

Source Book

Table of Contents

- Introduction 3
- Research design 4
 - Experimental designs..... 4
 - Quasi- and pre-experimental designs..... 5
- Choosing between designs 7
 - Figure 1. A decision tree to help policy makers and researchers choose between possible designs..... 8
- Ethical conduct..... 9
- Economic analysis..... 10
- Glossary..... 11

Introduction

This Source Book is a companion to the Translational Research Framework. It provides some additional information that might be useful when planning your research.

This Source Book includes:

- more information about types of research design, and control groups (page 4)
- a tool to help you start thinking about the kinds of research designs that might work for your project (page 8)
- a brief summary of ethical considerations, with a link to the NHMRC national guidelines that are worth reviewing when you are preparing your applications (page 9)
- a summary of methods for measuring and assessing costs and conducting economic analysis (page 10)
- a glossary of terms used in both the framework and this source book (page 11).

Research design

The term **research design** (also known as a *study design*) describes the set of tasks necessary to systematically examine the effects of an innovation. The purpose of a good research design is to be as confident as possible that the innovation *caused* any changes that were observed. To do this we need to ensure that:

- the innovation was optimally developed and planned (formative evaluation), implemented as intended, and reached the target audience (process evaluation)
- the processes of recruitment of people into the innovation are described (who they were, how they were selected)
- the best *measurements* possible were used to assess the impact and outcomes from the innovation (the results)
- the best possible *research design* was used to assess the effects of the innovation
- there are no alternative explanations for the results, so that we can be confident that the results observed are attributable to the innovation
- we can identify the individuals, population groups or sub-groups to whom these observed innovation effects do and do not apply
- we can identify how and why the program worked (or did not work) for the whole, or subsets of the target group.

Put simply, the better the research design and methods that we use in assessing impact and outcome, the more confident we can be that the observed effects of a program were caused by the innovation and did not occur by chance, or were not due to other factors or influences. The research design should be the 'best possible' in the context of the program, its implementation, and in meeting the expectations of the different stakeholders.

Experimental designs

Randomised controlled trials

There is a hierarchy of research designs, from the 'most scientific', which use *experimental* designs, and are commonly referred to as *randomised controlled trials* (RCTs). In this design, the people that receive the innovation are not pre-determined, with individuals randomly allocated

to receive the program, or not to receive the program. Every individual or group has an equal chance of being offered the program, or not. This *random allocation* of individuals makes it more likely that the differences (such as personal background, existing health status) between a population receiving an innovation and a population not receiving an innovation are minimised. This in turn minimises the possibility that observed changes in the innovation group are due to "chance" effects caused by pre-existing differences in the two populations, and any such changes were *caused* by the innovation.

Once the individuals have been randomly allocated to the *innovation or comparison (control) groups*, a baseline assessment is made of their characteristics (e.g. age, gender) and the objects of the innovation (e.g. clinical status, or personal behaviour such as smoking) to determine that the innovation and control groups are comparable. Measurements are then performed on the same individuals after the innovation has been completed to assess change in the objects of the innovation and to test that change for *statistical significance*. The quality of an experimental research design can be examined according to well established criteria, such as the comprehensive 25-item [CONSORT checklist](#) ([Consolidated Standards of Reporting of Trials](#)).

Cluster randomised controlled trials

In addition to individual-level randomisation, it is possible to randomise at the level of communities or groups. For example, randomising whole clinical populations by hospital sites or community groups in workplaces or schools to receive an innovation; this is known as a cluster RCT. For example, within the clinics, service delivery staffs are likely to share common influences on their health behaviour or beliefs, which mean we need to statistically take account of this clustering. These common influences on staff behaviours and beliefs make it appropriate to consider them as a "group" in a research study. This type of research design is well-suited to innovations that are intended to be delivered to whole groups (such as a whole clinic), or innovations based on a modification to the environment that might have an impact on a whole group (e.g. the introduction of an exercise program for all patients in a clinic).

Step-wedge design

A third RCT design is the step-wedge design, where units or groups are randomly allocated sequentially to an innovation, so that waiting-list groups can be compared as the innovation rolls out across a population. This is useful across a large region, where for financial or other practical reasons an implementation needs to be rolled out over time.

As randomised trials usually involve some individuals receiving an innovation, and some not, it is important to make sure that nobody fails to receive the care that they need. In these circumstances 'usual care' or minimal innovations are often provided. For example, in a clinic or primary care setting it may be possible to randomly allocate groups of patients with diabetes to receive a comprehensive education and skills development program and others to be allocated to a control group, consisting of their 'usual clinic care'.

It is also important to keep the people in innovation and control groups separated from each other as much as is practically possible. One of the challenges faced by evaluators is to ensure that there is no *contamination* of the control group. For example, sometimes those receiving an innovation can share information or program resources with control group participants, who are not intended to receive the innovation. This increases the chances that this non-innovation (control) group will make changes that are object of the innovation, and such *contamination* makes it (statistically) more difficult to detect the effects of a program. This is particularly challenging in public health innovations that are designed to reach whole populations – hence the use of cluster RCTs, randomising by whole settings or geographically discrete units, as a research design alternative.

Although the use of experimental designs is always preferable, for the reasons indicated above it is often impractical to evaluate a health system innovation using an RCT design. This kind of design may require substantial funding and good control over the innovation delivery. It is most necessary when there is a high need for 'generating evidence'. This might be when a program is being tested for the first time, or is expensive (and would be costly to reproduce widely), or may be controversial, or is considered risky.

Quasi- and pre-experimental designs

Alternative designs provide less rigorous evidence of program effectiveness, but may be the most feasible in many situations. 'Best practice' will always require consideration of the 'optimal against the possible' in research design. These non-RCT designs are categorised as 'quasi-experimental' and 'pre-experimental' designs.

Quasi-experimental designs

Quasi-experimental research designs have clearly defined *control* or *comparison* populations – a population who do not receive an innovation and against which innovation group effects could be compared. Here, the group receiving the innovation is pre-determined and is not randomly assigned, so there is a greater chance that any observed changes may be influenced by differences between innovation and control groups or communities, and not caused by the innovation. This is especially the case when the innovation is delivered to enthusiastic volunteers who are then compared with a less committed 'control' population.

As is the case with RCTs, the quality of the results from quasi-experimental studies is dependent upon the size and representativeness of the study population, the use of valid measures before and after the innovation, the implementation of the innovation as planned, and optimal approaches to analysis and interpretation. The analyses may need to be statistically adjusted for baseline differences between innovation and control groups or communities (e.g. differences in the age, gender or social background of participants).

Cross-sectional study

RCTs have the same individuals assessed before and after the program. Quasi-experimental studies may also involve the same people (cohort) followed up from pre to post program, but some population innovations are evaluated using different (independent) cross-section samples of people from the target population to assess changes over time. This is referred to as a repeat *cross sectional study*, and whilst feasible in many health system evaluations, it is not as methodologically strong as a cohort study for explaining how and why observed changes occurred.

Time series design

Another type of quasi-experimental research design is a *time series design*. In this research design there are multiple pre-innovation measurements, followed by the innovation, and then several post-innovation measurements. Here trends in an outcome of interest (positive change in clinical condition, screening rates, or smoking behaviour) can be observed, and the innovation effect is judged by changes to the outcomes of interest over time, and whether the innovation group showed significantly greater change than a comparison group. Time series designs may be useful in the evaluation of policy innovations, as they allow for structured observation of change in a population where a policy has been introduced with little consideration for the evaluation of its effects (e.g. the relocation of a service, or a ban on smoking in specific places).

The time series design approach is strengthened by the addition of one or more comparison groups or regions, which also have trend data. This is a quasi-experimental design as the population receiving the innovation is not-randomly assigned, so there is a risk that any observed changes may have been influenced by factors or events other than the innovation. This type of quasi-experimental design is particularly useful where there are routine data collections by local health districts or other agencies (such as

cervical cancer screening tests). In these circumstances for example, the effects of a mass media campaign encouraging screening in one area can be compared with another not holding such a campaign.

Pre-experimental designs

The last group of designs have been described as *pre-experimental*. These provide the weakest evidence and should only be after other possibilities have been considered. A *before and after* (pre-post, one group) is a relatively weak design as it does not provide compelling evidence that an innovation *caused* any observed changes. Nonetheless, this simple research design does give some estimate of change, and is often used in pilot studies to estimate the likely effect of an innovation.

The weakest design is the one group, 'post program only' evaluation. This is where people are only surveyed or assessed following the program. This design should never be used for assessing program effects as it is not possible to claim that self-reported changes were caused by the innovation. Such a design may be quite useful for collecting process evaluation measures, including as participants' assessment of their experience of an innovation's component parts.

Choosing between designs

The decision tree in **figure 1** is intended as a guide to only some of the most common research designs available. There are almost any number of variants of research that can be designed and appropriately applied to the roll-out of innovations and programs. Nevertheless, the decision tree is useful in providing a greater shared understanding between policy makers and researchers about what is possible.

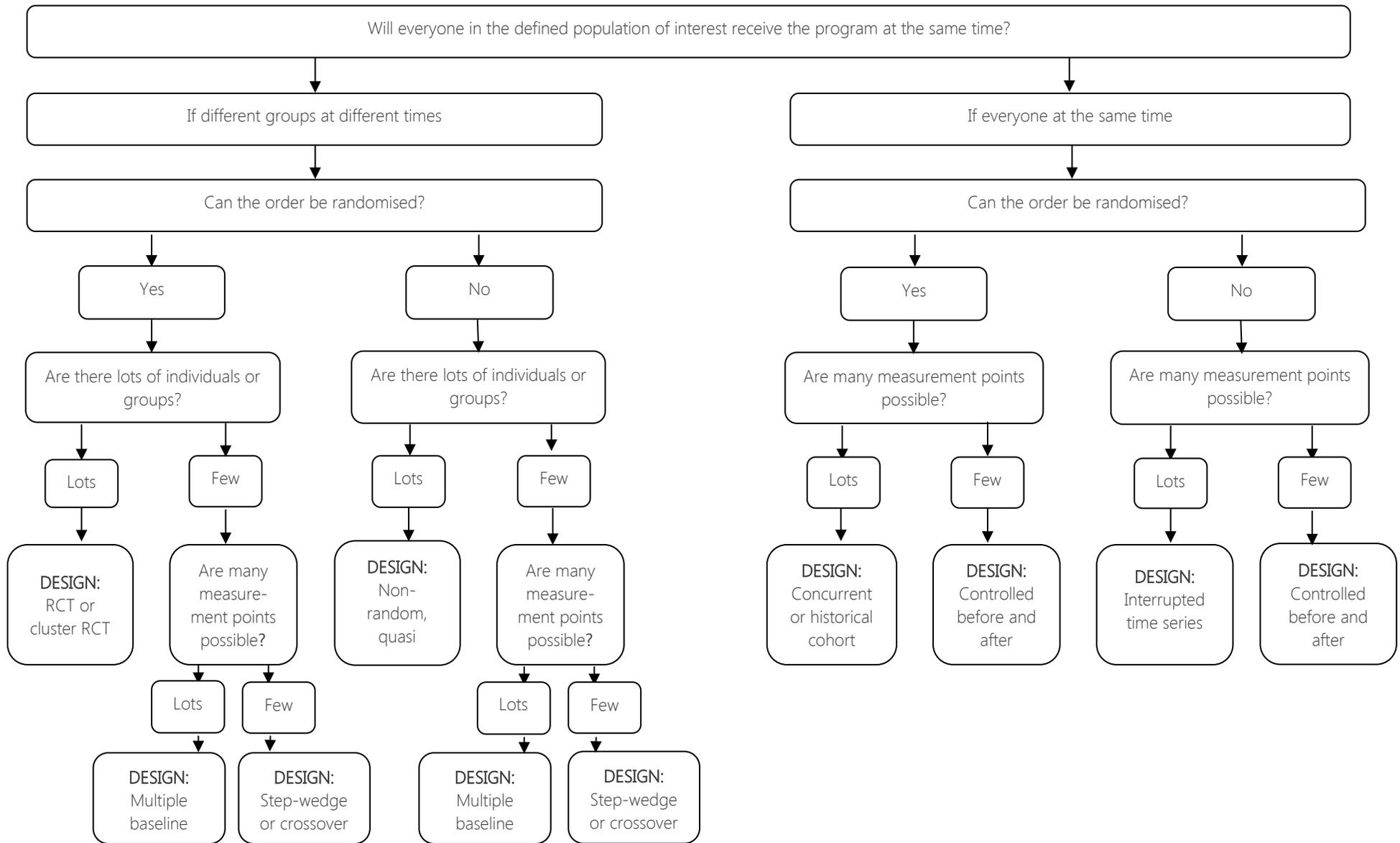
As indicated in the decision tree, typically the first decision point is about whether an innovation or program is to be rolled out across a whole population at once, or whether it might be possible to roll-out the innovation to different groups over time. In addition to the research design implications, the latter option may also be useful for practical considerations, such as resource requirements, budget implications or different settings/groups within a population being ready to implement the innovation at different times.

If the innovation can only be rolled out across a whole population at once, the next decision point is whether there are few or many measurement points available. If there are many data points available (e.g. hospital admission data), then an interrupted time series evaluation may be possible, or alternatively a cohort evaluation with concurrent or historical controls. If relatively few data or measurement points are available (e.g. pre-post innovation survey) this restricts the type of research design to a before and after study, with or without some sort of control group.

If it is possible to roll-out the innovation to different groups over time, the next decision point is whether the order in which the groups receive the innovation can be randomised. As indicated earlier, randomisation is the strongest tool we have to ensure non-random differences between groups (or individuals) are accounted for which, in turn, increases confidence that the observed outcomes are due to the innovation, rather than a systematic bias introduced by the evaluation. Consequently, it is highly desirable that the order in which groups receive the policy/program is randomised. In the case where randomisation is possible, if there are a relatively large number of groups (e.g. GPs or schools in NSW), then a RCT or cluster RCT is likely to be possible, which is regarded as the gold standard evaluation. If there are only a few groups, however (e.g. Area Health Services), then the type of research design depends upon the number of measurement points available. If there are many than a multiple-baseline design may be appropriate, but if there are few then a stepped-wedge or crossover design will be possible.

Even in the case where the order in which groups receive the innovation cannot be randomised, it will still be possible to do an evaluation. If there are a relatively large number of individuals or groups than a non-random (or quasi) experimental design can be used. If there are only a few groups, then a multiple baseline design could be used and if there are many groups, then a step wedge or crossover design is most appropriate.

Figure 1. A decision tree to help policy makers and researchers choose between possible designs



Ethical conduct

When developing your application, it is important to consider the ethical conduct and responsibility of your project. Research involving human populations – including studies that are about or with people, their data or tissue – need to ensure that potential risks or harm to participants have been accounted for and removed or minimised at all stages of the study (e.g. recruitment, innovation delivery, data collection, reporting and dissemination of findings). Thinking about this early is important because it has implications on research methods, design and timing. Early discussion with Human Research Ethics Committees (HREC) can be very useful.

The National Health and Medical Research Council's [National Statement on Ethical Conduct in Human Research](#) (updated 2015) provides guidance that will help you understand and describe the level of risk involved in your planned

research, how to minimise and manage that risk, and what level of ethical review is suitable (see chapters 2.1 and 5.1).

Additional reading:

The National Health and Medical Research Council. [Ethical considerations in quality assurance and evaluation activities](#). 2014, Canberra: Commonwealth of Australia.

NSW Health. *Research ethics and governance*. 2016, June 17. Retrieved from <http://www.health.nsw.gov.au/ethics/pages/default.aspx>

You can also contact your local Ethics Committee for advice. A list of HRECs in NSW can be found at: <http://www.health.nsw.gov.au/ethics/Pages/contacts-hreecs.aspx>

Economic analysis

Economic analysis is considered to be a powerful tool for decision making, priority setting and ultimately the allocation of scarce resources in health. Economic evaluation represents the core set of methods for assessing costs and benefits of innovations and health programs. Cost-effectiveness and other economic evaluation techniques address the question of efficient allocation of resources within budget constraints. The principle underlying economic evaluation is that the program that should be chosen is that which maximises benefit (usually defined as some measure of social welfare), within the available resources (as measured by budget constraint).

Costing is the collection of financial resource use associated with an intervention or a program. Costing of a program or intervention is important for two reasons. The first is to identify where resources are flowing within a program and whether the various arms are receiving money as was intended in the original plan. The second reason is that accurate and transparent costing analysis plays a significant role in economic evaluation of health and healthcare interventions. The way the analysis is undertaken depends largely on the reason that a costing exercise is being undertaken. The type of analysis may differ according to the depth of the analysis undertaken.

Other economic evaluations may involve more complex planning where the advice of a health economist may be needed. These can include:

- **Cost minimum analysis (CMA)**, which is used when the consequences of two or more health programs are judged to be *equivalent*. In this case, the comparison is only in terms of the costs of delivering the innovation or program.
- **Cost effective analysis (CEA)**, which looks at possible consequences of the innovation measured in terms of a single uni-dimensional unit considered to capture the relevant outcomes (e.g. lives saved, life years saved, cases detected or cases prevented).
- **Cost utility analysis (CUA)**, a specialised form of CEA developed for the health setting. In CUA the consequences of the innovation are measured in terms of an outcome that combines survival and quality of life, allowing for comparison across innovations with disparate outcomes, across different health care conditions and population groups. A common measure used in CUA is Quality Adjusted Life Years (QALYs).
- **Cost benefits analysis (CBA)**, where both the consequences and the costs of the innovation are measured in monetary units. This facilitates a direct comparison of innovation benefits to innovation costs.
- **Cost consequence analysis (CCA)**, which presents the full array of outcomes rather than summarising innovation consequences into a single measure to enable the user to form their own judgements.

Additional reading:

NSW Department of Health. [Issues in the costing of large projects in health and healthcare](#). 2008, Sydney: NSW Ministry of Health.

Glossaryⁱ

Adaptability: the degree to which the innovation can be changed while still maintaining effectiveness.

Adoption: the proportion of intended intermediary target settings, practices or organisations (examples may include schools and workplaces) that adopt an innovation before proceeding to implementation with the intended target group.

Comparability: refers to how consistent the context in which the original innovation was implemented is with that of the new environment or setting.

Compatibility: refers to how well the innovation fits with the systems, services and practices of the new environment or setting.

Contamination: the amount to which control, comparison groups or communities are exposed to elements of the innovation. For example, in a community-wide campaign, the control communities may share some media channels in common with communities receiving the innovation. If they are exposed to the campaign messages, it will be more difficult to show greater program effects in the community receiving the innovation.

Effectiveness: the extent to which an innovation is successful in 'real life' conditions in achieving the impact and outcomes that were predicted in the planning of the program.

Efficacy: the extent to which an innovation is success under controlled or 'best possible' conditions.

Evaluation: the process of judging the value of something. An evaluation can determine the extent to which an innovation or program has achieved its desired outcomes, as well as assess the different processes that led to these outcomes. It is important to note, that there is no standard, one size fits all approach to evaluation; it is context specific.

Evaluation design: the set of procedures and tasks that need to be carried out to examine the effects of the innovation. The purpose of good evaluation design is to enable us to be as confident as possible that the innovation caused any changes that were observed.

Feasibility: the viability, practicability, or workability of the study, program or innovation.

Fidelity: the extent to which delivery of an innovation adheres to the protocol or program model originally developed.

Formative evaluation: a set of activities designed to develop and pre-test program materials and methods. Formative evaluation occurs as part of program planning, and occurs before any elements of the program are implemented.

Generalisability: the extent to which findings from the study are likely to be reproduced in other groups or in the whole population.

Innovation: a set of actions intended to bring about change or produce outcomes.

Outcome: the intended change of the innovation. Outcomes will vary depending on the innovation, and might include health status, health behaviours, and clinical behaviours or systems factors.

Process evaluation: a set of activities designed to assess the success of program implementation. Process evaluation describes and explains what happens once the program has actually started, and the extent to which the program is implemented and delivered as planned.

Reach: the level of contact with an individual or participation of an intended target population in an intervention.

Replicability: the degree to which the results of the innovation can be repeated in a different setting, or different population or sub-group.

Sample: a group of individuals selected from a population for study, or to be the subjects for an innovation.

Sample size calculation: determines the number of people needed for an evaluation study using standard statistical formulas. To do this, it is necessary to specify what quantitative change is expected or hoped for in the intervention (e.g. a 10% increase in breast cancer screening from 70-80% following the innovation).

Scalability: the ability of an innovation shown to have been efficacious on a small scale and/or under controlled conditions to be expanded under real world conditions to reach a much

greater proportion of the eligible population, while retaining effectiveness.

Statistical significance: a measure of the extent to which the relationship between variables, or observed results, from a study might have occurred by chance. Statistical significance is assessed after the application of appropriate statistical tests.

ⁱ The definitions in this glossary have been based on Bauman AE, Nutbeam D. *Evaluation in a Nutshell: A practical guide to the evaluation of health promotion programs*. 2nd edition. 2014, North Ryde: McGraw-Hill.