



36.1 Definition of Complex Intervention

There is an increasing interest in evaluating complex interventions. This is because epidemiological changes increasingly call for composite interventions to address patients' needs and preferences. It is also because such interventions increasingly require explicit reimbursement decisions. That was not the case in the past, when these interventions often entered the benefit package automatically, once they were considered standard medical practice. Nowadays, payers as well as care providers are intrigued to know not just if a healthcare intervention works but also when, for whom, how, and under which circumstances. In addition, there is broad recognition in the research community that evaluating complex interventions is a challenging task that requires adequate methods and scientific approaches. One of the main points of discussion across all interested parties is what exactly a complex intervention is.

One of the first attempts to define complex interventions was undertaken by the Medical Research Council (MRC) in UK, which issued a guidance in 2000 for developing and evaluating complex interventions (Campbell et al. 2000). The guidance was updated and extended in 2008 to overcome limitations in the earlier guidance (Craig et al. 2008). The guidance was published in response to the challenges faced by those who develop complex interventions and evaluate their impact. MRC defines an intervention as being complex, if it includes one or more of the following characteristics:

A. Tsiachristas (✉)

Nuffield Department of Population Health, Health Economics Research Centre,
University of Oxford, Oxford, UK

e-mail: apostolos.tsiachristas@ndph.ox.ac.uk

M. P. M. H. Rutten-van Mólken

Institute of Medical Technology Assessment, Institute of Health Care Policy
and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

e-mail: m.rutten@eshpm.eur.nl

(a) various interacting components, (b) targeting groups or organizations rather than or in addition to individuals, (c) a variety of intended outcomes, (d) they are amendable to tailoring through adaptation and learning by feedback loops, and (e) effectiveness is impacted by behaviour of those delivering and receiving the intervention. In other words, the MRC argues that the greater the difficulty in defining precisely what exactly are the effective ingredients of an intervention and how they relate to each other, the greater the likelihood that a researcher is dealing with a complex intervention. Examples of complex interventions are presented in Box 36.1.

Box 36.1 Examples of complex interventions

Tele-health, e-health, and m-health interventions

- Online portal for diabetes patients to support self-management
- Home tele-monitoring.

Mobile phone-based system to facilitate management of heart failure

Interventions directed at individual patients:

- Cognitive behavioural therapy for depression
- Cardiac or pulmonary rehabilitation programmes
- Care pathways.

Motivational interviewing and lifestyle support to improve physical activity and a healthy diet.

Group interventions:

- Group psychotherapies or behavioural change strategies.

School-based interventions to reduce smoking and teenage pregnancy

Interventions directed at health professional behaviour:

- Implementation strategies to improve guideline adherence
- Computerized decision support systems.

Service delivery and organization:

- Stroke units
- Hospital at home

Community and primary care interventions:

- Community-based programmes to prevent heart disease
- Multi-disciplinary GP-based team to optimize health and social care

for frail elderly.

Population and public health interventions

- Strategies to increase uptake of cancer screening.
- Public health programmes to reduce addiction to smoking, alcohol, and drugs
- Integrated care programmes for chronic diseases.

- Could include all interventions above.

In the same line, other definitions also emphasize the degree of flexibility and non-standardization of complex interventions, which may have different forms in different contexts, while still conforming to specific theory-driven processes (Hawe

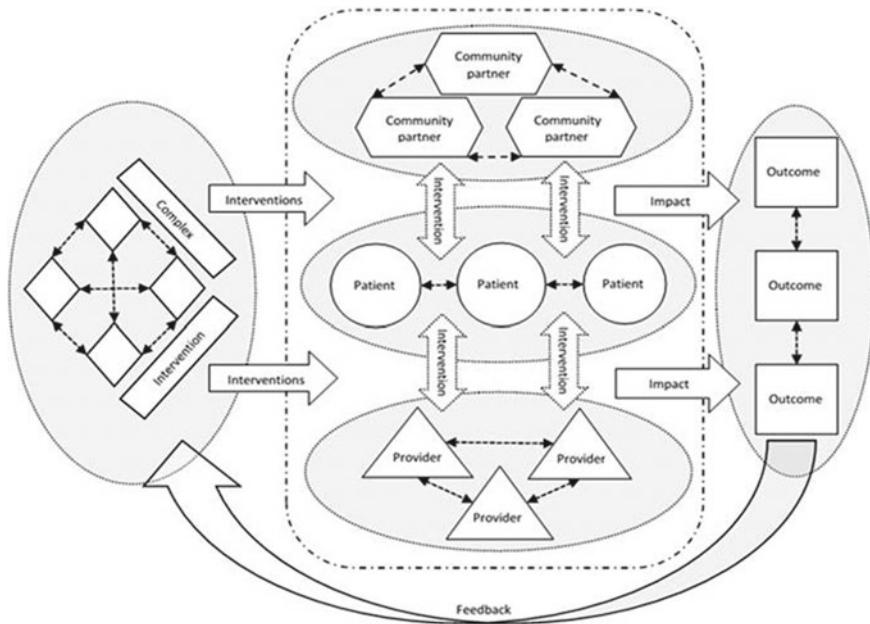


Fig. 36.1 Illustration of complex intervention

et al. 2004). Although there are many more definitions of complex interventions, they all tend to emphasize multiple interacting components and nonlinear causal pathways. Figure 36.1 illustrates how a complex intervention is diffused to different groups of recipients, interacts, and impacts different outcomes.

In contrast, health technologies such as medicines, diagnostic tests, medical devices, and surgical procedures are considered to be simple interventions because they are usually delivered by one care provider or provider organization and have mostly linear causal pathways linking the intervention with its outcome. However, the distinction between complex and simple interventions may be not entirely clear because after all simple interventions can also have a degree of complexity. Complexity is defined as ‘a scientific theory which asserts that some systems display behavioural phenomena that are completely inexplicable by any conventional analysis of the systems’ constituent parts (Hawe et al. 2004). Reducing a complex system to its components amounts to irretrievable loss of what makes it a system.

It has also been suggested that complexity is not necessarily a feature of an intervention but it is the complexity of the setting in which interventions are implemented. In other words, complexity is a property of the setting in which an intervention is being implemented not an inherent feature of the intervention itself (Shiell et al. 2008). For example, a vaccination programme for tuberculosis in a low-income country may be seen as a simple intervention implemented in a

complex setting because its implementation requires the interaction between primary care, hospitals, local community, and schools.

It has also been argued that the research question and the perspective from which that question is answered define the complexity of an intervention. Researchers often treat interventions as simple because it is convenient to answer simple research questions (Petticrew 2011). Addressing complexity requires studying synergies between components, phase changes and feedback loops, interactions between multiple health and non-health outcomes as well as processes. Alternatively, focusing on the effectiveness of the single most-important component of an intervention simplifies the research question considerably. The intervention is the same but the research questions are different, and therefore, the adopted research methods are different. Based on this argument, not every complex intervention requires complex analysis unless the research question demands it.

In any of the above arguments to define complex interventions, integrated care is a brilliant example of a complex intervention. The World Health Organization (WHO) defines it as ‘services that are managed and delivered in a way that ensures people receive a continuum of health promotion, disease prevention, diagnosis, treatment, disease management, rehabilitation and palliative care services, at the different levels and sites of care within the health system and according to their needs throughout their life course. It is an approach to care that consciously adopts the perspectives of individuals, families and communities and sees them as participants as well as beneficiaries of care’ (WHO 2015). Similar definitions of integrated care can be found elsewhere (Kodner and Spreeuwenberg 2002; Nolte and McKee 2008). Based on this definition, integrated care may be considered an ultra-complex intervention or according to Shiell et al. (2008) a complex system (Shiell et al. 2008) because it is composed of multiple complex interventions (e.g. shared decision-making and self-management support); it behaves in a nonlinear fashion (i.e. change in output is not proportional to change in input), and the interventions interact with the context in which they are implemented, and involved decision-makers are merely interested in complex research questions.

36.2 The Rationale for Evaluation

Although research and service innovation have not been always aligned, service leaders and managers are increasingly keen to assess the effects of changes in such a way that they can be causally attributed to the complex intervention. Policy-makers are also keen to ensure that they allocate scarce healthcare resources only to services that have proven value for money (i.e. to increase allocative efficiency). Some healthcare systems, such as Germany, do not allow process innovations without proof of efficiency. This is mainly driven by the notion that we cannot afford to make poor investments in times of tight budgets. Investing in any new interventions requires an increase in taxes, premiums, patients’ co-payments or takes away budget from other interventions. As a result, there is a rationale to evaluate complex

interventions already during their development and implementation. However, there are some questions to be addressed by researchers before pursuing an evaluation of a complex intervention, including (Lamont et al. 2016):

(a) why is it important to address the aims of the evaluation and what is already known about the intervention, (b) who are the main stakeholders and users of research at outset, (c) how will the evaluation be performed in terms of study design and research methods, (d) what to measure and which data to be used, and (e) when is the perfect timing to maximize the impact of the evaluation results.

Similarly, policy-makers may want to assess its evaluability to support more systematic resource allocation decisions depending on the knowledge generated by an evaluation of a complex intervention. An assessment of evaluability may include the following questions (Ogilvie et al. 2011): (a) where is a particular intervention situated in the evolutionary flowchart of an overall intervention program?, (b) how will an evaluative study of this intervention affect policy decisions?, (c) what are the plausible sizes and distribution of the intervention's hypothesized impacts?, (d) how will the findings of an evaluative study add value to the existing scientific evidence?, and (e) is it practical to evaluate the intervention in the time available?

36.3 Challenges in Evaluating Complex Interventions

Key challenges in the evaluation of complex interventions were identified in a recent review of 207 studies (Datta and Petticrew 2013). One of the main challenges was related to the content and standardization of interventions due to variation in the delivery of services in terms of frequency of interventions and lack of precise definition of the start of the treatment and a wide range of patients' diagnoses, stage of diseases, needs, and preferences. Other challenges were related to the people (healthcare providers and patients) involved in the delivery of complex interventions. On the provider side, time and resource limitations may obscure data collection for evaluation purposes. Data collection may also be challenged due to issues related to patient's preferences, patient/provider interaction, and recruitment and retention to trials.

Furthermore, the organizational context of implementation, such as hierarchies, professional boundaries, staffing arrangements, social, geographical and environmental barriers, and the impact of other simultaneous organizational changes may affect the implementation of a complex intervention. A deterrent organizational context alongside with lack of support from healthcare providers poses another major challenge in evaluating complex interventions. Considering the plural, multi-dimensional (bio-psychosocial-clinical aspects), and multi-level (patient/organizational/local level) outcomes of complex interventions and their time spanning (i.e. short, medium, and long term), researchers face difficulties in establishing 'hard' outcomes that capture all effects. Combining quantitative with qualitative methods may ease part of this challenge. However, to do that sufficiently, more resources should be committed to the evaluation. Furthermore, we

have seen an increase in the use of so-called composite endpoints (Hofman et al. 2014). Taking this step further, Datta and Petticrew suggested a departure from focusing on primary outcomes and a small number of secondary outcomes towards a much more multi-criteria form of assessment which acknowledges the multiple objectives of many complex interventions (Datta and Petticrew 2013).

Similar challenges were identified in a cross-national study that investigated barriers in the evaluation of chronic disease management programmes in Europe (Knai et al. 2013). The study found that lack of awareness for the need of evaluation and capacity to undertake sound evaluations, including experienced evaluators, deterred the development of an evaluation culture. Other reported barriers included the reluctance of payers to commit to evaluation in order to secure financial interests and the reluctance of providers to engage in evaluation due to perceived administrative burden and compromises their freedom. A more technical set of barriers to evaluate disease management programmes was related to low quality of routinely collected data, or the lack of, inaccessibility, fragmentation, and wide variety of information and communication technology (ICT). The authors argued that these barriers lie on the complexity of the intervention and current organizational, cultural, and political context.

The evaluation of a complex intervention may also be challenged at the policy-making level, where the decision to allocate substantial resources to implement and evaluate a complex intervention is often taken. Failing to convince policy-makers about the ‘evaluability’ of a complex intervention may hamper any action for evaluation.

36.4 Evaluation Frameworks

The increasing attention for complex interventions and urgent need to evaluate them boosted the development of evaluation frameworks in the last decade. One of these is May’s rational model, which focuses on the normalization of complex interventions. Normalization is defined as the embedding of a technique, technology or organizational change as a routine and taken-for-granted element of clinical practice (May 2006). In this model, four constructs of normalizing a complex intervention are distinguished. The first is interactional workability, referring to the immediate conditions in which professionals and patients encounter each other, and in which complex interventions are operationalized. The second construct is relational integration, which is the network of relations in which clinical encounters between professionals and patients are located, and through which knowledge and practice relating to a complex intervention are defined and mediated. Skill-set workability is the third construct and includes the formal and informal divisions of labour in healthcare settings and to the mechanisms by which knowledge and practice about complex interventions are distributed. Finally, the fourth construct is the contextual integration and refers to the capacity of an organization to understand and agree on the allocation of control and infrastructure resources to implementing

a complex intervention and to negotiating its integration into the existing patterns of activity. The model is argued to have face validity in assessing the potential of a complex intervention to be ‘normalized’ and evaluating the factors of its success or failure in practice.

The multiphase optimization strategy (MOST) is another framework for optimizing and evaluating complex interventions (Collins et al. 2005). It consists of the following three phases: (a) screening; in which randomized experimentation closely guided by theory is used to assess an array of programme and/or delivery components in order to select the components that merit further investigation; (b) refining; in which interactions among the identified set of components and their interrelationships with covariates are investigated in detail, again via randomized experiments. Optimal dosage levels and combinations of components are identified; and (c) confirming; in which the resulting optimized intervention is evaluated by means of a standard randomized intervention trial. To make the best use of available resources, MOST relies on design and analysis tools that help maximize efficiency, such as fractional factorial designs.

The MRC guidance is probably the most influential framework in developing and evaluating complex interventions. It is based on the following key elements (Craig et al. 2008): (a) development including the identification of evidence bases and theory as well as modelling of processes and outcome, (b) feasibility/piloting incorporating testing procedures, estimating recruitment, and determining sample size, (c) evaluation by assessing the effectiveness and cost-effectiveness as well as understanding the change processes, and (d) implementation including dissemination, surveillance and monitoring and long-term follow-up. Regarding evaluation, the MRC guidance is supportive of using experimental study designs when possible and combining process evaluation to understand process changes with formative and summative evaluation to estimate (cost-) effectiveness.

36.5 Process Evaluation

Process evaluation is as important as outcome evaluation, which can provide valuable insight not only within feasibility and pilot studies, but also within definitive evaluation studies and scale-up implementation studies. Process evaluations can examine how interventions are planned, delivered, and received by assessing fidelity and quality of implementation, clarifying causal mechanisms, and identifying contextual factors associated with variation in outcomes (Craig et al. 2008). It is particularly important in multi-site studies, where the ‘same’ intervention may be implemented and received in different ways (Datta and Petticrew 2013). The recognition that the MRC guidance elaborated poorly on guiding process evaluation (Moore et al. 2014) resulted into a separate MRC guidance on the process evaluation of complex interventions (Moore et al. 2015). This guidance provides key recommendations for planning, designing and conducting, analysing,

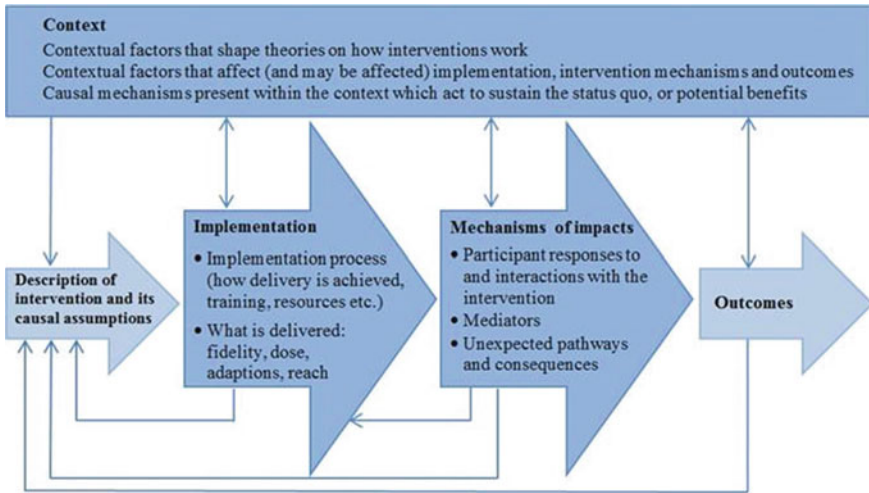


Fig. 36.2 Elements and relations of process evaluation. *Source* Moore et al. (2015)

and reporting process evaluations. Figure 36.2 shows the functions of process evaluation and relations among them as identified in the MRC guidance.

Following the MRC guidance and the earlier work of Steckler et al. (2002), the following subsections provide more details on the implementation, context, and causal mechanisms of complex interventions as the main components of their process evaluation. This is in accordance with an early case study of treating integrated care as complex intervention, Bradley et al. (1999) who suggested three levels of defining the intervention, including theory and evidence which inform the intervention, tasks, and processes involved in applying the theoretical principles, and people with whom and context within which the intervention is operationalized (Bradley et al. 1999).

36.5.1 Fidelity and Quality of Implementation

A complex intervention may be less effective as initially thought because of weak or incomplete implementation (Boland et al. 2015). This is because they often go through adaptations depending on the context, which might undermine intervention fidelity. Standardizing all components of an intervention to be the same in different sites would treat complex interventions as being simple interventions. According to Hawe et al. (2004), the function and process of a complex intervention should be standardized not the components themselves. This allows the intervention to be tailored to local conditions and could improve effectiveness. Intervention integrity would be defined as evidence of fit with the theory or principles of the hypothesized change process. However, others may argue otherwise and propose the

standardization of the components, while allowing flexible operationalization of these components based on the context.

Hence, the first stage in process evaluation focuses on the fidelity (the extent to which the intervention is delivered as intended), reach (whether an intervention is received by all those it targeted), dose delivered (the amount or number of units of intervention offered to participants, and dose received (the extent of participants' active engagement in the scheme). Steckler and Linnan conceive of evaluating intervention reach and dose, and participants' responses to an intervention largely in quantitative terms (Steckler et al. 2002). Reach and dose are commonly examined quantitatively using methods such as questionnaire surveys exploring participants' exposure to and satisfaction with an intervention. However, receipt can also be seen in qualitative terms as exploring participants' reports of an intervention in their own terms. Qualitative research can be useful in examining how participants perceive an intervention in unexpected ways which may not be fully captured by researcher-developed quantitative constructs. Qualitative research can also explore how providers or participants exert 'agency' (willed action) in engaging with the intervention rather than merely receiving it passively.

At this stage of the process evaluation, the RE-AIM framework developed by Glasgow et al. (1999) may be used to assess the reach, efficacy, adoption, implementation, and maintenance of a complex intervention at individual and organizational level. This framework provides also specific metrics on each of these five dimensions (Glasgow et al. 2006a) and has been used in the process evaluation of many complex interventions including diabetes self-management interventions (Glasgow et al. 2006b) and community-based interventions for people with dementia (Altpeter et al. 2015).

36.5.2 Context

Context is a critical aspect of process evaluation. Although there is no consistent definition, context refers to the social, political, and/or organizational setting in which an intervention is implemented (Rychetnik et al. 2002). In broader terms, this could include factors such as the needs of participants, the infrastructure within which interventions will be delivered, the skills and attitudes of providers, and the attitudes and cultural norms of potential participants. The context in which a complex intervention is implemented usually influences the intervention's implementation by supporting or hindering it (Steckler et al. 2002). For example, an intervention may be delivered poorly in some areas, but well in others, because of better provider capacity or more receptive community norms in some areas. Context can be measured quantitatively in order to inform 'moderator' analyses, but this occurs rarely and inconsistently between studies (Bonell et al. 2012). Qualitative research allows for a different understanding of the importance of context, for example, examining how intervention providers or recipients describe the

interaction between their context and their own agency in explaining their actions (Oakley et al. 2006).

Moreover, the context interacts with complex interventions and therefore influences outcomes. The interaction of context and interventions has two major implications (Rychetnik et al. 2002). Firstly, it is likely to affect the transferability of a complex intervention. Secondly, interactions greatly complicate attempts to pool the results of different interventions. Distinguishing between components of interventions that are highly context dependent (e.g. a self-management support programme) and those that may be less so (e.g. wearable health devices that support self-management) may be a way of scaling down these implications. A process evaluation should therefore determine whether interactions between the context and intervention have been sought, understood, and explained. Where such interactions seem to be strong, it may be preferred to explore and explain their effects, rather than pooling the findings. To do this, a combination of different qualitative methods, including interviews, focus groups, observations, and tick descriptions, should be used. Qualitative research can also enrich the understanding of intervention effects and guide systematic reviews. Standards for conducting qualitative investigations are widely available (Taylor et al. 2013).

36.5.3 Causal Mechanisms

Assessing an intervention's mechanisms of effects involves assessing whether the validity of the theory of change does indeed explain its operation. Such analysis can explain why an intervention is found to be effective or ineffective within an outcome evaluation. This might be critically important in refining an intervention found to be ineffective or in understanding the potential generalizability of interventions found to be effective. Quantitative data can be used to undertake mediator analyses to assess whether intervention outputs or intermediate outcomes appear to explain intervention effects on health outcomes (Ricklees 2009). Qualitative data can be used to examine such pathways, and this is particularly useful when the pathways in question have not been comprehensively examined using quantitative data, as well as when pathways are too complex (e.g. using multiple steps or feedback loops) to be assessed adequately using quantitative analyses. However, such analyses can be challenging. First, quantitative analyses require evaluators to have correctly anticipated what data is needed to examine causal pathways and to have collected these. A second challenge involves using qualitative alongside quantitative data to understand causal pathways. If qualitative data is analysed in order to explain quantitative findings, this may introduce confirmation bias. This may occur because the qualitative analysis will be used to confirm hypothesis of the quantitative analysis and focus disproportionately less to alternative possibilities. Furthermore, quantitative and qualitative methods originate from different research paradigms. Qualitative research is inductive, and generalizations are made from particular circumstances making the external validity of the findings somewhat uncertain.

36.6 Formative and Summative Evaluation

In a formative evaluation, complex interventions are typically assessed during their development or early implementation to provide information about how best to revise and modify for improvement. For these purposes, a pilot study can be designed to test that both the intervention and the evaluation can be implemented as intended. If the pilot is successful and no changes are made, then data from it can be incorporated into the main study. Moreover, a feasibility study can be used to indicate whether or not a definitive study is feasible and to examine important areas of uncertainty such as possibility and willingness for randomization, response rates to questionnaires collecting outcome data, or the standard deviation of the primary outcome measure required for the sample size calculation.

In summative evaluation, complex interventions are assessed for their definitive effectiveness and cost-effectiveness to support decisions-making of whether an intervention should be adopted, continued, or modified for improvement. The key statistical design issues alongside formative and summative evaluations of complex interventions are related to the study design and outcomes (Lancaster et al. 2010).

36.6.1 Study Design

The MRC guidance advocates the adoption of an experimental study design when evaluating complex interventions because it is the most robust method of preventing selection bias (Craig et al. 2008). Experimental designs include randomized controlled trials where individuals are randomly allocated to an intervention or a control group. These trials are sometimes considered to be inapplicable to complex interventions, but there are many flexible variants that can overcome technical and ethical issues associated with randomization such as randomized stepped wedge designs (Brown and Lilford 2006), preference trials (Brewin and Bradley 1989) and randomized consent designs (Zelen 1979; Torgerson and Sibbald 1998), and N-of-1 designs (Guyatt et al. 1990). When there is a risk of contamination (i.e. the control group is affected by the intervention), cluster randomized trials, in which groups of individuals (e.g. patients in a GP practice) are randomized instead of single individuals, are preferred.

Realist RCTs have also been suggested as adequate design in evaluating complex intervention because they emphasize the understanding of the individual and combined effects of intervention components and examination of change mechanisms (Bonell et al. 2012). Realist RCTs should be based on ‘logic models’ that define the components and mechanisms of specific interventions and combine qualitative and quantitative research methods. However, Marchal et al. (2013) objected the ‘realist’ nature of RCTs and proposed that the term ‘realist RCT’ should be replaced by ‘theory informed RCT’, which could include the use of logic model and mediation analysis that are entirely consistent with a positivist philosophy of science. Such an approach would be based on theory-based impact

evaluations for complex interventions and would be aligned with the approach suggested in the MRC guidelines. Irrespective of the terminology, both studies agree that experimental designs should be based on theories and incorporate methods adequate to evaluate complex interventions.

If an experimental approach is not feasible, for example, because the intervention is irreversible, necessarily applies to the whole population, or because large-scale implementation is already under way, non-experimental alternatives should be considered. Quasi-experimental designs or natural experiments may be the best alternatives when evaluating complex interventions because they involve the application of experimental thinking to non-experimental situations. They widen the range of interventions beyond those that are amendable to planned experimentation, and they encourage a rigorous approach to use observational data (Craig et al. 2012). Natural experiments are applicable when control groups are identifiable or when groups are exposed to different levels of intervention. Regression adjustment and propensity-score matching could reduce observed confounding between the comparators, while difference-in-differences, instrumental variables, and regression discontinuity could reduce the unobserved confounding between the comparators. A combination of these techniques is also possible in the evaluation (Stuart et al. 2014).

The selection of the study design could be informed by primary studies, literature reviews, and qualitative studies (Lancaster et al. 2010) and decided based on size and timing of the expected effects, the likelihood of the selection bias, the feasibility and acceptability of randomization, and the underlying costs (see Box 36.2).

Box 36.2 Choosing between randomised and non-randomised designs

Size and timing of effects: randomisation may be unnecessary if the effects of the intervention are so large or immediate that confounding or underlying trends are unlikely to explain differences in outcomes before and after exposure. Randomization may be inappropriate if the changes are very small, or take a very long time to appear. In these circumstances a non-randomised design may be the only feasible option, in which case firm conclusions about the impact of the intervention may be unattainable.

Likelihood of selection bias: randomisation is needed if exposure to the intervention is likely to be associated with other factors that influence outcomes. Post-hoc adjustment is a second-best solution, because it can only deal with known and measured confounders and its efficiency is limited by errors in the measurement of the confounding variables.

Feasibility and acceptability of experimentation: randomisation may be impractical if the intervention is already in widespread use, or if key decisions about how it will be implemented have already been taken, as is often the case with policy changes and interventions whose impact on health is secondary to their main purpose.

Cost: if an experimental study is feasible, and would provide more reliable information than an observational study, you need then to consider whether the additional cost would be justified by having better information.

Source Craig et al. (2008)

36.6.2 Outcomes

Given the nature of complex interventions, an appraisal of evidence should determine whether the outcome variables cover the interests of all the important stakeholders, not just those who conduct or appraise evaluative research. Important stakeholders include those with responsibility for implementation decisions as well as those affected by the intervention. Identification of the appropriate range of outcomes that should be included in a formative/summative evaluation requires a priori agreement about the relevant outcomes of an intervention from important stakeholders' perspectives, including agreement on the types of evidence deemed to be adequate to reach a conclusion on the value of an intervention, and the questions to be asked in evaluating the intervention (Rychetnik et al. 2002).

Outcomes can be measured using qualitative and quantitative research methods. Qualitative studies can also be used as a preliminary to a quantitative study to establish, for example, meaningful wording for a questionnaire. The selection of the outcome measures can be based on recommendations and evidence from the literature as well as practical issues in collecting or gathering the necessary data. The outcomes measures should extend over different dimensions (e.g. dimensions of quality of life), time scales (e.g. short, medium, and long term), and levels (e.g. patient, organizational, and local). For this reason, there is a need for a multi-criteria form of assessment which acknowledges the multiple objectives of many complex interventions (Datta and Petticrew 2013).

Costs should be included in an evaluation to make the results far more useful for decision-makers. Ideally, economic considerations should be taken fully into account in the design of the evaluation, to ensure that the cost of the study is justified by the potential benefit of the evidence it will generate, appropriate outcomes are measured, and the study has enough power to detect economically important differences.

36.7 Reporting and Reviewing Evaluation Results

It is of crucial importance to provide detailed reporting of the results from the process, formative, and summative evaluations for several reasons. First, information on the design, development, and delivery of interventions as well as its context is required to overcome the challenges of evaluating complex interventions (Datta and Petticrew 2013) and enable the transferability of the interventions to other settings (Rychetnik et al. 2002). Second, well-reported outcomes, knowledge of factors that influence the intervention's sustainability and dissemination, and information on the characteristics of people for whom the intervention was effective or less effective support evidence-based decision-making and practice. Third, poor reporting limits the ability to replicate interventions and synthesize evidence in systematic reviews.

The availability of such information from an evaluation study is a marker of the quality of evidence on a complex intervention. High-quality evidence should refer to evaluative research that was matched to the stage of development of the intervention; was able to detect important intervention effects; provided adequate process measures and contextual information, which are required for interpreting the findings; and addressed the needs of important stakeholders (Rychetnik et al. 2002).

Several instruments have been developed and reported in the literature to systematize the reporting of evaluation studies of complex interventions. Some of them are mentioned in the MRC guidance for developing and evaluating complex interventions (Craig et al. 2008) and included generic statements (i.e. not specifically applicable to complex interventions) such as the CONSORT statement for reporting clinical trials (Moher et al. 2010) and the STROBE statement for observational studies (von Elm et al. 2007). Extended versions of the CONSORT statement for cluster randomized trials (Campbell et al. 2012), pragmatic trial, (Zwarenstein et al. 2008) and complex social and psychological interventions have been issued (Montgomery et al. 2013a). Similarly, the Criteria for reporting the Development and Evaluation of Complex Interventions in healthcare (CREDECI 2) is a checklist based on the CONSORT statement and EQUATOR network to report 17 items related to the development, feasibility and piloting, and evaluation of complex interventions (Mohler et al. 2015).

Authors of systematic reviews are increasingly being asked to integrate assessments of the complexity of interventions into their reviews. The challenges involved are well recognized (Shepperd et al. 2009). Some studies attempted to contribute in overcoming these challenges by systematically classifying and describing complex interventions for a specific medical area (Lamb et al. 2011). A more comprehensive attempt towards that direction was the Oxford Implementation Index. This tool was developed to incorporate information in systematic literature reviews and meta-analyses about the intervention characteristics with regards to their design, delivery, and uptake as well as information about the contextual factors (Montgomery et al. 2013b). Furthermore, the Cochrane collaboration has published a series of methodological articles on how to consider complexity of interventions in systematic reviews (Anderson et al. 2013).

References

- Altpeter, M., Gwyther, L. P., Kennedy, S. R., Patterson, T. R., & Derence, K. (2015). From evidence to practice: Using the RE-AIM framework to adapt the REACHII caregiver intervention to the community. *Dementia (London)*, *14*, 104–113.
- Anderson, L. M., Petticrew, M., Chandler, J., Grimshaw, J., Tugwell, P., O'Neill, J., et al. (2013). Introducing a series of methodological articles on considering complexity in systematic reviews of interventions. *Journal of Clinical Epidemiology*, *66*, 1205–1208.
- Boland, M. R., Kruis, A. L., Huygens, S. A., Tsiachristas, A., Assendelft, W. J., Gussekloo, J., Blom, C. M., et al. (2015). Exploring the variation in implementation of a COPD disease management programme and its impact on health outcomes: A post hoc analysis of the RECODE cluster randomised trial. *NPJ Primary Care Respiratory Medicine*, *25*, 15071.
- Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine*, *75*, 2299–2306.
- Bradley, F., Wiles, R., Kinmonth, A. L., Mant, D., & Gantley, M. (1999). Development and evaluation of complex interventions in health services research: Case study of the Southampton heart integrated care project (SHIP). *the SHIP Collaborative Group. BMJ*, *318*, 711–715.
- Brewin, C. R., & Bradley, C. (1989). Patient preferences and randomised clinical trials. *BMJ*, *299*, 313–315.
- Brown, C. A., & Lilford, R. J. (2006). The stepped wedge trial design: A systematic review. *BMC Medical Research Methodology*, *6*, 54.
- Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *BMJ*, *321*, 694–696.
- Campbell, M. K., Piaggio, G., Elbourne, D. R., Altman, D. G., & Group, C. (2012). Consort 2010 statement: Extension to cluster randomised trials. *BMJ*, *345*, e5661.
- Collins, L. M., Murphy, S. A., Nair, V. N., & Strecher, V. J. (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine*, *30*, 65–73.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., Petticrew, M., & Medical Research Council Guidance. (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*, *337*, a1655.
- Craig, P., Cooper, C., Gunnell, D., Haw, S., Lawson, K., Macintyre, S., et al. (2012). Using natural experiments to evaluate population health interventions: New medical research council guidance. *Journal of Epidemiology and Community Health*, *66*, 1182–1186.
- Datta, J., & Petticrew, M. (2013). Challenges to evaluating complex interventions: A content analysis of published papers. *BMC Public Health*, *13*, 568.
- Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health*, *89*, 1322–1327.
- Glasgow, R. E., Klesges, L. M., Dzawaltowski, D. A., Estabrooks, P. A., & Vogt, T. M. (2006a). Evaluating the impact of health promotion programs: Using the RE-AIM framework to form summary measures for decision making involving complex issues. *Health Education Research*, *21*, 688–694.
- Glasgow, R. E., Nelson, C. C., Strycker, L. A., & King, D. K. (2006b). Using RE-AIM metrics to evaluate diabetes self-management support interventions. *American Journal of Preventive Medicine*, *30*, 67–73.
- Guyatt, G. H., Keller, J. L., Jaeschke, R., Rosenbloom, D., Adachi, J. D., & Newhouse, M. T. (1990). The n-of-1 randomized controlled trial: Clinical usefulness. Our three-year experience. *Annals of Internal Medicine*, *112*, 293–299.
- Hawe, P., Shiell, A., & Riley, T. (2004). Complex interventions: How “out of control” can a randomised controlled trial be? *BMJ*, *328*, 1561–1563.

- Hofman, C. S., Makai, P., Boter, H., Buurman, B. M., De Craen, A. J., Olde Rikkert, M. G., et al. (2014). Establishing a composite endpoint for measuring the effectiveness of geriatric interventions based on older persons' and informal caregivers' preference weights: A vignette study. *BMC Geriatrics*, *14*, 51.
- Knai, C., Nolte, E., Brunn, M., Elissen, A., Conklin, A., Pedersen, J. P., et al. (2013). Reported barriers to evaluation in chronic care: Experiences in six European countries. *Health Policy (Amsterdam)*, *110*, 220–228.
- Kodner, D. L., & Spreeuwenberg, C. (2002). Integrated care: Meaning, logic, applications, and implications—A discussion paper. *International Journal of Integrated Care*, *2*, e12.
- Lamb, S. E., Becker, C., Gillespie, L. D., Smith, J. L., Finnegan, S., Potter, R., et al. (2011). Reporting of complex interventions in clinical trials: Development of a taxonomy to classify and describe fall-prevention interventions. *Trials*, *12*, 125.
- Lamont, T., Barber, N., Pury, J., Fulop, N., Garfield-Birkbeck, S., Lilford, R., et al. (2016). New approaches to evaluating complex health and care systems. *BMJ*, *352*, i154.
- Lancaster, G. A., Campbell, M. J., Eldridge, S., Farrin, A., Marchant, M., Muller, S., et al. (2010). Trials in primary care: Statistical issues in the design, conduct and evaluation of complex interventions. *Statistical Methods in Medical Research*, *19*, 349–377.
- Marchal, B., Westhorp, G., Wong, G., Van Belle, S., Greenhalgh, T., Kegels, G., & Pawson, R. (2013). Realist RCTs of complex interventions—An oxymoron. *Social Science & Medicine*, *94*, 124–128.
- May, C. (2006). A rational model for assessing and evaluating complex interventions in health care. *BMC Health Services Research*, *6*, 86.
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C., Devereaux, P. J., et al. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*, c869.
- Mohler, R., Kopke, S., & Meyer, G. (2015). Criteria for reporting the development and evaluation of complex interventions in healthcare: Revised guideline (CRedeCI 2). *Trials*, *16*, 204.
- Montgomery, P., Grant, S., Hopewell, S., Macdonald, G., Moher, D., Michie, S., & Mayo-Wilson, E. (2013a). Protocol for CONSORT-SPI: An extension for social and psychological interventions. *Implementation Science*, *8*, 99.
- Montgomery, P., Underhill, K., Gardner, F., Operario, D., & Mayo-Wilson, E. (2013). The oxford implementation index: A new tool for incorporating implementation data into systematic reviews and meta-analyses. *Journal of Clinical Epidemiology*, *66*, 874–882.
- Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Cooper, C., et al. (2014). Process evaluation in complex public health intervention studies: The need for guidance. *Journal of Epidemiology and Community Health*, *68*, 101–102.
- Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., et al. (2015). Process evaluation of complex interventions: Medical research council guidance. *BMJ*, *350*, h1258.
- Nolte, E., & Mckee, M. (2008). In E. Nolte & M. McKee (Eds.), *Caring for people with chronic conditions: A health system perspective* (p. 259). Maidenhead: Open University Press/McGraw-Hill, (European observatory on health systems and policies series). ISBN: 978 0 335 23370 0.
- Oakley, A., Strange, V., Bonell, C., Allen, E., Stephenson, J., & Team, R. S. (2006). Process evaluation in randomised controlled trials of complex interventions. *BMJ*, *332*, 413–416.
- Ogilvie, D., Cummins, S., Petticrew, M., White, M., Jones, A., & Wheeler, K. (2011). Assessing the evaluability of complex public health interventions: Five questions for researchers, funders, and policymakers. *The Milbank Quarterly*, *89*, 206–225.
- Petticrew, M. (2011). When are complex interventions 'complex'? When are simple interventions 'simple'? *European Journal of Public Health*, *21*, 397–398.
- Rickles, D. (2009). Causality in complex interventions. *Medicine, Health Care, and Philosophy*, *12*, 77–790.
- Rychetnik, L., Frommer, M., Hawe, P., & Shiell, A. (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health*, *56*, 119–127.

- Shepperd, S., Lewin, S., Straus, S., Clarke, M., Eccles, M. P., Fitzpatrick, R., et al. (2009). Can we systematically review studies that evaluate complex interventions? *PLoS Medicine*, *6*, e1000086.
- Shiell, A., Hawe, P., & Gold, L. (2008). Complex interventions or complex systems? Implications for health economic evaluation. *BMJ*, *336*, 1281–1283.
- Steckler, A. B., Linnan, L., & Israel, B. A. (2002). *Process evaluation for public health interventions and research*. San Francisco, CA: Jossey-Bass.
- Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M., & Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health Services & Outcomes Research Methodology*, *14*, 166–182.
- Taylor, B. J., Francis, K., & Hegney, D. (2013). *Qualitative research in the health sciences, Methodologies methods and processes*. London: Routledge.
- Torgerson, D. J., & Sibbald, B. (1998). Understanding controlled trials. What is a patient preference trial? *BMJ*, *316*, 360.
- von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gotsche, P. C., Vandenbroucke, J. P., & Initiative, S. (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet*, *370*, 1453–1457.
- WHO. (2015). *WHO global strategy on people-centred and integrated health services: Interim report*. Geneva: World Health Organization.
- Zelen, M. (1979). A new design for randomized clinical trials. *The New England Journal of Medicine*, *300*, 1242–1245.
- Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Haynes, B., Oxman, A. D., Moher, D., Group, C., & Pragmatic Trials In Healthcare, Group. (2008). Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *BMJ*, *337*, a2390.