RUNNING RANDOMISED CONTROLLED TRIALS IN INNOVATION, ENTREPRENEURSHIP AND GROWTH:

AN INTRODUCTORY GUIDE

TRIIN EDOVALD
TEO FIRPO

Innovation Growth Lab

NESTA

Nesta is an innovation charity with a mission to help people and organisations bring great ideas to life.

We are dedicated to supporting ideas that can help improve all our lives,with activities ranging from early–stage investment to in–depth research and practical programmes.

WWW.NESTA.ORG.UK

INNOVATION GROWTH LAB

The Innovation Growth Lab (IGL) is a global collaboration that aims to enable, support, undertake and disseminate high impact research that uses randomised controlled trials to develop and test different approaches to support innovation, entrepreneurship and growth.

WWW.INNOVATIONGROWTHLAB.ORG

WE WOULD LIKE TO ACKNOWLEDGE THE SUPPORT RECEIVED FROM OUR PARTNERS:

Nesta...

Ewing Marion KAUFFMAN Foundation

argidius

ACCIÓ
Generalitat de Catalunya
Government of Catalonia

Australian Government
Department of Industry

Innovate UK
Technology Strategy Board

growth analysis

Tekes

Department for Business Innovation & Skills

CONTENTS

CONTENTS

INTRODUCTION TO THE GUIDE

Despite the importance of innovation and high-growth entrepreneurship for economic growth, there is still too little reliable evidence on how best to achieve this. Across many fields there has been a growing understanding of the value that randomised controlled trials (RCTs also called 'trials') can bring as an effective research method to find out 'what works' – but they are still not as widely understood or used as they should be, particularly in innovation, entrepreneurship and growth (IEG) policy.

We believe that we need a more experimental approach to innovation and growth policy, through both trialling new interventions but also evaluating their impact more rigorously. This involves making much more use of RCTs to find out what works and what doesn't, while learning from the successful experience in conducting RCTs in other fields, such as development economics, health or education.

This is a guide on why, when and how to do an RCT in IEG. The guide is divided into two parts:

Sections 1 to 3 form a primer that gives you an overview of the basic information about the underlying principles, methodology, and role of RCTs in policy development and decision-making.

Sections 4 to 6 introduce the stages in the practical process of getting from an initial research question to a completed trial, while going through a number of individuals steps in each stage.

The guide has been designed for policy-makers, researchers but also practitioners who:

• Want a better understanding of the underlying principles of RCTs

• Want to test and evaluate new programmes and policies

• Need to develop new programmes and policies

• Need to assess RCT proposals

• Are responsible for managing and steering ongoing RCTs

You do not need in-depth methodological expertise to use this guide. Some familiarity with issues and challenges that might come up in RCTs alongside an inquisitive mind are expected instead. The guide is intended to equip you with enough knowledge to allow you to have meaningful conversations with technical experts, but also to help you better design and plan RCTs in the field of IEG. For those who are interested, the guide also offers signposts to more specialist sources.

This is our first version of the guide and we welcome your feedback at innovationgrowthlab@nesta.org.uk.

WHERE DO I START?

TAKE ME THROUGH IT FROM THE TOP

The first half of this guide is a primer that gives you an overview of the underlying principles, methodology, and role of RCTs in policy development and decision-making.
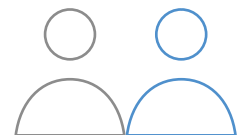
I'M READY TO GET STARTED

We recommend reading through the primer as a refresher, but if you agree with the following statements, you can skip the first part of the guide and go straight to the trial process:

✔ *I want to understand how RCTs are carried out in practice*

✔ *I have some experience with RCTs in IEG or other fields*
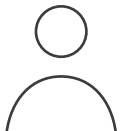
1

THE BASICS OF RCTS

What makes RCTs so valuable? Why and when do you need them? This section introduces the key features of an RCT and explains what distinguishes this type of study from others. It will help you understand in which contexts RCTs are suitable, when they are feasible and how they are managed, as well as laying out their limitations. Reading this section will equip you with the basic knowledge needed to understand RCTs.

1.1 WHAT IS AN RCT?

In essence, an RCT is an experiment carried out on two or more groups where participants are randomly assigned to receive an intervention or not. Participants are randomly assigned to either an intervention group (also called a 'treatment' group) who are given the intervention, or a control group who are not. The introduction of a randomly assigned control group enables you to compare the effectiveness of the new intervention against what would have happened if you had changed nothing.

RCTs are considered the 'gold standard' for establishing a causal link between an intervention and change. More specifically, RCTs are considered to be the most rigorous way of establishing if evidence resulting from an experiment shows that the outcomes have been caused by the programme or intervention – known as 'causal description' (or 'descriptive causation').

In RCTs, each group is tested at the end of the trial and the results from the groups are compared to see if the intervention has made a difference and achieved its desired outcome. If the randomised groups are large enough, you can be confident that differences observed are due to the intervention and not some other factor.

The people (or businesses, startup teams, business incubators etc.) who take part in RCTs (the study population) are called 'participants' or, less popularly, 'subjects'.

Randomised designs can take many forms. The figure below (and the accompanying example overleaf) which focuses on a straightforward two-group approach in order to highlight the key principles of RCTs, illustrates a simple randomised design. →
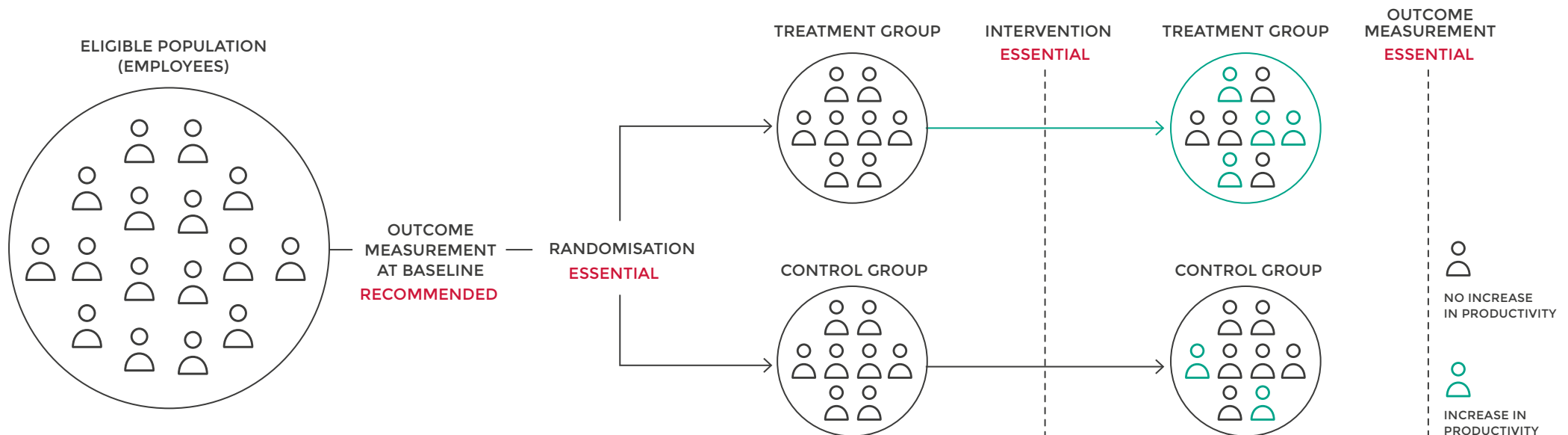
→ Imagine that you've introduced flexible working to a group of employees, allowing them to choose when to start and end work (within agreed limits) while working certain 'core hours' (e.g. 10am to 4pm every day). How will you know whether those practicing flexitime are showing increased productivity because of flexitime, and that they wouldn't have become more productive anyway?

And how can you know that it wasn't something else outside flexible working that improved their ability to be more productive? In RCTs you control for all other factors that could also affect the outcome, in this case factors (e.g. change in the immediate supervisor) that could impact the productivity of employees.

1.2 WHEN ARE RCTS NEEDED?

RCTs are usually needed to determine whether an intervention is achieving its aims and intended impact. In other words, they help you to determine whether an intervention works or not. There are a number of reasons for why RCTs should be considered as the first choice to establish the effects of an intervention:

• Due to the randomisation, they help to eliminate selection bias and offer a robust methodology that allows you to make causal conclusions. Selection bias occurs when the selected groups are not similar to begin with, which may lead any benefits of the new intervention being either exaggerated or underestimated due to external factors.

• A well-designed and executed RCT avoids potentially misleading results from non-experimental work (e.g. a cohort study), which has inadequately controlled for selection bias.

• They provide a concise and clear-cut conclusion of intervention effectiveness that avoids lengthy caveats.

• Their results can be incorporated into future systematic reviews and meta-analyses.

However, it is important to acknowledge that RCTs can still suffer from other considerable biases. Some of these can be controlled with careful planning, but others are inevitable in any RCT and must be considered when interpreting their results (for more details, see Section 1.7 ↰ on Limitations of RCTs). Sometimes, other approaches, such as quasi-experimental designs (QEDs) may be better suited to evaluate an intervention (see also Section 1.6 ↰).

1.3 WHY WOULD YOU RUN AN RCT?

RCTs can have multiple purposes. Their theoretical purpose in the domain of IEG is to improve understanding of the benefits or harms of one or more interventions. A well-conceived and well-executed RCT can inform, enhance, and sometimes change practice or policy:

• RCTs can help individual practitioners (e.g. innovation programme managers) or practitioner communities to guide or modify their practice.

• RCTs can provide those at the receiving end (e.g. entrepreneurs) with the information to help them choose what they would benefit from the most as individuals.

• The effects measured with RCTs can be used to assess the relative efficiency of interventions by studying an intervention's cost-effectiveness or undertaking a full cost-benefit analysis.

These features of RCTs mean they provide important information to policy-makers tasked with allocating resources to different interventions. Decisions regarding the funding of potential interventions take place within a context of limited resources. Increasingly, resource allocation decisions are being made against a backdrop of fiscal austerity. In this context, decision-makers need sound evidence of intervention impacts and cost-effectiveness so that they can use the available resources optimally.

Those responsible for interventions and concerned with ensuring their programmes continue to attract funding will have a keen interest in promoting RCTs in order to show that their programmes provide value for money and yield measureable benefits to participants, as well as to society as a whole.

1.4 HOW ARE RCTS MANAGED AND OVERSEEN?

A lot of attention is usually given to when and how RCTs are devised, designed and analysed. It is equally important, however, to pay attention to the actual on-going management and oversight of a trial.

Ideally, all activities within a trial must be guided by a protocol. A trial protocol is a document that sets out, in detail, the objectives, design and methodology of a trial. The protocol explains the purpose and function of the trial as well as how to carry it out, including the systems that must be set up for recruitment of participants, randomisation, data management and analysis. Protocols are described in detail in Section 4.3.2 ↰.

RCTs are typically conducted by research teams led by a principal investigator (PI), a person who is responsible and accountable for conducting the trial. The PI assumes full responsibility for the treatment and evaluation of trial participants, and for the integrity of the research data and results. Another key member of the research team is often the trial coordinator (also called a 'trial manager'), a person responsible for the day-to-day management of the trial who responds to issues that inevitably arise. In addition to the PI and the coordinator, research teams often include research assistants, statisticians, programmers, data managers and administrative staff. The final composition of the research team depends on the complexity and scale of the trial.

1.5 DOES MY INTERVENTION LEND ITSELF TO AN RCT?

The nature of the intervention is relevant when choosing any evaluation design, including an RCT. Some interventions lend themselves more readily to RCTs compared to others. The simpler, more linear and well defined the intervention, the more appealing it is as a subject for an RCT. Complex, multi-layered and highly flexible interventions with many actors involved, on the other hand, are less suited to an RCT.

In general, interventions that are more specific about the following four factors are more likely to be successful in improving IEG outcomes:

1) who they are trying to reach
2) what they are trying to achieve
3) what they actually consist of (i.e. what is delivered)
4) the rationale underpinning the intervention (e.g. theory of change)

Before carrying out any evaluation research, it is important to address questions regarding an intervention's focus, practicality and logic and this can be done in the design phase.

IT IS ALSO IMPORTANT TO CONSIDER:

• the scale of intervention implementation – imagine you are running an RCT on an intervention that is being offered around the country in different places; this would require the intervention to be delivered consistently across multiple sites as intended by its designers.

• how easy it is to control the intervention – imagine you are running an RCT on a business mentoring programme. Each mentor might have a different style, or might spend different amounts of time on each business according to their needs. However, in order to be able to properly estimate the effect of mentoring, the intervention's core aspects should be consistently delivered by all mentors according to an original design. In other words, interventions are to be delivered with what is called 'fidelity'. Resources and materials to promote fidelity include manuals, training materials, implementation procedures, technical support and fidelity protocols or checklists. Furthermore, it is important that the intervention fits seamlessly into existing practices and systems.

• what stage the intervention is at – namely, if the intervention is at an early stage of development and it is not sufficiently robust in its delivery methods, it may simply be too early for an RCT, which requires the intervention to be closely monitored and delivered in an agreed manner.

1.6 IS IT FEASIBLE TO RUN AN RCT?

It can be hard sometimes to decide in advance whether an RCT is the most feasible and robust way to evaluate an intervention, due to uncertainty regarding various trial aspects such as number of participants, likely intervention effects, and costs. A useful approach to establish whether an RCT can be successfully carried out, or whether a different evaluation approach should be taken, is to conduct a feasibility study. These focus on estimating important parameters and aspects of a trial, including:

• Identifying the number of eligible participants

• Establishing sample sizes that would need to be achieved for each group and the trial as a whole

• Understanding what outreach activities will be carried out to engage potential participants in the study

• Exploring the feasibility of randomisation and willingness of participants to be randomised

• Identifying feasible outcomes and how these might be measured, including designing any measurement instruments if necessary

• Ideally, estimating the standard deviations of outcome measures (this is important in order to decide sample size needed)

• Establishing effect sizes (for sampling purposes)

• Examining response rates (for example, in the case of surveys), follow-up rates, survival rates (for example, in the case of firms/small and medium-sized enterprises [SMEs]), etc.

• Estimating intraclass correlations in the case of cluster randomised trials (intraclass correlation is a measure of how similar members of a group, class or cluster are; clusters are described in Section 2.4 ↰ 'What can be randomised?')

• Exploring the availability of any potentially useful data sources (e.g. administrative data on firms)

• Identifying challenges in evaluating the effectiveness of an intervention

• Estimating the time and resources required to conduct the trial

Estimating these elements can help clarify whether a trial is likely to detect a certain effect in a robust manner, which can be helpful when deciding whether the RCT should be carried out. Usually, feasibility studies are not classified as RCTs. Sometimes, however, they can become small RCTs with a good enough sample size that should allow you to estimate some of the aforementioned aspects of a study with a sufficient degree of precision.

One of the key advantages of carrying out a piece of research to answer the question 'Can this study be done?' is that it allows you to understand and show beforehand if key elements of a study (e.g. ability to randomise selected units such as SMEs) are feasible before the main study starts. This becomes particularly important considering that it can take more effort to design RCTs and more money to run them when compared to some other research designs. Feasibility studies help you decide whether you should embark on a full trial to answer the questions you have about the intervention, or whether to take a different approach. Please see an example of a feasibility study overleaf.

EXAMPLE OF A FEASIBILITY STUDY

IGL has recently funded New Economy Manchester (UK) to carry out a feasibility study to assess which of Greater Manchester (GM) Business Growth Hub's programmes are most likely to make a growing contribution to the region's economic vitality as well as which are suited to a rigorous test of their impact through an RCT. In the first stage of the study, the research team will review the business support programmes and identify which lend themselves to an RCT. In the second stage, the research team will determine whether it will be feasible to conduct an RCT of the intervention. As part of the study, the research team will set out the theory underpinning the chosen intervention and the process by which the RCT will operate using a trial protocol as a guiding framework. A trial protocol will be developed alongside a briefing paper, aimed at senior GM officials, that will explain the proposed design of the trial and how the findings fit with the with the wider GM agenda of business support policy.

If an RCT is deemed unfeasible, QEDs can, to some extent, be used as an alternative design to estimate the causal effect of an intervention. QEDs involve the estimation of the counterfactual (i.e. comparing the observed results to those you would expect if the intervention had not been implemented) from a comparison group that has not been created at random.

QEDs are less robust than RCTs but are often more practical in application, particularly in retrospective studies or in prospective studies when randomisation is not possible and where ethical, political or logistical constraints rule out randomisation. Examples of QEDs include:

- Difference-in-differences: comparing the before-and-after difference for the group receiving the intervention (where they have not been randomly assigned) to the before-after difference for those who did not.

- Instrumental variables: estimating the causal effect of a programme by identifying instrumental variables, i.e. variables that impact on outcomes by affecting a key independent variable. This option can also be used to control for measurement errors.

- Propensity scores: statistically creating comparable groups based on an analysis of the factors that influenced people's propensity to participate in the programme.

- Regression discontinuity: assigning a cut-off or threshold above or below which an intervention is assigned and then comparing the outcomes of individuals or other units such as firms just below the cut-off point with those just above it.

Even though QEDs by definition lack random assignment, under certain assumptions and when properly designed they can provide useful insights. The quality of QEDs is primarily determined by the validity of the comparison group, particularly how similar it is to those treated on the basis of intervention measures.

1.7 LIMITATIONS OF RCTS

The main appeal of RCTs comes from their potential to reduce selection bias due to the randomisation, which gives us confidence that the effects measured are due to the intervention and not other differences between the treatment and control groups. Thus, RCTs have high internal validity. Still, random allocation does not protect RCTs against other types of bias such as those arising from poor concealment of treatment status, incorrect randomisation, or drop out of participants and other challenges. Below is a description of the main limitations of RCTs.

1.7.1 DIFFICULTY WITH GENERALISABILITY

One of the most frequent criticisms of RCTs is around their low generalisability, meaning that it can often be difficult to transport learnings from an RCT to different contexts. Although trials present the best evidence on the outcomes of an intervention, that evidence is specific to the context in which the intervention was set, and it is not always possible to infer that similar interventions would have the same effect in other environments, or even with an increased population.

The large numbers of people or other units (e.g. SMEs) required to power trials also means that the focus is on overall average effects and therefore they shed light on central tendencies. In turn, that means that you don't necessarily get evidence that can be applied to individuals. Moreover, participants in an RCT might be more willing to comply with the intervention. For instance, students in a programme aimed at encouraging entrepreneurship might feel more motivated to follow the course if they know they are part of a trial. This, coupled with the fact that each RCT comes with a specific set of inclusion/exclusion criteria, might mean that the participants in the trial may not be representative of the population that the intervention would eventually target. Therefore, scaling up the programme might not necessarily achieve the results one would expect based on trial results.

1.7.2 RCTS DON'T ALWAYS TELL YOU WHY

RCTs tend to tell you whether or not something works, and how well (though that is not always the case, see Section 4.1 ↰ for different types of trial questions). Why it does or doesn't work tends to be up to the interpretation of the researchers, whereas the reasons for 'why' are in fact crucial for the practitioners as they are the ones likely to adopt the intervention being tested.

Moreover, RCTs require researchers to settle on one specific design to ask a specific set of questions and, thus, RCTs offer answers to only the specific intervention designs tested. For example, you test a comprehensive bundle of services for SMEs that are new and inexperienced exporters. Your evaluation may show that the intervention – an in-depth capability assessment and a face-to-face skills-based programme – is effective in improving SMEs' readiness for international business, and helps them build international trade capacity. However, your study doesn't necessarily tell you whether the intervention would be just as effective with a more light-touch assessment or an online programme, or whether adding other components would have a greater impact on SMEs' understanding of the stages of export in relation to their own business.

Additional research, often qualitative, is needed to answer specific questions, including why the intervention worked better for some than others and what the mechanisms of change for a particular intervention could be. Further insights could also be acquired by conducting a process evaluation as part of the RCT to see if the intervention has been carried out as intended. Process evaluations often use a mixed method approach and focus on assessing implementation fidelity, which is crucial in interpreting the RCT results. Mixing other methods with RCTs is becoming increasingly common to aid with validation, contextualisation, triangulation and control. Overall, the analysis of an ordered sequence of events linking the causes of a problem with its effects (a causal chain) should not be overlooked since it is necessary to understand how an intervention was implemented, and how and why it worked for whom and where.

1.7.3 SAMPLE AND TIMELINE CHALLENGES

Not only do RCTs often need large number of participants to have adequate statistical power, it can also take a long time to achieve the required sample size and for the expected outcomes to come about. Furthermore, RCTs that require recruiting and retaining large samples of participants to detect effects of a specific size might be faced with the constraint that there are simply not enough eligible participants. Equally, RCTs can sometimes be faced with ethical constraints, such as the impossibility of denying an intervention to a subset of participants. There has to be sufficient doubt about the particular intervention being tested to allow the withholding of it from half the study participants.

RCTs tend to look at evidence in the short-term rather than the long-term. On the one hand, this means that trials are likely to capture only a subset of intervention results. On the other hand, RCTs that require large samples and long study periods to achieve the desired outcomes can be difficult and expensive to undertake. Moreover, the timescales that are required to yield rigorous evidence may not meet the needs of, for example, policy-makers who might need results in a shorter timescale than an RCT can provide.

KEY MESSAGES:

▸ RCTs are considered the gold standard for establishing causation between an intervention and outcomes. This means that RCTs help to inform us on what works.

▸ RCTs typically involve a trade-off between the ability to trace causal inferences to the intervention (internal validity) and the generalisability of results (external validity). The so-called perfect RCT is designed strictly with internal validity in mind.

▸ Even though there are a limitations with RCTs, criticism can equally be made of any other method since all do some things better than others.
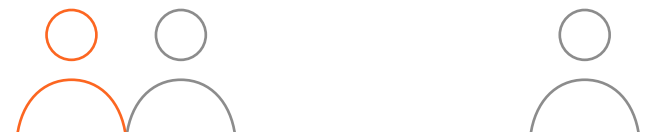
2

RANDOMISATION: THE CORNERSTONE OF THE RCT

Randomisation is the key element that separates RCTs from other study designs. The theory behind it is beautifully simple – though in practice it can sometimes be tricky to implement. This section explains the value of randomisation, and sets out how to achieve it. It also describes some of the practicalities of randomising participants. Reading this section will help you to better understand the concept of randomisation and how to implement it in practice.

2.1 WHAT DOES RANDOM ALLOCATION MEAN?

Random allocation means that participants are chosen at random to either receive the experimental intervention or not. More specifically, it means that all participants have a defined probability of being assigned to a particular intervention or trial arm. An arm of a trial is a group of participants receiving a specific intervention (or no intervention). In RCTs, neither the investigator, practitioners or participants determine allocation and it is not predictable based on a pattern.

It's important to note that there are a number of ways to assign participants into trial arms that may appear to be random, but in fact are not. This is the case when the allocation is based on a predictable characteristic or element of the participant, such as the entrepreneur or the firm. Examples include:

• Date of birth:
  e.g. all entrepreneurs born on odd days are assigned to one type of business training, while those born on even days receive a different type of training or no training at all.

• Some sort of record number:
  e.g. firms registered with a code ending in an even number are given vouchers, while those with odd last numbers are not.

• Day of enrolment:
  e.g. participants signing up on Monday receive the intervention; those signing up on Tuesday are assigned to the control groups.

• Alternating:
  e.g. the first startup to sign up for an accelerator gets an extra intervention, the second startup does not.

None of these methods described above should be considered as really generating random allocation sequences, but rather systematic occurrences. The generation of allocation sequences in randomised trials must be about chance, not choice.

2.2 WHAT'S SO SPECIAL ABOUT RANDOMISATION?

The key point is that randomisation ensures the two groups are statistically equivalent in all respects at the point they are randomised. By allocating the participants randomly, the characteristics of the participants are likely to be similar across groups at the start of the comparison (also called 'baseline'). By keeping the groups balanced at baseline – as similar as possible at the beginning of the study – the investigators will be better placed to isolate and quantify the impact of interventions they are investigating. They can do this while minimising effects from other factors that could influence the outcomes (these are called 'confounding factors', factors that are associated both with the outcome of interest and with the intervention of interest).

2.3 HOW TO ACHIEVE RANDOMISATION?

There are many ways to generate random allocation sequences, depending on the number of trial arms, the type of participants, etc. The most common randomisation methods include:

2.3.1 SIMPLE RANDOMISATION

This is the simplest method of random allocation, where participants are assigned to treatment or control groups through a single sequence of random assignment. The most common example is tossing a coin (for studies with two trial arms) although this is rarely used in practice as these days statistical software is commonly used (see Section 2.5.2 ↰ on tools for randomisation).

The strength of simple randomisation is its simplicity, which maintains complete randomness. However, when dealing with small samples there is a risk that simple randomisation will lead to groups of different sizes. For instance, when assigning a sample of 20 firms into two groups, simple randomisation could easily lead to a control group with five firms and a treatment group with 15 – all by chance. This can be very problematic for technical reasons (see Section 4.2.5 ↰ Sample sizes). In large samples, however, it can be trusted to produce groups with similar numbers.

STRENGTHS:

Easy to implement, ensures complete randomness.

WEAKNESSES:

Can lead to unbalanced groups when working with small samples, which poses threats to internal validity.

2.3.2 BLOCKED RANDOMISATION

One way to ensure that each group has similar numbers is to use restricted or blocked randomisation. This consists of assigning participants to their group by randomising them in blocks of sequences. This reduces the possibility of imbalances occurring through chance and thus ensures that the same number participants will be allocated to the trial arm within each block.

For example, imagine you have a sample of firms for a programme for which they sign up on a rolling basis (known as a 'trickle sample') and you want to allocate firms to receive the experimental intervention or the standard intervention (also known as 'business as usual'). Since there are two trial arms, each new firm must be assigned to either the intervention or the control group (with a 1:1 allocation ratio that allocates 50% of participants to the intervention group and 50% of participants to the control group).

A simple way to ensure that the allocation remains roughly stable as more firms sign up is to randomise them in blocks. The smallest block is usually a block of four (and the block size must be divisible by the number of trial arms – in this case divisible by two; arm A and an arm B). Given a block size of four, there are six possible ways to allocate participants (ABAB, AABB, BABA, BBAA, ABBA, BAAB). Allocation proceeds by randomly selecting one of the orderings and assigning the next block of firms to trial arms according to the specified sequence.

It is important to note that in blocked randomisation, particularly if the blocks are small, the allocation of participants may be predictable. This could lead to selection bias. However, this could be reduced by using random block sizes, longer blocks, and keeping the investigator blind to the size of each block.

STRENGTHS:

Helps to keep the numbers of participants in all trial arms as equal as possible, trial arms will tend to be uniformly distributed by key outcome-related characteristics.

WEAKNESSES:

Small block size increases the risk that the allocation process may be predictable, especially if the assignment is open or there is a chance of uncovering which participant will end up in which group beforehand.

2.3.3 STRATIFIED RANDOMISATION

Another form of restricted randomisation involves stratifying the sample by certain baseline characteristics. Stratification divides the original sample into groups according to characteristics that are known to be related to the outcome of the study (e.g. gender of participants, size of the firm). Using stratified randomisation helps to keep the characteristics of participants as similar as possible across the trial arms and thus achieve approximate balance of important characteristics.

Stratification is useful when the size of the sample is small enough that simple randomisation might yield groups with very different characteristics. For example, in a trial including 200 SMEs, all 30 micro-firms could end up in the control group, meaning that the intervention group would consist of relatively larger firms only. This would make it hard to estimate the true effects of the intervention and the results might be difficult to interpret. Stratification can also help increase statistical power in smaller trials and when the stratification factors have a large effect on expected outcomes.

Once the participants are stratified (with, for example, size being the stratification factor or 'stratum'), the next step is to produce a separate block randomisation list for each subgroup (stratum). In the example above, separate lists of random numbers would need to be constructed for micro-firms and larger firms to ensure that each type of firm is evenly spread between control and treatment groups.

In practice, stratified randomisation can get quickly unmanageable if participants are stratified by several factors across multiple trial arms since you may end up with too few participants with the stratified characteristics to ensure numerical balance across the different factors. This applies particularly to smaller studies where it is not advisable to stratify on more than one or two variables, as the number of strata can quickly approach the number of participants. When it is really important to achieve close similarity between different trial arms for several variables, minimisation can be used (see Section 2.3.6 ↩ on Minimisation below). For bigger studies, the benefits of stratification become very small once the sample size is large enough since randomisation itself tends to create balanced groups.

STRENGTHS:

Helps obtain balanced groups in terms of characteristics that are deemed to be important predictors of the outcome.

WEAKNESSES:

Cannot be achieved without knowing the full sample in advance (i.e. in the case of a trickle sample), and can be complex to implement in practice.

2.3.4 MATCHED RANDOMISATION

In matched randomisation (also called 'case-matched randomisation'), each participant is matched with one or more participants with similar characteristics. Each participant is then randomly allocated to one trial arm. In the basic example – with just one intervention group and one control group – participants are matched into pairs, with pairs being chosen so that participants (e.g. firms) are as similar as possible with respect to potential confounding variables. One member of each pair is then assigned at random to the intervention group and one to the control group. Similar matching procedures can be used when there are more than two intervention groups. For example, in the case of a three-arm trial, matched triplets would be used.

This is another way of achieving a balanced sample – matched randomisation can be considered a special case of stratified randomisation, in which the strata are each the size of two or three. However, it tends to come with certain challenges including some loss of statistical flexibility (since the matching must be taken into account in the analysis) and the fact that the sample must be known in advance. Moreover, if one member of the pair (or triplet) drops out during the trial, it requires for the other matched participant(s) to be dropped as well, which introduces a bias.

STRENGTHS:

Ideally, allows us to achieve numerical balance; can improve the power of the study when there are small samples and high variability in the outcome; requires and equal number of participants.

WEAKNESSES:

Loss of statistical flexibility, difficulty if one of the participants drops out, cannot be used for trickle recruitment.

2.3.5 PAIRWISE RANDOMISATION

In pairwise randomisation, participants are first paired regardless of their characteristics, and then randomised. This approach only helps keep a numerical balance between trial arms. This randomisation method is particularly helpful when the intervention can only be provided to a limited number of participants at any given moment. It shares many of the limitations of matched randomisation.

STRENGTHS:

Ideally, it allows us to achieve numerical balance; may be beneficial for logistical reasons, for example, when wanting to balance allocation or make it more predictable to allow better planning of resources for intervention delivery.

WEAKNESSES:

Does not allow stratification by an individual covariate, loss of statistical flexibility, difficulty if one of the participants drops out.

2.3.6 MINIMISATION

A further way to achieve a balanced allocation is to use minimisation. This method is a type of adaptive randomisation that involves altering the allocation schedule as the trial proceeds to ensure the groups are balanced in terms of participant numbers and characteristics. Strictly speaking, minimisation is a non-random method of forming comparable groups.

Usually, a computer programme is required to carry out minimisation. In practice, it is not unusual to allocate the first five or ten participants at random. Thereafter, as the groups build up, before assigning the next participant to a group, the programme checks the imbalances within groups and participants are deliberately allocated to a group depending on the characteristics of those already allocated. For example, suppose you have a three-arm trial with two intervention groups and a control group, and a firm is the unit of randomisation. If the control group already has five micro-firms but the two intervention groups only have one each, the programme will assign the next micro-firm to one of the latter. This can be done directly without any actual randomisation. Alternatively, a random component can be maintained by changing the likelihood of a participant ending up in a given group (in the example above, the programme will assign a very high probability to the firm being assigned to the treatment groups).

The advantage of minimisation is that it can achieve a balanced sample even with many overlapping strata; it can also achieve balanced groups even when the full sample is not known in advance, unlike stratification. The disadvantage is that it is not completely random, and sometimes the allocation could theoretically be uncovered before a participant is assigned.

STRENGTHS:

Can ensure balance between groups, even with many strata; unlike stratification, can be done without knowing the full sample in advance.

WEAKNESSES:

Not a random process and thus can lead to predictability (which is why a random element is sometimes introduced to minimisation).

2.4 WHAT CAN BE RANDOMISED IN RCTS?

Randomisation can be done using different units, such as individuals or clusters. For instance, you might be interested in the effect of an intervention on individual entrepreneurs, or on firms as a whole. The difference is consequential and you should be clear about which one makes more sense in your case.

The most frequent unit of randomisation across different fields is the individual (e.g. an individual entrepreneur). However, sometimes it will be impossible to direct an intervention towards a selected group of individuals. For instance, a trial within a large firm varying its employees' schedule flexibility to assess its impact on employee motivation may not be feasible. Even if some employees are randomly assigned to a group that can have flexible working hours and others to the control group, there is a threat of contamination (also called 'spill-over effects'), which occurs when participants assigned to the intervention group affect the outcomes of those in the control group (or vice-versa). This means that the difference in outcomes between the intervention and control group no longer represents the impact of the intervention as the effects of the intervention may be 'diluted'. Thus, contamination poses a great threat to the integrity of the RCT design.

To limit contamination of the control group in such cases, it may be more appropriate to randomise groups of individuals or clusters. Drawing on the examples above, instead of randomising individual employees, entire teams within a large firm could be randomised to either the intervention or the control group. Moreover, entire firms could be randomised into different trial arms.

Usually, you would aim to minimise spill-over effects, as they tend to complicate the interpretation of results. Sometimes, however, measuring spill-over effects can be part of the goals of the trial.

> **FOR DETAILS ON MEASURING SPILL-OVER EFFECTS, SEE THE FOLLOWING PAPER:**
>
> Baird, S., Bohren, A., McIntosh, C., & Özler, B. (2014). Designing experiments to measure spillover effects. 🔗 *Policy Research Working Paper 6824*, The World Bank.

Even though cluster allocation helps to avoid or reduce contamination – and may easily be the only feasible method of doing a trial – these trials require larger sample sizes of participants than individually randomised trials. Cluster RCTs also require more sophisticated statistical techniques to deal with the multilevel characteristics of the data.

2.5 **PRACTICALITIES OF RANDOMISATION**

Although the theory of randomisation is very simple, there are a number of practical elements that need to be taken into account in order to ensure proper randomisation. They mostly revolve around allocation concealment. The idea behind this is that people in charge of recruiting the participants, the participant themselves and the researchers allocating participants into different groups should not know the allocation sequence (i.e. which participant is destined to be assigned to which group) until after the randomisation is completed. This prevents any of the parties involved from influencing the randomisation process.

For example, researchers and practitioners might be tempted to assign certain participants to the treatment group, in order to increase the chances of showing that their intervention works. Similarly, participants might have a preference regarding whether they are assigned to the treatment or control groups. The advantage of allocation concealment is that it prevents selection bias by facilitating enrolment of comparable participants in each group.

Allocation concealment should not be confused with blinding (or masking), which consists of keeping the participants, researchers, trial managers, outcome assessors and even those analysing the data unaware of whether each participant is in the treatment or control groups. Blinding is used more often in medical trials, where it is usually possible to provide a placebo to patients in the control group. In the field of IEG, however, blinding of participants is rarely achievable.

2.5.1 **WHO SHOULD GENERATE AND IMPLEMENT THE ALLOCATION SEQUENCE?**

To achieve a concealed allocation, it helps to break the randomisation process down into two steps:

STEP 1: GENERATING THE ALLOCATION SEQUENCE

The first step is the sequence generation process. Researchers in the trial team can carry out this step by specifying the method of generating the allocation sequence, and a list of any factors for stratification. To keep the sequence as unpredictable as possible, it is good practice to provide details of any planned restrictions (e.g. blocking) in a separate document that is unavailable to those who recruit participants or assign interventions.

STEP 2: RANDOMISING THE PARTICIPANTS

The second step consists of carrying out the randomisation. In order to minimise bias, it is beneficial to separate individuals involved in the sequence generation process and allocation concealment mechanism from those involved in the implementation of study group assignments. A third party – such as an independent statistician, a data manager, or a researcher who is not otherwise involved in the trial →

$\rightarrow$ – should execute this step. For instance, the list of participants can be entered into a web-based system by an independent statistician, who would then run the randomisation process and provide the final allocation to the trial coordinator. The benefit of computer-based randomisation is that it 'locks' participants into a trial arm, without either the participants, the intervention managers or the researchers being able to change it.

When such separation is not possible, it is important for the investigators to ensure that the assignment schedule is unpredictable and locked away from even the person who generated it.

2.5.2 WHAT TOOLS TO USE FOR RANDOMISATION?

As indicated in the section on simple randomisation, simple 'analogue' techniques for randomisation are technically correct, but they can lead to practical difficulties. For instance, tossing a coin can be both burdensome and potentially open to tampering (as the person in charge of the randomisation can easily alter the process). The same issue exists with similar methods such as shuffling envelopes with allocation cards or using random number tables. The drawback of these methods is that none of them leave an audit trail that would allow the generation of the allocation sequence to be checked.

Therefore, most investigators use software packages to perform randomisation. There is a wide range of software programmes available that can be used, including Excel, STATA or R (see also a prepared STATA code by J-PAL 𝒫). The choice depends on the user's preferences, the complexity of the randomisation, and the ease of use.

2.5.3 WHEN SHOULD YOU RANDOMISE?

Randomisation should always be carried out after participants have been checked for eligibility criteria in order to ensure a balance of participant characteristics across trial arms. Participant informed consent to be randomised into one of the trial arms should also be secured before randomisation.

Additionally, if you plan to measure outcomes at baseline so that this information can be added to the final statistical analysis, then these data should be collected prior to randomisation as participant knowledge of which group they belong to may affect their responses.
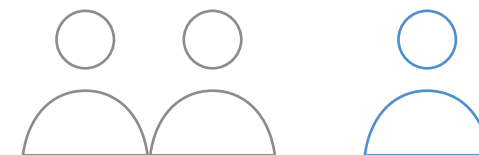
2.5.4 WHAT ALLOCATION RATIO TO USE WHEN ASSIGNING PARTICIPANTS TO EACH GROUP?

In general, participants are typically allocated in equal numbers to different trial arms. For example, in a two-armed trial the most efficient approach (from a statistical perspective) is to randomise participants using a 1:1 allocation ratio. In other words, 50% of participants go to the treatment group and the other 50% to the control group. However, there are a number of practical reasons for which you may choose to have a different allocation ratio, i.e. an unequal allocation.

One of the common reasons for using unequal allocation is resource constraints or costs. For instance, the organisation in charge of the intervention might have a fixed quota of participants it needs to assign to the treatment group (this is often the case when a government programme has a fixed allocated budget that must be spent in providing the intervention). In that case, the number of participants in the treatment group is fixed but the number in the control group could be different. For example, an allocation ratio of 3:1, where the intervention group has three times as many participants as the control group, would ensure that the intervention programme runs at full capacity while making the most efficient use of the study budget.

Alternatively, there might be limited resources to provide the intervention (including the availability of staff delivering it), or one treatment is significantly more expensive than the other, in which case unequal allocation may help to reduce overall trial costs. If providing the standard intervention for the control group is relatively inexpensive when compared to the experimental intervention, this can lead to a larger control group. For example, an allocation ratio of 1:2 (twice as many participants in the control group compared to the intervention group) could be used to minimise costs while increasing the total sample size. This may also increase the power of the study. However, if the total sample size of the trial is fixed, unequal allocation reduces statistical power. Whatever you choose, the allocation ratio needs to be taken into account in sample size estimations (see Section 4.2.5' ↰ on Sample sizes).

KEY MESSAGES:

▸ Random allocation or minimisation creates comparable groups.

▸ Randomisation ensures internal validity but not external validity.

▸ Individuals and clusters (including geographic areas) tend to be the most common units of randomisation.

▸ As far as possible, the allocation sequence should be concealed from those assigning participants to intervention groups, until the moment of assignment.

▸ A third party, independent of the research team, should execute the randomisation process.

▸ To create an audit trail and ensure transparency of randomisation, it is better to use software packages to perform randomisation.

▸ If the total sample size is fixed, equal allocation is best for statistical power. However, unequal allocation can provide a more powerful trial if there are resource constraints for the intervention.

3

TYPES OF RCTS

There are a number of ways to conduct an RCT, depending on your intervention, your objectives, the type of participants and the method of randomisation. Some RCTs will have one intervention and one control group, while others will have a more complicated structure. This section covers the main types of RCTs, describing in detail how they are structured and when they are best used. Reading this section will help you understand what type of trial is best suited to address your research questions.

3.1 WHAT ARE EXPLANATORY TRIALS?

Over the years, various jargon terms have been used to describe different types of RCTs. These terms tend to vary across different fields and can depend on the type of interventions tested in different populations of participants, in different settings, and for different purposes.

In this chapter, we will describe the terms most frequently used to describe different type of RCTs, such as explanatory and pragmatic trials. We will also cover other terms used to categorise RCTs according to different aspects of trials.

Explanatory RCTs focus on determining how and why an intervention produces an effect under ideal or controlled conditions. To create these ideal conditions, explanatory trials are often characterised by having tightly defined inclusion and exclusion criteria for participants, a strictly defined and tightly controlled intervention delivered fully and precisely by highly trained practitioners, and all trial arms receiving the same management except for the intervention under investigation.

It is common in education and social policy for the people who design the intervention to be involved in the trial. This is less common in the field of IEG policy. Explanatory trials are also known as 'efficacy trials'. Many trials, for example, in healthcare are explanatory studies as the primary question of interest is often 'can it work?', which would then be followed by 'does it work?' (as routine practice) once the questions of efficacy are answered.

Explanatory trials tend to be very useful in the early stages of developing an intervention as they can help you to demonstrate the effectiveness of an intervention under ideal circumstances. However, they are of less help when informing routine policy and practice since their lack of a pragmatic setting can limit their wider application. Thus, explanatory trials maximise internal validity but have poorer external validity.

EXAMPLE

Imagine, for instance, a trial designed to understand the impact of a training programme for SMEs rich in intangible assets (e.g. trademarks). It is thought that training key employees in a firm that holds many of these assets on how to best leverage them can help the firms' grow and access more financing. In an efficacy trial, firms would be recruited according to strict eligibility criteria. For example, only firms in a specific geographic region that have been audited and shown to have a certain (high) level of intangible assets might be eligible to participate in the study. Furthermore, eligible firms would not receive other relevant interventions. The training programme would be delivered by highly trained specialists focusing on increasing firms' knowledge of how to exploit their intangibles, which in turn could potentially lead to positive outcomes such as an increase in access to financing and growth. The training programme under examination would then be compared to business as usual.

3.2 WHAT ARE PRAGMATIC TRIALS?

Pragmatic RCTs (sometimes also called 'practical trials') focus on determining how and why an intervention produces an effect under real-world conditions. Pragmatic trials are often delivered in actual real-life practice conditions, where many variables are not held constant and where adherence to protocol can be somewhat incomplete. Pragmatic trials are also known as 'effectiveness trials'.

Similarly to explanatory trials, pragmatic trials incorporate random allocation of participants to the treatment group, thereby promoting internal validity. However, because of the range of study characteristics – such as broader participant selection criteria, intervention delivery by regular staff with standard training, and more flexibility and ultimately more variation in the management of trial groups – pragmatic trials tend to be more generalisable and thus help to make policy and practice more effective. It is important to note, however, that both types of trials are needed. Namely, if a pragmatic trial fails to demonstrate any effects, you cannot be sure without a preceding efficacy trial whether the lack of positive results is due to a lack of efficacy (that can be established via explanatory trials) or a lack of effectiveness.

In reality, most trials fall somewhere on the explanatory-pragmatic continuum. It is more commonly trials of clinical interventions that can be distinctively classified as explanatory trials due to certain features that are easier to apply in clinical settings and thus make the conditions highly controlled (e.g. using a placebo or a sham intervention when it comes to the control group, blinding, 'one size fits all' interventions as opposed to tailored ones, etc.). Furthermore, policy-makers are highly interested in pragmatic trials since these are designed to inform them not only about the comparative effectiveness of interventions but also on cost implications of various options in real-life situations.

EXAMPLE

Elaborating on the previous example on page 29, where a training programme is offered to intangible-rich SMEs, a pragmatic trial would test the impact of the training programme on the firms' outcomes on a wider range of firms in a real-life setting in order to maximise applicability and generalisability. For instance, the inclusion criteria for firms to participate in the trial would be less strict. The eligibility assessment may not include an audit to ensure they are intangible-rich, or a light-touch approach to audit could be applied. Practitioners delivering the training programme would not be the people who developed it, they would not necessarily need to adhere to strict guidelines, and therefore you might expect more variation in the quality of delivery. The best alternative intervention would be used for comparison with no restrictions on its application. Furthermore, to reflect real-world conditions there would be no particular plans to improve or alter compliance for the experimental or the comparative treatment. The pragmatic trial would therefore evaluate the intervention's effectiveness in real life.

3.3 WHAT ARE PILOT TRIALS?

A pilot RCT is a small-scale version of the main study that helps to test if all the main parts of the study work together. Often the terms 'pilot study' and 'feasibility study' are used interchangeably. However, what tends to set pilot studies apart from feasibility studies (described in Section 1.6 ↰) is that pilot studies may also help answer the research question.

Sometimes the pilot study forms the first part of the main study (also known as 'internal pilot study'), and researchers use data from the pilot study when they analyse the results of the main study. In other cases, data from the pilot study may be analysed but not used in the main study results (also known as 'external pilot study').

EXAMPLE

For instance, running a full-scale pragmatic trial on the effects of personalised business coaching for startup founders might be very costly, while the impacts on desired outcomes are unknown. A pilot trial could be conducted to provide a first estimate of potential effects. If successful, it could help convince the funders to commission a full trial.

KEY MESSAGES:

▸ Explanatory RCTs are well suited to inform understanding of intervention effects and mechanisms.

▸ Pragmatic RCTs are well suited to inform practice and policy as they evaluate an intervention in a intervention in a real-world setting.

▸ Explanatory RCTs maximise internal validity whereas pragmatic RCTs maximise external validity.

▸ Explanatory and pragmatic trials represent ends of a continuum rather than distinct entities.

▸ Explanatory RCTs tend to be better suited for some fields than others, such as healthcare.

▸ Pilot RCTs are extremely useful in for informing the definite study design.
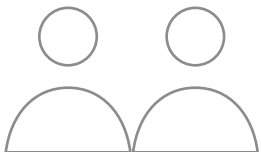
3.4 WHAT OTHER TYPES OF TRIALS ARE THERE?

There is a range of terms that are used to describe different types of trials. Some terms are mutually exclusive, some overlap considerably, and some complement each other.

RCTs can be classified according to various different attributes. As indicated in the sections above, trials can be categorised based on the aspects of an intervention they explore (e.g. explanatory trials versus pragmatic trials), their randomisation or analysis unit (e.g. individual-level versus cluster randomised trials – see Section 2.4 ↩) or whether the researchers and participants know which intervention is being assessed (e.g. open trials versus blinded trials - see Section 2.5 ↩).

Another way to look at trials is according to how participants are exposed to the intervention (see Section 3.4.1 to 3.4.3 ↩), or how participants' preferences are taken into account (see Section 3.4.4 to 3.4.6 ↩).

3.4.1 PARALLEL TRIALS

Parallel trials (also called 'RCTs with parallel group design') are the most frequently used of all trial designs, where each group of participants is exposed to only one of the study arms.

EXAMPLE:

In a parallel trial designed to assess the effects of a new entrepreneurship programme targeted at potential entrepreneurs on their ability to set up their own business, the investigators would randomly assign potential entrepreneurs to either receive the training programme or be exposed to business as usual.

STRENGTHS:

A simple and commonly used design that is easiest to implement. It tends to result in simple statistical analysis that often boils down to a simple t-test of the between group difference in the outcome, which is usually a mean or a proportion.

WEAKNESSES:

May be less efficient than other designs as these studies generally require large number of participants for the analysis.

3.4.2 FACTORIAL DESIGN TRIALS

Factorial design trials (also called 'factorial trials') compare two or more experimental interventions in combination as well as individually in a single experiment. It also allows you to explore whether or not there is an interaction between two interventions. Interactions occur when an intervention works more effectively in the presence of another intervention, or conversely is less effective. Unfortunately factorial trials rarely have sufficient statistical power to detect interaction effects. Unwanted interactions can also be a drawback as they lead to a reduction in the power to detect the main effects of an intervention. The simplest factorial design is a 2x2 factorial where there are four groups rather than two.

EXAMPLE:

A trial might be set up to test the effects of two interventions to incentivise scientific researchers to collaborate with the private sector and commercialise their products – such as organising meetings with companies in their sector and giving researchers a day off a week to work on a commercialisation project. In a 2x2 factorial design the trial would have four groups: one group receiving no intervention, one group exposed to only the meetings with companies, another with only time off for projects, and a fourth group receiving both interventions. The advantage of this approach (other than testing two interventions at the same time) is that it allows investigators to test the benefits of receiving both interventions together.

STRENGTHS:
Can be very cost-effective as they allow the evaluation of multiple interventions for the 'price' of evaluating a single intervention (in terms of sample size). Has the ability to detect interaction effects, as long as the sample size is large enough (see Section 4.2.5 ↰ on Sample sizes)

WEAKNESSES:
Can be complex to implement and the recruitment of participants can be confusing depending on the design complexity. Statistical analysis of factorial designs is more complex when compared to parallel trials.

3.4.3 CROSS-OVER TRIALS

Cross-over trials (also called 'cross-over designs' or 'change-over trials') are trials where each of the participants is given all the study interventions in successive periods. In other words, the participants act as their own control. The order in which they receive each of the study interventions is determined randomly. Cross-over trials are well-suited for investigating stable conditions, and therefore they are not used equally across different fields. For example, in medical clinical trials they are used to explore chronic and stable conditions such as angina, asthma, or arthritis, where the treatments tested focus on alleviating symptoms and not curing participants. Namely, if treatment A cures the patient during the first period, then treatment B will not have the opportunity to show its effectiveness when the patient crosses over to treatment B in the second period.

EXAMPLE:

In the field of IEG, imagine a trial around manufacturing processes that compares two methods: either never stopping the production line, or stopping the entire production line any time an error occurs. The goal of the trial would be to measure which method leads to better product quality. The manufacturing process would be randomised to produce non-stop for four months, followed by a four month period with the other method, or the reverse sequence. In order to ensure that the effects of the first intervention are not carried over, a 'wash-out' (with no intervention) period would be used avoid the effects carrying over.

STRENGTHS:

Require fewer participants compared to parallel trials, as each participant acts as his or her own control.

Well suited when the condition under study is (reasonably) stable/constant. The effects of the intervention must be reversible as the effects of one intervention cannot be present during the evaluation of another.

WEAKNESSES:

The principal drawback is that the effects of one intervention may 'carry over' and alter the response to subsequent interventions.

3.4.4 ZELEN'S DESIGN

If participants do not receive their preferred intervention in an RCT, there may be difficulties with participant recruitment and scientific problems with bias (for example, bias may occur when participants are aware of a new intervention not available to them and don't take up the standard intervention well). Zelen's design (also called 'randomised consent design') tries to address such difficulties by randomising participants before consent to participate has been sought.

In its simplest form, eligible individuals are randomised to a control group or to an intervention group before they give consent to participate in the RCT. Those in the control group receive business as usual and are not informed that they are part of a trial. Those who are allocated to the intervention group are offered the intervention and are informed that they are part of a trial. If those allocated to the intervention group refuse to participate in the trial, they are given the standard intervention that the control group is exposed to. However, they are analysed based on their original group assignment.

EXAMPLE:

Consider a trial investigating an intervention involving a certification programme to promote entrepreneurship amongst young people, with a special focus on the use of digital media. Participants would be unemployed 18-30-year-olds who have a business idea to explore. The control group would be exposed to business as usual, which might include volunteer mentors, online videos and promotional material about helping them become self-employed. The intervention group would receive three modules such as a creative enterprise workshop, digital promotion and presentation skills. Consent for participation in the trial would be sought only from those who were allocated to the intervention group and only after randomisation. Participants allocated to the intervention could refuse their allocated intervention and be exposed only to business as usual.

STRENGTHS:

Allows us to include almost all eligible individuals in the trial and estimate the true effect of offering the experimental intervention to participants as those in the intervention group can choose whether they want to take up the intervention or cross over to the control group.

WEAKNESSES:

Provokes strong reactions among researchers as participants are not asked for consent to be randomised and this is considered ethically controversial (thus not suited, for example, for therapeutic interventions). Large sample sizes may be required due to control group participants becoming aware of the study and/or intervention group participants refusing their original assigned intervention.

3.4.5 COMPREHENSIVE COHORT DESIGN

In comprehensive cohort designs (also called 'Brewin-Bradley design'), eligible participants with strong intervention preferences are allowed their desired intervention without randomisation. Those who do not have particular preferences are individually randomised in the usual way. All participants are followed up regardless of their randomisation status. At the end, the outcomes of those who participate in the RCT are compared with those who did not agree to be randomised and received their preferred intervention to assess their similarities and differences.

EXAMPLE:

This design could be applied in a trial designed to compare the provision of personalised mentoring with receiving business advice for 18-30-year-olds running startups that are less than two years old. Eligible participants would be asked to accept random assignment to either the mentoring programme or business advice service. Those who agree to be randomised would be randomised. Those young entrepreneurs who were 'randomisable' but refuse to be randomised would be given their preferred intervention and followed up as if they were part of a cohort study.

STRENGTHS:
Allows us to estimate the representativeness of the randomised sample. If randomised participants resemble non-randomised ones, the trial provides greater evidence of the external validity of the trial results. At the same time, any comparison that uses non-randomised groups is unreliable because of the presence of unknown and uncontrolled confounding factors.

WEAKNESSES:
The sample size must be larger than, for example, in a parallel trial. This is because the number of non-randomised participants must be sufficient to allow comparison of the effect of each intervention for individuals who express a preference for it with the effect for those who do not, and also for the comparison of individuals who are willing to be randomised and those who are not.

3.4.6 WENNBERG'S DESIGN

In trials using [Wennberg's design](), eligible participants are randomised to either a preference group or a randomisation group. Participants in the preference group are offered the opportunity to receive the intervention that they choose, and participants in the other group are assigned any of the study interventions based on randomisation. At the end of the study, all groups are analysed to assess the impact of participants' preferences on outcomes.

EXAMPLE:

Imagine a trial where people over the age of 50 who would like to set up their own business, and who have consented to participate in the study, are randomised to either a preference group or a randomisation group. Those in the preference group choose whether they want to attend a four-day interactive workshop on starting and running a business, or whether they would prefer one-on-one mentoring. Those in the other group are randomly assigned to either attend the workshop or receive mentoring. Analysis across groups allows researchers to assess the impact of participant preference on outcomes.

STRENGTHS:

Randomisation to either a preference group or a randomisation group of a trial overcomes a significant limitation in comprehensive cohort designs, which is the bias of refusing randomisation in the first place.

Overcomes the ethical problems of Zelen's design as the consent is acquired prior to randomisation.

WEAKNESSES:

Can never be entirely confident that it removes the preference bias as the participants still have to consent to be randomised and those randomised into the RCT arm still have preferences, which can affect the outcomes across groups.

KEY MESSAGES:

▶ Trial designs are selected by the characteristics of the interventions under study, the available resources, and the academic, political or practice-driven motivations behind the study.

▶ Even though participant preference designs are not used widely, more prominence could be given to such trials, which recognise the effects that choice may have on outcome.

▶ At the same time, incorporating participant preferences into trial design comes with specific challenges – such as the potential for additional differences between study groups other than preference, and increased sample size requirements or cost to complete a trial. A design without limitations is yet to be identified.

4

PLANNING AND DESIGN STAGE

In the following three sections, we describe a process of how to get from an initial research question to a completed trial. We have divided this process into three main stages, each of which include a number of individual steps that take you through the trial stages. The diagram on the following page outlines each of these three stages and the nine steps that comprise them. The rest of this section introduces the first stage, which involves designing and planning the trial. Reading this section will equip you with knowledge on how to best prepare for the trial implementation.

TRIAL PROCESS DIAGRAM

1. PLANNING AND DESIGN STAGE

STEP 1	STEP 2	STEP 3
RESEARCH QUESTION	TRIAL DESIGN	PRE-IMPLEMENTATION PREPARATION

2. IMPLEMENTATION STAGE

STEP 4	STEP 5	STEP 6
RECRUITMENT	RANDOMISATION	OUTCOME DATA COLLECTION

3. ANALYSIS AND REPORTING STAGE

STEP 7	STEP 8	STEP 9
ANALYSIS	REPORTING THE RCT	DATA MANAGEMENT AND STORAGE

4.1 **RESEARCH QUESTION**

Some questions would benefit from an RCT whereas others would be better addressed with different methodologies. So, the first question you have to ask yourself is do you really need to do an RCT?

4.1.1 **ASKING THE RIGHT TYPE OF QUESTION**

As a first step it's always a good idea to check whether the answer to the question you are planning to study already exists. If a thorough literature search on the subject indicates that it does, then there is no need to proceed with your study. However, if your literature search indicates that the answer to your question isn't already out there, you need to ensure that you are asking the type of question that requires an RCT. There are many different types of questions that you might be interested in. For example, you might be interested in strategic questions such as 'What areas of innovation policy should we focus our attention on?'. You might be interested in descriptive questions such as 'What are the key challenges that SMEs in certain geographic areas are facing?'. Equally, you might have process-related questions such as 'Is the young entrepreneurs mentoring programme using the resources dedicated for programme delivery?'. These questions are all crucial for a successful programme design and evaluation – but they don't need an RCT to be answered.

RCTs can answer questions about impact: did the programme or policy work? However, RCTs can also go beyond looking at what works and explore questions such as:

• Which components of the intervention were most crucial for achieving impact?

• Which version of alternative interventions produces the highest impact?

• Is the impact of a multifaceted intervention programme, which addresses a number of problems simultaneously, greater than the impact of the sum of the individual components?

• Are the results achieved in one context replicable in another context?

• What are the underlying processes for achieving (or not achieving) impact?

STAGE IN TRIAL

1. PLANNING AND DESIGN STAGE

STEP ONE
RESEARCH QUESTION

STEP TWO
TRIAL DESIGN

STEP THREE
PRE-IMPLEMENTATION PREPARATION

2. IMPLEMENTATION STAGE

STEP FOUR
RECRUITMENT

STEP FIVE
RANDOMISATION

STEP SIX
OUTCOME DATA COLLECTION

3. ANALYSIS AND REPORTING STAGE

STEP SEVEN
ANALYSIS

STEP EIGHT
REPORTING THE RCT

STEP NINE
DATA MANAGEMENT AND STORAGE

4.1.2 PRIORITISING THE QUESTIONS

If you decide to carry out an RCT, it is common to have multiple questions of interest, in which case you are likely to need to prioritise them.

On the one hand, you need to contemplate how influential the results generated by the RCT would be in informing decision-making, and thus policy and practice. On the other hand, you need to consider which of your questions of interest could be answered:

• well (e.g. can the outcomes be measured?)

• with precision (e.g. do you have a sufficient sample size?)

• in a representative context
  (e.g. can the intervention be scaled up later?)

• in an ethical way

4.1.3 ASKING A CLEAR STUDY QUESTION

Once you've prioritised the question(s) under study, you need to ensure that those question(s) are clearly defined and well formulated. Formulating a researchable question is a critical step for facilitating good research. In clinical research, there are tools to assist researchers by providing step-by-step guidance on the formulation of a research question. Such tools include PICO 🔗 (Population, Intervention, Control and Outcomes) framework for formulating a research question, and it has been increasingly used in other fields.

> A PICO question must include all four elements and an example in the field of IEG might be: For SMEs (the population), is a business support service through which businesses can receive bespoke advice via telephone and digital channels (the intervention) as helpful in providing advice and signposting (the outcome) as face-to-face support (the control)?

Even if you have clearly stated your study questions, you still need to be careful and avoid overloading the study with too many research questions and too much data collection – this is a rule of thumb that applies to most research studies and not only RCTs. Ideally, trials should have a single primary question around which to focus the development of the study design and sample size estimates (as described in Section 4.2 ↩). There can then be more than one secondary research question, which might be related to the primary question or to other hypotheses.

Once you have decided on and prioritised your research questions, you will be able to move on Step 2: Trial design ↩.

4.2 TRIAL DESIGN

The word 'design' is often used in two ways. It has a narrower interpretation referring more specifically to a trial type (see Section 3 ↰). However, it is also used to refer to all aspects of how a trial is set up. In this guide, we refer to a range of specific aspects of the broader trial design, including details on the trial type, randomisation, outcomes, sample size and analytical methods.

A robust trial design is essential to ensure a successful outcome. It will help you to identify all necessary practical requirements for the trial. This in turn will give you an adequate picture of resources required to carry out the trial.

4.2.1 SPECIFYING THE TRIAL TYPE

As a starting point, it's important to define the trial type as part of your design. As indicated in Section 3 ↰, there are many different types of trials though perhaps the most common type includes a trial where participants are randomised to one of two parallel groups. You should also specify the trial framework – whether it is an exploratory trial or a pragmatic trial, and any other details. This might include whether your trial is designed to identify the superiority of a new intervention, or to determine if an experimental intervention is no worse, or whether an experimental intervention is no better or worse than an existing standard intervention if applicable.

4.2.2 SPECIFYING ELIGIBILITY CRITERIA FOR PARTICIPANTS

You need to consider carefully the population that the intervention you are testing is targeted at, and specify both inclusion and exclusion criteria (if applicable) for study participants. This is important for various reasons:

• Eligibility criteria that are applied before randomisation do not affect the internal validity of a trial, but they are central to its external validity.

• Having a clear idea of who the intervention is meant for helps you to decide how you will recruit a representative sample into your study.

• Clarity around the study population and eligibility criteria allows you to make judgements about a trial's applicability (i.e. to whom the results of a trial apply) and thus relevance to policy and practice.

• It also allows you to ensure that the study meets legal and ethical norms. →

STAGE IN TRIAL

1. PLANNING AND DESIGN STAGE

STEP ONE
RESEARCH QUESTION

STEP TWO
TRIAL DESIGN

STEP THREE
PRE-IMPLEMENTATION PREPARATION

2. IMPLEMENTATION STAGE

STEP FOUR
RECRUITMENT

STEP FIVE
RANDOMISATION

STEP SIX
OUTCOME DATA COLLECTION

3. ANALYSIS AND REPORTING STAGE

STEP SEVEN
ANALYSIS

STEP EIGHT
REPORTING THE RCT

STEP NINE
DATA MANAGEMENT AND STORAGE

→ You also need to consider getting [informed consent](#) from study participants (see Section 4.3.3 ↩ Ethical considerations). Thinking through the method of recruitment allows you to consider how to acquire informed consent as part of the recruitment process.

4.2.3 SPECIFYING THE INTERVENTION

A central element of any trial is the intervention(s) you want to test. Reaching this step of the trial design assumes that you have a clear understanding of the intervention(s) you want to investigate, so you should be able to describe each intervention thoroughly, including control interventions. As best practice, the description should allow a practitioner or a policy-maker who may want to use the intervention in the future to know exactly how to deliver the intervention that was evaluated in the trial. If the control group is to receive business as usual, it is important to describe thoroughly what this constitutes.

4.2.4 SPECIFYING OUTCOMES

Having defined the intervention, you need to specify the outcomes of interest. An outcome – sometimes called 'end point' – is the change or impact caused by the programme being evaluated. Outcomes are driven by the research question; it can often be helpful to develop a logic model, or a theory of change, to spell out how it is thought the intervention will lead to the outcomes, and what assumptions underpin the logic (Please see examples of some helpful resources in the following box).

→

FOR FURTHER INFORMATION ON LOGIC MODEL OR THEORY OF CHANGE DEVELOPMENT, SEE THE FOLLOWING RESOURCES:

Anderson, A. (2007). *The community builder's approach to theory of change: A practical guide to theory development.* 🔗 The Aspen Institute.

W. K. Kellogg Foundation. (2004). *Logic model development guide.* 🔗 W. K. Kellogg Foundation.

It is also important to distinguish between primary and secondary outcomes. Before the programme is implemented, you might expect it to have impact in a number of ways. Nevertheless, it is best practice to concentrate on one primary outcome – although it is not uncommon in social science RCTs to have up to three primary outcomes. By narrowing down the number of primary outcomes, you allow the trial to answer your primary research question unambiguously. These key outcomes will also determine your sample size calculations (see Section 4.2.5 ↩). Choosing primary outcomes does not preclude identifying secondary outcomes, i.e. additional changes caused by the programme that you might be interested in observing.

In order to measure these outcomes, indicators and instruments should be specified. Indicators, or outcome measures, are observable signals used →

→ to measure outcomes. For instance, you might be interested in measuring a firm's level of innovation (outcome), by looking at the number of new products developed in the past 12 months, or the amount of revenue spent on R&D (indicators). Instruments are the tools used to measure the indicators. They can be direct survey questions, tests (e.g. in the case of teaching teenagers entrepreneurship methods), direct observation records (e.g. visiting manufacturing facilities to observe productivity) or administrative data (e.g. firms' turnover records).

4.2.5 SAMPLE SIZES

Another key step in the design of a trial involves estimating how many participants should be recruited into the study to ensure that the research questions can be assessed. A trial with too few participants is called 'underpowered', meaning that it lacks the statistical power (i.e. the ability to detect a certain impact) to answer the research questions. There are a number of techniques to estimate how large a sample size will be sufficient, and they can be carried out with statistical software. (Please see some examples of free tools for sample size calculations in the following box). Wherever possible, it is preferable to involve a statistician for these purposes.

FREE TOOLS FOR SAMPLE SIZE CALCULATIONS:

G*Power ⬀ is a free software developed by researchers at Düsseldorf University which allows simple means comparisons.

3ie ⬀ has developed a calculator for sample sizes and with an accompanying manual. It includes calculations for cluster RCTs.

Optimal Design ⬀ developed by a group of researchers funded by the William T. Grant Foundation.

Sample size calculations should be focused on the primary outcome, i.e. they should ensure that the study will be able to detect changes in the primary outcome. For instance, a study on business support services might have a primary outcome (increase in profits), but also a number of secondary ones (better management, higher employment, etc.); the sample calculations should first focus on ensuring there are enough participants to detect significant changes in profit, and only then worry about the other outcomes. Studies with many questions, and trial designs with many trial arms (e.g. factorial designs), require bigger sample sizes.

Sample size calculations in the design stage are based on a number of assumptions. The most important one regards the expected effect size, i.e. a measure of the difference in outcome between trial arms we anticipate the intervention will cause. More specifically, it is common to consider the →

→ Minimum Detectable Effect Size (MDES), which is the smallest true effect size that has a good chance of being found to be statistically significant. The more participants take part in the study, the smaller effect it will be able to detect.

Often, however, resource constraints will dictate how many participants can be recruited into the trial. For instance, there might only be budget to provide the intervention to 100 SMEs. In these cases, another calculation (known as 'power calculation') is carried out instead to estimate the probability that a trial with the specified number of participants will detect a statistically significant intervention effect of a certain size.

There are a number of things to keep in mind when estimating sample sizes:

Intraclass correlation (ICC): when designing a cluster RCT (see Section 2.4 ↰), ICC should be taken into account. This is because changes in outcomes for participants in the same cluster will be correlated. For instance, imagine a trial assessing the intervention to teach secondary school pupils about entrepreneurship and encourage them to launch their own startups, in which randomisation is done at the class level. The pupils in any given class are more likely to be similar to each other than to those in another class. This is because pupils of similar characteristics are often selected into a class, and the same teachers in the same environment teach them. As a result their outcomes are not independent of each other, but they correlate. ICC measures the degree to which outcomes of participants in a cluster are correlated. This is then used to adjust the sample size calculations (the higher ICC, the more participants are required to detect the same effect).

Choice of significance and power levels: when conducting the sample size calculations, you must choose the minimum levels of statistical significance (the confidence that your result did not arise by chance) and statistical power. These are conventionally 95% and 80%, respectively, although it is up to the investigators to decide which levels to use. It is important, however, to indicate explicitly which significance and power level are used for the sample size calculations.

Allocation ratio: the proportion of participants allocated to treatment or control will be determined by a variety of factors (see Section 2.5.4 ↰ on allocation ratios), and will have an impact on the sample size calculations. In general, allocating an equal amount to each trial arm is the most efficient strategy, i.e. it minimises the number of participants needed. Whichever allocation you decide, this should be taken into account in the sample size calculations (most calculators will allow you to do this).

Intervention compliance: not all participants will take up the trial intervention as planned, as some will not actively participate in the intervention at all or will switch from the intervention group to the control group. However, non-compliance will lead to the loss of power, which is why you should account for this in your sample size calculations.

R-squared: in the context of sample size calculations, r-squared is a measure of how much of the difference in outcomes between participants is explained not by the difference in interventions, but by differences in other characteristics. For instance, one might expect mature SMEs to have better outcomes than younger ones, regardless of treatment, because of their accumulated experience. This fact can be used to achieve higher power in a study – the higher the expected r-squared due to covariates (variables that might be predictive of the outcome other than the intervention itself), the lower the number of participants needed.

4.2.6 RANDOMISATION

The next thing to do is to decide how the randomisation will be conducted. This step is dependent on the type of trial, the intervention, and participants. In particular, it is important to determine whether participants will all be recruited into the programme (and randomised) at the same time, in several batches, or as a trickle sample. This will help to determine what the best type of randomisation is (e.g. simple or blocked randomisation), the method to generate the random allocation sequence (e.g. random-number tables, computerised random-number generator, etc.), and the details of any restriction (e.g. blocking and block size) if needed (see also Section 2 ↰).

When designing the randomisation process it is important to consider what the allocation concealment mechanism will be, i.e. what steps need to be taken to ensure the allocation status of any participant is unknown and not predictable until the randomisation has been completed. Similarly, you should determine who will implement the allocation sequence in order to ensure that the process is carried out correctly.

4.2.7 SPECIFYING METHODS

You should have a clear idea of how you will analyse the data once the intervention has taken place and outcome data has been collected. The methods used to compare groups for primary and secondary outcomes should be clearly specified to avoid accusations of data-dredging or data-fishing, which occurs when researchers search for any statistically significant relationship in a data set. Equally, methods for additional

analyses, such as subgroup analyses, should also be clearly set out. There are a number of methods that can be used, depending on whether outcomes are continuous or categorical.

Ideally, all participants should be included in the final analysis and retained in the group to which they were assigned. For instance, imagine a business assigned to the treatment group that decided not to participate after being randomised. This business should still be included in the final analysis as a 'treated' firm. This is known as 'intention-to-treat' (ITT) analysis, which should always be the main analytical strategy for the trial. The reason for doing this is so that you preserve fully the huge benefit of randomisation, which avoids bias when allocating interventions to participants.

Even though an ITT approach should be the preferred approach due to its robustness, other approaches could be used for secondary analyses. This is particularly as strict ITT analysis is often hard to achieve due to missing outcomes for some participants and non-adherence to the trial protocol (also referred to as 'deviations from the protocol'), for example, if they did not receive the minimum amount of intervention or no intervention at all.

An alternative approach is to only analyse participants that actually received the intervention as intended (and exclude those who didn't adhere to the assigned intervention), known as 'per-protocol analysis' or 'modified intention-to-treat' or 'on-treatment' analysis. This can be carried out as a secondary analysis, and the results should be interpreted as the best case treatment results. However, it should be clearly indicated as a non-randomised, observational comparison, as any exclusion of participants from the analysis compromises the randomisation and →

→ may lead to biased results. If you decide to use per-protocol approach, you should provide a clear description of exactly who will be included in each analysis.

Special consideration should be given to the issue of missing outcome data and how to handle this. As discussed above (see Section 4.2.5 ↩ on Sample sizes), you will often end up without some outcome data for some participants, either because they did not respond to surveys, or because they dropped out of the trial. It's not uncommon to exclude participants without an observed outcome. This may be reasonable, but exclusion of randomised participants means that the analysis is not strictly ITT analysis. Furthermore, you will lose power by reducing the sample size, and bias may be introduced if the missing data is related to a participant's response to intervention. Therefore, you must often choose whether to omit the participants without final outcome data or find ways to deal with their missing outcome data. Unfortunately, there is no methodological approach for handling missing values that is universally accepted in all situations.

Missing data is usually imputed, and there are a number of techniques to do so. The choice of technique depends on the reasons for the missing data. So, as a first step you should set out to explore the reason for the missing data. Depending on the type of 'missingness' – missing completely at random, missing at random or missing not at random– you would need to consider the implications this has for likely differences in the outcomes between those missing compared to those who have observed outcome data. If it is safe to assume that the data is missing completely at random, analysing only those with observed data gives sensible results.

However, often there will be a specific pattern to the missing data – such as more data missing in one intervention arm, or among one type of participants. Imagine, for instance, a trial assessing the effect of different types of accelerators on startup survival, i.e. whether a startup continues to operate after a given amount of time. At the end of the trial, you notice that there is some outcome data missing, especially in one treatment group. In this case you would need to consider imputing the missing observations, as there is probably a reason they are missing – perhaps many of the startups in that trial arm failed because the treatment was not helpful, and they were therefore unable to respond to surveys.

Once you've considered the implications that the type of 'missingness' has for likely differences in the outcomes between those missing compared to those who have observed outcome data, you can decide whether to proceed with your analysis using only those participants with observed data or use multiple imputation (MI). MI involves estimating participants' missing outcomes from other information that has been collected. It is currently considered to be the preferred method for handling missing data. Simpler imputation methods may be are appealing because of their simplicity. However, these methods may introduce bias and should be avoided.

FOR MORE INFORMATION ON MISSING DATA AND MULTIPLE IMPUTATION PLEASE VISIT:

www.missingdata.org.uk 🔗

Furthermore, sometimes additional analyses will be conducted to study the effects of the intervention for specific subgroups (e.g. small firms vs. larger firms, or female vs. male entrepreneurs). These should also have been specified in the design stage. These analyses require larger sample sizes within each subgroup in order to detect significant effects. Unless the subgroup analysis is planned for, i.e. it is decided before the intervention begins and enough participants are recruited into each subgroup, the results should be considered exploratory. This means that even significant results should be treated with caution, as they might have arisen by chance. Such analyses can be used to generate further hypotheses (e.g. the intervention is more effective for smaller firms), but findings require replication studies to ascertain their significance.

4.3 PRE-TRIAL PREPARATION

In addition to specifying the research question(s) and trial design, there are other practical aspects that you need to consider in the planning and design stage. These may or may not happen in parallel to the other steps in this stage, but we have grouped these under a 'pre-trial preparation' step.

4.3.1 TRIAL REGISTRATION

Registering your trial means making all relevant information about a trial (both administrative and scientific) available on a publicly accessible database so that it can be searched for when it is on-going or completed. This is now a common practice not only in medical trials, but increasingly in the social sciences and other fields as well.

WHY SHOULD YOU REGISTER YOUR TRIAL?

Registering a trial and making this information publicly available fulfils a number of purposes and benefits:

• It allows editors, peer reviewers and ultimately readers to access a summary of the planned trial design and to compare it with the paper that later reports the trial results. This is important in order to prevent selective reporting, i.e. the bias that derives from the exclusion of negative or undesirable results.

For instance, if a trial was conducted to estimate the impact of an innovation voucher programme for SMEs, but →

STAGE IN TRIAL

1. PLANNING AND DESIGN STAGE

STEP ONE
RESEARCH QUESTION

STEP TWO
TRIAL DESIGN

STEP THREE
PRE-IMPLEMENTATION PREPARATION

2. IMPLEMENTATION STAGE

STEP FOUR
RECRUITMENT

STEP FIVE
RANDOMISATION

STEP SIX
OUTCOME DATA COLLECTION

3. ANALYSIS AND REPORTING STAGE

STEP SEVEN
ANALYSIS

STEP EIGHT
REPORTING THE RCT

STEP NINE
DATA MANAGEMENT AND STORAGE

→ only the positive results were reported while leaving aside negative results, the evidence presented would be misleading. Comparing the published results to the information presented in the registry would uncover this discrepancy and give a more accurate picture of the real impact of the programme.

• Registering trials helps to prevent publication bias, which arises when trials with negative or null results are not published at all.

Drawing on the example of the innovation voucher programme for SMEs above, imagine that five trials were conducted on the vouchers, but only one – which found a positive impact on SMEs' outcomes – was published. Registering all five trials would make them all available to interested parties regardless of their results. If policy-makers could access only the published results, this would introduce a biased view on what the evidence tells us about the programme effectiveness. Drawing on the example of the innovation voucher programme for SMEs above, imagine that five trials were conducted on the vouchers, but only one – which found a positive impact on SMEs' outcomes – was published. Registering all five trials would make them all available to interested parties regardless of their results. If policy-makers could access only the published results, this would introduce a biased view on what the evidence tells us about the programme effectiveness.

• It helps to promote collaboration and reduce duplication of efforts.

For example, enabling researchers to identify trials they may be interested in could result in more effective collaboration among researchers. This type of collaboration may include prospective meta-analysis, which involves identifying studies that are eligible for meta-analysis before the results of are known. Also,

improving awareness of similar or identical trials will allow researchers and funding agencies to avoid unnecessary duplication and thus wasteful allocation of research funding.

• It allows researchers to conduct systematic reviews and meta-analyses that include all registered studies, and therefore ensures that decisions about policy and practice are informed by all of the available evidence. Systematic reviews and meta-analyses are essential for translating evidence into actionable lessons.

Even though trial registration is not always compulsory, it is strongly recommended. Additionally, a growing number of medical journals insist on trial registration before they'll consider the submission of a paper about a study protocol and/or results. This is likely to expand into other fields. Furthermore, an increasing number of funding agencies and official bodies (e.g. the Education Endowment Foundation in the UK) require trials to be registered.

WHEN SHOULD A TRIAL BE REGISTERED?

Registration should take place as soon as possible once it is agreed that the trial will be undertaken. Ideally, this should occur before any participants are recruited into the study.

WHERE SHOULD A TRIAL BE REGISTERED?

The American Economic Association (AEA) has made available a registry specifically for trials in the social sciences (available here ⊘). RCTs can be registered by creating an online account. Registration is free and you do not need to be a member of the AEA to register. Even though all new trials should be registered at their outset, the AEA registry allows also past studies to be registered given the backlog of existing trials.

4.3.2 PROTOCOL DEVELOPMENT

Trial registries only have space for a limited amount of key information. Therefore, it is strongly recommended that you develop a document called a trial protocol with more details.

WHAT IS A TRIAL PROTOCOL?

A trial protocol is a document that sets out, in detail, the objectives, design and methodology of the trial. While a trial registry will focus on the outcome measures selected before data collection, as well as general information on the design and analysis methods, a trial protocol also describes in detail how the trial will be carried out. This includes the precise system for participant recruitment, randomisation, data management and analysis.

For instance, a trial registry might only specify that a trial will randomise participants into two groups. A trial protocol, however, would describe in detail what method will be used to generate the random allocation sequence (e.g. computer software programme), specify the type of randomisation (e.g. stratified randomisation) and allocation ratio, provide details on the stratification variables (e.g. firm size, geographic location) and why these were chosen, and describe the randomisation implementation (e.g. who generates the random allocation sequence, who enrols participants, and who assigns participants to interventions). This would offer transparency about the trial and thus allow readers to assess the methods used and likelihood of bias, and for other independent researchers to replicate the randomisation process in the future if needed.

WHY IS A PROTOCOL NEEDED?

As stated above, trial registration tends to provide a summary of key aspects of the trial. A trial protocol increases the transparent reporting of research even further and is helpful for the investigators as well as for external researchers.

Writing the trial protocol allows the investigators to better understand the implications of their trial design before implementation starts and any participants are recruited. In particular, developing a protocol allows them to assess whether the trial is feasible, ethical, and if it is safe for participants to take part. Moreover, writing a protocol involves making a number of assumptions about your study explicit, for example, how many participants are expected to take up the intervention, how many may drop out before outcome data collection, etc. Therefore, writing protocols can help investigators plan ahead to ensure the trial is successfully implemented as well as dividing trial-related tasks within the research and implementation team.

A trial protocol is crucial for adequate peer review once the trial has been completed and published. Namely, interpreting trial results and assessing potential bias is difficult without fully understanding how the trial was planned, designed and conducted. While trial registration facilitates that to some extent, it is common for methods to be reported selectively, partly, or not at all and registration cannot fully prevent selective reporting of methods. Equally, pre-specified outcomes are often not reported, while others are added and reported in analyses that may not be valid – known as 'outcome switching'. Developing trial protocols at the design stage of the →

→ trial and making them available when trial results are reported allows us to optimise the evidence base for interventions.

Yet another reason for developing trial protocols is that they help to ensure successful replication of the trial (and thus results) by other (independent) researchers. Replication studies either reproduce the trial in a similar context to check intervention transportability and external validity (external replication) or use the data from the original study to check the validity and robustness of the estimations and recommendations (internal replication).

WHAT SHOULD A PROTOCOL INCLUDE?

There are many resources available to help write a protocol, including a number of protocol templates with pre-defined protocol elements. IGL has developed a template specifically for trials in the field of IEG, which can be found in Appendix A ↩. The items are based on the SPIRIT statement that sets out the essential items for the study conduct, reviewing, reporting, and interpretation.

VIEW THE SPIRIT STATEMENT:
www.spirit-statement.org 🔗

4.3.3 ETHICAL CONSIDERATIONS

Research ethics is a part of research studies that we take for granted today. RCTs, particularly clinical RCTs, have received a lot of attention in regards to ethical principles as the majority of trials include some element of human experimentation.

The extent to which they are ethical depends on how and why they are used. They are certainly considered ethical if they help to answer important questions about healthcare, education, crime, social policy and other domains as long as the rights of all individuals concerned are safeguarded.

Sometimes policy-makers, practitioners or even potential participants may feel that it is unfair to give a new intervention to some people or businesses but not to others. However, it is often the case that we do not know which of possible interventions is best. In fact, we do not even know if any intervention is better than doing nothing at all. This is when we are in a position of equipoise, a state of equal balance. Equipoise is the key ethical requirement as it demonstrates that the researcher is genuinely uncertain which of two options works best, and this provides good justification to proceed with a trial as it can be argued that it is unethical not to attempt to establish which is more effective.

As for potential participants, it is important to explain that an evaluation of a new possible intervention is taking place, not an evaluation of a better intervention. If we already knew that one intervention was better than another, we should be using it already rather than studying it. Also, it is often possible to evaluate interventions without necessarily withholding them from trial participants. For example, sometimes a trial is about trying out different ways →

→ of delivering an intervention to see which way works the best. Furthermore, in most trials all participants get some form of intervention (e.g. business as usual).

All RCTs involving human participants are likely to have an element of risk, which is why these studies should undergo an appropriate ethics review. Such reviews should assess the likelihood and magnitude of risks, considering both the minimal risk of serious harm and moderate risk of minimal harm, as ethical considerations are different in each situation. (For further information on ethical consideration and informed consent please see Appendix B ↰.)

It is the responsibility of the principal investigator to exercise appropriate professional judgment in determining the ethics review required. For example, some trials could be subject to a light-touch ethics review whereas others would require full ethics review. Different research organisations have their own policies and procedures for light-touch, expedited and full reviews as well as procedures in place for submitting proposals to their Research Ethics Committees.

4.3.4 PLANNING FOR TRIAL IMPLEMENTATION

The implementation of a trial tends to require extensive administrative planning before participant recruitment can take place. In addition to registering your trial, developing a protocol and initiating the ethics review process, thorough planning of other aspects of your study is essential for effective trial management.

The planning may contain a range of activities that vary depending on the study characteristics and could include, but are not limited to, the following examples:

• Assessing potential strategies for the optimal way to deliver the intervention and capture the target population.

• Developing standard operating procedures for those involved in the trial implementation, including details on quality control.

• Developing consent form(s).

• Developing a recruitment plan and materials for eligible participants.

• Developing a detailed project timeline and budget for conducting and completing the trial, including preparation of a final study report.

• Identifying collaborators and study site(s), which may include negotiating sub-contracts in some circumstances.

• Developing training materials and training/ certification plans for study staff who will carry out the study.

• Negotiating agreements with industry, as needed, to provide equipment or other resources.

• Designing data collection methods and tools.

• Developing a plan for baseline data collection.
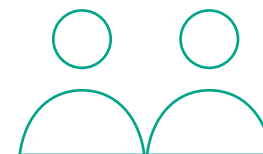
KEY MESSAGES:

▸ The research questions need to be relevant, clearly formulated and prioritised.

▸ As part of the trial design, it is essential to clearly specify which trial type you will use, who the eligible participants for your trial are, what the interventions to be tested are and what the primary and secondary outcomes are.

▸ Guided by the primary outcome(s), you need to ensure that you plan to recruit a big enough sample to detect significant differences between the trial arms.

▸ A detailed randomisation process and a pre-specified analysis plan need to be set out in the design stage.

▸ As part of best practice, it is important to register your trial, develop a detailed trial protocol and give thought to the ethical considerations of conducting the trial.

5

TRIAL IMPLEMENTATION STAGE

This section focuses on the trial implementation, which involves recruiting participants, carrying out the intervention and collecting the data. Before the implementation stage can take place, all the necessary planning steps outlined in Section 4 should be completed. In particular, a trial protocol with all its components should be prepared and the trial should be registered. The implementation will follow the instructions set out in the protocol, while ensuring the details of the intervention are on course. Reading this section will help you to understand some of the challenges that you may face in the implementation stage.

5.1 **RECRUITMENT**

Getting the recruitment and thus participant numbers right is crucial to running a successful trial. Under-recruitment may lead to an underpowered trial, which in turn could affect the RCT's ability to produce reliable findings.

Sometimes you may have access to an existing pool of participants that have already undergone a selection process, for instance a group of firms identified for their high growth potential and chosen to receive an intervention as part of a government business support programme. However, if your study design requires actively recruiting a new sample, it is important that you recruit your participants – whether they are individuals or firms or other units – not just to the intervention but also to the study in general. Getting their 'buy-in' to produce evidence and explaining the wider benefits of the trial will help to increase their commitment to the study irrelevant of their allocation status, i.e. whether they get the intervention or not.

It is important to factor in the time that might be required to recruit participants into a trial, as it tends to be significantly more lengthy and time-consuming than expected. It is also essential to consider how you are going to explain the study to your potential participants – including the fact that the control group

is as valuable to the evaluation as the intervention group. Namely, the better they understand the demands of the study (e.g. participating in the intervention, data collection), the more likely they are to be committed and less likely to drop out.

In some cases, when you are recruiting clusters (e.g. large firms) rather than individual participants, it is worth ensuring that you get a Memorandum of Understanding (MoU) in place. A light-touch contract such as this that explains the terms of the study and outlines everyone's roles and responsibilities can sometimes reduce problems you may encounter when running a trial.

ATTRITION AND RETENTION

Successful recruitment is crucial for minimising attrition, which occurs when participants drop out of the study (or the intervention, although they might continue with the study). Some attrition is unavoidable, but high levels of attrition could bias the results of the study. Therefore, once the participants have been recruited to the study, it's equally as important to keep them in it (i.e. retention).

Again, different strategies may work in keeping participants in the trial depending on the target →

STAGE IN TRIAL

1. PLANNING AND DESIGN STAGE

STEP ONE
RESEARCH QUESTION

STEP TWO
TRIAL DESIGN

STEP THREE
PRE-IMPLEMENTATION PREPARATION

2. IMPLEMENTATION STAGE

STEP FOUR
RECRUITMENT

STEP FIVE
RANDOMISATION

STEP SIX
OUTCOME DATA COLLECTION

3. ANALYSIS AND REPORTING STAGE

STEP SEVEN
ANALYSIS

STEP EIGHT
REPORTING THE RCT

STEP NINE
DATA MANAGEMENT AND STORAGE

→ population, the nature of intervention, and other characteristics. Even though it may sound self-explanatory, it's important to develop a good working relationship with participants in both study groups (particularly if you are dealing with larger clusters, such as large firms, where individual employees are your study participants) and keep them up-to-date on the study's progress if applicable.

OUTCOMES AT BASELINE

If your study includes measuring outcomes at baseline and you therefore need to collect initial data, then you should ensure that this takes place after participants have been recruited into the study but before randomisation has been carried out. This is because if data collection takes place after randomisation, then knowledge of the group allocation may influence participants' responses. Please note that the measurement of outcomes at baseline is not a pre-requisite of an RCT; particularly as sometimes it is not either feasible or desirable.

MONITORING

If your study includes active recruitment of, for example, large firms, then it is important to keep a recruitment and retention log. This allows you to monitor aspects such as:

• How many firms were assessed to be eligible for your trial.

• How many firms you approached, how many of those refused to take part in the trial and their reasons for refusing.

• How many firms signed up for the trial, how many dropped out after signing up and their reasons for dropping out.

• How many firms were allocated to the intervention or control, how many dropped out from the intervention or control and their reasons for dropping out.

• How many firms were followed through to post-intervention outcome data collection, how many did not continue with the trial and their reasons for dropping out.

Keeping track of this information will help you to understand who the results of your study apply to. This information also helps you to populate a participant flow diagram, which is part of standard reporting guidelines in many fields and forms a part of the final trial report (see Section 6 ↰ on analysis and reporting).

5.2 **RANDOMISATION**

Having recruited the participants, the next step is to proceed with randomisation. It is essential that all eligibility checks are completed before randomisation, and no other type of exclusion of participants based on characteristics should take place after randomisation.

Ideally, you should follow the randomisation process as it was set out in the planning stage. However, sometimes there will be some amendments to the randomisation process, due to unforeseen circumstances, such as an obvious numerical imbalance between the participants in the trial arms (which might require minimisation to rebalance the groups). Changes to the randomisation process should be avoided however, and this can usually be achieved with careful planning. If they do arise, any amendments should be recorded and the trial protocol updated accordingly.

STAGE IN TRIAL

1. PLANNING AND DESIGN STAGE

STEP ONE
RESEARCH QUESTION

STEP TWO
TRIAL DESIGN

STEP THREE
PRE-IMPLEMENTATION PREPARATION

2. IMPLEMENTATION STAGE

STEP FOUR
RECRUITMENT

STEP FIVE
RANDOMISATION

STEP SIX
OUTCOME DATA COLLECTION

3. ANALYSIS AND REPORTING STAGE

STEP SEVEN
ANALYSIS

STEP EIGHT
REPORTING THE RCT

STEP NINE
DATA MANAGEMENT AND STORAGE

5.3 OUTCOME DATA COLLECTION

You will have identified your expected outcomes in the trial design stage and ideally will also have decided which measures you are going to use to assess those outcomes(see Section 4.2.4 ↩ on outcomes). The choice of outcome measure is crucial in a trial and will be guided by the research question(s) (for further details, see Section 4.1 ↩ Research question).

If you need to revisit your choice of outcome measure prior to data collection, then you should bear in mind that as a general principle the measure needs to be sensitive enough to detect important effects. Also, the measure needs to be reliable. In other words, it should produce the same findings when participants are measured again under the same conditions. It's also important for the measure to be valid, so that it actually measures the outcome you want to measure. These principles apply regardless of whether you use administrative data (e.g. business records) or standardised assessment tools (e.g. SAT tests) as measurement instruments.

If your trial includes an element of primary data collection and you are planning to use a standardised measurement instrument or another tool specifically designed for the trial, it is important to consider the following:

• All groups (e.g. intervention and control group) should have their outcome data collected at the same time and under the same conditions. This also applies to trickle samples where participants get recruited into a study at different time points and thus receive the intervention at different times. In such cases you would need to ensure that the timing of post-intervention data collection is the same for all participants in all groups.

• Those carrying out data collection should be blinded to the group membership of participants; otherwise they may influence the participants' responses.

• If your data needs to be scored, those doing the scoring should also be blind to the participants' group membership. This will help them to avoid consciously or unconsciously awarding higher scores to one group compared with the other.

It is not uncommon for the researcher or intervention developer to design the measurement tool (e.g. design a survey with a range of indicators to measure your outcomes). Such measurement tools can come with an extra risk however, in that researchers/intervention developers may unintentionally create a tool that favours the experimental group. This is worth taking into a consideration when choosing the outcome measurement tools.

STAGE IN TRIAL

1. PLANNING AND DESIGN STAGE

STEP ONE
RESEARCH QUESTION

STEP TWO
TRIAL DESIGN

STEP THREE
PRE-IMPLEMENTATION PREPARATION

2. IMPLEMENTATION STAGE

STEP FOUR
RECRUITMENT

STEP FIVE
RANDOMISATION

STEP SIX
OUTCOME DATA COLLECTION

3. ANALYSIS AND REPORTING STAGE

STEP SEVEN
ANALYSIS

STEP EIGHT
REPORTING THE RCT

STEP NINE
DATA MANAGEMENT AND STORAGE

KEY MESSAGES:

▸ Successful recruitment of participants is crucial to prevent high levels of attrition, which helps to ensure study validity.

▸ Randomisation should happen after eligibility checks are completed. No further exclusion criteria should be introduced after randomisation.

▸ If the study design involves a baseline outcome measurement, this needs to be carried out before randomisation.

▸ It is important to log recruitment and retention efforts in order to make judgements about to who the study results are applicable.

▸ Primary outcome data collection should be done for all groups at the same time and under the same circumstances.

6

ANALYSIS AND REPORTING STAGE

The final stage, analysing the trial data and reporting the results, may seem self-explanatory. However, RCTs can still be influenced by a number of factors that introduce bias during both the analysis and reporting of the trial. This section sets out the key principles of data analysis and reporting. It also discusses the importance of managing, storing and archiving trial data and making it available to the wider research community. Reading this section will equip you with knowledge on how to best approach analysis and reporting.

6.1 **ANALYSIS**

The first step in this stage is to analyse the data collected as set out in the protocol. This includes carrying out the primary and secondary analysis as originally intended and specified in the trial protocol. As covered in Section 4 ↰ , secondary analyses based on subgroups can be carried out but it should be clear that the results are only exploratory and findings must be confirmed in replication studies.

At this stage, you should also carry out power analysis to estimate MDES that your achieved sample is realistically able to detect, and compare these results to the calculations set out in the pre-trial stage (for further details on MDES, see Section 4.2.5 ↰ on Sample sizes).

STAGE IN TRIAL

1. PLANNING AND DESIGN STAGE

STEP ONE
RESEARCH QUESTION

STEP TWO
TRIAL DESIGN

STEP THREE
PRE-IMPLEMENTATION PREPARATION

2. IMPLEMENTATION STAGE

STEP FOUR
RECRUITMENT

STEP FIVE
RANDOMISATION

STEP SIX
OUTCOME DATA COLLECTION

3. ANALYSIS AND REPORTING STAGE

STEP SEVEN
ANALYSIS

STEP EIGHT
REPORTING THE RCT

STEP NINE
DATA MANAGEMENT AND STORAGE

6.2 **REPORTING THE RCT**

Transparent trial reporting is crucial in allowing others to assess your trial quality. Those interested in trial results often have to rely on what is available in written reports and thus don't have the ability to assess the actual quality of trials, particularly if they have no access to trial protocols.

In an effort to improve reporting standards in medical sciences, a statement called CONSORT (Consolidation of the Standards of Reporting Trials) was first published in 1996 and has been further developed since then to assist the reporting of RCTs.

VIEW CONSORT STATEMENT:
www.consort-statement.org 🔗

The latest version of the CONSORT statement (2010) consists of 25 items and a participant flow diagram. The checklist items focus on reporting how the trial was designed, analysed and interpreted, while the flow diagram shows the progress of all participants through the trial. Reporting according to CONSORT guidelines is made easier by the fact that the SPIRIT checklist designed for protocol development (see Section 4.3.2 ↰ on Protocol development) closely mirrors the CONSORT statement, so a well-developed protocol forms a good basis for trial reporting later on.

The CONSORT statement, its extensions or modified versions are being increasingly used in psychology and education and could also be adapted for other areas of public policy such as crime and justice and social welfare. The modified CONSORT criteria could easily be applied to RCTs in IEG as well in order to increase the transparency of trial reporting.

STAGE IN TRIAL

1. PLANNING AND DESIGN STAGE

STEP ONE
RESEARCH QUESTION

STEP TWO
TRIAL DESIGN

STEP THREE
PRE-IMPLEMENTATION PREPARATION

2. IMPLEMENTATION STAGE

STEP FOUR
RECRUITMENT

STEP FIVE
RANDOMISATION

STEP SIX
OUTCOME DATA COLLECTION

3. ANALYSIS AND REPORTING STAGE

STEP SEVEN
ANALYSIS

STEP EIGHT
REPORTING THE RCT

STEP NINE
DATA MANAGEMENT AND STORAGE

6.3 DATA MANAGEMENT AND STORAGE

It is important to follow the principles of transparency beyond trial reporting into the trial data's management and storage. This is particularly as funders, journals and host organisations are likely to have their own policies and procedures covering storage, sharing and archiving requirements.

As part of good research practice you should give consideration to:

• How to record and securely hold all primary/raw data and related materials that are retained, stored or archived, so that they can be understood and used by others in the future (including Do or Syntax files, which are used to record the logic and process used for analysis and to create any derived variables to support replication of previous analysis).

• How to best back up data that is electronically held and how to keep duplicate copies in a secure and accessible format.

• How to ensure the digital continuity and future accessibility of electronic records and data.

Making the data available to the wider research community will enable it to use the data to:

• Verify the results of the studies.

• Conduct secondary analysis, such as on particular subgroups or types of intervention.

• Link to other datasets for research purposes.

STAGE IN TRIAL

1. PLANNING AND DESIGN STAGE

STEP ONE
RESEARCH QUESTION

STEP TWO
TRIAL DESIGN

STEP THREE
PRE-IMPLEMENTATION PREPARATION

2. IMPLEMENTATION STAGE

STEP FOUR
RECRUITMENT

STEP FIVE
RANDOMISATION

STEP SIX
OUTCOME DATA COLLECTION

3. ANALYSIS AND REPORTING STAGE

STEP SEVEN
ANALYSIS

STEP EIGHT
REPORTING THE RCT

STEP NINE
DATA MANAGEMENT AND STORAGE

KEY MESSAGES:

▶ Primary and secondary analyses should be carried out as planned in the trial design.

▶ Subgroup and other exploratory analyses may be conducted, but the results should be interpreted with caution and all final conclusions left for future replication studies.

▶ The ability to assess the quality of trials depends on the transparency of reporting.

▶ The CONSORT statement could be adapted for the reporting of trials in IEG.

▶ It is important to consider how to best manage, store and archive trial data and make it available to the wider research community.
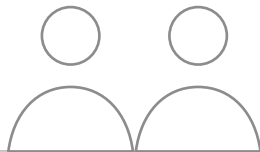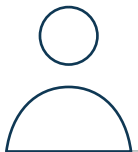
FINAL MUSINGS

Whilst RCTs are not new, and many fields such as healthcare research have capitalised on the development of the methodology and methods of RCTs, relatively few studies using an RCT design have been undertaken in the field of IEG. However, there is an increasing interest in using RCTs in this field. This is particularly the case given the increasing interest in questions such as what works, when, for whom and why, due to resource constraints and political demands for more accountability and transparency.

In this guide we have attempted to introduce the basics of RCTs and set out the steps in getting from the initial research questions to reporting the trial results. The trial design process is particularly important because many trials are rightly criticised for being too small or poorly designed. This is why we encourage you to dedicate sufficient time to designing and planning your trial in order to avoid poor implementation and thus a weak study. This is particularly as interventions in IEG are often very complex and less straightforward to implement when compared to, for example, clinical interventions. The reality is that there is often a trade off between the scope of an intervention and the strength of causal claims emerging from an RCT.

While running trials strictly by the book would be ideal, we acknowledge that there are plenty of pitfalls to watch out for in order to design and implement RCTs that lead to better policies and better outcomes. In fact, many of the implementation challenges sometimes just simply can't be avoided – you may fail to recruit the numbers needed for your trial, struggle to get participants to take up the experimental intervention, face challenges with data collection, or need more complicated analysis to account for a number of limitations. However, these challenges are not unique to RCTs but apply to many other research designs as well. The key is to set out a good implementation plan that minimises the various challenges you are likely to encounter. Equally, we urge you to be as transparent as possible about your study so that informed judgments can be made later on about the quality of evidence.

Considering the breadth of complexity when it comes to interventions in IEG, we encourage you to not overlook the role of theory in developing your RCT. A well-defined theoretical rationale for how an intervention may or may not work will help you to polish up your research questions, choose your primary and secondary trial outcomes and inform your process evaluation. At the same time however, you also shouldn't be too reliant on the theory to guide RCT development –

simply because your theory and assumptions may be wrong. But even if they are, this doesn't always stop you from identifying the correct solution, and you may find that the intervention works even though your thoughts about why or how might have been incorrect.

RCTs are not an everyday commission. They are often perceived as costly in terms of time and money, particularly if we want them to be thorough and rigorous. However, RCTs don't have to be difficult and expensive to carry out. What you really need is a clear rationale for deciding to undertake one to ensure that you don't end up wasting the time and money of funders, researchers and the intended beneficiaries.

We hope that this guide can help provide you with the starting steps for developing your own RCT.

APPENDIX A

IGL TRIAL PROTOCOL TEMPLATE

Version 1

TITLE

Descriptive title identifying the study design, population, intervention, and if applicable, trial acronym.

TRIAL REGISTRATION

Trial identifier and registry name. If not yet registered, name of the intended registry.

PROTOCOL VERSION

Date and version identifier.

ROLES AND RESPONSIBILITIES

Names, affiliations, and roles of trial personnel.

1. INTRODUCTION

Background and rationale: Scientific background and justification for undertaking the trial

Objectives:
Specific objectives or hypothesis[1]

2. METHODS

Trial design:
- Description of trial design (such as parallel, factorial) including the unit of randomisation (e.g. individual or another unit such as startup, SME, class, school), number of trial arms and allocation ratio.

- Description of methods used to generate the allocation sequence including details of any pairing or stratification.

Participants:
Description of who is eligible and how they will be identified and recruited; description of exclusion criteria for participants if applicable.

Interventions:
Details of the interventions for each group with sufficient detail to allow replication.

Outcomes:
Clear definition of primary and secondary outcomes[2], including the specific measurement variable, analysis metric which corresponds to the format of the outcome data that will be used from each trial participant for analysis (e.g. change from baseline, final value, time to event), method of aggregation which refers to the summary measure format for each study group (e.g. mean, proportion with score > 2), and time point of interest for analysis for each outcome. If some of your outcomes will be constructed, e.g. "women empowerment", please provide a description of how the outcome will be constructed from the main variables. $\rightarrow$

---

[1] Please ensure that all study objectives that you set out to meet are included whether it's to assess the effectiveness of the intervention in improving outcomes (direct effects), identify on whom and under what circumstances intervention has different effects (moderation) or to explore why and how intervention has effects (mediation).

[2] Please note that reporting more than one outcome in trials in the field of innovation, entrepreneurship and economic growth may be appropriate because a single measure may not sufficiently characterise the effect of an intervention on a broad set of domains. Please also note that there may be a need for the use of corrections for multiplicity in the analysis.

→ Sample size:

Description of estimated number of participants needed to achieve study objectives and how sample size is determined, including assumptions supporting any sample size calculations alongside the minimum detectable effect size for main outcomes. (Please see Table 1 for examples of assumptions to consider.)

TABLE 1. POTENTIAL ASSUMPTIONS RELEVANT TO SAMPLE SIZE CALCULATIONS

ASSUMPTIONS TO CONSIDER	VALUE
Criterion for statistical significance (probability level; typically 0.05)	
Power against alternative hypothesis (conventionally 80%)	
Proportion of randomisation units assigned to treatment (e.g. 50% of the total sample assigned to treatment in a two-arm trial)	
Treatment compliance	
Number of individuals per randomisation unit (applicable to cluster randomised trials)	
Intraclass correlation (ICC) coefficient (rho) (applicable to cluster randomised trials)	
Proportion of variance in the outcome explained by covariates (r-squared)	
Whether the test is 2-tailed or 1-tailed (if applicable)	

3. DATA COLLECTION AND ANALYSIS

Data collection methods:

- Plans for assessment and collection of baseline, outcome and other trial data (incl. how and when).

- Description of data collection instruments (e.g. questionnaire, test, scale, rating, or tool) along with their reliability and validity, if known.

Analysis plan:

- Description of the statistical methods to be used to compare the groups on the primary and secondary outcome measures.

- Description of methods for any additional analysis (e.g. subgroup and adjusted analyses or mediation analysis).[3]

- (Reference to where other details of the statistical analysis plan can be found, if not in the protocol.)

4. PROCESS EVALUATION (IF APPLICABLE)

- Description of methods used in the data collection (incl. why, how and when).[4]

- Description of methods used in the data analysis (both quantitative and qualitative if applicable).

5. ETHICS

- Description of the process for ethical approval.

- Description of the level of consent from participants (if applicable).

6. RISKS

- Description of risks to the trial and how they might be addressed (see Table 2 on the next page).

---

[3] Please ensure that the analysis plan addresses all research objectives set out in the 'Objectives' sections above.

[4] Process evaluation can be crucial for understanding the effects and exploring potential causal mechanisms of complex interventions or for assessing programme fidelity.

TABLE 2. TRIAL RISK REGISTER WITH EXAMPLES

RISK	ASSESSMENT	COUNTERMEASURES AND CONTINGENCIES
Venture attrition	Likelihood: moderate Impact: moderate	Clear information / initial meeting with the Providers explaining the principles of the trial and expectations. Both intention-to-treat and per-protocol analysis will be used. Attrition will be monitored and reported according to CONSORT guidelines.
Interventions are not implemented well	Likelihood: low Impact: moderate	Clear information / initial meeting with the Providers explaining the principles of the trial and expectations. Both intention-to-treat and per-protocol analysis will be used. Process evaluation will monitor this.
Failure in recruiting ventures	Likelihood: low Impact: high	Project team will make use of their research operations unit at their organisation to recruit more businesses. Timescale could be revised.
The Provider does not follow correct trial protocols	Likelihood: moderate Impact: high	Meetings with the Providers at start of project. Provision of clear guidance describing protocols for distribution to all Providers.
Etc.		

7. TIMELINE

Description of a timetable (including specification of who completes each task if possible).

APPENDIX B

ETHICAL ISSUES WITH INFORMED CONSENT

WHAT IS INFORMED CONSENT?

Whether people are motivated to participate in an RCT because they want to benefit others and advance scientific knowledge, or whether because of self-interest to, for example, receive free services, they would need to be willing to participate in a trial without the possibility of coercion. Therefore, it is crucial that a potential participant understands why the trial has been proposed, and why they are being asked to participate. This way prospective participants can make an informed and free decision on their possible involvement, otherwise known as 'informed consent'. Thus, the primary purpose of informed consent is to safeguard the potential participant.

The most ethical approach to informed consent is considered to be the opt-in approach, where the participant signs a consent form to take part in the study. Sometimes, however, an opt-out approach is used where the assumption is that participants would be willing to take part unless they indicate otherwise by signing a consent form to not be in the study. Even though the 'opting out' approach can result in a higher recruitment rate and may generate a more representative sample of the population of interest, it should be only used when the risk is very low (e.g. if the intervention was providing participants with some reading material on a specific topic).

WHAT ABOUT CLUSTER RCTS AND CONSENT?

Even though cluster RCTs can have several advantages when compared to individual level trials (see Section 2.4 ↰), discussions in the literature surrounding the ethical implications of carrying out cluster RCTs is relatively thin on the ground. Nevertheless, the questions such as from whom, when, and how informed consent must be obtained in cluster RCTs would need to be considered before embarking upon a trial.

Cluster RCTs tend to involve two levels of consent: for the involvement of the group and the individual. Commonly, head teachers (when a school is a cluster) or general practitioners (when a GP surgery is a cluster) or community leaders (when a village is a cluster) act as cluster guardians who consent to participation. Group consent is not always considered to be a substitute for individual consent, which should follow similar lines to individual level trials.

However, there can be different reasons for not seeking individual consent. For example, there may be logistical reasons when an intervention is delivered to a defined community or large geographic area. In such cases only a cluster guardian's consent is sought for participation and randomisation.

There may also be cases where the goal of full disclosure may undermine the objective of a trial irrelevant of the level of randomisation. This applies particularly to behavioural interventions when participants cannot →

$\rightarrow$ necessarily be blinded. Their knowledge of the treatment comparison can introduce bias through differential changes in the behaviour or attitudes of the intervention groups. This is common, for example, in trials where the intervention under study is educational. In such trials the unit of randomisation is the cluster to avoid contamination but the intervention is implemented at the level of the individual.

Knowledge of the intervention group amongst control group members can result in resentful demoralisation, where those in the control group become resentful of not receiving the intervention that the experimental group receives. This in turn can lead to other biases such as performance or dilution bias. The former refers to the tendency for participants to change their behaviour or responses to questions because they are aware of being in a trial and of the treatment allocation. The latter refers to circumstances where those in the control group seek out some form of alternative intervention (which may dilute the effects of the experimental intervention under study).

Sometimes, in situations like this, randomisation is carried out prior to seeking consent and consent is usually sought only for the treatment to which an individual is allocated – that is, without revealing the comparison condition. Such an approach, however, comes with its own limitations (see Section 3.4.4 ↰ on Zelen's design).

Also, there are circumstances where a study proceeds when informed consent is not possible. In such circumstances Research Ethics Committee (see also Section 4.3.3 ↰) may grant a waiver of consent when it is not feasible to obtain consent and study participation poses only minimal risk.

Furthermore, there may be circumstances where there may be no obligation to obtain participants' informed consent and secure a waiver for consent for the trial. Namely, when participants are not directly intervened upon by an investigator, they are not deliberately intervened upon via manipulation of their environment and there is no interaction with participants for the purposes of collecting data, the participants would not need to be considered as human research subjects and thus it could be argued that there is no obligation to obtain their informed consent.

GLOSSARY

ALLOCATION CONCEALMENT: A process used to prevent selection bias by concealing advanced knowledge of which intervention group participants have been assigned to in an RCT, until the moment of assignment. Allocation concealment prevents researchers from (unconsciously or otherwise) influencing which participants are assigned to a given intervention group.

ALLOCATION SEQUENCE: A list of trial groups that is randomly ordered and used to assign sequentially enrolled participants to a group. Also termed the 'assignment schedule', 'randomisation schedule', or 'randomisation list'.

ALLOCATION RATIO: The ratio of the number of participants in each of the comparison groups. For two-group trials, the allocation ratio is usually 1:1, but unequal allocation (such as 1:2) is sometimes used.

BASELINE: The set of measurements at the beginning of a study, with which subsequent results are often compared.

BLINDING: The practice of keeping the trial participants, intervention providers, data collectors and sometimes even those analysing data unaware of which intervention is being administered to which participant. Blinding is intended to prevent bias on the part of the study personnel. The term 'masking' may be used instead of blinding.

CAUSAL DESCRIPTION: Establishing if evidence resulting from an experiment shows that the outcomes have been caused by the intervention. In other words, it is a demonstration of whether an intervention, in its entirety (i.e. taken as a whole package with multiple components), produces a particular outcome with a particular group of interest.

COMPREHENSIVE COHORT DESIGN: A study design whereby participants who do not consent to be randomised, or who cannot be randomised, are followed up alongside the randomised groups.

CONFOUNDING FACTORS: Factors that are associated with cause and predict the outcome of interest, thus obscuring the effects of another variable on the outcome. Namely, any time we observe a relationship among two variables such as cause (i.e. an intervention) and an outcome, there's always the possibility that some third variable which we don't know about is responsible for (confounding) the relationship.

CONTAMINATION: When two elements in a trial that should be kept separate, such as the intervention and control groups, come into contact and interact, potentially affecting outcomes. For example, if startups in the same accelerator randomised into two groups came into contact and those in the intervention group shared information they've been provided as part of the intervention with those in the control group.

CONTROL GROUP: A group of people in a study who do not receive the intervention being studied. Instead, they may receive the standard intervention (sometimes called 'business as usual') or a dummy intervention (e.g. a placebo; mainly used in healthcare research). The results from the control group are compared with those from a group that has received the intervention being tested. The purpose is to check for any differences between the two to see if they can be attributed to the intervention.

COST-BENEFIT ANALYSIS: A technique used to compare the total costs of an intervention with its benefits, using a common metric (most often monetary units) to see whether the benefits exceed the costs. This enables the calculation of the net cost or benefit associated with the programme. If the costs are lower than the benefits, then the intervention would normally be adopted.

COST-EFFECTIVENESS: The relationship between monetary inputs and the desired outcome. Cost-effectiveness is established by a cost-effectiveness analysis.

COST-EFFECTIVENESS ANALYSIS: A technique that compares the relative costs to the outcomes (effects) of two or more interventions. When comparing two mutually exclusive alternatives, the intervention with the lowest cost-effectiveness ratio would normally be adopted.

COUNTERFACTUAL: A conditional statement of what 'would have' happened had something occurred that did not in fact occur. Counterfactuals have a central role in theorising about cause. In a trial, the control group performance is used to estimate what would have happened in the intervention group had they not received the intervention.

COVARIATE: A variable other than the independent or predictor variable that correlates with the dependent variable (also known as 'outcome variable'). Typically, researchers seek to control for the covariate by using techniques that statistically subtract the effects of the covariate.

CROSS-OVER TRIAL: A type of trial in which subjects receive different interventions at different times. Interventions are allocated randomly.

DATA-DREDGING OR DATA-FISHING: This occurs when a researcher hunts in a dataset for any statistically significant relationship. For example, a researcher may run many statistical tests when he or she has no theory or hypothesis. Reporting the 'significant relationships' that were found due to data-dredging is considered to be deceptive and unethical if the lack of theory or hypothesis has been concealed from the reader of the research.

DEPENDENT VARIABLE: A variable where its values are predicted or explained by the independent variable, whether or not it is caused by it. Also called 'outcome variable', particularly in trials where the focus is on testing whether an intervention (independent variable) causes changes in the outcomes.

GLOSSARY

DESCRIPTIVE CAUSATION: See causal description.

DILUTION BIAS: A bias that occurs when participants in a group receive one of the other/comparator interventions after randomisation. For example, a proportion of the control group participants may receive the experimental intervention, and a large number of intervention group participants may fail to receive the intervention, thus tending to 'dilute' any observable treatment effects.

DIFFERENCE-IN-DIFFERENCES: An estimation that assesses the likely impact of an intervention by comparing the pre- and post-intervention differences in the outcome of those who received it with those who did not receive it. The simplest difference-in-differences model measures two groups twice - before and after. The difference in the before-after difference is the estimation of the effect. This is a quasi-experimental alternative when randomisation is not possible.

EFFECT SIZE: The observed association between interventions and outcomes, or a statistic to summarise the strength of the observed association. In a trial, the difference between two groups (effect size) is described in standard deviation units (i.e. difference divided by the standard deviation).

EFFECTIVENESS TRIALS: See pragmatic trials.

EFFICACY TRIALS: See explanatory trials.

ELIGIBILITY CRITERIA: The characteristics (e.g. demographic or condition related) that define those participants eligible to be enrolled in the trial.

EQUIPOISE: A situation in which researchers do not know which intervention will work better (e.g. a new intervention versus a control). In this situation, it is ethical to conduct an RCT.

EXCLUSION CRITERIA: Exclusion criteria is a list of characteristics that prevent research participants from being eligible to take part in a trial.

EXPERIMENTAL INTERVENTION: A new intervention that is being studied to see whether it has an effect on the outcome.

EXPLANATORY TRIALS: Trials in which as homogeneous a population as possible is recruited to determine how and why an intervention produces an effect under ideal or controlled conditions.

EXTERNAL PILOT STUDY: A type of study that precedes a definite trial, but where the data from the pilot gets analysed alongside the results of the main trial.

EXTERNAL VALIDITY: The extent to which the findings of a study apply beyond that study. In other words, the extent to which the findings are generalisable or applicable to other circumstances.

FACTORIAL DESIGN: Trials that use a design where two or more different interventions are evaluated using the same participant sample. The main advantage is that it allows two trials to be undertaken for the price of one.

FIDELITY: The degree to which an intervention is implemented exactly as specified and is consistent with procedures in the intervention manual.

INCLUSION CRITERIA: Inclusion criteria is a list of characteristics (e.g. factors or reasons) that research participants must have in order to be eligible to take part in a trial.

INDEPENDENT VARIABLE: The presumed cause in a study. In other words, a variable manipulated by a researcher (i.e. an intervention) who predicts that the manipulation will have an effect on another variable such as dependent or outcome variable.

INFORMED CONSENT: A process by which a subject voluntarily confirms his or her willingness to participate in a particular trial, after having been informed of all aspects of the trial that are relevant to the participant's decision to participate.

INTENTION-TO-TREAT ANALYSIS (ITT): An analysis in which participants are included in the group to which they were randomised irrespective of whether or not they dropped out of the trial, fully complied with an intervention, switched to an alternative intervention or were subject to administrative errors (e.g. an error in eligibility). ITT analysis is the preferred type of analysis as it maintains the benefits of randomisation and mirrors real-life practice, where not everyone will adhere to their allocated intervention.

INTERACTION: The situation in which the effect of one explanatory variable on the outcome is affected by the value of a second explanatory variable. In a trial, a test of interaction examines whether the treatment effect varies across subgroups of participants.

INTERACTION EFFECTS: An effect that occurs when the relation between two variables differs depending on the value of another variable.

INTERNAL VALIDITY: The extent to which the results of a study can be attributed to the intervention(s) rather than to flaws in the research design. In other words, the extent to which you are able to say that no other variables except the one you're studying (i.e. the intervention) caused the result.

INTERNAL PILOT STUDY: A type of study that precedes a full trial. Data from the pilot study may be analysed but not used in the main study results. Intervention: Programme, policy, project or any other course of action that aims to change outcomes for participants.

GLOSSARY

INTERVENTION GROUP: The group in a trial that receives the intervention being tested. Also called 'treatment group' or 'experimental group'.

INTRACLASS CORRELATION: A measure of how similar members of a cluster (e.g. pupils within a class) are. The degree of within-cluster correlation plays an important role in estimating sample sizes for cluster randomised trials as we can't assume that outcomes for participants within a cluster have no relationship with the outcomes of others in the same cluster.

INSTRUMENTAL VARIABLES: A method of estimation that is an important way of compensating for the lack of random assignment to experimental and control groups in quasi-experimental designs. An instrumental variables approach uses an instrument (a type of variable) to isolate exogenous variation in the receipt of an intervention. An instrument needs to be related to the intervention receipt and affect results only through intervention receipt and should not be correlated with any other determinants of results. In practice it is often difficult to find a convincing instrument.

MASKING: See blinding.

MEASUREMENT ERROR: The difference between what the value obtained by a measurement indicates (e.g. a student's test score) and the actual value (e.g. a student's knowledge and abilities) that appears due to flaws in a measuring instrument, due to mistakes of those using it, or simply due to random or chance factors. Since perfect precision is impossible, measurement error is inevitable. It can be an important source of bias.

META-ANALYSIS: A method often used in systematic reviews to combine results from several studies of the same intervention to estimate the overall effect of the intervention.

MINIMUM DETECTABLE EFFECT SIZE (MDES): The smallest true effect (in standard deviations of the outcome) that is detectable for a given level of power, statistical significance and number of participants. The more participants that take part in a study, the smaller effect it will be able to detect. See also effect size.

MODIFIED INTENTION-TO-TREAT: See per-protocol analysis.

ON-TREATMENT ANALYSIS: See per-protocol analysis.

OPT-IN/OPT-OUT: Recruitment processes used by researchers to recruit the study samples. 'Opt-in' samples are those where participants are asked to actively to 'opt-in' to a study – to volunteer to take part. 'Opt-out' samples are those where participants are contacted without volunteering to take part in the research and excluded only when they say they are unwilling to participate. The latter approach is seen as problematic by many ethics committees because it undermines the principle that consent should be freely given.

OUTCOMES: The impact that a test, treatment, policy, programme or other intervention has on a person, group or population.

PARALLEL TRIALS: A type of trial design in which two or more groups of participants receive different interventions. For example, a two-arm parallel design involves two groups of participants. One group receives intervention A, and the other group receives intervention B. During the trial, participants in one group receive intervention A 'in parallel' with participants in the other group, who receive intervention B.

PARTICIPANTS: A term used to refer to a research subject or respondent that has been selected to take part in a trial.

PER-PROTOCOL ANALYSIS: An analysis in which participants are included in the group corresponding to the treatment they actually received, as opposed to the one they were originally allocated to. Thus, participant compliance and 'switchovers' to alternative interventions are considered in the analysis. Typically, in a per protocol analysis, participants who do not meet all of the eligibility criteria or did not complete the intervention they were originally allocated to are excluded. If done alone, this analysis leads to bias.

PERFORMANCE BIAS: A bias that occurs when participants' response to the intervention is affected by knowledge of the group to which they are assigned, or when practitioners deliver interventions differently between treatment arms.

PRAGMATIC TRIALS: Trials that attempt to establish effectiveness in actual routine practice, rather than under strictly controlled trial conditions. In other words, trials that focus on determining how and why an intervention produces an effect under real-world conditions.

PROPENSITY SCORES: An estimated probability of being treated given all of the background (covariate) information about intervention selection. In other words, it is the propensity or inclination to participate in a programme. It is used in propensity score matching, when units that were treated are matched to untreated units that are comparable in all relevant covariates. Propensity scores are typically used when randomisation is not possible.

PROSPECTIVE STUDIES: Research studies in which participants are followed for a period of time to see what happens to them. This contrasts with retrospective studies.

PROTOCOL: A plan or set of steps that defines how research will be carried out. Before conducting a research study, for example, the research protocol sets out what question is to be answered and how information will be collected and analysed.

GLOSSARY

PUBLICATION BIAS: The bias that arises from the fact that some studies are more likely to get published than others and thus be more available for, for example, systematic reviews and meta-analyses. The most common source of publication bias is the tendency for studies showing statistically significant results to be published more frequently than those that do not have statistically significant findings.

QUASI-EXPERIMENTAL DESIGNS (QEDS): Research designs for measuring intervention effects, when subjects cannot be randomly assigned to control and experimental groups, but the intervention that participants in different groups receive can still be manipulated.

RANDOMISED CONTROLLED TRIAL (RCT): An experiment in which participants are randomly assigned to intervention and control groups in order to equate the groups on all known and unknown variables. In most trials one intervention is assigned to each individual, but sometimes assignment is to defined groups of individuals (for example, businesses or geographic regions).

R-SQUARED (ALSO R²): A proportion of unexplained variation in the outcome within experimental groups explained or predicted by covariates.

REGRESSION DISCONTINUITY: A research design whereby participants are not allocated in experimental and control groups by random assignment but rather according to a cutoff score (e.g. score on a test). It is used as a quasi-experimental alternative to an RCT when randomisation is not possible.

RESENTFUL DEMORALISATION: A threat to internal validity that may occur in trials in which comparison groups not receiving a desirable treatment become discouraged or retaliatory and, as a result, perform worse on the outcome measures. This is likely to exaggerate differences in outcomes between groups, making the intervention look more effective than it actually is.

RETROSPECTIVE STUDIES: Studies that use information from the past to draw conclusions.

SELECTION BIAS: Systematic error that arises when the researcher is unable to randomly assign participants to intervention and control groups. If selection bias occurs, the groups are likely to differ in measured or unmeasured baseline characteristics because of the way participants were selected or assigned.

SELECTIVE REPORTING: Selecting methods and outcomes to include in the publication of a trial based on results. The concern is that trial design elements and statistically non-significant results might be selectively withheld from publication.

SPILL-OVER EFFECT: See contamination.

STATISTICAL POWER: The probability that a trial will detect, as statistically significant, an intervention effect of a specified size given the particular sample size of a study. The prespecified trial size is often chosen to give the trial the desired power.

STATISTICAL SIGNIFICANCE: The likelihood that any differences observed between groups being studied are 'real' or simply due to chance. If it is unlikely enough that the difference in outcomes occurred by chance alone, the difference is determined to be 'statistically significant'. It is important to note that statistical significance does not always imply substantive or practical significance.

SYSTEMATIC REVIEWS: A type of research that reviews studies answering a defined research question by collecting and summarising all empirical evidence that fits pre-specified eligibility criteria. Systematic reviews are characterised by clearly defined and replicable procedures for selecting studies to be reviewed and for drawing conclusions from them.

THEORY OF CHANGE: A theory of change explains how activities are understood to produce a series of results that contribute to achieving the final intended impacts. It can be developed for any level of intervention – an event, a project, a programme, a policy, a strategy or an organisation.

TREATMENT EFFECT: See effect size.

TRIAL ARM: A subsection of people within a study who receive a particular intervention (for example, the 'business as usual' arm and the 'experimental' arm).

WENNBERG'S DESIGN: A trial design where eligible participants are randomised to a 'preference group' or an 'RCT group'. Those in the preference group are given the opportunity to receive the intervention that they choose, whereas those in the RCT group are allocated randomly to receive any of the study interventions, regardless of their preference.

ZELEN'S DESIGN: A trial design where participants are randomised before consent to take part in the study is obtained. In this guide we present the single consent method is where consent is only sought from those allocated to the intervention group.

RUNNING RANDOMISED CONTROLLED TRIALS IN INNOVATION, ENTREPRENEURSHIP AND GROWTH:

AN INTRODUCTORY GUIDE

IGL Innovation **Growth** Lab