Deliverable 3.6 Learning Report

Version 1.0

(External Version)

[Written by James Phipps, Hugo Cuello Diaz & Stella Ishack under service contract n° EASME/H2020/2018/005 with EASME]
[*May – 2020*]

EUROPEAN COMMISSION

Executive Agency for Small and Medium-sized Enterprises (EASME)
Unit A.1.5 - COSME

E-mail: easme-innosup@ec.europa.eu

European Commission
B-1049 Brussels

Disclaimer

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of EASME or of the Commission. Neither EASME, nor the Commission can guarantee the accuracy of the data included in this study. Neither EASME, nor the Commission or any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Table of Contents

Introduction

As part of Horizon 2020, the European Commission launched a Call for Proposals (INNOSUP-06-2018), aimed at innovation agencies, to design new or improved policy schemes supporting SME innovation, and test these using randomised controlled trials (RCTs). In the autumn of 2018, the Innovation Growth Lab (IGL), based at Nesta, was selected by EASME to deliver support to these projects, with many beneficiaries set to undertake their first experiment. This learning report has been produced as part of IGL's support for the INNOSUP-06-2018 programme.

By making the initial learnings from this programme available to those who design and run experiments, this report seeks to outline the initial set of challenges and lessons encountered when setting up trials of this nature and how these could be addressed, sharing what does and doesn't work for future innovation policy.

The novelty of the approach in this area means that project teams are treading new ground. Designing a trial is a complex undertaking and adjustments often have to be made as plans develop. Though challenges are an inevitable part of running trials, we also provide suggestions for how future programmes supporting innovation policy experimentation could be designed to help avoid such issues or lessen their impact on the feasibility and insights from the chosen experiments.

Challenges

Challenge 1 - Identifying the right question to yield the right answer

Randomised controlled trials (RCTs) are very good at answering specific impact questions, such as, "does offering intervention 'x' to a sample of SMEs cause a change in outcome 'y'?". We found that many projects were motivated by policy questions that were too broad or complex for a trial. For example, many of the projects focused on descriptive questions (e.g. what challenges do SMEs face when adopting technology?) or strategic questions (e.g. how do we best provide advice to SMEs?) Trials can help explore these topics but only when combined with other research approaches.

Establishing a clear research question is vital for shaping trial design and ensuring all stakeholders understand what they are set to learn from a trial.

We encouraged project teams to use the PICO (Population, Intervention, Control and Outcomes) framework[1] to help refine their research questions. These will then provide the basis for the hypothesis that they would test through their analysis.

Research questions may need to be adjusted as trial designs are further developed. For instance, a lack of statistical power and limited resources to boost sample sizes can mean that a trial has to focus on testing the overall impact of a single intervention rather than comparing the relative benefits of different approaches.

Challenge 2 - Moving from broad outcome measure to specific indicators and data

At first, outcome measures (i.e. the change or impact that projects were seeking to evaluate) are typically described in general terms (or 'topic' area). In many cases it proves a challenge to

[1] Further discussed in IGL's Introductory Guide to RCTs.

determine the actual indicators that will be used to track the outcomes and provide data for the analysis.

For instance, we have worked with projects to progress from a broad initial expectation that the intervention would increase 'levels of innovation amongst SMEs', to the specific nature of that innovation activity and the survey questions that would allow them to extract the information from participating SMEs. The table below shows the steps that need to be considered as progress from a general topic area of policy interest through to the data that will be used in analysis.

Topic	The general policy area that you are exploring?
Outcomes	What is the key outcome that you are looking to see change? Are there any changes that need to occur before this in the logic model, what would come after it?
Indicator(s)	Do you know what indicators you can use to measure changes in these outcomes?
Instrument	How are you planning to collect information for the indicators from SMEs?
Data	What will this data look like and will you need to undertake any changes for the analysis? - e.g. combine several indicators into a single variable

This is not straightforward as trials require indicators that are:

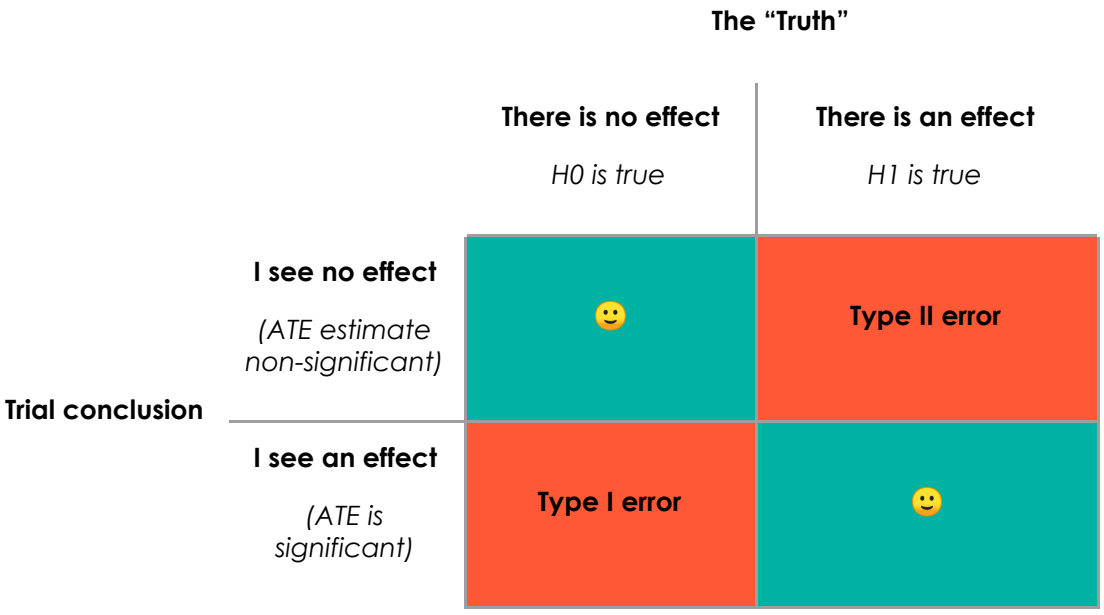
- **Valid** - They need to actually measure the selected outcome.
- **Sensitive to change** - They should be something that the intervention can impact and generate detectable changes.
- **Reliable** - They must produce the same findings if participants are measured again in similar conditions.
- **Clear and unambiguous** - i.e. all SMEs who experience benefits from the intervention will see changes in the same direction.
- **Unbiased/Independent of allocation** - You want to make sure that measurement is only affected by true changes in the outcomes you are seeking to measure.
- **Policy relevant** - A finding from the evaluation that the intervention did (or did not) lead to changes in the chosen outcome would influence policy decisions.

Ideally projects would be able to adopt tried and tested approaches to measuring impacts. However, it is often not possible to identify proven indicators for the chosen outcome measure or context. In these cases we encourage at least piloting the measures first.

We believe there is a more general need for a review and collation of potential indicators that could be used within trials evaluating innovation policies. This is something that we intend to pick up through wider activity at IGL.

Challenge 3 - Ensuring sufficient statistical power

It is important that a trial is designed with sufficient statistical power (a high probability that a trial will yield a significant, meaningful result) so that if the intervention has a true impact, we can be confident that we would detect it. This guards against a Type 2 error - when the intervention has a true Average Treatment Effect (ATE) but it cannot be detected in the trial and so we falsely conclude that there is no impact.



In our experience statistical power is often overestimated or not adequately considered. This can have very consequential implications for the project. Overestimation can stem from not considering factors such as:

- The level of compliance - i.e. will businesses assigned to the intervention use it as intended? Will those assigned to the control stay there?
- Attrition - i.e. will participants remain in the trial and for how many will the required data be collected?
- Lack of sensitivity in outcome measures - e.g. a business that has already adopted a technology by baseline cannot do so again.

We have also found that once underway, trials are very likely to encounter further issues with recruitment, compliance and survey response rates. Therefore, anticipating how to potentially respond if these are to be lower than expected is also crucial.

We discussed a range of measures that project teams could take to maximise statistical power and make sure they retain it. These strategies include:

- Increasing the sample size;
- Ensuring high take-up of tested interventions and survey response rates;
- Reducing 'noise' - Use stratified randomisation; measure outcomes more than once; collect pre-treatment values of outcome variables or other relevant variables that are correlated with it;
- Reduce number of comparisons - limiting treatment arms and subgroup analysis;
- Increase detectable impacts - increase the sensitivity of chosen outcome measures; make the treatment and control conditions very different or ensure high fidelity of implementation.

Sometimes, the statistical power required to answer the intended research question is not achievable. These experiments can still be very valuable with the evaluations enabling teams to:

- Assess the potential feasibility and effectiveness of a novel intervention;
- Test and further develop their theory of change;
- Gauge reception from the target innovators and observe take-up & compliance;
- Gather information that will enable them to design a future trial;
- Improve monitoring & evaluation processes and data systems.

On the other hand, for projects that have the potential power to draw robust conclusions, they should refine their power calculations, maximise their statistical power and include strategies to make sure they can retain it. Some of these strategies include increasing the sample size, ensuring high take-up, reducing noise and the number of comparisons, as well as to increase detectable impacts.

Challenge 4 - Identifying the best method to implement randomisation

Randomisation is the cornerstone of an RCT. The theory is simple but in practice there are many factors to consider to ensure that it is implemented in a way that will provide unbiased results.

Where possible, randomisation should take place after both eligibility checks and the collection of baseline data - otherwise a significant amount of statistical power through participants who cannot use the interventions and cannot be used in the data analysis.

After that, there are additional aspects of the randomisation process to consider, in particular the method of randomisation and how to generate the random allocation sequence.

There are many ways to generate random allocation sequences - see IGL 's guide to randomised controlled trials for discussion of the main approaches. Which is most appropriate depends on several factors such as the size of the population; number of trial arms; the type of participants and how they are to be recruited and assigned. For instance, stratification can help avoid imbalances between groups but is much harder when the sample is recruited and randomised in batches than when the whole sample is to be recruited and randomised in one go.

Determining the correct approach can be complex, especially if projects lack existing data that could have helped inform decisions - e.g. to identify factors that are important predictors of outcomes and suitable for use in stratification.

Challenge 5 - Knowing how long it will take to observe impacts

In many cases projects may have little existing evidence to determine what time period they should leave between the delivery of the intervention and measuring outcomes. In theory, taking repeated measures would be useful but this has budgetary implications and in practice, benefits could be lost through survey fatigue.

Some projects may also leave insufficient time in their project plans for all outcome data to be collected and analysed, which is important for yielding robust results. In those cases, projects may last longer than originally expected.

Challenge 6 - Keeping delivery consistent

When there are several partners involved in the design and implementation of the trial it is important to ensure consistency in the approach.

Divergence in how participants are recruited, the delivery of the intervention and measurement of outcomes can make it much harder to detect impacts and to understand what has or hasn't worked.

We've observed that sometimes heterogeneity across time and place can reduce the effectiveness of the intervention, or make the results more noisy. It can also make the results more difficult to interpret, especially if they were null - did all of the approaches not work, just some or all worked but in different ways?

It is therefore crucial to standardise the delivery across locations. For instance, the lead partner could plan a workshop with all delivery partners as well as regular follow ups with them to ensure they follow the same approach to recruitment, intervention delivery and data collection. This sort of consistency amongst those delivering the project is a crucial part of yielding useful, balanced results.

Challenge 7 - Alignment with wider activities and programmes

Field experiments are not being run in laboratories under controlled conditions. This creates challenges with how the intervention aligns with other activities conducted by the innovation agencies or in the areas covered by their trial.

Some projects may have to consider how their trial will interact with other available public programmes for similar target groups in their location. On the one hand, there are arguments to avoid areas where other support is available, due to concerns such as the control group receiving other subsidised support, which could affect one's ability to draw conclusions about the additionality and impact of subsidised interventions. On the other hand, there might be sufficient differences between target populations and objectives, meaning this contamination won't take place and the different organisations can combine recruitment activities and help ensure they meet recruitment targets.

Challenge 8 - Recruiting the right type and number of SMEs

Recruitment is normally one of the toughest challenges, and projects need to consider how they might respond if recruitment numbers are different from what they expected. For example, they may need to consider whether they would need to reduce the number of comparisons that they intend to make.

As well as achieving targets for the number of participants, it is also important that projects recruit the right type of participant. For example, if the intention of the intervention is to encourage SMEs to adopt a new innovation method, then it may be better to exclude those who are already doing so given the limited scope for them to benefit further.

In addition, a project may want to consider the benefits of using baseline knowledge of a specific innovation activity in their selection criteria. If they have a relatively small sample, it would be better to recruit SMEs who have the most potential to benefit from their intervention. If it doesn't work for the 'ideal candidates' it is unlikely to work for others, but a positive impact for those most placed to benefit would support the assumption that it can also work for more marginal cases.

Challenge 9 - Expecting the unexpected: dealing with the impact of the COVID-19 pandemic

No matter how thoroughly one tries to identify the factors that might affect the successful delivery of an experiment, there will always be something that is overlooked. At the time of writing, there is a common unprecedented factor that is having profound impacts on project plans: the COVID-19 pandemic. Health concerns and social restrictions have made it impossible to deliver many of the planned interventions. Business needs and the priorities of innovation agencies are also changing dramatically.

In cases where unexpected external shocks appear, it's important for those funding such projects to talk to individual teams to learn how they have been affected and discuss how they should respond.

Teams should investigate whether they need to change their project and if so, how. In this particular case, groups were asked to consider whether it was feasible to continue as planned; if the project could be adapted to meet urgent business needs, what the implications of delaying the project were and whether there may need to be a change in their approach regardless (e.g. adjustments to expectations about impacts).

The main changes to plans in an extraordinary situation such as this one, are found across three areas:

- Delays to project activity - As the scale of an extraordinary situation becomes clear, some projects might be ready to start or in the early stages of recruitment. For some projects it may not be safe to continue their planned interventions at this time. Even if online interventions were possible, the feasibility of recruiting SMEs at this time could also be compromised.
- Implementation plans - Depending on the nature of the situation, many small businesses may be forced to close or to focus all resources on ensuring their immediate survival. As the economy recovers, businesses may still have reduced appetite and capacity to undertake innovation than was the case before the extraordinary situation and assumed by project teams. Therefore although projects may be set to pause their trials, they could still find a drop or change in demand for innovation support.
- Changes to delivery approach and research focus - Interventions may require changes if they are to continue (e.g. in the case of COVID-19, we expect many interventions to change from in-person support to virtual settings). Helping businesses through a crisis period may require a more agile approach that may not be well suited to an RCT. Expectations about when and how benefits are realised by businesses could change and agencies might be more reluctant to conduct intensive data collection.

Lessons for future funding calls for policy experiments

With regards to the overall funding call, perhaps the main lesson that we can draw from the process is the **importance of helping innovation agencies understand the experimental methodology** as they develop their proposals.

For most innovation agencies, calls like INNOSUP-06-2018 pose a great opportunity to apply an experimental approach to policy development and within this, learn how to use RCTs. There are many factors that agencies need to consider when they decide if, when and how to run an RCT. The approach also requires a change in the way agencies may typically undertake evaluation as

it needs to be planned in detail at the outset of the project, with a specific research question in mind and integrated into delivery.

Initial project reviews and engagement with the teams is likely to lead to substantial changes in projects. Therefore, it's useful to consider that similar support is provided in future calls and to **be aware that projects would also benefit from this additional engagement earlier in the process**. For instance, in addition to webinars before final stage proposals, it could be beneficial for potential participants to join a more intensive workshop (or series of webinars) before project selection starts. This could further improve the range and quality of projects coming forward. Being fully aware of the demands for running an RCT (e.g. sample size demands) will also ensure that projects include sufficient time and resources for trial development into their proposals.

We also believe in **the importance of having an evaluation partner from the design stage of the project.** Some agencies may consider designing their experiment with their internal staff, and without the necessary technical support on the design side. However, in our experience this can lead to problems once the projects start, and then make them face some delays as they would need to change the design of the programme because it was not suitable for an RCT.

When it comes to supporting the trial design and implementation, requiring project teams to include dedicated research and statistical expertise is highly beneficial. While project teams may be highly motivated and responsive to feedback, lack of familiarity with robust evaluations could make running this type of project more difficult. In particular, this skills gap may inhibit their ability to respond effectively to feedback and to make informed decisions about the trial.

At the same time, it is crucial to **make sure the project selection and objectives reflect the current status of the intervention and the technical feasibility of the trial.** When interventions are at an early stage of development often too little is known about how effectively it can be delivered and how a trial to evaluate its impact can be designed. An RCT is unlikely to be beneficial for the evaluations of policy experiments.

When the intention is to run an impact evaluation then it is important to select projects where this will be feasible. That does not necessarily mean that only perfectly designed trials should be selected. There are many challenges that can be addressed during the development of the project. But the selection process could be developed to include specific questions to help assessors gauge the feasibility of their proposal (e.g. the inclusion and justification of sample size calculations).

Even if it's more resource-intensive, opportunities to learn and **share experiences together in the same room** brings very positive results and should be encouraged as far as the programme allows. Where there are project teams that are not expected to be fully familiar with the methodology, having a crash course on policy experimentation with peers would allow them to share some questions and concerns that may not be easily presented during webinars or online chats.

Once selected projects start designing and developing the experiment, it would be helpful to **have a clear timetable to foresee the needs and resources for each stage.** For instance, early design stages are more intensive than implementation ones, as several implementation details need to be clarified before recruitment starts. A clear timetable and setting expectations about the time required to review and refine evaluation plans would ease programme management and avoid rushing in times where detailed planning is needed. We would also suggest that those managing a trial run pilots of the intervention and data collection before proceeding to the full trial.

Overall, the INNOSUP-06-2018 programme has so far provided various lessons to improve project selection and development of proposals; consistency, adequate and informed

preparation, while employing the right methodology at the right time, allows for successful trials and robust outcomes.