaluation or an early stage decision model it is important that a logic model (often known as a conceptual model) be developed to help economists understand the complexity of the real system that the model will attempt to represent. An economic logic model will help inform the choices available for translating this complexity into a credible conceptual and mathematical structure (Tappenden, 2012; Squires 2016).

(vii) **Model process and outcomes:** Following development of programme theory and a logic model, and prior to a full-scale economic evaluation, an early stage decision analytic model of the complex intervention can be undertaken – comparing the intervention to one or more relevant comparators. This will establish potential cost-effectiveness of the planned intervention and highlight areas of uncertainty which should be included or focussed on in forthcoming research. Value of Information (VOI) analyses can then be undertaken (Fenwick, 2005) to estimate a monetary value on further research activities, based on the expected gain from reducing uncertainty surrounding the current cost-effectiveness decision.

(viii) **Consider implementation:** Implementation is covered in more detail within section 4.4. However, it is important to highlight that preparing for implementation during the development stage can help with refining and optimising the intervention. Exploring the best methods of delivering intervention components, both to individuals, and by intervention providers, is a key part of successful implementation. Importantly, this links back to the use and understanding of context. Normalisation Process Theory (Murray et al., 2010a) provides guidance on how to improve this process (see Signpost box below) and approaches like the RE-AIM framework (Glasgow et al., 1999b) can help evaluators consider adoption and sustainability of implementation. Questions that should be considered include: Could this intervention be used? By whom? In what population? In what setting? What barriers might there
be to implementation? Thinking about implementation at this stage also helps identify who will need to know about the outcomes, and what data will they need to inform their decisions. Therefore, early economic considerations and development of the research question should also have considered implementation issues. There are logistical, structural and political issues which can hinder implementation.

(ix) **Involve stakeholders:** Types of stakeholders to consider are outlined in section 3.1. Including relevant stakeholders in the development process can improve the quality of the resulting intervention, and also help identify considerations for future implementation. In particular, co-production alongside experienced service providers can promote an interventions feasibility and potential for implementation within a specific context. Co-producing with the right stakeholders provides opportunity to share their knowledge of the setting and target population (Hawkins et al., 2017). In addition to qualitative methods, such as stakeholder interviews and focus groups, soft system methodologies can be a helpful approach (Box 6). For example, systems mapping workshops can be facilitated to engage groups of multidisciplinary stakeholders. These workshops can: map the system; develop a shared understanding of the system from different perspectives; agree the system boundary and the system of interest; identify the problem(s); and co-produce solutions (Allender et al., 2015, Wittenborn et al., 2016). It is important to have this dialogue and generation of shared understandings about the system and the problem (Buckle et al., 2010).

(x) **Modification:** The end product of development should be an intervention that is optimal given the resources and information available. The development phase should therefore be considered flexible, iterative and adaptable. There may be iterative cycles of development, followed by stakeholder feedback and further iteration. This cycle will be repeated until an intervention is produced that is considered ready for feasibility testing/evaluation.
Box 6: Examples of systems tools that can be used in stakeholder workshops

[DETAIL TO BE ADDED]

Brief overview of tools that can be used to define the problem, take a whole systems approach, explore different perspectives etc.

E.g. Rich pictures, context diagrams, multiperspective diagrams, laddering, concept maps, cognitive maps, and causal loop diagrams.

Intervention development approaches should be tailored to the needs of the team and the context including resources. It may not be possible or necessary that each of these is addressed but each should be considered with regard to relevance and importance. The steps identified above are not intended to be linear, they are dynamic, overlapping and iterative and different activities will occur in parallel and inform each other. Teams involved in developing interventions are encouraged to publish their development approaches. This can help others improve their own interventions.

4.1.2. Identifying an intervention

Researchers are not always involved in the design of an intervention. Often, interventions are policy-driven and/or designed in clinical practice. Researchers therefore can have different starting points in the process of developing/identifying and evaluating an intervention.

Although researchers may not be involved in the design stage they should still employ the six overarching considerations (Table 3). In particular, they should undertake the development of a programme theory. This is an important step, even if the programme theory and accompanying logic model is developed after the intervention has been rolled out. This helps identify mechanisms, important contextual factors and appropriate outcome measures for data collection (i.e. what do you expect this intervention to impact on, positively or negatively?).
The role of Natural Experiments

Natural experiments can be identified from the withdrawal and scaling back of interventions, as well as the introduction of new ones. Evaluation of a natural experiment may be appropriate when: an RCT is not practical or ethical; where a significant health impact is expected but there is uncertainty about the nature and size of the effect; or the intervention has potential for being replicated and scaled-up (Craig et al., 2017). Other uses include producing evidence quickly and at lower cost; with high external validity while minimising threats to internal validity; and producing evidence on long term and non-health outcomes (Bärnighausen et al., 2017). Further detail is provided by the MRC’s guidance on natural experiments (Craig et al., 2017).

Evaluability assessment

Evaluability assessment is a process for determining whether a policy or practice driven intervention should be evaluated, and if so, which methods are the most appropriate. Its main purpose is to minimise research waste on poorly designed and/or implemented interventions, or those cases in which suitable data sources to permit outcomes evaluation are not available. Evaluability assessment involves a systematic approach, with a range of stakeholders collaborating towards a decision based on what the expected outcome of the intervention is, what data could be collected to assess processes and outcomes, and the options for approaching the evaluation (Craig and Campbell, 2015). Although they share some elements (for example, developing programme theory), it is important to distinguish that an evaluability assessment is a form of evaluation design, rather than intervention design. The end result is a recommendation on whether an evaluation: is feasible; should be undertaken; can be carried out at a reasonable cost; and by which methods (Ogilvie et al., 2011). Further details on when and how to conduct an evaluability assessment can be found in the signpost box below.

Table 3: Overarching considerations for developing or identifying the intervention

<table>
<thead>
<tr>
<th>Overarching considerations</th>
<th>Issues to consider</th>
<th>Risk of not considering them</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Stakeholders
- Include relevant stakeholders in the development process, in particular for: Developing theory of the problem; considering system and context; developing programme theory; providing feedback to help modify and improve the intervention; collaborating on evaluability assessments; and identifying potential issues with future implementation.
- Choose different approaches to engage different stakeholders e.g. interviews/focus groups versus soft system methodologies.
- Resulting intervention that fails to achieve its proposed effect.
- Effective intervention that fails to be adopted and implemented.
- Resulting evaluation that fails to generate useful data on the processes, mechanisms and/or outcomes of the intervention.
- Reduced chance of impact on policy and practice.

### Programme theory
- What is the underlying theory, or combination of theories?
- What are the proposed mechanisms of change?
- What are the proposed short, medium and long-term outcomes?
- What are the wider impacts of the intervention on the system?
- Explore the current evidence base, in particular, for information to suggest the proposed intervention may not work as intended.
- An intervention that has little chance of working even in idealised settings.
- Intervention that is unlikely to be sensitive to different contexts ‘in the real world’.
- Failure to measure impacts of the intervention that might be important.
- Unable to advance theory.

### Context
- What are the immediate contextual influences on the intervention?
- What is the environment in which the evaluation will take place?
- Where is the system boundary?
- Lack of understanding context may result in an intervention that works in one setting, but may be ineffective, or even harmful, when delivered elsewhere.

### Economic
- What is the existing economic evidence?
- Remaining uncertainty not identified
| **considerations** | • What is the appropriate economic question – taking account of the system and program theory/economic logic (conceptual) model developed?  
• What perspective(s) is/are appropriate?  
• What type of economic evaluation is most appropriate to address the research question?  
• Should spillover impacts be included? | • Inappropriate question with limited focus  
• Economic results which under/over estimate potential cost-effectiveness.  
• Does not capture the full range of costs and outcomes. |
| **Modification** | • What aspects of the *intervention* could be improved prior to delivery? | • Delivery of a sub-optimal intervention and/or evaluation.  
• Failure to effectively gather meaningful data on potential mechanisms and outcomes of the intervention.  
• Waste of resources. |
| **Uncertainty** | • Where is the uncertainty? Is it in the nature of the problem; the nature of the intervention and its components; the interaction of the intervention with the system; or the feasibility of evaluation? | • Aims of the intervention and the methods of evaluation do not match up.  
• Failure to generate meaningful data. |
Key points

- Start by identifying or developing theory of the problem and proposed solutions.
- Consider existing scientific evidence.
- Establish research questions.
- Establish theory-led research questions to drive choice of evaluation methods.
- Involve stakeholders, including PPI.
- Consider the value of the information generated from the proposed research.
- Consider the existing economic evidence.
- Use an economic logic (conceptual) model to develop the appropriate economic question.

Specific considerations for intervention development

- Model process and outcomes.
- Develop and refine programme theory.

Specific considerations for intervention identification

- Undertake evaluability assessment to determine whether and how an evaluation should be undertaken.
- Identify your starting point within the research cycle most applicable to the current status of the intervention.
- Consider which natural experimental methods could be used to evaluate the intervention.
- Develop and refine programme theory if one does not exist.
Signposts to further reading and guidance

- Six steps in quality intervention development (6SQuID) (Wight et al., 2016).
- Taxonomy of approaches to developing interventions to improve health: a systematic methods overview (part of the INDEX study) (O’Cathain et al., 2019).
- Normalisation process theory: a framework for developing, evaluating and implementing complex interventions (Murray et al., 2010a).
- Evaluating the public health impact of health promotion interventions: the RE-AIM framework (Glasgow et al., 1999b)
- Taking account of context in population health intervention research: guidance for producers, users and funders of research (Craig et al., 2018).
- Normalisation process theory: a framework for developing, evaluating and implementing complex interventions (Murray et al., 2010a).
- Taxonomy of approaches to developing interventions to improve health: a systematic methods overview (part of the INDEX study) (O’Cathain et al., 2019).
- Evaluating the public health impact of health promotion interventions: the RE-AIM framework (Glasgow et al., 1999b)
- Normalisation process theory: a framework for developing, evaluating and implementing complex interventions (Murray et al., 2010a).
- Evaluating the public health impact of health promotion interventions: the RE-AIM framework (Glasgow et al., 1999b)
- Taking account of context in population health intervention research: guidance for producers, users and funders of research (Craig et al., 2018).
- Natural Experiments: An overview of methods, approaches, and contributions to public health intervention research (Craig et al., 2017).
- Evaluability Assessment: (i) A Systematic Approach to Deciding Whether and How to Evaluate Programmes and Policies (Craig and Campbell, 2015); (ii) Assessing the evaluability of complex public health interventions: five questions for researchers, funders, and policymakers (Ogilvie et al., 2011).
- Guidance on conducting economic evaluations alongside natural experiments (Deidda et al., 2019).
- Economic evaluation of public health interventions (Tudor-Edwards and McIntosh, 2019).

4.2. FEASIBILITY

Note re: terminology: We use the term feasibility in this guidance, acknowledging that this type of work may also be referred to as an exploratory or pilot study. The term feasibility is consistent with the recent CONSORT extension for feasibility and pilot studies and the newly published MRC guidance on feasibility studies (in press).
4.2.1. Why is a feasibility stage needed?

The feasibility stage explores the feasibility and acceptability of the intervention and also the evaluation design in order to underpin decisions about progressing to a full effectiveness study. Methodological research suggests that this vital preparatory work is often rushed (Eldridge et al., 2004). The negative impact of this is seen in evaluations which are often undermined by problems related to delivery of the intervention including compliance and acceptability, issues with recruitment and retention, and smaller-than-expected effect sizes. In the field of public health, interventions are often developed and tested within timeframes that do not allow for sufficient exploration of their feasibility (Kessler and Glasgow, 2011, Sanson-Fisher et al., 2007). This increases the likelihood of failed implementation and failure to generate evidence that is useful. It is therefore important to test certain processes to identify the feasibility and/or adaptations needed to optimise both the:

(i) intervention design (e.g., mode of delivery);
(ii) evaluation design (e.g., choice of outcome measures).

A feasibility study is not necessarily a ‘scale model’ of the proposed main evaluation, but designed to explore the uncertainties that have been identified in the development work. A mixture of qualitative and quantitative methods will likely be needed, for example to understand barriers to participation and to estimate response rates. Process evaluations have an important role here in exploring issues of context, mechanisms of impact and implementation (Moore et al., 2015). Recent developments in the design and evaluation of feasibility trials are reflected in the new MRC Guidance for fEasibility STudies (GUEST) (in press).

**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” (Moore et al., 2014).
4.2.2. Feasibility of the intervention design

Following development of the intervention, it should be tested to optimally address the key uncertainties that exist prior to considering an effectiveness study (see Table 5). These are likely to include uncertainties related to: optimal intervention content and mode of delivery; acceptability of the intervention to participants and providers; ability to collect appropriate input and outcomes data for assessment; likelihood of cost-effectiveness, and capacity of providers to deliver the intervention and fit within the setting.

Adaptation of the intervention should be undertaken, as required, to result in an optimised intervention being taken forward for full evaluation (see Table 5). As discussed in Section 3.5 any modification should have agreed limits and be reported transparently. Depending on the results of the feasibility study, further work may be required to progressively refine the intervention, before embarking on a full-scale evaluation. Ideally incremental adaptations should each be tested separately, but in practice adaptations can be made simultaneously if sufficiently rich data are collected to enable judgements about which adaptations are helpful and which are not. Practical constraints eventually force the decision that the intervention is ‘good enough’ to go to the next step.

Ultimately the goal of a feasibility study is to produce a refined intervention and programme theory ready for a full evaluation (see Table 4). The programme theory should be refined in on an on-going way throughout the feasibility study and indeed all phases of the cycle (see Figure 2), as useful data emerges, for example, understanding related to the context in which the intervention is being delivered.

4.2.3. Feasibility of the evaluation design

In addition to testing the feasibility of the intervention, it is important to assess the feasibility of the evaluation design. The primary purpose of a feasibility study is not to assess effect size but to generate useful data to improve the quality of a future full scale evaluation (Cook et al., 2018). The evaluation should be designed to optimally address the key uncertainties that exist prior to an effectiveness study. These are likely to include uncertainties related to: recruitment, retention, sample size; acceptability of randomisation; duration of follow-ups;
choice of outcomes measures; floor or ceiling effects; potential harms; reasons for attrition; or the impact of the intervention on inequalities.

4.2.4. Economic considerations for feasibility phase

There are multiple reasons for incorporating economic considerations and economic modelling at this stage. We outline some of the benefits below:

- Checking through the conceptual model and adapting it to outline/define what data we (i) want to collect given the appropriate perspective(s) and type of evaluation, and (ii) are able to collect and from what sources. Data to inform early economic analysis at the feasibility stage may come from multiple sources (ref).

- Qualitative and process evaluation feed valuable information and parameters into the economic evaluation, whether that be an economic analysis/model to feed into the feasibility intervention design or the evaluation design – use of a decision model to synthesise data from multiple sources as well as evaluation using all the data from the feasibility study is extremely helpful at this stage. [Include examples] Economic (financial, non-financial and patient perspective) questions can be added into qualitative research (interview guide, questionnaire, phone interviews) to elicit a sample of patient perspective responses. To incorporate this adequately requires economic involvement early on with the qualitative and process evaluation teams.

- Development of data collection tools and systems in the feasibility study is of great importance, requiring time and careful consideration, followed by testing of these for acceptance – both by the ‘patient’ completing the form/tool and by those delivering the intervention. This type of data may not be routinely collected in current system or may not be in an appropriate form for easy incorporation into the study dataset. In these circumstances, development of specific data collection tools and a process for prospective data collection is suitable and acceptable to all those involved, would be required. [Add example] In addition, with complex interventions the standard care or comparator arm may differ in mode of delivery by site - again requiring careful consideration of data collection forms and processes at each site.

- Economic evaluation within a feasibility study is undertaken in exactly the same way as a full-scale trial, although the aim is not necessarily to show statistical significance.
However, the study findings may well be significant and provide useful information on completion rates, resource use data collection, patient acceptability, appropriate outcome measures and political issues. This information can be used in economic modelling to indicate the likelihood of cost-effectiveness at the feasibility stage and help decision makers decide whether it is worthwhile proceeding to a more definitive study. It should be noted that the analyses may also increase rather than reduce uncertainty, which in itself is a helpful outcome. For example if a policymaker is convinced that something works an evaluation might usefully reduce that certainty where it demonstrates the certainty is misplaced. (EXAMPLE TO BE ADDED).

- Early stage economic modelling either prior to and/or during the feasibility stage can then be followed by Value of Information analyses. Such analyses consider the current uncertainty in the cost-effectiveness decision (given current information) and estimate a monetary value on further research activities, based on the expected gain from reducing uncertainty surrounding the current cost-effectiveness decision. Expected value of perfect parameter information (Claxton) can also be undertaken, which looks at identifying specific parameters or groups of parameters which drive the uncertainty in the economic model, aiding the focus of further research activities to focus on these key parameters with high uncertainty. [Add examples].

4.2.5. Progression criteria

Data generated from a feasibility study should be assessed against predetermined progression criteria to assess whether the intervention is feasible and acceptable and whether the evaluation design is suitable for measuring effectiveness in a future study, and thus whether the research should progress to a full evaluation. Progression criteria should be developed with input from an independent steering group and relevant stakeholders. To ensure transparency the criteria should be approved prior to the collection of data.

Assessment of progression criteria will typically require both qualitative and quantitative data. Examples include: (i) qualitative data such as acceptability of the intervention content and mode of delivery for participants and providers; and (ii) quantitative data such recruitment and retention rates (Simpson et al, 2019 in press).
Progression criteria should be used to guide the decision to proceed or not proceed to a full effectiveness study or to complete further work, which might involve revisions to the intervention and further feasibility work. Criteria should be worded in a way that provides scope for discussion with the steering committee rather than rigid stop-go criteria. Use of a traffic light system for progression criteria is a useful approach (green=go, where there are no issues identified; red=stop, where issues identified cannot be resolved; and amber=amend, where issues can be resolved). Ultimately the decision to proceed lies with the study team (funding and ethical and governance approvals permitting).

Involving stakeholders in the design of the feasibility study and the progression criteria is critical in ensuring that sample sizes and outcomes are generating the data needed for future decisions (see Table 5).
Table 4: Overarching considerations for the feasibility phase.

<table>
<thead>
<tr>
<th>Overarching considerations</th>
<th>Issues to consider</th>
<th>Risk of not considering them</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stakeholders</td>
<td>• Input is critical to ensure correct data is collected to inform future delivery, evaluation, implementation, and impact.</td>
<td>• Lack of useful and convincing data to inform progression to evaluation and/or implementation.</td>
</tr>
</tbody>
</table>
| Programme theory           | • Begin with a proposed programme theory.  
• Refine the programme theory as data is gathered to inform how the intervention operates.  
• Aim to produce a strong programme theory by the end of study.  
• Incorporate aspects of process evaluation to explore potential mechanisms of impact. | • Key context dependencies are neglected resulting in an intervention and evaluation that fails to operate as designed. |
| Context                    | • Determine the context dependencies for the intervention and evaluation.  
• Consider how these contextual factors may change for a future evaluation or implementation within various settings.  
• Incorporate aspects of process evaluation. | • An intervention and/or evaluation that does not demonstrate feasibility and/or acceptability, with no understanding why.  
• An intervention and/or evaluation that fails to meet progression criteria. |
<table>
<thead>
<tr>
<th>Economic considerations</th>
<th>Resource use and categories of costs are needed to inform a full economic evaluation.</th>
<th>Under estimation of time and resources required to collect resource use and outcome data - could ultimately lead to ‘failure’ of trial or ‘ineffective intervention’ due to poor data collection.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Logistical issues with resource use data collection (practical, logical etc. establishing process).</td>
<td>Data collection tools or systems which are unacceptable to patients/clients or those delivering the intervention.</td>
</tr>
<tr>
<td></td>
<td>Development of data collection tools and systems. Testing acceptability of collection tools.</td>
<td>Inappropriately designed evaluation for definitive trial.</td>
</tr>
<tr>
<td></td>
<td>Qualitative and process evaluation can feed valuable information into the economic evaluation alongside the feasibility study or economic model.</td>
<td>Expense of proceeding with full scale definitive trial for an intervention that is highly unlikely to be cost-effective or implementable in practice.</td>
</tr>
<tr>
<td></td>
<td>Establishing appropriate economic outcomes and endpoints.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VOI – potentially useful for identifying areas of particular uncertainty for further research and identifying a monetary value limit for the costs of any further research activities.</td>
<td></td>
</tr>
</tbody>
</table>
### Modification

- Agree boundaries and limits on how much the intervention or evaluation design can be modified.
- Report all modifications transparently.

| Risk of proceeding to full evaluation with a sub-optimal intervention and/or evaluation design. |

### Uncertainty

- Is the intervention and evaluation acceptable to participants and providers?
- Can participants be recruited, randomised (if appropriate), and retained for follow-up?

| Failure to understand the interventions potential to progress to full evaluation/implementation. |
| Insufficient information on benefits and harms. |

### Key points

- Assess the feasibility of the intervention design and the evaluation design.
- Use qualitative and quantitative methods.
- Avoid using data to assess effectiveness.
- Involve stakeholders to ensure relevant data are collected for future decision-making.
- If needed, adapt the intervention and/or evaluation design (within agreed boundaries) and report changes transparently.
- Use pre-defined progression criteria (where appropriate) to guide decisions about proceeding to a full evaluation.
- Identify and measure key resource use and outcomes for the purposes of designing a full economic evaluation.
- Consider whether it is appropriate and relevant to undertake early economic evaluation and Value of Information analyses.
Signposts to further reading/guidance

- Taking account of context in population health intervention research: guidance for producers, users and funders of research (Craig et al., 2018).
- Applied Methods of Cost-Benefit Analysis in Health Care (McIntosh et al., 2010).
- Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement (Husereau et al., 2013).
- Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report (Ramsey et al., 2005).

4.3. EVALUATION

4.3.1. Effectiveness and usefulness

Evaluation enables judgments to be made about the value of an intervention. This does not solely mean whether an intervention ‘works’, but what other impacts it has, theorising how it works, taking account of how it interacts with the system, its context, and how the evidence can be used to support real-world decision making. The choice of evaluation design should be the most appropriate to answer the research questions posed, which should have been carefully generated from consideration of the uncertainty surrounding the intervention and informed by theory (see sections 2 & 3).

A key change in the field of evaluation has been a shift towards assessing the ‘usefulness’ of information for decision-making, as opposed to focusing exclusively on obtaining unbiased estimates of effectiveness (Deaton and Cartwright, 2018). Randomisation, coupled with other elements of trial design, such as concealment of the allocation sequence and blinding of participants and researchers, can minimise biases associated with selective exposure to the intervention. To maximise the usefulness of the evidence for decision-making, other elements are needed, such as the use of programme theory and consultation with stakeholders to identify outcomes that matter, and a process evaluation to assess elements...
of implementation, context, and mechanisms of impact (Moore et al., 2014). A more fundamental challenge to the conventional idea that precise and unbiased estimates of effectiveness are the goal of evaluation is the proposal that improving theories about how to intervene should be the ambition. In this view, effect estimates are inherently context-bound, so that average effects are not a useful guide to decision-makers working in a particular context. More useful may be contextualised understandings of how an intervention works, and what are the most important enablers and constraints on its delivery across a range of settings.

Often, questions about the delivery, reach and maintenance of an intervention are assigned to ‘implementation’, and explored separately from questions of effectiveness. Some frameworks, such as RE-AIM, try to pull them together (Glasgow et al., 1999a). A key recommendation of this guidance is that implementation questions should be considered alongside effectiveness ones from the outset. Moving away from a narrow ‘effectiveness’ focus should increase the relevance of the evidence produced by evaluation studies, and increase the speed with which effective interventions can be implemented within policy and practice.

In section 4.3.2, we revisit and expand on the research perspectives introduced in section 2. In section 4.3.3 we describe the main options for evaluation design and then in section 4.3.4 we consider the choice of outcome measures.

4.3.2. Research perspectives & evaluation

Table 5 sets out the main research perspectives used in evaluation: (i) efficacy (ii) effectiveness; (iii) realist; and (iv) systems. In section 2 we provided guidance as to how to make a decision about the research perspective. Here, we focus on the study designs available for evaluation.
Table 5: Research perspectives & evaluation

<table>
<thead>
<tr>
<th>Research perspective</th>
<th>What the research perspective means for evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Efficacy</strong>&lt;br&gt;What works in ideal circumstances?</td>
<td>Efficacy research focuses on answering research questions about whether the intervention works in experimental settings. It is a proof of concept stage, does not take account of complexity and real world change. This is not to say that there is not a place for efficacy studies; the uncertainty surrounding the intervention may be such that a proof of concept type study in controlled settings is an important first step. However, this type of evaluation is unlikely to be suitable for complicated interventions because the features of complicated interventions e.g. numerous components, are likely to mean that there is interaction between features of the intervention itself, and system properties such as feedback loops are possible.</td>
</tr>
<tr>
<td><strong>Effectiveness</strong>&lt;br&gt;What works in the real world?</td>
<td>Most commonly, evaluations focus on identifying whether or not the intervention ‘works’: does the intervention lead to a change in hypothesised outcomes? This question may be addressed in idealised circumstances to establish efficacy (as above), or in a more pragmatic real-world evaluation to establish effectiveness. In both cases, the principal focus is to identify an unbiased estimate of the average effect of the intervention, either with precision in a homogenous context (efficacy) or with less precision but among a more heterogenous context that is more representative of the scenarios in which the intervention will be applied in practice.</td>
</tr>
<tr>
<td><strong>Realist</strong>&lt;br&gt;What works, for whom, under what circumstances and why?</td>
<td>A realist research perspective emphasises the importance of understanding how intervention mechanisms may generate different outcomes in different contexts. This approach places less value on the average effect estimate and more value on understanding the interplay of mechanism and context, how change is being brought about and how this may vary across different contexts and recipients.</td>
</tr>
<tr>
<td><strong>Systems</strong>&lt;br&gt;What contributes to improving the system in a positive way?</td>
<td>Systems thinking treats the intervention as a change or disruption to a complex system structure and/or to relationships within the system. For example, it focuses on the way that a system’s constituent parts interrelate and how systems work over time and within the context of larger systems. It recognises the dynamic nature of systems and specific properties of that dynamic interdependence, such as emergence and feedback.</td>
</tr>
</tbody>
</table>
4.3.3. Types of evaluation design

There are many study designs to choose from, and different designs suit different questions and different circumstances (McKee et al., 1999). There is no clear mapping of research perspective to evaluation designs, though a purely quantitative approach, using experimental design, is rarely appropriate. This is particularly true in complexity-informed research where answering research questions using a realist or systems perspective is necessary; qualitative and mixed method designs are suited to answering questions beyond effectiveness. A design may rarely be used in a particular field, but that does not mean it cannot be used, and you should make your choice on the basis of specific needs of your evaluation, in particular the research questions that you aim to answer. Awareness of the range of experimental and non-experimental approaches should lead to more appropriate methodological choices. Examples of evaluation designs are presented in Box 7 and elaborated on below.
Box 7: Experimental and non-experimental evaluation designs

### Experimental designs

- Individually randomised controlled trials [CASE STUDY TO BE ADDED]
- Cluster randomised trials [CASE STUDY TO BE ADDED]
- Randomised Stepped wedge designs [CASE STUDY TO BE ADDED]
- Preference trial and randomised consent designs [CASE STUDY TO BE ADDED]
- N-of-1 designs [CASE STUDY TO BE ADDED]

### Natural experimental designs

- Differences in exposure within populations [CASE STUDY TO BE ADDED], e.g. instrumental variables, propensity scoring, regression discontinuity designs (Moscoe et al., 2015)
- Differences in exposure between populations or over time [CASE STUDY TO BE ADDED], e.g. interrupted time series, difference in differences, synthetic controls

### Systems designs

- Modelling e.g. agent-based modelling [CASE STUDY TO BE ADDED]; system dynamics modelling [CASE STUDY TO BE ADDED]; social network [CASE STUDY TO BE ADDED]
- Case study e.g. qualitative comparative analysis, realist case study [CASE STUDY TO BE ADDED]

### Experimental and natural experimental designs

These designs seek to focus on maximising internal validity, but other design elements are needed to increase external validity. Therefore, on their own, experimental and natural experimental studies are limited to answering questions that take an effectiveness perspective. However, these studies can be conducted in combination with methods that explore the process or context and are appropriate for realist or systems perspectives.

Experimental designs randomise individuals or groups to either receive the intervention or be part of the control group, which does not receive the intervention. Natural Experiment studies use intervention and control groups which are constructed by a processes outside
the control of the researcher such as implementation of policy or legislation e.g. means testing for social security benefit, or a diagnostic or educational test score. Randomisation is not a necessary, or usual, feature of natural experiment studies, but is sometimes available e.g. Brazil’s Minha Casa, Minha Vida housing programme, which is allocated by lottery [REF TO BE ADDED]. Natural Experiment studies may have higher external validity than other experimental designs, because they are evaluating the intervention as it is implemented. However, internal validity depends on a good understanding of the process that divides groups or individuals into comparator and intervention groups.

Beyond the list of established evaluation designs, with linked case studies (as linked to in Box 7), to avoid simply reiterating the previous guidance or the subsequent Natural Experiment guidance (Craig et al., 2012b), we focus here on the evaluation designs that have come to fore since those documents were published. There are a number of evaluation designs that have come to fore since previous guidance was published:

**Randomised stepped wedge designs**: This is a variant of the cluster randomised controlled trial design. It may be used to overcome practical or ethical objections to experimentally evaluating an intervention for which there is some initial evidence of effectiveness, or which cannot be made available to the whole population at once. It allows a randomised controlled trial to be conducted without delaying roll-out of the intervention. Eventually, the whole population receives the intervention, but with randomisation built into the phasing of implementation. Randomised stepped wedge designs are also increasingly being used and provide an alternative to regular cluster randomised controlled trials, where it is not possible to randomise individuals to the intervention (Hemming et al., 2015). Given that the roll out of the intervention in a stepped wedge design has a temporal element, particular consideration should be given to any changes to the context that concurrently occur.

**Adaptive trial designs**: This trial design involves advance planning for modification to the trial design throughout the study period, based on interim data (Thorlund et al., 2018). Adapations to trials could include, for example, the decision to drop inferior trial arms, reassessment of sample size requirements, and reassessment of eligibility criteria. Adaption of the trial is based on careful planning, simulation of potential scenarios, and resulting
decision rules. Advantages over fixed trial design include improvements in efficiency and trial success.

**Preference trials and randomised consent designs:** Practical or ethical obstacles to randomisation can sometimes be overcome by the use of non-standard designs. Where patients have very strong preferences among treatments, basing treatment allocation on patients’ preferences, or randomising patients before seeking consent, may be appropriate (Case study x). **Expertise-based randomised controlled trials** are applicable for the delivery of interventions which are substantially skill-based and in which individual practitioners may be familiar with only a particular approach or have a strong preference for one intervention over another (Cook et al., 2015). Patients are randomised to a one of two or more practitioners, who only provide the intervention for which they have expertise.

**Natural experiment designs:** This is a growing field, with increasing recognition of the value of using these designs and as such there has been considerable methodological development (Bärnighausen et al., 2017, Basu et al., 2017, Craig et al., 2017). There are some new methods, such as the use synthetic controls, where experiences are accumulating very rapidly [CASE STUDY] (Bouttell et al., 2018).

**Evaluation of digital interventions:** Considerations specific to evaluating digital interventions have been summarised by Murray et al (2016), who have suggested that although traditional evaluation methods e.g. RCTs, are appropriate they should be done following a more iterative approach to development and optimization. To allow for multiple versions of web-based or mobile interventions to be deployed simultaneously, continuous evaluation of evolving behavioural intervention technologies or CEEBIT trials have developed (Mohr et al., 2013). The strategy allows interventions to be continuously updated and evaluated, promoting ongoing development and assessment of the e-mental health interventions after deployment.

**Systems methodologies for evaluation**

As yet, systems methods for evaluation in health and social care/policy have largely been limited to describing and modelling systems rather than for actually evaluating (Carey et al., 2015, Rutter et al., 2017). Systems modelling designs are not strictly evaluative methods, but
they may support the understanding of potential effects of interventions, which may not be directly observable, by simulation. Agent-based and system dynamics modelling methods are computational simulations that draw upon objectively measured and experiential data from a range of perspectives. The starting point for these methods is often other soft-systems methodologies as touched upon in the development phase (section 4.1). These models can be used to evaluate different scenarios, e.g. different variants of an intervention, or to explore longer-term effects, which may occur outside of the time-frame of an empirical evaluation, or spill over effects that may not be captured in a small scale trial. Agent-based modelling involves virtual representation of real world situations that models individuals and their interactions with their environment, allowing for system properties such as feedback and emergence (Bruch and Atwell, 2013). System dynamics modelling is an approach that aims to understand the behaviour of complex systems over time, where the starting point is the structures that trigger the behaviours of the system itself, rather than the entities within it (as in agent-based models).

Network analysis is a tool to understand structures of systems by studying relationships within a system e.g. of individuals, organisations. This method can provide information on the agents that are central to the network, or who may be ‘influencers’ within the network. System change can be analysed using this method. [CASE STUDIES TO BE ADDED]

Qualitative comparative analysis is a case study method that involves comparing whole cases, rather than reducing them to independent variables. [MORE DETAIL AND CASE STUDIES TO BE ADDED].

**4.3.4. Choice of evaluation outcomes**

A crucial aspect of the design of an evaluation is the choice of outcome measures or evidence of change. You need to think about which outcomes are most important for decision makers and how you will deal with multiple outcomes in the analysis. A sharp distinction between a single primary outcome and a number of secondary outcomes is not necessarily appropriate, particularly where the programme theory identifies outcomes across a range of domains. A good theoretical understanding of the intervention, derived from careful development work, is key to choosing suitable outcome measures. Rather than choosing one outcome and judging the effectiveness based only on that we should look at
effects over a range of outcomes, all theoretically motivated and justified. These outcomes should be identified prospectively in the development of programme theory, and shown in a logic model, and should all be reported on in the final evaluation. A more broadly based judgement on effectiveness can then be made. Sample size decisions need to be based on one outcome measure, and one can be selected from the range of outcomes and reported as such (e.g. the one that requires the largest sample), without this necessarily having to be the primary outcome on which the effectiveness is solely based.

No matter how thorough your development work, you should remain alert to the possibility of unintended and possibly adverse consequences. Pre-identified sources of variation in outcomes are important, and appropriate subgroup analyses should be carried out and reported on. In the case of health interventions which are expected to impact on inequalities in health, analyses stratified by socio-economic position may be needed. Even where this analysis is underpowered there is value in including it e.g. for input into meta-analysis, and for developing hypotheses for testing in further research.

**Refined theory as an outcome of an evaluation:** One key outcome of an evaluation is refined/developed theory. Improved theory helps us to understand transferability across settings and to produce evidence and understanding that is more applicable to informing future decisions by policymakers. Theory is used to inform evaluations, but the purpose of evaluation should also be the gradual incremental testing and refinement of theory. Whereas interventions are likely to require modification to be successful in different contexts, the theory produced from the study can be a generalizable outcome. It is important that results from experimental trials are not over-interpreted; unbiased estimates may not provide evidence on ways forward in different contexts (Deaton and Cartwright, 2018). Decision makers, for example, if operating in a context different to the evaluation context may gain more from the theory that is produced than from the effect estimate, which is likely bound up in the evaluation context.

**Understanding processes:** A process evaluation can provide insight into why an intervention fails unexpectedly or has unanticipated consequences or why a successful intervention works and how it can be optimised. It can answer questions around fidelity and quality of implementation e.g. what is implemented, and how?; mechanisms of change, e.g. How does
the delivered intervention produce change?; and context, e.g. how does context affect implementation and outcomes? Process evaluations should be conducted to the same high methodological standards and reported just as thoroughly as evaluation of outcomes.

4.3.5. Economic Evaluations / Analyses

This section aims to (i) highlight strengths and limitations of each type of economic evaluation in a complex interventions and health sciences context and where and when they might be appropriate, (ii) signpost readers to the most recent methods guidance reports and new and re-emerging methodologies which can aid economic evaluations for complex interventions, and (iii) provide a take home message and recommendations for economic evaluations for complex interventions.

1. Types of economic evaluation frameworks which can be used for evaluating CIs

- Cost-effectiveness analysis (CEA) is the most commonly used form of economic evaluation but as outcomes are measured in natural units relevant to the health care area there are inherent issues with the ‘short term’ nature of these outcomes which are often indicators that do not measure or relate to health outcomes directly. [E.g. use child mental health and SDQ as an outcome example - improvements on a non-preference based scale and clinical meaningfulness, which has implications for the meaningfulness of this in an incremental cost-effectiveness ratio (ICER).] In the absence of a decision threshold for these interim health outcome measures, the next step would be to value this, for example, using willingness to pay for improvements in the SDQ scale.

- Cost-utility Analysis (CUA): QALY - well known advantages – summaries and ref but quite narrow terms. Limitation in context of CI and PH interventions, especially where there are cross-sectoral shifts in resource use or outcomes arising from the intervention.

- Cost-Consequence Analysis (CCA)- has been much less frequently used than CEA and CUA, probably because it does not formally synthesise costs and outcomes (so it is often not considered to be full economic evaluation), and it does not include any preference weighting of outcomes. However, CCA is the most straightforward methodologically and has the intuitive appeal of presenting outcome (consequence)
data across the range of factors decision makers are likely to be concerned with in making a policy decision. In this way it can be considered as a first step towards CBA. However, without valuation of these outcomes (e.g. preference weighting as in QALYs or monetary valuation in CBA) decision making in the absence of known thresholds requires judgement. Further, if some outcomes are better for some interventions and other outcomes are better for others, the policy implications of a CCA can be ambiguous. The broad nature of a CCA allows the analyst to populate both a public sector and societal perspectives for the economic evaluation and is therefore an extremely important tool for assessing complex interventions, as well as useful in explorative analyses at very early or feasibility study stage.

- Cost-Benefit Analysis (CBA) – involves comparison of costs and health and non-health benefits valued in monetary terms. It is more suited than CEA or CUA to the evaluation of complex interventions because it can involve the measurement of non-health benefits across different sectors (McIntosh E, 2010 book). However, it is also much more expensive and time consuming than the other types of evaluation tools and ascribing monetary values to all the dimensions of benefit in a CBA is challenging.

For complex interventions, the broad range of potential outcomes and the populations and organisations affected means that narrower forms of economic evaluation, such as CEA or CUA, are unlikely to fully reflect the full range of costs and consequences of the intervention of interest. Nor are they useful to decision makers seeking to strike a balance between different policy goals and different potential beneficiaries. For example, in Scotland, minimum unit pricing for alcohol is being evaluated in relation to its main intended goal of reducing alcohol-related harms, in particular for those drinking at harmful levels. The evaluation is also considering a range of potential additional effects. These include effects on drinkers themselves, for example, if they substitute alcohol with other harmful products; on family members; on health and social care providers; on the criminal justice system, and on the alcohol industry in Scotland. CEA and CUA in this context would not provide an adequate framework for a comprehensive assessment of the costs and benefits of the policy. CBA and CCA are the ‘broader’ forms of economic evaluation likely to be best suited to the evaluation of complex intervention, although they have limitations of their own – as detailed above.
II. Recent methods guidance and new, existing and re-emerging methodologies which can aid EE for CIs

- Recent Guidance updates: NICE Public Health guidance 2012 (highlight key amendments – e.g. lowered discount rate to 1.5%), then NICE social care guidance 2013 (highlight key changes here). The most recent NICE guidance update was 2017 where the new NICE manual is inclusive of NHS, public health and social care settings, with criteria adapted to context and perspective. Allows wider perspectives (beyond NHS) and alternative approaches such as CCA, CBA as appropriate. (Highlight key elements)

- Further reading and advances in economic evaluation methods (Drummond 2015 – key changes), (Diedda, et al 2018; Krieff et al. 2015) economic modelling (Boyd & Squires 2019) conceptual modelling frameworks and applied methods for research in economics of public health (Tudor-Edwards and McIntosh, 2019), reporting guidelines (Husereau et al., 2013).

- Newer emerging methodologies that can lend themselves to economic evaluations of complex interventions, include Social Return on Investment (SROI) and Multi-Criteria Decision Analysis (MCDA). These are not usually considered as tools for economic evaluation, but they are becoming more commonly used, particularly in the area of complex interventions. SROI (Charles, J. et al Chapter 12 in Edwards R and McIntosh E, 2019) is favoured by industry but less favoured by economists due to theoretical weaknesses and handling of outcomes. Regardless, it may be helpful to decision makers. Additionally, Multi-criteria decision analysis (MCDA) can be used where decision making involves trade-offs between multiple, often conflicting objectives. The use of MCDA involves a structured, explicit approach to decisions involving multiple criteria and can improve the quality of decision making (Marsh et al., 2016). This is particularly helpful following a CCA.

- Conceptual and decision analytic modelling. Include the role of conceptual modelling followed by decision analytic modelling for complex and public health interventions (Tappenden, 2012; Squires 2016; Squires & Boyd 2019), the role modelling plays and importance to complex interventions (Sculpher 2005; Boyd 2011). Highlight more recent advances in modelling techniques such as discrete event simulation and Agent based modelling [example of Itamar Megiddo work],
which are recommended for CI and systems approaches (Squires 2016), but more simple models can also be appropriate.

III. Economic Considerations for complex interventions - Take home message

The key message for economic considerations in complex interventions is the emphasis on use of broader economic evaluation frameworks (such as CCA and CBA) and the associated move away from powering studies on a single outcome. The recent NICE 2017 guidance for public health and social care interventions illustrates this move. CIs are about multiple outcomes of importance to different people. Therefore, CCA and CBA offer us tools to do this. Some suggestions going forward

- CCA and CBA – can look at costs in relation to educational attainments, mental health etc. We can power a trial on any of these outcomes, but these outcomes are valuable to a host of others. [EXAMPLES TO BE ADDED]. We recommend CCA as a first ‘identification’ step to CBA – and encourage that these be undertaken. CASE STUDY TO BE ADDED: examples of recent and ongoing CI studies attempting CCA and or CBA.

- If undertaking CCA you can show the intervention is cost-effective on (one of) the primary outcomes (via CEA), and illustrate the other outcomes on which it performs well and those for which it performs less well.

- The standard application is to report an ICER from a CEA or CUA but this may not be feasible or becomes irrelevant where a standard threshold does not apply. We suggest the reader considers a different approach to reporting and presenting results - not just in economic results but in the overall evaluation - in a more pragmatic way, e.g the Incredible years study (Tudor-Edwards et al) shows a cut-off point of significance at which the intervention/policy would need to reach to be cost-effective. This is a very useful way of reporting in CIs where ICER thresholds become limited or irrelevant.

- When using CCA, decision makers may then want to consider MCDA. Or, in absence of a threshold the next stage is to value an outcome – e.g. willingness to pay. It is possible to use DCE and then the CCA becomes a CBA [EXAMPLE]. This is ideal for multisector interventions.
4.3.6. Approaching complexity in systematic reviews

[TO BE ADDED]

4.3.7. Reporting guidelines

Where possible evaluations should be reported according to established guidelines as this will help to ensure that the key information is available for replication studies, systematic reviews and guideline development. Again, rather than repeat the previous version of this guidance we focus on reporting guidelines that have been published since 2006. The CONSORT statement is a well-established standard for reporting randomised trials, relevant extensions are published for: social and psychological interventions (CONSORT-SPI 2018) (Montgomery et al., 2018); randomised pilot and feasibility trials (Eldridge et al., 2016); cluster randomised controlled trials (Campbell et al., 2012); pragmatic trials (Zwarenstein et al., 2008); N of 1 trials (Vohra et al., 2015); and cluster stepped wedge trials (Hemming et al., 2018). Checklists have also been produced for intervention description and replication (Hoffmann et al., 2014) and have been extended to population health and policy interventions (Campbell et al., 2018); and for reporting behaviour change techniques (Michie et al., 2015).

The overarching considerations set out at the beginning of this guidance document have been touched upon throughout the discussion on evaluation, and are summarised in Table 6.
Table 6: Overarching considerations for the evaluation phase

<table>
<thead>
<tr>
<th>Overarching considerations</th>
<th>Issues to consider</th>
<th>Risk of not considering them</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Uncertainty</strong></td>
<td>The evaluation should respond to current uncertainties of evidence surrounding the intervention, and this should be generated by engaging with stakeholders and informed by theory.</td>
<td>There is no point in answering research questions that do not address areas of uncertainty, therefore the utility of the evaluation will be limited. The research questions may not address the most important issues.</td>
</tr>
<tr>
<td><strong>Stakeholders</strong></td>
<td>Gather appropriate input from a range of relevant stakeholders to ensure that evaluation methods and outcomes meet their requirements. Example to be added.</td>
<td>Not collecting the correct data to inform changes in policy and practice negatively impacts the utility of the evaluation and future implementation of the intervention.</td>
</tr>
<tr>
<td><strong>Programme theory</strong></td>
<td>Has the programme theory changed? Example to be added.</td>
<td>Programme theory is an important development of an evaluation and changes should be reported, otherwise potential learning will be missed.</td>
</tr>
<tr>
<td><strong>Context</strong></td>
<td>Consider the wider contextual factors that may influence the evaluation, either in relation to influences on either the ‘intervention’ or the ‘study processes’. Example to be added.</td>
<td>Questions related to external validity will not be answered.</td>
</tr>
<tr>
<td><strong>Economic considerations</strong></td>
<td>Importance of perspective Sensitivity analysis will be crucial Identify preferred economic</td>
<td>Too narrow a perspective will not reveal the full economic impact Cost-effectiveness of alternative scenarios will not be identified</td>
</tr>
<tr>
<td><strong>evaluation framework (other frameworks can be embedded)</strong></td>
<td>Incorrect frameworks may not identify all relevant costs and impacts</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>-------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Consider short and long term time horizon for cost-effectiveness</td>
<td>Ignoring long term costs and outcomes may result in an incorrect cost-effectiveness conclusion</td>
<td></td>
</tr>
</tbody>
</table>

**Intervention modification**

| **Decide beforehand what type of modification is appropriate to allow for a robust evaluation.** | **Not allowing for appropriate changes to the intervention or study processes may lead to outcomes that are not useful or meaningful. Validity of the evaluation may be compromised by failing to consider what kind of intervention modification is acceptable.** |
| Example to be added. |  |

**Key points**

- Consider assessing ‘usefulness’ of data versus only effectiveness outcomes.
- Choose from a range of experimental or non-experimental designs focused on answering the defined research questions.
- Explore real world outcomes that matter to recipients of an intervention, and to other stakeholders, such as practitioners and policy- or other decision-makers.
- Include methods of evaluation to explore the change process (i.e., mechanisms, process, context, theory).
- Conduct economic evaluation using a broad perspective with an understanding that the appropriate time horizon will impact the cost-effectiveness result.
- Involve stakeholders in the choice of evaluation design and outcomes to ensure useful data are produced for policy/practice/industry.
- The main outcome of evaluation is the collation of useful data to inform future iteration and implementation.
- Clearly report the evaluation, including updated programme theory, following recognised reporting standards if available.
4.4. IMPLEMENTATION

The intended outcome of implementation is the sustainable delivery of an effective intervention in different contexts. Successful implementation of interventions, even after they have been shown to be effective, remains low (Grimshaw et al., 2004). Reasons include limited engagement with stakeholders, poor system fit, professionals who are resistant to change, time constraints and participant preferences (Presseau et al., 2009, Petticrew et al., 2004, May, 2006, Pick, 2008).

Complexity-informed research means that decision makers do not have to guess the transferability of evidence from different contexts, rather they can refer to the evidence, often in the form of theory or models, that has taken the real-world uncertainties into account.

4.4.1. Designing research with implementation in mind

Although implementation is often considered ‘phase 4’, this is misleading - it should be addressed in the earlier development and feasibility stages. Earlier consideration of implementation increases the potential of its success.

To optimise the chances of an intervention being successfully implemented, the research itself should address the uncertainties around the uptake of an intervention. Often research focusses on demonstrating effectiveness and cost-effectiveness, where these are unknown. However, it may instead be the case that there is already reasonable evidence of effectiveness, and further evidence on this point would not help implementation. Instead, feasibility issues such as acceptability to the public or to the healthcare providers that would deliver the intervention are more responsible for a lack of implementation (Diepeveen et al., 2013).
It is useful to identify the precise uncertainties or barriers to implementation of an intervention through the engagement of stakeholders. This can ensure that research addresses these uncertainties, and produces interventions that are likely to be implemented. The rise of “very brief interventions” that fit with routine practice can be effective but also be implemented at scale (Aveyard et al., 2016), and hence have more impact than interventions that would be more effective in principle but would not be taken up in practice. It also can prevent the development and evaluation of interventions that demonstrate effectiveness, but are not “implementable” for a variety of reasons, such as lack of fit to healthcare practitioners’ usual practice (Murray et al., 2010b).

4.4.2. Implementation as an intervention

A range of approaches to implementation have been characterised (Greenhalgh et al., 2004). The most passive has been described as “letting it happen” or diffusion of information, through e.g. publishing evidence of study effectiveness. A more active has been described as “helping it happen” or dissemination regarding the intervention through e.g. clinical guidelines and training. The most active concerns active attempts at implementation, with people actively promoting the use of a particular programme, with support for the effective use of effective implementation. The first two approaches are more common, but are not sufficient for producing implementation (Greenhalgh et al., 2004).

As a minimum, users should develop an implementation plan, alongside relevant stakeholders relevance. Such a plan should take into account: appropriate dissemination of the findings; the specific context of intervention delivery; resource requirements e.g., funding, staff time, staff training, equipment; barriers and solutions to adoption; potential unintended outcomes and reduction of harm, such as research inequity (Macintyre, 2007). In addition to publication, sufficient details of the intervention and its underlying theory should be reported (Pinnock et al., 2015). This improves synthesis of evidence that may be more useful to decision-makers (e.g., theoretical outputs compared with effect sizes). Evidence also needs to be disseminated actively to decision-makers via methods that are accessible and convincing (Kelly et al., 2004). This includes making recommendations as specific as possible (Michie and Johnston, 2004).
Given that an active approach is typically required for uptake of effective interventions, the field of implementation science proposes that it is helpful to empirically consider the issue of how to implement such interventions as a field in its own right, rather than an “add-on” to effectiveness studies. This field has utilised a number of theories, models and frameworks to help guide different aspects of implementation (Nilsen, 2015). Some of these theories (process models) aim to facilitate implementation by describing or guiding the process of translating research into practice. For instance, the Quality Implementation Framework provides 14 specific actions that can be taken to facilitate implementation (Meyers et al., 2012). Others aim to understand and explain the process of implementation. Normalisation Process Theory (NPT) is a useful tool for promoting effective implementation (Murray et al., 2010b). This approach considers the extent to which an intervention “fits” within routine practice as a criterion for it being taken up. The theory proposes that, e.g. where those affected by a new intervention see the point of it, believe it will work, and it will not disrupt currently working practices in a major way then they will invest effort into making it work. This theory emphasises the importance of the context, particularly healthcare context, in which a intervention may be embedded, for influencing the extent to which new interventions are taken up (May et al., 2016). A third group are concerned with evaluating the extent to which to which successful implementation has occurred, e.g. the RE-AIM framework (Glasgow et al., 1999b).

### 4.4.3. Understanding implementation

It is often useful to understand how interventions have been implemented. Some slightly different issues arise, depending on whether the intervention was delivered by the research team, and when the interventions were not designed by researchers, e.g. was driven by policy.

A key issue in understanding the impact of interventions is the extent to which it has been delivered or received with fidelity (Bellg et al., 2004). Interventions can be delivered with fidelity to form, such as intervention manuals, or hypothesised mechanism of action/function (Hawe et al., 2004). It is generally agreed that where an intervention is not delivered with high fidelity to mechanism of action, this is problematic for interpreting the effects of that intervention. For example, where it is not clear what is being delivered by people who
are supposed to be delivering an intervention with a specified mechanisms of action, then ineffective interventions may be genuinely ineffective, ineffective in that context, or potentially effective, but there may be a failure of training. Where adaptations have been made to an intervention in a particular context, it is important to understand exactly what has been adapted (Stirman et al., 2013).

The above point relates to the transferability of an intervention into different contexts – an important aspect of long-term implementation. Developing and refining programme theory and considering wide-ranging contextual factors can help identify issues related to transferability (Villeval et al., 2016). Stakeholder involvement is then a helpful process to identify potential solutions to overcome identified issues.

Few experimental trials are powered to detect rare adverse events. Effects are likely to be smaller and more variable once the intervention becomes implemented more widely, and unanticipated consequences may begin to emerge (Jabeen, 2017, Bonell et al., 2015). Long-term follow-up may be needed to determine whether short-term changes persist, and whether benefits demonstrated from the original study do in fact occur. Although long-term follow-up of complex interventions is uncommon, such studies can be highly informative. It is worth thinking about how to measure rare or long-term impacts, for example through routine data sources and record linkage, or by re-contacting study participants. Plans for the collection of appropriate outcome data, and obtaining appropriate consents, should be built into the study design at the outset.

Other key considerations in examining the implementation of an intervention relate to issues other than efficacy/effectiveness (Glasgow et al., 1999b). Interventions which are effective, but which do not reach many people are likely to have a smaller health impact than interventions that are less effective but achieve wide population coverage (Rose et al., 2018).

### 4.4.4. Economic considerations

Issues with implementation can also be addressed through economic considerations. This has not traditionally formed part of guidance on economic evaluation, but for a fully informed economic evaluation of a complex intervention, economic considerations for implementation can and should be considered: (i) at the early stages of stakeholder
engagement intervention and study development; and (ii) in reporting of the economic results.

(I) Logistical issues, structural and political issues can be brought to light through stakeholder engagement and using a systems approach at the outset and early consideration of economic aspects alongside these stages will help with the appropriate design of the economic components. If so, the economic analyses can adopt the appropriate perspectives and type of evaluation, so as to address the concerns and answer the questions relevant to the decision makers and those who will be implementing the intervention.

(ii) How the results of economic analyses are presented to decision makers can affect whether and how they act on the results. It can also help to overcome implementation issues. In reporting the economic results of a complex intervention, economists should ask themselves ‘How can we present outcomes to decision makers to facilitate implementation?’ A key issues is avoiding and anticipating misconceptions in reporting of economic analyses results. Interventions described as ‘cost savings’ may well be potentially cost saving, but these savings may be difficult to realise during implementation. The ‘cost saving’ may not refer to a ‘financial’ saving but rather a release of resources for other potential uses [EXAMPLE], such as a freeing up of beds, or transfer of resources from the the NHS into another sector such as social services (EXAMPLE). Another important issue in reporting economic results of complex interventions is to deal explicitly with multi-sectoral issues.

Cost of implementation studies can also be undertaken post-evaluation, to help aid the implementation process. This could be particularly important where an intervention is proven to be effective and cost-effective, yet implementation requires investment from central government before local authorities can adopt it [EXAMPLE].
Table 7: Overarching considerations of the implementation phase

<table>
<thead>
<tr>
<th>Overarching considerations</th>
<th>Issues to consider</th>
<th>Risk of not considering them</th>
</tr>
</thead>
</table>
| Stakeholders               | • Have you involved stakeholders in formulation of the problem and proposed solution?  
• Wider societal and contextual barriers are also present, in particular engaging policymakers in the research process, where research often does not proceed within a timeline suited to that of the policy agenda. | • Failure to collect the appropriate data needed to influence decision-makers for policy and practice.                                                                                                                                 |
| Programme theory           | • Determine the potential mechanisms and outcomes that should be measured.  
• Also important for interventions where the researcher wasn’t involved in the design to understand how the intervention is thought to work.  
• Continue to update programme theory throughout any implementation phases.                                                                                                           | • Failure to measure the most appropriate mechanisms, influences and outcomes that affect present and/or future implementation.                                                                                                  |
| Context                    | • Consider the wide ranging contextual factors that could influence the intervention in this and other settings.  
• Where is the boundary of the system within which the intervention is being implemented?                                                                                                                                           | • Failure to understand how the intervention interacts and/or interrupts the system in which it is located.                                                                                                                   |
| Economic consideration     | • Highlight and work towards multi-sectoral issues from outset. Establish who are payers and who receives benefits. Co-payment options                                                                                                           | • Misconceptions of reported economic results  
• Decision makers misunderstanding                                                                                                                                                                                                                                                   |
<table>
<thead>
<tr>
<th><strong>Updated Guidance: Developing and Evaluating Complex Interventions [DRAFT FOR CONSULTATION]</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost of implementation analyses – overcome logistical barriers</strong></td>
</tr>
<tr>
<td><strong>Cost-effective interventions not implemented in practice</strong></td>
</tr>
<tr>
<td><strong>Modification</strong></td>
</tr>
<tr>
<td><strong>Uncertainty</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Key points

- Consider implementation during the early development and feasibility phases.
- Involve stakeholders from the early stage of development.
- Consider a wide range of methods to capture useful data.
- Incorporate theories, models or frameworks of implementation.
- Actively present evidence that is useful, accessible and convincing for decision-makers.
- Learn from where others have identified a ‘failure of implementation’.
- Incorporate economic considerations alongside implementation issues from outset. These should be considered in the conceptual framework developed at the design stage.

Signposts to further reading and guidance.

- Diffusion of innovations in service organizations: systematic review and recommendations (Greenhalgh et al., 2004).
- Unintended outcomes evaluation approach: A plausible way to evaluate unintended outcomes of social development programmes (Jabeen, 2017).
- The quality implementation framework: a synthesis of critical steps in the implementation process (Meyers et al., 2012).
- Normalisation process theory: a framework for developing, evaluating and implementing complex interventions (Murray et al., 2010a).
- Making sense of implementation theories, models and frameworks (Nilsen, 2015).
- Development of a framework and coding system for modifications and adaptations of evidence-based interventions (Stirman et al., 2013).
- [To include refs on Economics around implementation (e.g. vaccination work)]
5. Case studies

[For draft only: We have collated numerous case studies to demonstrate various elements of the updated guidance. We will include summaries here. Please add to your response if you have suggestions for relevant case studies in the following areas]:

- Modifying interventions to improve their intervention design and/or evaluation design
- The use of programme theory throughout the research process
- Involvement of stakeholders throughout the research process
- Economic considerations throughout the research process
- The exploration and use of context throughout the research process
- Addressing uncertainties throughout the research process
- Complexity informed approaches to intervention research
6. Appendix 1. Glossary of terms

**Agent:** “An agent is any entity whose behaviour is autonomous yet complying with implicit or explicit behavioural rules” (Williams and Hummelbrunner, 2010).

**Complex interventions:** “Conventionally defined as interventions with several interacting components” (definition from 2006 guidance). However, thinking has moved on since the previous guidance was published; the complexity of complex interventions relates not only to components of the intervention itself, but how these components interact with aspects of the context in which they are located, as well as system-level properties such as: emergent effects, feedback, adaptation, and self-organisation (further detail on these properties in Table 1 p12). These additional aspects of complexity are typically outside the researchers’ control and emerge from the interaction of the parts of a system in ways that cannot be predicted from the properties of the individual. These properties of systems are argued to be critical in determining the impact of interventions and highlight the importance of (i) combining multiple contributory interventions and (ii) intervening at macro and meso levels.

**Complicated interventions:** There are always elements of complexity, however, whether an intervention is treated as such is a matter of perspective, and this is reflected in the research questions. To treat an intervention as simple/complicated is to ignore dimensions of complexity, and in some cases this may be appropriate i.e. where the context of the wider system is not important to the research questions that are being asked. Examples of simple or complicated interventions could be those where the purpose of the research is to answer simple efficacy questions, to take a proof of concept approach. This is usually done within a restricted context before consideration of contextual properties. The difference between simple and complicated is about the intervention itself e.g. number of interactions between components; number of groups targeted; number and variability of outcomes; number and difficulty of behaviours required by those delivering or receiving the intervention; degree of flexibility or tailoring of the intervention permitted.

**Conceptual model:** The abstraction and representation of complex phenomena of interest in some readily expressible form, such that individual stakeholders’ understanding of the parts of the actual system, and the mathematical/graphical/conceptual representation of that
system, may be shared, questioned, tested and ultimately agreed (Tappenden, OP. 2015). It is the first part of a modelling project and is used to guide the all other stages.

**Context:** “any feature of the circumstances in which an intervention is conceived, developed, implemented and evaluated” e.g. social, political, economic, geographical context (Craig et al., 2018).

**Complexity-informed research:** research that attends to complexity rather than focuses on controlling for it. Being aware of system properties and how they may play a role in the how the intervention affects change.

**Cost / Opportunity Cost:** The economic definition of cost (also known as opportunity cost) is the value of opportunity forgone, strictly the best opportunity forgone, as a result of engaging resources in an activity. Note that there can be a cost without the exchange of money. Also the economists’ notion of cost extends beyond the cost falling on the health service alone (e.g., includes costs falling on other services and on patients themselves).

**Cost benefit analysis (CBA):** A form of economic evaluation which expresses all gains and costs in monetary terms, allowing a judgement to be made of whether, or to what extent, an objective should be pursued.

**Cost consequences analysis (CCA):** A form of economic evaluation where the whole array of outputs are presented alongside the costs, without any attempt to aggregate the outputs.

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

**Cost minimisation analysis (CMA):** A form of economic evaluation where the consequences of competing interventions are identical, so comparison can be made on the basis of resource costs alone. The aim is to determine the lowest-cost way of achieving the same outcome.

**Cost of illness study:** Aims to identify and measure the total costs attributable to a particular disease. These are not a type of economic evaluation as they are not used to assess the costs and benefits of alternative interventions or programmes. They may provide
useful information that can be used in the context of an economic evaluation of interventions related to the disease category, although care must be taken as not all costs included in a cost of illness study represent resource costs, and not all costs will be avoided in an intervention is effective. Cost of illness studies may also be utilised in the estimation of the economic burden of disease.

Cost utility analysis (CUA): A form of cost-effectiveness analysis where benefits are measured in terms of a utility measure such as the quality-adjusted life year (QALY).

Decision analysis: The application of quantitative techniques to compare alternative strategies, taking account of preferences, costs, and consequences, with the aim of providing decision makers with the best available evidence on which to make a decision.

Economic evaluation: To determine whether an intervention is an efficient use of resources; “the comparative analysis of alternative courses of action in terms of both their costs and consequences” (Donaldson, 1990).

Effectiveness: The degree to which something is successful in producing a desired result in practice.

Efficacy: The ability to produce a desired or intended result (dictionary definition). Efficacy is about the ability to produce an intended outcome in experimental settings (explanatory), which is distinct to effectiveness, which is about the ability to produce the intended outcome in real world settings (pragmatic) (Flay et al., 2005).

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

Equality: Equal shares of some good or service.

Equity: Fair distribution of resources or benefits among different individuals or groups.

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

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

**Incremental cost-effectiveness ratio (ICER):** Obtained by dividing the difference between the costs of the two interventions by the difference in the outcomes (i.e., the extra cost per extra unit of effect).

**Implementation:** The act of carrying an intention into effect, in the current context the process through which interventions are delivered, and what is delivered in practice (Moore, Audrey et al. 2014).

**Intervention:** Changes to or interruptions in a system; this could be something developed and implemented by the researcher e.g. an exercise programme, or something beyond the researchers’ control e.g. a change in welfare policy, such as the implementation of Universal Credit. “An action or programme that aims to bring about identifiable outcomes” (Rychetnik et al., 2004).

**Logic model:** A visual representation of the programme theory, typically presenting mechanisms by which an intervention, and the context in which it operates, is thought to influence outcomes of interest (Moore et al., 2014).

**Markov model:** A particular type of decision analysis that allows for the transfer between different health states over a period of time.

**Natural experiment:** “Events, interventions or policies which are not under the control of researchers, but which are amenable to research which uses the variation in exposure that they generate to analyse their impact” (Craig et al., 2012a). A natural experimental study, is the methodological approach to evaluating the impact of a natural experiment on health or other outcomes.

**Process evaluation:** “A study which aims to understand the functioning of an intervention, by examining implementation, mechanisms of impact, and contextual factors. Process evaluation is complementary to, but not a substitute for, high quality outcomes evaluation” (Moore et al., 2014).
**Programme theory:** Describes how an intervention is expected to lead to its effects and under what conditions. This can illustrate what mechanisms are expected to generate the outcomes and what features of the context are expected to affect whether or not those mechanisms operate. A causal model, a theory of change, is used to guide the evaluation of the intervention. Theory of change and programme theory are used interchangeably (Rogers, 2008).

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

**Realist evaluation:** Theory-driven evaluation whereby researchers develop mid-range theories to take account of how context and mechanisms of interventions interact to produce outcomes i.e. causal mechanisms of an intervention are context specific. Complexity theory and realist philosophy are not necessarily incompatible, rather can be complementary, and researchers can draw on concepts from each (Westhorp, 2012).

**Resources:** Things that contribute to the production of output. Money gives a command over resources but is not a resource per se.

**Scarcity:** There will never be enough resources to satisfy human wants completely.

**Sensitivity analysis:** A process through which the robustness of an economic model is assessed by examining the changes in results of the analysis when key variables are varied over a specified range.

**Stakeholders:** those who are targeted by the intervention or policy, involved in its development or delivery, or more broadly those whose personal or professional interests are affected i.e. who have a stake in the topic.

**System:** A set of things that are interconnected in a way that they produce their own pattern of behaviour over time (Meadows, 2008). The properties of a system cannot be fully
explained by an understanding of each of the system’s individual parts (Gallagher and Appenzeller, 1999).

**Systems approach:** A systems approach takes a holistic way of thinking about complex systems, focusing on the interactions between the entities and between entities and their environment, rather than assuming that a system can be understood by breaking it down into its individual entities and studying each part separately (Miller and Page, 2007).

**System boundary:** It is not always feasible, necessary or desirable, to consider the whole system in evaluative research, so boundaries may need to be set to distinguish what does and does not belong to the set of interest in order to make the research tractable. Nonetheless, it is important to be aware of the wider system and to justify the choice of particular boundaries relevant to the research question. Boundaries are dependent on the problem and research question, and can be set in different ways e.g. based on geography or concept.

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

**Systems science study designs** have been developed for studying complex systems and include quantitative systems modelling methods such as agent based modelling, social network analysis, Markov modelling, discrete event analysis, system dynamic modelling; and qualitative action-based methodologies such as critical systems heuristics.

**Theory:** Beliefs or assumptions underlying action; explanations of the phenomena of interest. There is a distinction between grand (high-level, generalisations that can be applied across domains); mid-range (limited to being applied within a specific area); and programme.

**Theory of the problem:** Clarification of the problem with stakeholders, using the existing research evidence. Once the problem is defined it should be established how the problem is socially and spatially distributed, including who is most/least likely to benefit from an intervention. Also, to understand the immediate and underlying influences that give rise to the specified problem (Wight et al., 2016). This is relevant not only in intervention
development, but in evaluation of natural experiments as well, where it is important to understand the problem that the policy change/intervention was attempting to alter.

**Time preference**: Individuals are not indifferent to the timing of costs and benefits, preferring benefits sooner and costs later.

**Utility**: A measure of the ‘satisfaction’ (benefit) obtained from consuming goods and services.

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

**Willingness to pay**: This technique asks people to state explicitly the maximum amount they would be willing to pay to receive a particular benefit. It is based on the premise that the maximum amount of money an individual is willing to pay for a commodity is an indicator of the value to them of that commodity.
7. Appendix 2. Checklist for developing and evaluating complex interventions

[For draft only: We are developing a checklist that outlines the potential steps required throughout the research process. Example elements of the checklist are provided below. Please highlight alternatives or additions]

<table>
<thead>
<tr>
<th>Development stage</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>• [Examples only. Checklist to be finalised]</td>
<td></td>
</tr>
<tr>
<td>• Have you involved stakeholders and made plans to keep them engaged throughout the process?</td>
<td></td>
</tr>
<tr>
<td>• Have you defined the problem, from multiple perspectives?</td>
<td></td>
</tr>
<tr>
<td>• Have you defined the research questions?</td>
<td></td>
</tr>
<tr>
<td>• Have you explored current scientific evidence?</td>
<td></td>
</tr>
<tr>
<td>• Have you considered whether the intervention could be implemented in a real-world setting?</td>
<td></td>
</tr>
<tr>
<td>• Are you using a recognised methodology of intervention development?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feasibility stage</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>• [Examples only. Checklist to be finalised]</td>
<td></td>
</tr>
<tr>
<td>• Have you considered how you will measure if the intervention and evaluation design is acceptable and feasible for participants and providers?</td>
<td></td>
</tr>
<tr>
<td>• Have you involved stakeholders in this process?</td>
<td></td>
</tr>
<tr>
<td>• Have you identified elements of the intervention to be refined prior to further testing?</td>
<td></td>
</tr>
<tr>
<td>• Do you have clear progression criteria?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluation stage</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>• [Examples only. Checklist to be finalised]</td>
<td></td>
</tr>
<tr>
<td>• Have you considered the focus of your evaluation e.g. ‘What</td>
<td></td>
</tr>
</tbody>
</table>
happened?' versus ‘Did it work?’

- Who wants to know the answer to your research question? I.e. which policy or practice stakeholder?
- Have you considered outcomes which assess the ‘usefulness’ compared with only ‘effectiveness’ outcomes?
- Are you measuring all of the important outcomes as defined by the programme theory?
- Have you considered elements of evaluation which help refine the programme theory?
- Will the design provide evidence about how the intervention works?

**Implementation stage**

- **[Examples only. Checklist to be finalised]**
- Have you considered issues related to implementation at the earlier development and feasibility phases of research?
- Have you involved a range of stakeholders in this process?
- Do you have an explicit dissemination strategy which identifies the key audiences and considered the ways in which this study needs to be presented?
- Have you thought about who is going to use the evidence?
- Are you presenting the findings in ways that are accessible and usable?
8. References

[N.B. References to be finalised]


EVANS, R. E., CRAIG, P., HODDINOTT, P. & ET AL 2019. When and how do ‘effective’ interventions need to be adapted and/or re-evaluated in new contexts? The need for guidance. *J Epidemiol Community Health* online.


GREENHALGH, T. & PAPOUTSI, C. 2018 Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BM C Med*, 16, 95.


