Experimental innovation policy for SMEs: findings and recommendations

Final Findings Report EASME/H2020/2018/005

Period covered by the report: October 2019 – September 2021

Written by James Phipps, Rob Fuller and Hugo Cuello under a service contract the European Innovation Council and SMEs Executive Agency (EISMEA) and Nesta
September – 2021

EUROPEAN COMMISSION

European Innovation Council and SMEs Executive Agency (EISMEA)

I.01.1 EU and placed-based innovation ecosystems

Contact: Daniel Gassmann

E-mail: EISMEA-EU-ECOSYSTEMS@ec.europa.eu

European Commission
B-1049 Brussels

Experimental innovation policy for SMEs: findings and recommendations

Final Findings Report EASME/H2020/2018/005, results to September 2021

European Innovation Council and SMEs Executive Agency (EISMEA)
Financed by the Horizon 2020 - Work Programme 2018-2020
Action 7. Innovation in small and medium-sized enterprises
Support to design and running of randomized control trials under INNOSUP-06-2018

2021

EN PDF

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Luxembourg: Publications Office of the European Union, 2021

ISBN 978-92-9460-879-6
doi: 10.2826/43979

Abstract

This report from the Innovation Growth Lab brings together analysis and key findings from thirteen projects who were chosen to break new ground in many agencies by applying experimental approaches to innovation policy.

Randomised Controlled Trials (RCTs) are a powerful way to determine whether an intervention causes the expected outcomes. As part of a broader experimental approach, RCTs can help innovation agencies to explore new ideas, optimise actions and find out what works. RCTs are, however, rarely used.

Within the Horizon 2020 Work Programme, 2018-2020 - 7. Innovation in small and medium-sized enterprises (INNOSUP), the European Commission launched a call to directly incentivise innovation agencies to engage in policy experimentation and use RCTs to evaluate their support schemes for SMEs.

The selected projects include feasibility studies of new support, such as training to encourage SMEs to adopt new innovation methods and technologies. Others seek to optimise programme delivery, for instance how best to offer SMEs feedback on their grant applications.

Results from three completed projects are presented alongside lessons learnt by all agencies during experiment design and implementation. Recommendations and tools are also provided for innovation agencies wishing to follow the approaches of the experimental pioneers and undertake their own experiments.

Résumé

Ce rapport du Innovation Growth Lab(IGL) rassemble l'analyse et les principales conclusions de treize projets qui ont été choisis pour innover dans l'application d'approches expérimentales à la politique d'innovation.

Les essais randomisés contrôlés (ERC) constituent un moyen efficace de déterminer si une intervention produit les résultats escomptés. Dans le cadre d’une approche expérimentale générale, les ERC pourraient aider les agences d'innovation à explorer de nouvelles idées, optimiser les actions et découvrir ce qui fonctionne.

Sous le programme de travail Horizon 2020, 2018-2020 - 7. Innovation dans les petites et moyennes entreprises (INNOSUP), la Commission européenne a lancé son appel afin d’inciter de l’expérimentation de politiques d’innovation et utiliser des ERC pour évaluer leurs programmes.

Les projets comprennent des études de faisabilité d'un nouveau soutien, tel que des formations visant à encourager les PME à adopter de nouvelles méthodes et d'autres cherchent à optimiser l'exécution du programme, par exemple la meilleure façon d'offrir aux PME un retour d'information sur leurs demandes de subvention.

Les résultats de trois projets achevés sont présentés, ainsi que tous les enseignements tirés lors de la conception et l'exécution des expériences. Des recommandations et des outils sont également fournis pour ceux qui souhaitent entreprendre leurs propres expériences.

Table of Contents

Executive Summary

By introducing a dedicated funding call "Supporting experimentation in innovation agencies" ("INNOSUP-06") within the Horizon 2020 programme, the European Commission has found that innovation agencies will engage in policy experimentation if provided with the means and freedom to do so. There has also been a whole action (INNOSUP-05) since 2014 dedicated to the peer learning of innovation agencies to learn from each other and apply good practices in innovation support services.

In this report, we look at the rationale for encouraging innovation agencies to become experimental and the objectives of the thirteen experimental projects that have been funded. We bring together the key findings from three completed projects and from the experiences of all participants as they designed and implemented their experiments. Recommendations are provided for other innovation agencies who may wish to become more experimental and replicate the approaches being tested across the thirteen projects.

Innovation agencies face continual demands to innovate and adapt their support for Small and Medium-Sized Enterprises (SMEs). Knowing what changes to make and how best to use the limited resources available is a perennial challenge: there is a lack of firm evidence on what can work and complex innovation systems make it hard to diagnose problems and know how the effects of any intervention will play out.

One way to address these challenges is for innovation agencies to become more experimental - introducing new ideas but also putting in place systems to learn whether they are working. Within this experimental approach, Randomised Controlled Trials (RCTs) can be one of the most powerful ways to determine whether an intervention can achieve its intended impacts. The idea that agencies should become more experimental and adopt RCTs is not new, but one that has been hard to put into operation.

The power of RCTs to deliver clear findings on the causal impact of an intervention relies on the seemingly simple premise of using randomisation to allocate support and thereby remove the selection bias which undermines and complicates other methodologies. As the thirteen projects teams have found, however, the concept might be simple but running a successful policy experiment often is not.

Fully embracing policy experimentation requires substantial changes in how innovation agencies typically approach evaluation. The approach demands much greater upfront planning of the evaluation design, a detailed review of policy objectives and careful consideration of how to integrate the trial into programme delivery. Also, while the methodology addresses the key challenge of creating a reliable comparison, other common problems remain, such as maximizing survey response and delivering results in time to influence policy decisions.

To date only one of the full-scale RCTs in the portfolio has been completed. This project, the 200SMEchallenge, tested the use of collaborative 'design sprints' to foster greater adoption of design thinking and user-centered design among SMEs. The evaluation has produced clear evidence that the intervention was successful in building participants' understanding and 'know-how' about the design sprint approach, the first step towards having an impact on SMEs' innovation capacity and competitiveness.

Many of the projects were not of sufficient scale to generate clear findings. For them, randomisation allowed for more qualitative comparisons of outcomes and to understand the feasibility of their intervention. It also allowed some to identify at an early stage that

their intervention was not ready for scaling, however for others, this came too late. Finding out what doesn't work is as vital to experimentation as learning what does. Nevertheless, difficulties with recruitment and the uptake of support can be more efficiently identified without the need for an RCT designed to evaluate impacts. Trials could then be used to understand the reasons behind these problems, to test different solutions or for an impact evaluation once these issues are resolved.

All teams have had to contend with the global pandemic. This has caused project delays and made it harder for projects to achieve their objectives. That being said, it also inevitably led to further innovation and generated ideas for future experiments - e.g. how does the effectiveness of online workshops compare to in-person as originally intended.

Despite the challenges, the experiments have generated a large number of findings and an appetite amongst the agencies to further apply the approach.

Résumé exécutif

En introduisant un appel de fonds spécifique « soutenant l'expérimentation dans les agences d'innovation » (« INNOSUP-6 ») dans le cadre du programme Horizon 2020, la Commission européenne a constaté que les agences s'engageront dans l'expérimentation politique si elles en ont les moyens et la liberté. Il existe également depuis 2014 une action entière (INNOSUP-05) consacrée à l'apprentissage par les pairs des agences d'innovation pour apprendre les unes des autres et appliquer les bonnes pratiques dans les services de soutien à l'innovation.

Dans ce rapport, nous examinons les raisons d'encourager les agences d'innovation à devenir expérimentales et les objectifs des treize projets expérimentaux qui ont été financés. Nous rassemblons les principaux résultats de trois projets achevés et des expériences de tous les participants lors de la conception et de la mise en œuvre de leurs expériences. Des recommandations sont formulées à l'intention d'autres agences d'innovation qui pourraient souhaiter devenir plus expérimentales et reproduire les approches testées dans le cadre des treize projets.

Les agences pour l'innovation doivent innover sans cesse et adapter leur soutien aux petites et moyennes entreprises (PME). Savoir quels changements apporter et comment utiliser au mieux les ressources limitées disponibles est un défi permanent : il y a un manque de preuves solides sur ce qui peut fonctionner et les systèmes d'innovation complexes rendent difficile le diagnostic des problèmes et la connaissance des effets de toute intervention.

Pour relever ces défis, les agences d'innovation doivent notamment devenir plus expérimentales, c'est-à-dire introduire de nouvelles idées mais aussi mettre en place des systèmes permettant de vérifier si elles fonctionnent. Dans le cadre de cette approche expérimentale, les essais randomisés contrôlés (ERC) peuvent être l'un des moyens les plus puissants de déterminer si une intervention peut avoir les effets escomptés. L'idée selon laquelle les agences devraient devenir plus expérimentales et adopter les ERC n'est pas nouvelle, mais elle a été difficile à mettre en œuvre.

La capacité des ERC à fournir des résultats clairs sur l'impact causal d'une intervention repose sur le principe apparemment simple de l'utilisation de la randomisation pour allouer le soutien et ainsi éliminer le biais de sélection qui mine et complique les autres méthodologies. Toutefois, comme l'ont constaté les équipes des treize projets, le concept peut être simple, mais la mise en œuvre d'une expérience politique réussie ne l'est souvent pas.

L'adoption pleine et entière de l'expérimentation politique exige des changements substantiels dans la manière dont les agences d'innovation abordent généralement l'évaluation. Cette approche exige une planification initiale beaucoup plus importante de la conception de l'évaluation, un examen détaillé des objectifs politiques et une réflexion approfondie sur la manière d'intégrer l'essai dans la mise en œuvre du programme. De plus, si la méthodologie permet de relever le défi majeur que représente la création d'une comparaison fiable, d'autres problèmes courants subsistent, comme la maximisation du taux de réponse à l'enquête et la livraison des résultats à temps pour influencer les décisions politiques.

À ce jour, un seul des ERC à grande échelle de la gamme a été achevé. Ce projet, le 200SMEchallenge, a testé l'utilisation de « sprints de conception » collaboratifs pour favoriser une plus grande adoption de la pensée conceptuelle et de la conception centrée sur l'utilisateur au sein des PME. L'évaluation a clairement démontré que l'intervention a réussi à renforcer la compréhension et le « savoir-faire » des participants concernant l'approche « sprint de conception », la première étape pour avoir un impact sur la capacité d'innovation et la compétitivité des PME.

De nombreux projets n'étaient pas d'une ampleur suffisante pour générer des résultats clairs. Pour eux, la randomisation a permis des comparaisons plus qualitatives des résultats et de comprendre la faisabilité de leur intervention. Cela a également permis à certains d'identifier à un stade précoce que leur intervention n'était pas prête à être mise à l'échelle, mais pour d'autres, cela est arrivé trop tard. Découvrir ce qui ne fonctionne pas est aussi vital pour l'expérimentation que d'apprendre ce qui fonctionne. Néanmoins, les difficultés de recrutement et de prise en charge peuvent être identifiées plus efficacement sans qu'il soit nécessaire de recourir à un ERC conçu pour évaluer les impacts. Les essais pourraient alors être utilisés pour comprendre les raisons de ces problèmes, pour tester différentes solutions ou pour une évaluation d'impact une fois ces questions résolues.

Toutes les équipes ont dû faire face à la pandémie mondiale. Cela a entraîné des retards dans les projets et rendu plus difficile la réalisation de leurs objectifs. Cela dit, cela a aussi inévitablement conduit à d'autres innovations et généré des idées pour de futures expériences - par exemple, comment l'efficacité des ateliers en ligne se compare-t-elle à celle des ateliers en personne, comme prévu à l'origine.

Malgré les difficultés, les expériences ont permis de dégager un grand nombre de résultats et de susciter l'envie des agences de poursuivre l'application de cette approche.

1) Introduction: Why innovation policy needs more experimentation

Innovation is about finding new ideas that work. Across Europe, policymakers invest billions each year to enable businesses to develop and test new products and services, to adopt new technologies and ways of working. This accelerates the development and diffusion of proven ideas, delivering wide-ranging benefits to society. Paradoxically the approach to policy development itself is not very experimental. Policymakers that direct funding to scientific and business experiments, rarely experiment with their own programs and activities.

Innovation support agencies, both regional and national, are responsible for the focus, design and delivery mechanisms for much of the innovation support programmes provided for Small and Medium-Sized Enterprises (SMEs). The many choices they make will in a large part determine the economic impact of the actions. But what are the right choices? Are they making the most of each investment? Are there more effective or inclusive ways of using their funding? How would they ever know? Answering these questions is difficult.

Firstly, innovation systems are complex and continuously evolving. Shifting a policy lever may have unanticipated consequences due to new or previously unknown interdependencies. Innovation policymakers are being asked to address new challenges, such as in the areas of climate change or the transformation of work. These require imaginative solutions and encourage policymakers to become more agile and continuously search for new ideas. This was never more true than during the COVID-19 pandemic. Agencies created new interventions to help businesses faced with urgent needs to adapt business models and had to overhaul how they worked, such as streamlining assessment processes and switching them to virtual formats.

New tools are being developed but too rarely applied with the intention of learning what works - at least not in a structured and rigorous way.

Secondly, we lack much of the evidence that can guide policy decisions. This is a longstanding issue. The Compendium of Evidence on the Effectiveness of Innovation Policy Interventions and What Works Centre for Growth both reached similar conclusions about the lack of robust evidence. Whilst full of useful insights, all but a few evaluations provided evidence that was sufficiently robust about the causal impact of programmes to change minds.

This is not to say that we should aspire or expect all innovation policies to be evaluated in the same way. There are many relevant questions that counterfactual evaluation methods cannot answer and many important effects that cannot be easily quantified. However, there is undoubtedly much more that can be done to rigorously measure the impact of interventions. Ultimately, a lack of this evidence leads to using scarce resources on policies that are less effective (or potentially even counter-productive).

These two issues can be addressed jointly through a more experimental approach to policymaking. Recognising that policymakers are often compelled to act before they have all the answers, they can still build into their plans a structured and rigorous approach to learning as they act. In this way, they can create an evidence base to inform their own

decisions about what should be adjusted, scaled up or discontinued, but which can also be shared more widely.

Policy experiments can be used in different contexts and with different objectives. The experiments themselves can be focused either on exploration and discovery (understanding how the world works) or framed around evaluation (finding out what works).

In the first category the experiments are used to explore the feasibility and potential of a new intervention: Can it be delivered? What types of outcomes are likely to emerge? How do people or businesses respond to it? Do the underlying assumptions about the problem and the way change will occur hold?

The second group of policy experiments are focused on evaluation, although from two different perspectives: impact evaluations that estimate the ultimate impact of an intervention on outcomes, and process optimisation experiments that measure intermediate impacts of changes in the process.

Improving the evidence base is one motivation for experimentation, but an often overlooked benefit is in encouraging organisations to become more agile and innovative, continuously searching for new ideas to test rather than defaulting to the status quo.

Ideally, policy experiments start at a small scale, not being larger than what is required to answer the question or validate the hypothesis being tested. With entirely new programmes there is often high uncertainty and limited prior knowledge to build on. Because of that, there are clear benefits from setting up prototypes and continuously iterating and adapting designs to improve through trial and error, and only then advancing to full impact evaluations where costs and timeframes can be substantial.

The Context: Innosup-06-2018 - outline and objectives

In recent years, there has been growing recognition of the value Randomised Controlled Trials (RCTs) can provide to public programmes.

RCTs are a methodology to determine whether an intervention is achieving its aims and intended impacts. In the most simple form, participants are randomly assigned to either an intervention group, who are given the intervention, or a control group who are not. Randomly assigning participants to a control group removes selection bias, and enables you to compare the effectiveness of the new intervention against what would have happened if you had changed nothing. They can therefore provide a concise and clear-cut conclusion of the effectiveness of an intervention that avoids lengthy caveats and complex analysis.

The European Commission launched the first dedicated call for randomised trials as part of the Horizon 2020 Work Programme (INNOSUP-06-2018: Supporting experimentation in innovation agencies of their innovation support schemes for SMEs. The European Commission[1] and others have sought to encourage the use of trials with programme

[1] https://ec.europa.eu/competition/state_aid/modernisation/state_aid_evaluation_methodology_en.pdf

evaluations but with little progress. There has also been a whole action (INNOSUP-05) since 2014 dedicated to the peer learning of innovation agencies to learn from each other and apply good practices in innovation support services. The reasoning behind the creation of a dedicated call for innovation policy experiments was to go beyond the INNOSUP-05 action and to overcome inertia amongst agencies due to a 'lack of funds, time pressure to deliver new support, and the fear of a backlash against 'money wasting'.[2]

The expected impact of the call was that:

1. The number of innovation agencies engaged in policy experimentation significantly increases;

2. The use of RCTs in the design and testing of innovation support schemes significantly increases;

3. A broad range of new or significantly improved SME innovation support schemes are investigated and developed and their impact is rigorously tested. Pilot agencies scale up these new schemes.

To achieve this 'INNOSUP-06-2018: Supporting experimentation in innovation agencies' offered funding for two levels of experiments:

- Small grants of up to €60,000 were available for small-scale experimental pilots of novel innovation support ideas.

- Larger grants (€300.000-€500.000) for more substantial RCTs that evaluated scalable SME innovation support schemes whose feasibility had already been proven.

Thirteen projects received funding, with new forms of innovation support being provided for SMEs across 14 countries with 27 agencies involved in delivering these policy experiments. These included national agencies such as the Austrian Research Promotion Agency (FFG) and the Business and Cultural Development Centre (KEPA), also regional agencies such as Torino Wireless Foundation from Piedmont in Italy and the Institute for Business Competitiveness (ICE) of Castilla y León in Spain. A range of innovative interventions are being tested including schemes to teach SMEs innovation methods such as user design, to encourage innovation through co-creation workshops and to help SME become investment ready and access external funding (See Section 2).

The novelty of running trials within innovation policy has meant that project teams were to tread new ground. In the autumn of 2018, the Innovation Growth Lab (IGL), based at Nesta, was selected by EISMEA to deliver support to these projects, with most beneficiaries set to undertake their first experiment. IGL was also tasked with capturing and sharing findings, so that other agencies could learn from the findings and be encouraged to engage in their own experiments.

This Final Findings report has been produced under EASME/H2020/2018/005 'Support to design and running of randomized control trials' under INNOSUP-06-2018. It presents

[2] Horizon 2020. Work Programme 2018-2020: Innovation in small and medium-sized enterprises

the context and set-up of all thirteen projects; bringing together key findings that are emerging from each project; the challenges encountered as agencies designed and implemented their experiments; and sharing recommendations for innovation agencies wishing to undertake similar approaches in future.

As of September 2021, only three projects had been completed and were able to share their final findings with IGL for inclusion in this report. IGL is set to continue its support for projects that remain underway. As part of this we will produce future outputs to cover findings as other projects complete, culminating in an updated and more complete version of this report where there will be an opportunity to go deeper into cross-cutting themes and extend recommendations.

2) Details of the thirteen supported experiments

In this section we present an overview of the context and the RCT set-ups of the thirteen projects funded through 'INNOSUP-06-2018'. For each project we set out key information such as organisations involved, the intervention provided, and the current status. Where applicable we have also included references to the publications and results of each project.

DINNOS: Diversity Innovation Support Scheme for SMEs	
This collaboration between researchers at Trinity College Dublin, Aston Business School and the University of Wuppertal, together with the Kienbaum Institute and the Greater Birmingham Chambers of Commerce, involves scaling up a pilot programme that has shown promising initial results in preventing and even reversing the adverse impact of age diversity on innovation. The programme consists of cognitive training for older employees as well as leadership training for entrepreneurs that seek to reduce age stereotypes and associated conflicts and enhance appreciation of age diversity. The trial will measure the impact of this innovation support scheme on individual and organisation-level measures of various determinants of innovative behaviour among SMEs in the West Midlands Region of the UK and the Rhine-Ruhr region of Germany.	
Research question	Among SMEs in the West Midlands (UK) and Rhine-Ruhr (Germany) regions, does access to an online cognitive training programme for older employees as well as leadership training for owners increase individual cognitive ability and creativity and firm innovation capacity?
Innovation Agencies	Trinity College Dublin, Aston Business School, University of Wuppertal, Greater Birmingham Chambers of Commerce, Kienbaum Institute (Ireland, UK, Germany)
Tested intervention	Leadership training for entrepreneurs, cognitive training for employees
Current Status	Recruitment/intervention delivery
Key Deliverables to date	D5.1 – Online tool to benchmark innovation capacity at the organisational and individual level Trial registration
Further information	Dinnos diversity innovation support scheme for SMEs demographic change

	Diversity innovation support scheme for SMEs | DINNOS Project | Fact Sheet | H2020 | CORDIS | European Commission

RCT4MANU: Testing an innovative support scheme for manufacturing SMEs and accelerating the use of RCTs in innovation agencies	
This trial, developed by Innovate UK and KTN, aims to evaluate the effectiveness and impact of the 4Manufacturing tool and extra support offered by experts. 4Manufacturing is an innovative support programme based on diagnosis and one-to-one consultant advice, which aims to accelerate the adoption of industrial digital technologies for manufacturing SMEs. The project has designed useful and original ways of measuring the attainment levels and observing the increase in the number and speed of technology adoptions. This trial will be used as an exemplar within UK Research and Innovation regarding the benefits of using RCTs, and the results of this research may inform future funding decisions with regards to the manufacturing industry.	
Research question (PICO[3] statement)	P: UK Manufacturing SMEs who are familiar with technology adoption; I: Will be offered 4Manufacturing® support which assists manufacturing companies in identifying enabling technologies; C: Compared with the control group who will receive no support; O: With an outcome of an increase in the number and speed of technology adoptions, leading to increased productivity.
Innovation Agencies	Innovate UK (UK) and Knowledge Transfer Network (KTN)
Tested intervention	Innovative tool for manufacturing businesses to support tech adoptions
Current Status	Intervention delivery
Key Deliverables to date	Trial registration
Further information	Homepage | KTN 4Manufacturing

[3] Project teams were encouraged to use the PICO (Population, Intervention, Control and Outcomes) framework as they developed their research question. See 'Recommendation 2' in Section 4 for further discussion.

	KTN plays a part in European research and innovation projects - KTN Test an innovative support scheme for manufacturing SMEs and accelerate the use of RCTs in innovation agencies | RCT4MANU Project | Fact Sheet | H2020 | CORDIS | European Commission

DIHnamic: Dynamic facilitation and thrust from regional innovation agencies

The project is developed by the Institute for Business Competitiveness, the regional agency in charge of innovation activities in the Spanish region of Castilla y León. The intervention is supported by Digital Innovation Hubs in the area: centres that support companies to become more competitive by enhancing their digital transformation process. The Institute aims to test how much support is needed to ensure that businesses adopt new technologies to improve business productivity. Sometimes, adding an extra layer of support based on trying and testing the technology is necessary. The agency is interested in exploring whether the effect size is large enough to be cost-effective.

Innovation Agencies	Institute for Business Competitiveness (Spain)
Tested intervention	Specialised support scheme provided by the Digital Innovation Hubs
Current Status	Design stage
Key Deliverables to date	-
Further information	Digital Innovation Hubs: dynamic facilitation and thrust from regional innovation agencies | DIHnamic Project | H2020 | CORDIS | European Commission

InDemandRCT: Demand-driven and business co-creation for a new innovation business model

The Institute for Development of the Murcia Region in Spain aims to explore and test a new demand-driven business model based on co-creation patterns between customers and SMEs. Co-creation is an innovative approach to improve products and services, and SMEs may benefit highly from it in certain conditions. The intervention of this trial is based on providing co-creation opportunities (with larger

businesses or third sector organisations) and specialised business support to SMEs. The aim is for this to provide valuable information to improve the quality and usefulness of their products and make funding calls more impactful.

Research question (PICO statement)	P: Innovative SMEs: R&D at regional, national and EU level. Some 2,500 firms in the region from which we will recruit 72 SMES; I: 36 SMEs will be selected to use the InDemand methodology including business support and co-creation; C: The remaining 36 SMEs will receive the traditional ERDF funding without co-creation and business support; O: Comparison between the 2 groups, measure of how intervention has improved SMEs' competitiveness (satisfaction, job creation, revenues, absorption capacity, awareness). Better use for money, widening the basis of users.
Innovation Agencies	The Institute for Development of the Murcia Region (Spain)
Tested intervention	Co-creation and business support service scheme to develop innovative solutions
Current Status	Recruitment
Key Deliverables to date	-
Further information	Demand-driven and business co-creation for a new innovation business model | inDemand-RCT Project | H2020 | CORDIS | European Commission

D3T: Assisting SMEs with digital transformation

Torino Wireless Foundation in Northern Italy leads this project that aims to improve their Digital Transformation support scheme. The scheme allows SMEs from the Piedmont region to change their business and organisational activities, processes, competencies and products by taking advantage of a mixture of digital technologies. Through this scheme, the agency wants to explore if introducing a data-driven approach could help companies assess their digitalisation needs more efficiently. This approach would provide information on how the agency could support businesses in more cost-effective ways; allocating resources where

necessary, and considering what support can be provided using online materials to alleviate pressure on the delivery agency while also providing effective support.	
Research question	Does offering a data driven approach to the Digital Transformation (DT) support services (the intervention) improve SMEs (the trial population) readiness (outcome) and the timeliness (outcome) for the DT implementation?
Innovation Agencies	Torino Wireless Foundation (Italy)
Tested intervention	A data driven approach to delivery of Digital Transformation support services
Current Status	Completed
Key Deliverables to date	Key findings and lessons learned
Further information	D3T Homepage [Webinar March 9th] D3T final event Data Driven Digital Transformation | D3T Project | H2020 | CORDIS | European Commission

DepoSIt: Development and testing of the European Innovation Audit tool for Social Innovation	
This trial brings together six organisations across Europe (the Steinbeis-Europa-Zentrum, the Croatian Chamber of Economy, the South Muntenia Regional Development Agency, the Fomento San Sebastian, the Business Development Friesland and the Friuli Innovazione) to develop and test a novel version of the European Innovation Audit Tool that includes a strong social innovation element. By increasing awareness and knowledge of social innovation in the private sector, the trial aims to unlock opportunities for collaboration between the civil society and small- and medium-sized companies operating in the smart health, smart mobility and smart living sectors in the six European regions involved in the project.	
Research question	For innovation-driven SMEs (the population), does the exposure to a set of specific questions on social innovation potential during an Innovation Audit (the intervention) motivate them to consider business opportunities related to social challenges (the

	outcome) more than similar firms who did not take the Innovation audit (the control)?
Innovation Agencies	The Steinbeis-Europa-Zentrum, The Croatian Chamber of Economy, the South Muntenia Regional Development Agency, the Fomento San Sebastian, the Business Development Friesland and Friuli Innovazione (Germany, Croatia, Romania, Spain, Netherlands, Italy)
Tested intervention	Innovation Audit Tool, with a focus on social innovation
Current Status	Post-intervention data collection
Key Deliverables to date	D2.1 – Joint report on innovations and new solutions/trends for the innovation audits
Further information	DepoSIt Project - Homepage DepoSIt Project - LinkedIn DepoSIt Project - Facebook Homepage DepoSIt Project (@deposit_project) Development and testing of the European Innovation Audit tool for Social Innovation | DepoSIt Project | H2020 | CORDIS | European Commission

Create4Value: Creative collaboration to provide value for first time innovators - effective engagement of stakeholders and users in co-creation processes in SMEs

The Create4Value project is looking at whether a process of co-creation can be used to encourage first time innovators. Poznański Park Naukowo Technologiczny (PPNT), who are undertaking this pilot, will adapt current methods of co-creation to better meet the needs of first-time innovators in Poland. The small-scale experiment will take several SMEs through the co-creation process and compare their experience with a group of SMEs who are provided with more traditional support for first time innovators. PPNT will also explore how to engage users and stakeholders in the process. Findings will inform PPNT's future policy offer; provide insights for other agencies on the use of co-creation and help to elaborate the method of assessing such support schemes in future experiments.

Research question	For SMEs who receive our innovation support based on Business Model Canvas - does providing them with access to a sophisticated process of co-creation rather than just the BMC itself (the control) lead them to unlock their innovation potential (the outcome)?
Innovation Agencies	Poznań Science and Technology Park - Adam Mickiewicz University Foundation (Poland)
Tested intervention	Co-creation workshops to unlock innovation potential of SMEs
Current Status	Intervention Delivery
Key Deliverables to date	-
Further information	Create4value Homepage Creative collaboration to provide value for first time innovators - effective engagement of stakeholders and users in co-creation processes in SMEs. | Create4value Project | H2020 | CORDIS | European Commission

DCS-iSMEs: Design Customised Support for Innovative SMEs

This pilot is exploring the potential of a new innovation support scheme to enhance SME utilisation of 'Design thinking'. The Business and Cultural Development Centre (KEPA) in Greece will set up a brand-new innovation support service to help SMEs further develop their operations by enhancing the use of Design. The intervention consists of a 'Design Clinic' for initial assessment; introductory workshops and customised mentoring. Through a small-scale experiment, the experiences of SMEs taken through this new programme will be compared with a comparison group who, during the pilot period, are only provided with a DIY guidebook. Findings will be used to come up with a feasibility study for expanding the programme, which may include a larger impact evaluation experiment.

Research question	Would providing SMEs who have received support from KEPA with a new package of support to build capacity to apply a Design Thinking Method ('Design Clinic'), introductory workshop and mentoring) lead them to adopt this approach and boost business performance compared to if only offered existing support?

Innovation Agencies	Business and Cultural Development Centre (KEPA) (Greece)
Tested intervention	Customised mentoring in design thinking
Current Status	Completed
Key Deliverables to date	Feasibility study
Further information	KEPA Homepage - Design Customized Support for Innovative SMEs Design Customized Support for Innovative SMEs | KEPA Project | H2020 | CORDIS | European Commission

InReady: Designing a service to improve start-ups' investor readiness

The project, led by the Lithuanian Innovation Center, aims to design a service that supports startups dealing with investors more effectively. For some early entrepreneurs, pitching their idea to investors can be challenging. Even if the business plan is promising, they may lack the skills to pitch effectively, which could ultimately lead to less funding and lower rates of survival. The proposed intervention is a web-tool that improves the readiness of startups for investment pitching. The project aims to bring insights from three different agencies across Europe that would provide valuable lessons about the different startup readiness levels in various European countries.

Research question	Does the use of the InReady tool improve the quality (or success rate) of entrepreneurs' pitches to VC funds?
Innovation Agencies	The Lithuanian Innovation Centre, Foundation for Research and Technology Hellas, Agenzia per la Promozione della Ricerca Europea (Lithuania, Greece, Italy)
Tested intervention	Online tool and expert support to prepare pitches for funding
Current Status	Completed

Key Deliverables to date	D1.1 – User needs, service portfolio and futures plans D5.1 – D5.1. Final report, including assessment of project results and impacts
Further information	InReady Homepage Designing the Service to Improve the Investor Readiness of Start-ups | InReady Project | H2020 | CORDIS | European Commission

200SMEchallenge: UX Design for digital SMEs

This project is a collaboration of seven regional innovation agencies across Europe, led by Hub Innovazione Trentino (Italy), and evaluated by Fondazione Bruno Kessler (Italy). The trial tests the impact of a 'UX Challenge' - a two-day design sprint hackathon for SMEs and young talents operating in the digital industry sector. The rationale behind the trial is that user-centered design techniques inspired by design thinking, such as 'design sprints', have the potential to substantially improve the quality of digital products design; and yet there is little evidence on whether SMEs can benefit from these techniques. The project will evaluate whether the UX Challenge can increase knowledge, awareness and intention to adopt innovation approaches in the design of digital products. The results will help inform innovation agencies interested in supporting digital SMEs to increase their user-centered design capabilities.

Research question	For SMEs who operate in the digital industry sector (the population), does participating in a two-day UX Challenge focused on digital product design (the intervention), rather than not participating (the control), enhance their readiness and awareness about innovative approaches in the design of digital products (the outcome)?
Innovation Agencies	Hub Innovazione Trentino Fondazione, Steinbeis Innovation gGmbH, Fundació General De La Universitat Jaume I Fundació De La Comunitat Valenciana, Lithuanian Innovation Centre, Oulun Kaupunki, Design Society Fond, Sihtasutus Tallinna Teaduspark Tehnopol (Italy, Germany, Spain, Lithuania, Finland, Denmark, Estonia)
Tested intervention	'Design sprint' events

Current Status	Data analysis and reporting
Key Deliverables to date	D2.1 – Partners' handbook on implementing the intervention D2.2 – UX Challenge playbook D4.3 – Report on the results of the large-scale pilot and guidelines for improvement D4.4 – Evaluation of scheme impact through RCT
Further information	200SMEChallenge Homepage 200SMEChallenge Twitter (@2Echallenge) 200SMEChallenge LinkedIn Homepage Design-driven Open Innovation Challenge for 200 SMEs | 200SMEchallenge Project | H2020 | CORDIS | European Commission

FeedFirst: Providing feedback to successful grant applicants	
This trial, conducted by the Austrian research promotion agency (FFG), aims to evaluate the impact of providing feedback generated from funding application evaluation scores to firms that successfully applied for R&D grants. The rationale is that the agency currently does not provide feedback to successful applicants; and the evaluations of each proposal contain project-specific information that could be helpful to the projects. While there is some evidence from other fields that feedback generally helps innovative projects, it is less clear on what the feedback should focus on (in terms of content, format, etc.). This trial will seek to answer this question by testing different formats of feedback.	
Research question	What type of feedback from the evaluation assessment (the intervention) - 'absolute' vs 'relative' feedback - to firms that successfully applied for grant funding (the population) is most helpful to improve on their project delivery and eventual outcomes (the outcome)?
Innovation Agencies	Austrian Research Promotion Agency (FFG) (Austria)

Tested intervention	Variations in feedback provided to grant applicants
Current Status	Recruitment
Key Deliverables to date	Trial Registration
Further information	Feedback for Small Companies and Firsttimers (FeedS First) | FeedS First Project | H2020 | CORDIS | European Commission

SIM Crowd: Improving the chances of crowdfunded projects through public endorsements	
Crowdfunding is increasingly used to finance innovation projects, especially those with a social component. Sim Crowd, led by the Austrian research promotion agency (FFG), focuses on introducing the crowdfunding campaigns of social innovation projects to the public through a targeted email campaign to potential crowd investors. In this three-arm trial, FFG will test whether publicly acknowledging either seed funding or matched funding from the FFG can help them obtain additional funding from the crowd.	
Research question (PICO statement)	P: SMEs and social innovators with social innovation projects, recipients in FFG's email database; I: Test 'seal of approval' from FFG (either seed or matched funding) in email campaigns, support in crowdfunding campaign; C: No seal of approval; O: Effect of FFG seal (positive/neutral/negative) on projects' success in raising crowdfunding on CF platforms.
Innovation Agencies	Austrian Research Promotion Agency (FFG) (Austria)
Tested intervention	Different ways of framing FFG support in communications to potential investors
Current Status	Design stage
Key Deliverables to date	Trial Registration

Further information	Social Innovation Matched Crowdfunding (SIM Crowd) | SIM Crowd Project | Fact Sheet | H2020 | CORDIS | European Commission

InnoCAP: Increasing the innovation capacity of SMEs	
This project, delivered by the Austrian research promotion agency (FFG), tests an intervention offering innovation support tools to SMEs. The participants are SMEs that successfully applied to one of FFG's innovation funding schemes; the agency aims to find out whether two supplementary tools – a digital innovation management tool, and a platform providing expert mentoring – can help improve the firms' projects. The rationale is that, especially for companies innovating for the first time, funding might not be enough, as many lack the core capabilities needed for innovation.	
Research question (PICO statement)	P: SMEs in receiving funding from the 'Impact Innovation: Using Innovation Methods' grant scheme; I: Innovation management software and a voucher for online mentoring (worth €1,000) (Digital support tools to stick to a process); C: Funded projects with the possibility to purchase the tools but are not pushed or nudged towards doing so; O: Use of Innovation Methods (iteration and inclusion of users), achievement of project goals, project outcome (turnover, employee growth), use of tools.
Innovation Agencies	Austrian Research Promotion Agency (FFG) (Austria)
Tested intervention	Access to online innovation tools and mentoring
Current Status	Intervention delivery
Key Deliverables to date	Trial Registration
Further information	Innovation Capacity Building in SMEs (InnoCAP) | InnoCAP Project | H2020 | CORDIS | European Commission

Connection to wider innovation policy

Randomised trials can be used to evaluate policies and programmes which have a targeted population that can be randomised into different groups and where the agency can determine the treatment or intervention that participants in each group will be subject to. Trials are therefore well suited to targeted programmes such as entrepreneurship training, R&D grants, science funding, or tech transfer schemes.

In contrast, it will be very difficult to use a randomised trial to evaluate the overall impact of national regulation and tax policy, or large infrastructure investments. However, when the main motivation to experiment is to optimise rather than evaluate a policy the opportunities for experimentation are much larger. For example, designing an experiment to test the overall impact of changing intellectual property regulation may not be feasible but there would be much greater scope when it comes to testing ways to ensure SMEs understand and respond to any changes.

Potential uses of randomised trials in innovation policy

	Mechanism experiments	**Optimisation experiments**	**Evaluation experiments**
Framework conditions (e.g., tax, regulation)	Medium	Medium	Low
Ecosystem (e.g., clusters, infrastructure)	Medium	Medium	Low (overall) Medium (tools)
Targeted programmes (e.g., grants, advice)	High	High	High

See: Bravo-Biosca (2019) 'Experimental Innovation Policy', for further discussion

Each of the thirteen projects uses an experiment to test the overall impact or optimise the design of targeted direct support to SMEs, and can be grouped into four themes. Firstly, projects that are seeking to boost the innovation capacity of SMEs by providing training on innovation practices or access to tools. Create4Value, creates a second group by itself, it also engages SMEs with new innovation methods (co-creation) but now with a specific objective of doing so to encourage new innovators. A third group is hoping to encourage SMEs to adopt new technologies testing new interventions that are themselves based around new digital tools. The fourth group are interventions that connect to the provision of finance for innovative SMEs but each tackling this in very different ways. SIM Crowd is exploring how an innovation agency's involvement could motivate private investors, while Feedfirst are investigating the benefits of providing

feedback during funding assessments and InReady are building an SME's capability to attract finance.

Policy Objective	Projects
Enhancing the innovation capacity of SMEs	DCS-iSMEs InDemandRCT 200SMEChallenge InnoCap DINNOS DepoSIt DIHnamic
Encouraging SMEs to innovate	Create4Value
Encourage SMEs to adopt or better use technology	RCT4MANU D3T
SMEs access to innovation funding	SIM Crowd FeedFirst InReady

The thirteen projects only begin to touch on the policy questions that could be answered in these areas through an experimental approach. There are also many other innovation policy objectives where an experimental approach would be applicable, such as interventions to:

- Increase collaborations, such as connections between businesses and external researchers ;
- Encourage the next generation of innovators and address observed gaps in who goes on to become inventors and entrepreneurs;
- Increase the effectiveness, efficiency and fairness of funding processes;
- Improve the equality, diversity and inclusion of those who access and benefit from innovation policy.

In later sections, we look at what the experience of beneficiaries tells us about running innovation policy experiments with recommendations for those wanting to run their own trials that will be relevant regardless of the policy area. In the next section, we detail findings from the three completed projects. Results from other projects will be covered in

future outputs but agencies seeking examples can explore a wide range of experiments on the IGL website.

3) Analysis and key findings from completed projects

In this section we provide analysis and assessment of the three completed INNOSUP-06-2018 projects, with particular attention to the key findings produced by each. For each project we delve deeper into the rationale and logic behind the project, the intervention itself as well as the evaluation design, or RCT set-up. We also analyse the impacts of the intervention and provide lessons and wider learning for organisations interested in experimenting with similar approaches.

200SMEchallenge – Design-driven Open Innovation Challenge for 200 SMEs

Coordinator: Hub Innovazione Trentino

Participants: Steinbeis-Europa-Zentrum, FUGEN – Espaitec, Lithuanian Innovation Centre, Business Oulu, Dansk Design Centre, Tehnopol, Fondazione Bruno Kessler

Further details: Design-driven Open Innovation Challenge for 200 SMEs | 200SMEchallenge Project | H2020 | CORDIS | European Commission

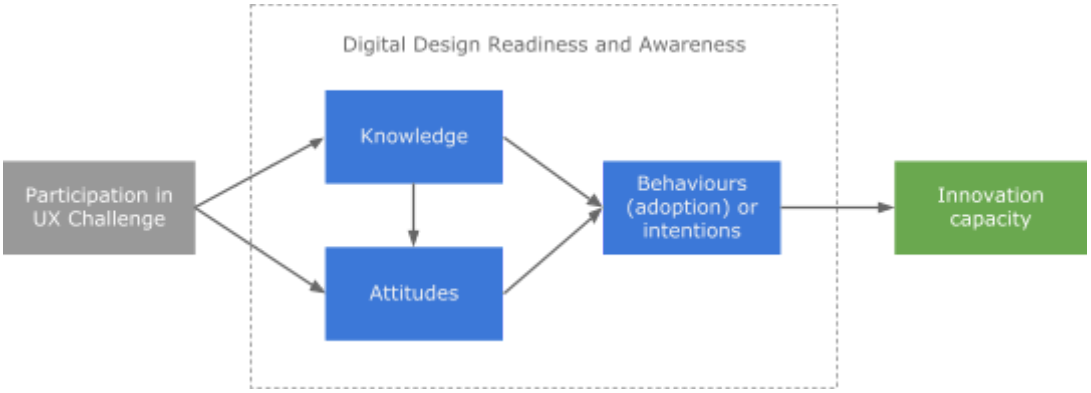
200SMEchallenge Homepage

Research question: For SMEs who operate in the digital industry sector or other SMEs who develop products with digital interfaces, does participating in the UX Challenge enhance knowledge, awareness and intention to adopt the innovative approaches in the design of digital products, compared to not participating?

Rationale and project logic

The 200SMEchallenge project is based on the hypothesis that the use of design thinking and user-centred design has the potential to improve the design and user experience of digital products and services provided by SMEs. An improved user experience leads in turn to growth in the user base and/or market share and ultimately to higher productivity and profitability, as well as building the business's capacity to innovate further. However, many SMEs have little awareness of design thinking and user-centred design, or knowledge of how to put it into practice. By giving them an opportunity to participate in a facilitated 'design sprint', the 200SMEchallenge project sought to improve SME managers' awareness of the potential benefits of design thinking and user-centred design and their knowledge of the design process, to enable and encourage them to use design techniques for themselves.

Figure 1: Logic model for the 200SMEchallenge project



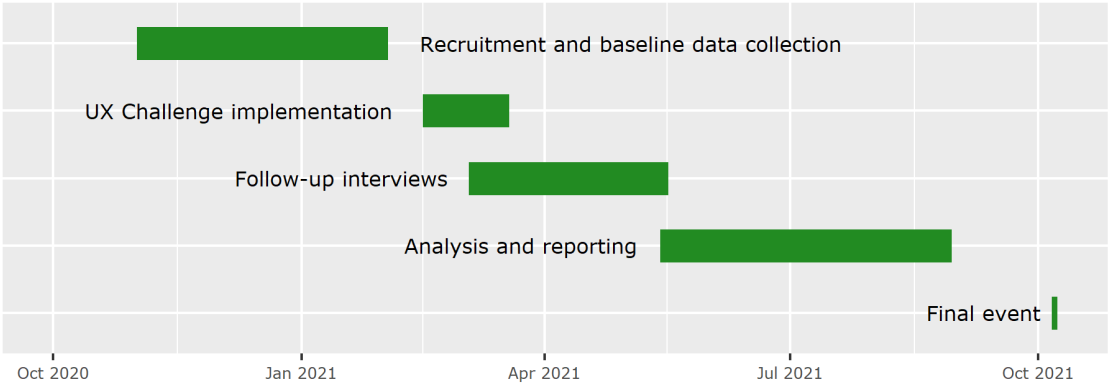
Intervention

The User Experience Challenge (UX Challenge) is an intensive 'design sprint' in which SME managers work together with user experience (UX) design professionals and teams of students or recent graduates to develop and test approaches to improving the user interfaces of products and services, according to an open innovation paradigm. This approach is modelled on the format used by GV (formerly Google Ventures), and has been refined and tested by Hub Innovazione Trentino in recent years.

The design sprint consists of five phases, beginning with mapping out the problem, followed by sketching out potential solutions and selecting one to take forward, before developing a prototype and testing it with target customers. These events are normally held as in-person events over two days, but this was adapted as a result of the COVID-19 pandemic to be held online over five days.

Under the 200SMEchallenge project, the UX Challenge was carried out by seven innovation agencies in seven countries across Europe, under the guidance of Hub Innovazione Trentino. Participation was open to SMEs with any products or services that make use of a digital user interface, including those providing services through a website, app or other software, as well as companies that produce devices or equipment with a digital interface.

Project timeline



Evaluation design

The 200SMEchallenge project was implemented as an RCT with a strong design.

A total of 190 eligible SMEs were recruited to participate in the project, just short of the target of 200. Of these, 60 were randomly selected (stratified by country) to participate in the UX Challenge. The remaining 130 SMEs acted as a control group, and were not given any support during the project's lifetime.[4]

The aim of the evaluation was to assess the impact of the UX Challenge on SMEs' Digital Design Readiness and Awareness. Specifically, data was collected on six outcome measures, including three indices of participants' knowledge about design principles and the design sprint process, a measure of their attitudes towards the use of design, and indicators of their aspirations and expectations for adopting the use of design thinking and user-centred design techniques in the next 6–12 months. Data was collected from all trial participants at baseline (prior to randomisation and to the implementation of the UX Challenge), and again in a follow-up survey carried out three weeks after the UX Challenge. The outcomes were all measured on 10-point scales.[5] The statistical power of the trial is such that an impact on any of the outcome measures would need to be approximately one point or larger on the 10-point scale in order to be reasonably (80%) confident of being able to detect it.

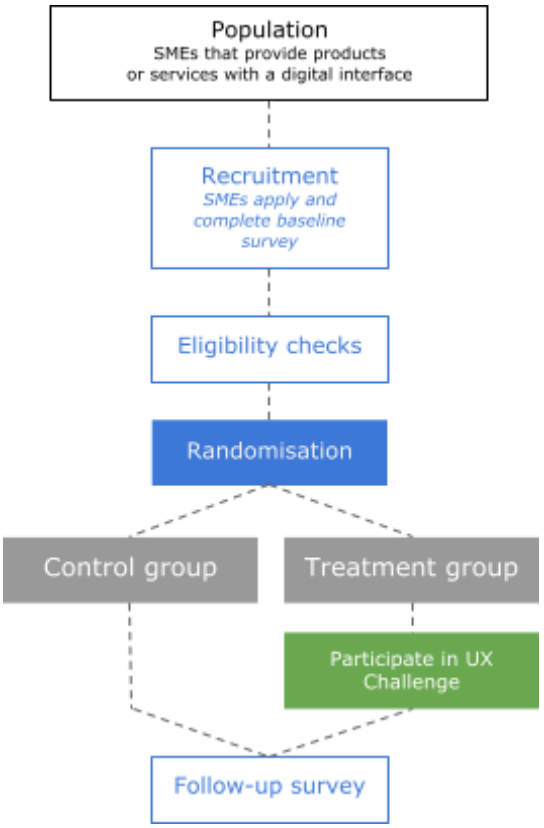
The most significant challenge to the robustness of the evaluation is that only 63% of the control group responded to the follow-up survey (against 95% of the treatment group). While there are no clear differences in the characteristics of those who responded to the follow-up survey and those who did not, it is possible that there are unobservable differences between them (e.g. in their motivations) that could bias the results of the evaluation. The evaluators have attempted to account for this by using

[4] Two of the companies allocated to the treatment group were not able to participate in the intervention, and were replaced with two randomly-selected companies from the control group. For the purposes of the analysis, those companies continue to be treated as members of the control group, with the consequence that the impacts of the intervention may be slightly underestimated.

[5] The specific methods for constructing the six outcome measures varied: refer to Section 2.2 of Deliverable 4.4 for the 200SMEchallenge project for full details.

alternative statistical models and by calculating 'bounds' for the estimates of the impact of the intervention.

Figure 2: Trial diagram for the 200SMEchallenge project



Impacts

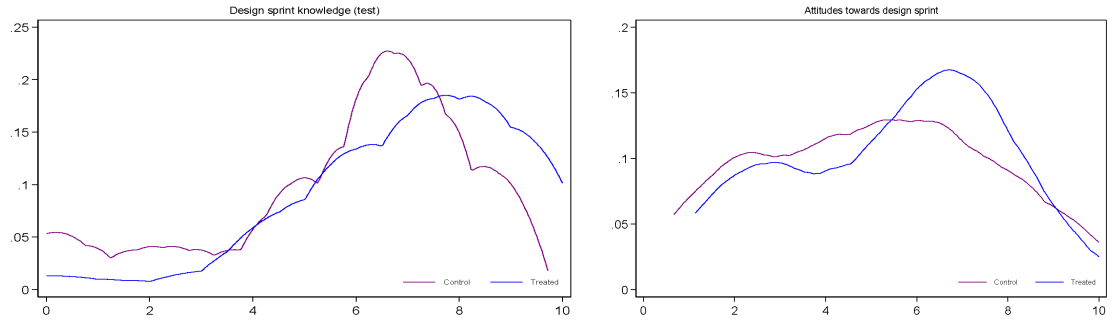
The evaluation results suggest that the UX Challenge had positive impacts on participants' knowledge about design sprints and their ability to implement design sprints. In the post-intervention survey, the treatment group on average scored 1.1 points more than the control group on the 10-point scale for design sprint knowledge. This is the best estimate of impact, though the data, once taking account of uncertainty, are consistent with an impact ranging from 0.3 to 1.8 points. The treatment group also scored 0.8 points more on average than the control group on the 10-point scale for knowledge about implementation of design sprints, with a range from 0.1 to 1.4 points.[6] (There was no indication of an impact on self-assessed general knowledge about design.) There was also some indication that there may have been a positive effect on attitudes towards the use of design techniques: those in the intervention group scored on average 0.4 points higher than those in the control group (again on a 10-point scale), though the range of results consistent with the data ranges from a negative effect of -0.5 points to a positive effect of up to 1.3 points.

There is no indication of an impact from the intervention on aspirations or expectations of the adoption of the design techniques in participants' companies, with the difference between the treatment and control groups in these respects being close to zero.

[6] These estimates are taken from Table AIII.2 in Appendix III of Deliverable 4.4 for the 200SMEchallenge project: IGL believes that the estimates in the appendix are more accurate than those presented in the main body of the report, though they are very similar in magnitude.

However, the small sample size of this trial limits the potential to identify such an impact. For example, the data are compatible with changes in expectations about adoption ranging from a decrease of 0.7 points to an increase of 0.6 points (again when considered on a 10-point scale).

Figure 3: Distributions of two of the six outcome measures, by treatment/control group



Potential for further testing or scale up

The evaluation has provided support for the initial steps in the project's logic model: participation in the UX Challenge led to significantly increased knowledge and 'know-how' about the design sprint method, and possibly (though less clearly) to improved attitudes towards design thinking. Because of both the limited sample size and the limited time frame of this experiment, it is not yet known whether the UX Challenge will result in greater adoption of design thinking and user-centred design techniques within participants' businesses, or whether this will also lead to a more general increase in SMEs' capacity to absorb and apply knowledge from outside the organisation. The link between adoption of design thinking and improvements in SMEs' innovation capacity, competitiveness and growth also remains to be tested.[7] We would therefore encourage the project team to seek opportunities to test the UX Challenge at a larger scale, doing follow-up data collection after sufficient time has elapsed to observe whether companies have begun to use design techniques. Collecting more than one round of follow-up data would increase the power of the experiment, and also create the potential to track longer-term outcomes, such as impacts on competitiveness and productivity. The most significant challenge would be to manage the survey attrition rate, perhaps by investigating ways to motivate the control group to keep engaging with the project and providing data.

It is also important to note that the way in which the UX Challenge events were implemented under this project were altered considerably from the original plan, as a result of the COVID-19 pandemic. It is possible that switching back from online to in-person events and to the original format of a two-day workshop would result in a more impactful learning experience. A future test should ideally be conducted once it is possible to resume in-person events at the right scale. It may even be of interest to test the impacts of online delivery against in-person delivery of the intervention.

[7] A study in the UK found that businesses that had received support from the Design Council between 2005 and 2015 had higher survival rates and experienced higher growth in turnover and employment, compared to a matched control group (Bonner, K., Hart, M., Heery, L. (2017) 'Design Council Support and Business Survival and Growth'). However, given the limitations of an observational study like this one, these findings cannot be treated as definitive.

Finally, the 200SMEchallenge project sought to evaluate the impact specifically on the SMEs themselves; it would be of interest in future replications to assess the impacts on the design professionals and on the students and recent graduates who also participated in the design sprints.

Wider learning

Survey attrition

The response rate to the follow-up survey among the control group, at 63%, was higher than in some other trials of business-support programmes (particularly among a control group that was not receiving any form of support), but even so it represents the weakest point of the 200SMEchallenge evaluation. One interesting approach adopted in this project was to invite participants (including the control group) to join a webinar about the design sprint in return for completing the follow-up survey. However, it is not clear that access to this event was a significant motivation: the number of respondents who completed the survey after being sent a reminder of this opportunity was small.

Policy application

Two other INNOSUP-06 beneficiaries are piloting interventions related to design thinking and open innovation. Create4Value is piloting a new approach for engaging SMEs who are first time innovators in co-creation with stakeholders and end-users, whilst DCS-iSMEs is exploring the feasibility of a new innovation support scheme to encourage SMEs to enhance their operations by adopting design thinking approaches. Results from these projects will be incorporated into future outputs.

For agencies wishing to experiment with a similar intervention, the project team have produced a playbook[8] with detailed guidance for running the UX Challenge. As noted above, this trial represents an extension of other work by Hub Innovazione Trentino on supporting SME innovation through design thinking, user-centred design and open innovation. Innovation agencies considering creating innovation support schemes for SMEs based on the open innovation paradigm may benefit from the actionable guide delivered by the INNOSUP-05 funded INNOCHALLENGE project (coordinated by Hub Innovazione Trentino). The guide builds on reasoning and data developed during the peer-learning activity, and also forms the basis of a recently published peer-reviewed research study[9].

[8] https://www.200smechallenge.eu/deliverables/
[9] Doppio, N., Väinämö, S., Haukipuro, L., (2021). Design elements of innovation contests supporting Open Innovation in SMEs – An action research study. Journal of Innovation Management 8(4):26-56. https://doi.org/10.24840/2183-0606_008.004_0003

D3T – Data-Driven Digital Transformation

Coordinator: Torino Wireless Foundation (TOWL)

Further details: Data Driven Digital Transformation | D3T Project | H2020 | CORDIS | European Commission

D3T Homepage

Research question: Does offering a data driven approach to the Digital Transformation support services improve SMEs' readiness and the timeliness of the DT implementation? Does the data-driven approach allow the innovation agency to deliver support services more efficiently?

Rationale and project logic

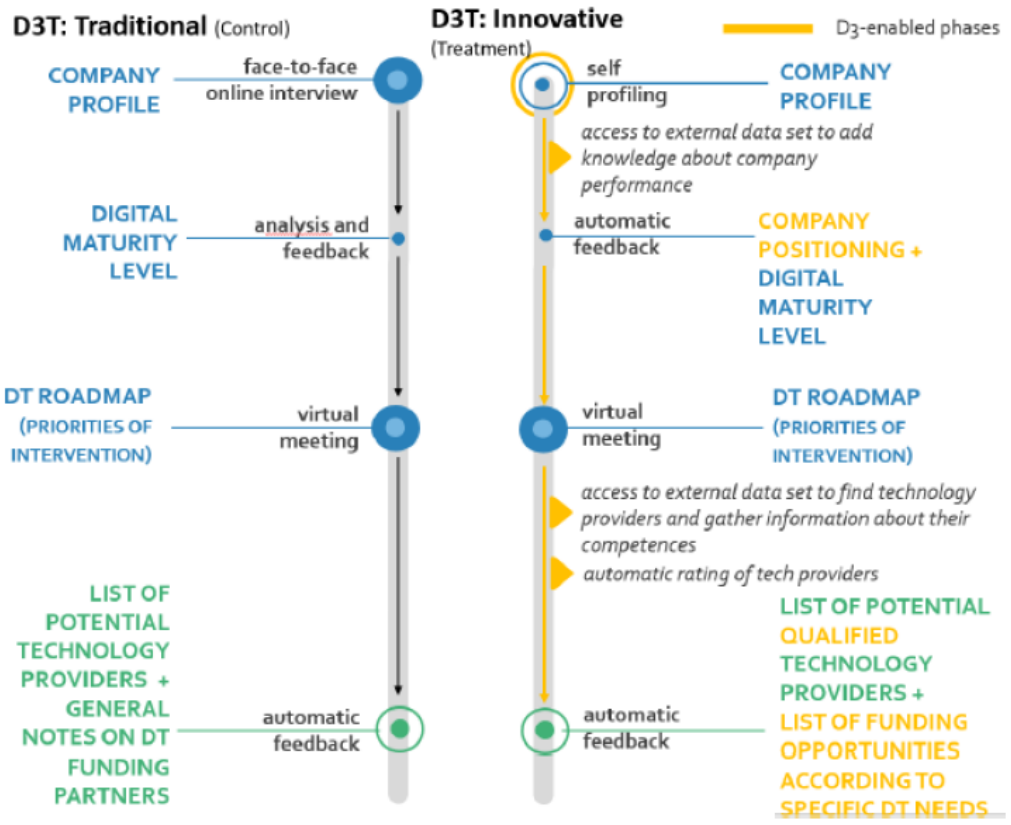
The D3T project was prompted by the recognition that many companies are not taking full advantage of the advanced technologies or the innovative business models offered by the digital economy. Transformation offers new business opportunities for SMEs and can fundamentally change business models, but it affects the whole value chain from product development to sales and therefore is challenging for companies to apply it successfully. The process is very demanding for the agency as well, as the support provided to achieve the digital transformation is intense and ambitious.

TOWL aims to deliver data transformation support using a different approach from their traditional service scheme. In particular, by implementing a data-driven approach to exploit the value that information and knowledge can generate, it makes innovation services more effective and the support process more efficient. At the same time, small businesses would benefit from an online and data-driven approach without demanding so many resources from the agency. The idea underpinning the D3T project is to test the capability of a data-driven approach in providing more effective support to companies and delivering services more efficiently as it reduces the resources demanded to deliver the support.

Intervention

The Digital Transformation Support Service provides a first assessment phase, where companies evaluate their digital readiness levels. This is followed by a planning phase where companies are provided with information on the actions needed to implement the roadmap, technology partners and providers, investment, and funding opportunities, as well as expected changes at organisational and business levels. The treatment group received an online and data-driven version of the service through the D3 platform. They autonomously assess their digital level through the platform and have access to external data sets, where the scouting range for providers and opportunities can be enlarged out of the existing TOWL network.

Figure 4: D3T traditional and innovative delivery



Evaluation design

The original policy question that motivated the policy experiment was: 'Does offering a data-driven approach to the Digital Transformation Support Services improve SMEs' readiness and timeliness to implement digital transformation?'.

Recruitment of small businesses located in the Piedmont region (Italy) was carried out through an open call published online and using direct contact with companies that were part of the local network of TOWL. Those recruited were already aware of their need for digital transformation but were seeking support to help them assess if and how their business would benefit.

To perform the evaluation, TOWL ran a small-scale pilot based on the principles of an RCT approach. 26 companies were randomly allocated into two groups: The treatment group received the D3 online data-driven version of the support service, while the control group received the traditional support service provided by TOWL. A set of primary and secondary outcome measures were selected to assess the feasibility and effectiveness of the support provided, such as readiness or time reduction in digital transformation implementation as well as number of collaborations and levels of adoption of the solutions provided. In addition to the main outcomes, TOWL also gathered some monitoring information and qualitative feedback based on measures of time demand to the analysts providing the service and the companies involved.

Baseline and follow-up surveys collected the information from both groups enabling comparisons of progress. However, it is not possible to draw robust conclusions about the impacts of the D3 intervention. Firstly, due to logistical problems, treatment and control groups received the support services at different points in time. This delay in treatment creates some complications for comparisons, as differences in the wait to receive support, gap between support use and outcome surveys,and any changes in the economic context when it is delivered[10], could influence the effectiveness of the intervention alongside any difference caused by the D3 element itself. Secondly, this was a small pilot meaning that groups were not fully comparable after randomisation and it wasn't possible to generate sufficient statistical power to provide confidence in the findings. Nevertheless, through the additional analysis, the evaluation provided useful information about the feasibility of the D3 intervention.

Impacts

While the project did not produce robust evidence about the impact of the D3 data-driven support service, there seems to be a positive evolution in the willingness of companies to invest more in the adoption of digital technologies after receiving the service for both the treatment and control groups. Feedback provided by the companies states that most of them found the services very useful (77%). Including other insights from qualitative research, the data-driven support seemed to have some positive effects on readiness to implement digital technologies. Results indicate that potential benefits may be provided on awareness of new opportunities, as well as on willingness to invest more in digital technologies. Although the effects for the data-driven approach were unclear on reaching new collaborations and improving the time to reach data transformation.

A relevant variable to consider seems to be the companies' self-awareness of their needs, and their openness to adopting new solutions and ways of thinking. In this intervention, companies who were more aware of their needs and more open, were better able to receive support. Nevertheless, the intervention seemed to be more effective for companies that did not have much technological knowledge, as they are in need of an assessment and matching of solutions (and therefore found the support service more useful). Some companies also reported that the roadmap including solutions and potential suppliers was particularly helpful. Therefore, companies who face a lack of time and the resources to carry out the analysis of potential solutions could also benefit from the support provided.

In addition to the main results, monitoring information and qualitative feedback suggested that the intervention had a strong effect on reducing the amount of time spent by TOWL supporting each company compared to the control group - with support through the traditional format demanding around two and half times as much input from the analysts. With companies receiving the digital support also spending less time using the support. If the D3 online version could demand less resources and time then it would only need to deliver comparable outcomes to the traditional support to be preferable - i.e. would be more efficient even if not more effective.

As above it is difficult to draw clear conclusions about the relative effectiveness of support. Based on the full range of evidence collected, the project team felt that a

[10] Implementation of the traditional intervention started in April 2020, with online support commencing several months later, after the immediate shocks from the COVID-19 pandemic.

complete digitisation of support would not be effective. The experience of those in the treatment group indicates that additional personalised support would be required at the stages of understanding needs and prioritising actions.

Potential for further testing or scale-up

This project has functioned as a useful proof of concept where TOWL provided two different approaches to delivery of the support scheme. The data-driven approach is a trend that innovation agencies can take advantage of to create value and enhance their operations, from the optimisation of the support services provided to the creation of more efficient plans for their members and networks. However, TOWL learned that the data-driven support offer is generic and not very personalised, and is therefore appreciated more by companies with low digital maturity. Companies that are at the beginning of their digitalisation processes and not directly related to the ICT sector may benefit more. TOWL has taken this into consideration as they design the support service that will be provided in the future.

From the analyst's feedback, a mix of both services would make the best support for the assessment and planning phase of digital transformation support services. The digitalisation of the full service is not as efficient as they first thought, as the human factor is really appreciated and makes a significant difference in some stages of the support provided, for instance not only in the interaction with the companies but also in the decision making and analysis processes. Therefore TOWL used these lessons to develop a new service merging elements from both groups.

InReady – Designing the Service to Improve the Investor Readiness of Start-ups

Coordinator: Lithuanian Innovation Centre

Participants: Foundation for Research and Technology Hellas, Agenzia per la Promozione della Ricerca Europea

Further details: Designing the Service to Improve the Investor Readiness of Start-ups | InReady Project | H2020 | CORDIS | European Commission

InReady Homepage

Research question: Does the use of the InReady tool improve the quality (or success rate) of entrepreneurs' pitches to venture capital investors?

Rationale and project logic

For small and innovative businesses, growth can often be constrained by difficulties accessing suitable finance. Many interventions seek to address concerns about the limited supply of finance, leading to for example, decisions to establish 'hybrid' venture capital schemes where public investments are used to draw private finance into the equity gap. However, effective policy solutions are also needed on the demand side with evidence that the growth of innovative small businesses can be constrained by a reluctance to seek external finance or difficulties to develop opportunities to a stage that can readily attract outside funding.

The InReady project was prompted by the observation that many SME founders are unable to deliver effective pitches to potential investors. When seeking venture capital or participating in competitive funding processes – such as the EIC Accelerator – a convincing pitch can make a major difference to the outcome. By providing founders with support in preparing an effective pitch, the InReady tool aims to ensure that more funding is directed to the start-ups with the best business ideas and the greatest potential for growing into successful businesses.

Intervention

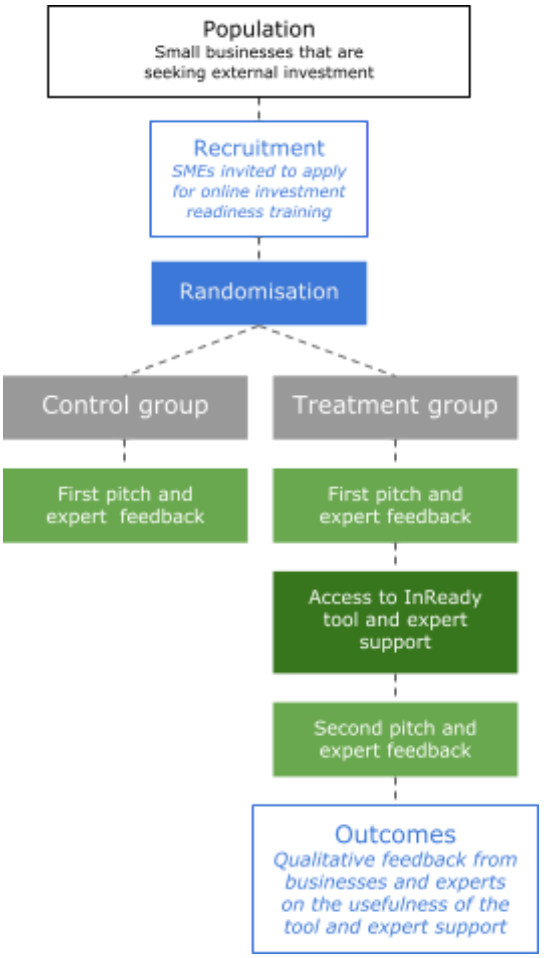
InReady is an automated web tool that uses a standardised set of information provided by the user to generate a slide deck and a 'state of play' document to support founders in their pitches. Data is provided on the business and the team, the problem that they are seeking to address and the solution or product, the market and competitors, their financials, and information about intellectual property and other resources. The interface is designed with gamified elements, to show users what progress they have made and to motivate them to continue. Upon completion, the slide deck and the 'state of play' document are generated automatically and can be downloaded immediately.

The tool also includes a function for users to send their information to an expert reviewer, who will then provide detailed feedback.

Evaluation design

27 businesses were recruited into the trial across the three countries. The businesses were at various stages of development, but most were providers of a new technology or an online service. These businesses were randomly allocated into a treatment group and control group, stratified by country. All participants were initially invited to present their pitch to a group of experts, who then provided feedback on areas for improvement. The treatment group were subsequently given access to the InReady tool, which they used to generate a slide deck and a 'state of play', and were also provided with more detailed comments by the experts. The treatment group were then given a second opportunity to pitch to the group of experts. The participants and the experts completed feedback forms at the end of the process, asking how useful they found the InReady tool and the expert support, as well as about any suggestions for improvement.

Figure 5: Diagram of the evaluation of the InReady project



It is not possible to assess the effects of the InReady intervention by comparing the experience of the treatment and control groups for two reasons. Firstly, the treatment group were given two opportunities to pitch, and could naturally be expected to improve with more preparation time and more practice, whereas the control group were given only one opportunity. In effect, with outcomes only measured after both pitches, this extra opportunity to practise a pitch and receive feedback also forms part of the

intervention provided to the treatment group. Secondly, there was no data collected on outcomes that would have enabled a comparison to be made between the two groups. For these reasons, the evaluation relies on the perceptions of the treatment group and the experts about how the InReady tool enabled them to improve over time.

Had these two issues been addressed it would have provided richer information on how the new tool affected the outcomes of participants. Nevertheless, it still wouldn't have made it possible to draw confident conclusions about the tool's causal impacts on outcomes as the sample size in this small pilot was not sufficient to generate the necessary statistical power.

Impacts

While the project did not produce robust evidence about the impact of the InReady tool and the associated support, feedback received from participants was highly positive. Users commented that the InReady online tool enabled them to improve the structure and focus of their pitches, with the market and financial analysis being identified by several participants as particularly valuable. The support from the experts was appreciated for guiding users through the use of the tool. The experts themselves also saw value in the InReady tool, though they noted that in its current state it is particularly useful to start-ups with little or no experience in pitching: additional content would be required to add value for more advanced users.

Potential for further testing or scale up

This project has functioned as a useful proof of concept: the project team have demonstrated that it is possible to develop an automated tool that (in conjunction with guidance from experts) SMEs find valuable in preparing pitches. Recognising the limitations of the existing InReady tool, the project team are seeking funding to further develop it, to add additional features (such as benchmarking) and to make it available in languages other than English. If such amendments can be made, IGL would then recommend another phase of testing the tool, to understand how much value it adds to start-ups' pitches before deciding whether to invest in rolling it out more widely. Such a test could follow the approach adopted in this project of inviting start-ups to practice pitching to a panel of experts, but with a random subset of the start-ups being provided with access to InReady in advance of the pitch session, and with the experts being asked to rate the quality of the pitches without knowing which of the companies had received access to InReady. To assess the scalability of the intervention, it would also be interesting to test whether the expert review and detailed feedback is a necessary component of the system (and if so, whether this can continue to be delivered online or is more effective if delivered in person), or whether the automated outputs alone are of value to start-ups.

Those seeking to replicate the approach or conduct their own experiment with training to make SMEs may be interested in the findings from an earlier randomised experiment undertaken in the Western Balkans[11]. Here the treatment intervention provided a combination of training, mentoring, master classes, and networking, with the control group receiving an online package. The more intensive combination of support led to an

[11] "McKenzie, D., Cusolito, A. P., Dautovic, E. (2018). ' Can Government Intervention make firms more investment-ready? A randomized experiment in the Western Balkans'. World Bank Policy Research Working Paper."

improvement in the judges' assessment of investment readiness and some evidence of later benefits to their ability to attract media and investor interest. An interesting question for further exploration is whether an online tool such as InReady can be as powerful, or at least as cost effective, as intensive and less scalable training.

4) Findings and recommendations when running an RCT

In this section we set out the challenges, anticipated and otherwise, that the innovation agencies involved in INNOSUP-06-2018 have faced when designing and running their experiments, and that IGL has provided them with support for. Building on the learnings from these challenges, we also pick out the recommendations for other agencies that seek to replicate the approaches and run their own trials.

Recommendation 1 - Allocate sufficient time and resources for trial design

An experiment is more than simply 'trying something new' and later looking back to identify any learning. Instead, the focus is on intentional learning. A policy experiment has a clearly structured approach to learning - defined before the experiment starts rather than afterwards - and it generates new information, evidence or data. A rigorous policy experiment will have a theory of change, systems and processes in place to capture learning, and a clear timeframe with limits or checkpoints where results will be assessed and decisions made about whether it should be adjusted, scaled up or discontinued.

For an experiment to work, therefore, you need to plan a lot of things in advance and this is especially true with a randomised trial. Many INNOSUP-06 beneficiaries, used to evaluation being considered much later, found that they hadn't allowed the required time and resources at the start of their project plans to complete the design of their experiment. This led to project delays or only finding issues too late, such as gaps in outcome measures and data collection.

Agencies looking to run their own experiments should ensure that they set aside sufficient time and resources to fully develop their own trial designs. INNOSUP-06-2018 beneficiaries were encouraged to complete a Trial Protocol before commencing their experiments.

A Trial Protocol sets how an RCT will be conducted and results reported, and it is produced in advance to:

- Ensure all elements are in place and are aligned;
- Provide confidence that the trial will be able to answer the key questions of interest;
- Establish credibility.

A copy of the template IGL provided to beneficiaries can be found in Annex A. However, the process of running a successful experiment will often start even earlier. Section 5 sets out the process that agencies can follow when advancing from the initial policy idea or problem identification through to responding to the findings from a completed evaluation.

Recommendation 2 - Take time on the question if you want a useful answer

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, or indeed if they really need to do an RCT at that time.

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'?'. RCTs can go beyond this but we found that many projects were attempting to design their trials around a policy question that was too broad or complex. 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. Other methods (e.g. qualitative surveys) are used to gather insights on the wider policy and implementation questions that cannot be directly answered by results from the trial itself.

Formulating a researchable question is a critical step for facilitating good research. Ideally, trials should have a single primary question around which to focus the development of the study design and sample size estimates. We encouraged project teams to use the PICO (Population, Intervention, Control and Outcomes) framework[12] when refining their research question. This approach helps you to be clear about who is being compared; what you are comparing in your study and what the outcome that you are trying to change is. Additional questions can then be added making sure not to overload the study with complexity and data collection.

On occasion, research questions needed to be adjusted as trials were further developed and it was determined that answering the question wouldn't be feasible or useful. For instance, one project had hoped to determine not just the overall impact of an intervention but also to use their trial to answer questions about the best approach to delivering support. They planned to do this by allocating participants to trial arms that would receive different forms of support but lacked statistical power. Without means to address this by increasing their sample meant that they instead had to focus their research question on the overall impact of a single intervention.

Recommendation 3 - Find outcomes that are meaningful and observable

An outcome is the change or impact caused by the programme being evaluated. At first, outcome measures chosen by the projects were typically described in very general terms (or 'topic' area). They were meaningful from a policy perspective but not sufficiently specific to be converted into indicators for which data could be collected.

For instance, we worked with teams to progress from a broad objective to increase levels of innovation amongst SMEs to a set of specific survey questions connected to the intervention being evaluated.

The table below shows the steps that need to be considered in order to progress from a general topic area of policy interest through to the data that will be used in the analysis.

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

Topic	What is 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)	What indicators can you 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. combining several indicators into a single variable

It is important for the success of a trial that indicators 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.

Some other points to consider are:

- Objective indicators are typically more meaningful and easier to interpret than subjective indicators - 'Are you good at setting targets?' vs 'How many targets does your business have? How often are they reviewed? Are they shared with employees?'
- Outcomes measures can be combined - Rather than pick one innovation practice out of 20, why not create a measure for how many of the 20 are used? This approach has been used effectively by projects such as the 200SMEchallenge to reduce the number of measures being assessed and improve their sensitivity to change.

- Likert scales are best avoided - All businesses may not interpret the scale in the same way, so changes can be difficult to interpret.

Ideally, projects would be able to adopt tried and tested approaches to measuring impacts. For example, the DINNOS project used a range of survey measures that have been validated in academic studies[13]. However, most found it wasn't possible to identify proven indicators that aligned with their outcome measure, context and the recommendations above. In these cases we encouraged projects to pilot the use of their measures. We believe there is a more general need for a review and collation of potential indicators that could be used within trials for evaluating innovation policies. This is something that we intend to pick up through wider activity at IGL.

Recommendation 4 - Consider when to measure as well as what

When choosing outcome measures it can 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. This can highlight the different dependencies and timings of outcomes.

In many cases projects had little existing evidence to determine what time period they should leave between the delivery of the intervention and measuring outcomes. Taking repeated measures would be useful but this has budgetary implications and careful management to avoid survey fatigue.

Timing data collection too soon may mean that a study misses outcomes that occur after measurements have been taken. On the other hand, waiting too long adds unnecessary time until results can inform policy decisions but also risks causing attrition, as participants become harder to track and less willing to respond to surveys.

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. If timescales cannot be extended then it may be possible to review logic models and related research to identify alternative outcome measures.

Recommendation 5 - Get comfortable with power

If we want clear guidance on whether an intervention works as intended it is important that the trial is designed with sufficient statistical power[14]. A well-designed trial will ensure that, if the intervention has an impact at a scale that is meaningful for policy decisions, we can be confident that the trial will detect it. If the trial then produces no evidence of impact, we can be assured in our assessment that it is the programme that failed (in that the impacts are too small to be relevant for policy), rather than the evaluation.

[13] See DINNOS project deliverable 5.1 https://cordis.europa.eu/project/id/824217/results

[14] Edovald and Firpo (2016) 'Running Randomised Controlled Trials In Innovation, Entrepreneurship And Growth: An Introductory Guide'

Statistical power was often overestimated or not adequately considered. This can have very consequential implications for the project. Overestimation can stem from being over-optimistic about factors such as:

- Level of compliance - i.e. will businesses assigned to the intervention use it as intended? Will those assigned to the control group not get access to the treatment?
- Attrition - i.e. will participants remain in the trial? How many will respond to follow-up surveys?
- Sensitivity of the outcome measures - e.g. if many businesses are already using a technology at baseline, a yes/no measure of adoption of that technology will have little potential for demonstrating any change.

In the course of supporting the INNOSUP-06-2018 project teams, we have discussed a range of measures that they could take to maximise statistical power and make sure that they retain it. These strategies include:

- Increasing the sample size[15];
- Ensuring high take-up of tested interventions and survey response rates;
- Reducing 'noise' - Using stratified randomisation; measuring outcomes more than once; collecting pre-treatment values of outcome variables or other relevant variables that are correlated with the outcome;
- Reducing the number of comparisons - limiting treatment arms and subgroup analysis;
- Increasing detectable impacts - increasing the sensitivity of chosen outcome measures; making the treatment and control conditions very different or ensure high fidelity of implementation.

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

- Assess the potential demand for and feasibility of implementing 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 and systems for handling randomisation;
- Improve monitoring & evaluation processes and data systems;
- Design and develop plans for carrying out a larger-scale trial in the future.

Recommendation 6 - Plan the recruitment and randomisation journey

Randomisation is the cornerstone of an RCT. Allocating support with the 'toss of a coin' sounds simple but when it came to planning its implementation project teams found there were many factors to consider as they sought to balance what was best for the research with what was most practical.

[15] The sample size that is required will depend on a number of factors. The following provides an accessible discussion of the key factors: https://www.povertyactionlab.org/blog/5-21-18/six-rules-thumb-understanding-statistical-power

Most importantly, from a delivery perspective, randomisation creates complications when recruiting SMEs to participate in a trial.[16]

Randomisation should only take place after a business has committed to participate in and has provided baseline data. However, this means that it is not possible to inform businesses about exactly what support they will receive in the trial – or, in some cases, whether they will receive any support at all. Some projects (such as the 200SMEchallenge) have dealt with this by informing potential participants openly that this is a research project with random selection and that there is a chance that they will not get to participate. This, however, can understandably deter many potential participants and gives those who are allocated to the control group little incentive to respond to later surveys. Alternatively, in the case of trials in which the treatment and control groups will receive alternative forms of support, the description of the support to be provided can be kept quite generic. Another interesting approach is that adopted by the DINNOS project, in which the treatment and control groups are being provided with the same forms of support, but in a different sequence: this should increase the motivation for businesses to participate, but with the consequence that it is only possible to assess short-term impacts of the programmes being tested.

A further complication can arise if eligibility checks need to be carried out. Again it is preferable to check eligibility before randomising - otherwise statistical power is lost with participants who cannot use the interventions having to remain as allocated within the analysis. However, project teams found this was not always possible. For instance, for RCT4MANU full eligibility checks relied on data that was only to be collected once intervention status was allocated.

Most projects sought to recruit the full sample of SMEs for their trial before randomly allocating them between treatment and control groups. As well as being administratively simpler, this also makes it easier to use stratified randomisation, which helps avoid imbalance between the groups in smaller trials. However, this approach led to difficulties in some cases (notably the DINNOS project), as slower than expected progress with recruitment meant that SMEs that signed up early in the process were kept waiting and eventually lost interest in the offer. This problem can be avoided by randomising on a continuous basis or in cohorts/batches, at the expense of making it more difficult to use stratification.

Recommendation 7 - Keeping delivery consistent

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

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

In the past, IGL has found that sometimes heterogeneity across time and place can reduce the effectiveness of the intervention, or make the results noisier and therefore leaving the trial less likely to provide clear findings. It can also make the results more

[16] See IGL's guide to randomised controlled trials for discussion of the concept of randomisation, the main approaches for how to implement it in practice and their implications for your research.

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. We are aware that the DepoSIt and 200SMEchallenge projects, in particular, have involved a great deal of coordination between the implementation partners in multiple countries, to ensure that they were following the same approaches 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.

Recommendation 8 - Check interactions with wider innovation support

Field experiments like those run by INNOSUP-06-2018 beneficiaries are not being run in laboratories under controlled conditions. This creates challenges as in the field the delivering of the tested intervention can interact with other activities undertaken by the innovation agencies or other stakeholders.

For example, the RCT4MANU team had to determine how the recruitment and delivery of their intervention would affect and be affected by a number of other programmes, with public funding, that were also offering tech adoption support to UK manufacturers. In their case, the complications proved to be of limited consequence with only marginal overlap in both the types of SMEs that were being targeted and the forms of technology being encouraged. They may, however, have been less fortunate and found that these other programmes had much more in common. This would have raised concerns that the control group could have received comparable support elsewhere; if everyone is able to receive subsidised support by accessing other sources, the trial wouldn’t have provided any results on the interventions additionality and impact of subsidised interventions.

Recommendation 9 - Don’t assume that just because you built it they will come

Never underestimate the challenge of recruiting the right type and number of SMEs. Achieving recruitment targets has been one of the most common and toughest challenges that projects have faced, particularly those running the largest experiments. 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.

It is not a good use of resources to design a trial to evaluate the impact of an intervention only to learn that there is little demand for the intervention from the intended participants. Uncertainties about the ability to engage participants should be resolved before a full trial is undertaken – ideally through testing recruitment in a pilot.

That being said, difficulties can occur for unexpected reasons and so even when confident about recruitment, projects should consider how they would respond if recruitment numbers are different from what they expected. For example, it may be

necessary to reduce the number of 'arms' of the trial, so as to focus on the single comparison of greatest interest.

Recruitment is not just a matter of numbers. When only a relatively small sample is achievable, 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 amongst those best placed to benefit would be more supportive of investing greater resources to extend the trial to test the benefits for more marginal cases.

Recommendation 10 - Don't assume just because they come they will use it

Recruitment struggles were not the end of the implementation challenges faced by project teams, with several finding that uptake of the intervention fell short of expectations.

There were a number of explanations for this:

- Over-estimating the attractiveness of the intervention - even without the economic fluctuations, businesses were found to be more circumspect than expected about the benefits of investing their limited time using the support.

- Implications of trial design - with elements of support being randomised, many projects were not able to provide full details of the support being provided, making the offer appear less attractive to potential participants and harder to market.

- People were kept waiting - Interest eroded when there was a long pause between when businesses had submitted an application and when they were able to start using the support.

- Rapidly changing conditions - The disruption and uncertainty caused by the global pandemic accentuated the preceding issues. Businesses that signed up to participate in a project with the hope of boosting long term growth may have found that they were suddenly overwhelmed with more immediate issues - e.g. a rapid bounce back in orders as economies reopened.

Based on these experiences, we would make these recommendations to agencies in the future:

- Test assumptions about user-need and conversions from application to use.

- If running a randomised trial, consider whether it is possible to randomise after the intent to actually use the intervention has been tested.

- Limit the time between recruitment and delivery, especially when user-need is deemed time-sensitive. When running a randomised trial, consider implementing randomisation on a continuing basis or in cohorts, rather than waiting until the whole sample has been recruited.

Recommendation 11 - Expect the unexpected

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.

All thirteen projects were hit by the unprecedented shock of the COVID-19 pandemic. Health concerns and social restrictions made it impossible to deliver many of the planned interventions. Business needs and the priorities of innovation agencies were also changed dramatically.

IGL worked with project teams to consider; whether it was feasible to continue the delivery of the intervention as planned; ways to make the delivery of the intervention feasible (e.g. switch to online rather than in-person workshops); the potential to meet urgent business needs; and whether there may need to be a change the approach to measuring impacts.

Ultimately all project teams were able to move ahead, with several, such as 200SMEChallenge and DINNOS, having to rapidly develop new forms of interventions. Unfortunately despite best efforts, many will not be able to generate the depth of evidence they had planned for, due to difficulties recruiting and delivering support to the intended population.

Hopefully, those replicating the approach for their own experiments will not encounter anything like the same scale of shock. That being said, it may be worth at least considering how decisions, including to end the trial, would be made about the future of the experiment if something unexpected occurs.

Recommendation 12 - Invest to optimise survey response and data collection

A perennial challenge in trials of business-support interventions is in achieving good response rates to surveys. High levels of survey attrition undermine the results of a trial by reducing the sample size available for analysis and hence the ability to detect differences in outcomes between the treatment and control groups. Even worse, if levels of attrition are unequal between the treatment and control groups (as has happened in cases like the 200SMEchallenge, in which the control group had little motivation to respond to the follow-up survey), this may introduce bias into the estimates of impacts. It is therefore crucial to maximise survey response rates, and to take action if rates seem to be declining.

There are no simple solutions to the problem of attrition, and IGL will be looking to learn from the experience of INNOSUP-06-2018 projects over the coming year. Two points that seem to be important are to consider the motivations of businesses to answer surveys, and to minimise the length of questions. It also appears that SMEs can be more receptive when they have a sustained personal contact at the agency carrying out the survey. For example, the DepoSIt project achieved a high response rate to their first follow-up survey, even among the control group who have received no direct benefit.

Another way of approaching this challenge is to assess outcomes using sources of data other than surveys[17]. The DepoSIt project is experimenting with an interesting approach, observing whether businesses in their trial have begun posting about social innovation on their websites or social-media accounts, as a complement to their survey-based measures. Another option is to observe how trial participants respond when they are offered further opportunities. For instance, when evaluating an innovation training programme a relevant outcome could be that it makes them more receptive to future support. This could be tested through a rapid-fire trial, where all participants in the experiment are emailed information about another new support programme, with responses to the email tracked to see if those who had been assigned to the treatment group the main training programme show a more positive response than those in the control - e.g. are they more likely click through for more information and start an application. Finally, there may also be opportunities (if the data-protection agreements allow) to make use of administrative data that is routinely collected by service providers in their course of work, or to measure productivity by matching participants to tax records collected by governments.

Recommendation 13 - Plan your analysis early and try to keep it simple

IGL encourages project teams to prepare a comprehensive trial protocol before launching, but it is often not possible to specify all the details of the analysis at that stage. However, it is still important for the robustness of the trial results to prepare a statistical analysis plan before final outcome data is collected. Having a clear analysis plan in place ensures that evaluators do not engage (even unconsciously) in specification search or 'fishing' by attempting various approaches to the analysis and only reporting those that produce positive results. Although the use of statistical analysis plans is a relatively recent innovation, it is quickly becoming standard practice among quantitative researchers, and a trial's findings may be called into question if the analysis was not prespecified. A further advantage of planning the analysis clearly beforehand is that it will enable the evaluators to produce results rapidly once the outcome data becomes available.

Recognising that most researchers do not yet have experience in preparing a statistical analysis plan, IGL has prepared a template for use by project teams. This template has been shared with a selection of INNOSUP-06-2018 teams, and will be revised based on their feedback (See Annex A).

One of the key points that IGL has been keen to communicate to trial teams – particularly to researchers who are used to other kinds of quantitative evaluation – is that the analysis is often simpler than they may expect. The beauty of RCTs is that randomisation takes care of a task that causes much of the complexity in other evaluations – creating a counterfactual. We therefore recommend that evaluators keep the analysis as simple as possible and prioritise transparency by, for example, including graphs of their data. We are gratified to see that the first of the large-scale projects to reach this stage – the 200SMEchallenge – has followed these principles and produced a very clear and comprehensible findings report.

[17] For example, http://www.eurito.eu/ is investigating new research and innovation analytics.

Recommendation 14 - Pilot, as there is no substitute for experience

The last recommendation is perhaps one of the most important. There is often no substitute for experience - use pilots to test key assumptions and learn more quickly and cheaply.

Many of the most critical challenges faced by project teams were the result of finding that untested assumptions did not hold once the intervention was being delivered - e.g. the ability to recruit, how businesses would use support or the survey response rates. While these issues often became apparent quickly, given the way project plans were structured and resourced made it too late to make changes.

With a new programme there is always a risk of failure but small scale pilots would have helped projects to ensure feasibility and identify immediate issues, making it possible to 'fail early and learn fast'.

5) Lessons and recommendations for becoming an experimental agency

Beneficiaries INNOSUP-06-2018 were motivated by learning how to improve SME innovation support schemes. Some had completely novel ideas, whilst others aimed to determine and optimise the effectiveness of existing programmes. Alongside evidence for specific interventions, agencies also saw INNOSUP-06-2018 as an opportunity to adopt a more systematic approach to evaluation and to generate evidence that can facilitate exchanges of best practices.

Participating agencies report that they now better understand how the approach can offer a robust method to think through and then test the logic underpinning their support schemes. However, this shift to becoming experimental has not always been easy and is still a work in progress. Some agencies reported significant barriers, including a lack of buy-in from senior leaders to run RCTs or limited expertise for design and implementation. The COVID-19 pandemic posed an additional challenge, as it forced many agencies to change their focus and quickly adjust their experimental projects. The above mentioned barriers relate to the agency's culture and capabilities.

The following section brings together some of the learnings from agencies' adoption of experimental approaches, in particular randomised trials. We have collated findings about the enabling conditions for experimentation within two factors: openness to experiment and capabilities to experiment.

These are broad and non-comprehensive factors, and units within a single agency can be at different levels. We hope that they serve as a useful guideline for agencies to recognise their strengths and weaknesses and to assess their experimental readiness. Further details can be found in IGL's earlier report '[Boosting experimental innovation policy in Europe]', where these findings were first presented.

Openness to experiment

An agency's openness to experiment refers to its willingness to learn, [being open about uncertainty] and the use of randomised experimentation as a form of policy design and evaluation. Agencies with high degrees of openness have a tolerance for risk-taking and recognise that failure can be inevitable when exploring new ideas.

An agency's degree of openness is influenced by a mix of internal (e.g. senior leaders' attitudes towards experimentation or the wider organisational culture) and external forces (e.g. demand from businesses or political pressures to prove impact). Based on the interviews with innovation agencies, we identified some characteristics that organisations with high degrees of openness have in common:

- **Experimentation champions:** More open organisations typically have champions at mid- and senior levels who advocate for experimentation. Some even develop communities of employees who work across areas to develop and test ways to improve programme design and evaluation.

- **Flexibility to try new things:** Legal or institutional constraints (e.g. strong hierarchical decision-making) can limit openness, as can a culture of risk aversion

and fear of failure. Agencies with a high degree of openness instead encourage testing new ideas even if to show they don't work.

- **Informing decisions with evidence rather than intuition:** Agencies with developed processes for identifying and using evidence to make decisions are more likely to be open to experiment. They will be more aware of evidence gaps and have expectations about what evidence is most relevant and reliable.

- **Embracing the benefits of randomisation:** Some organisations, although willing to test new ideas and use evidence, are not yet convinced by the value of RCTs. This is perhaps as they are focused on high-level ecosystem building and yet consider immediate questions regarding the effectiveness of specific schemes.

Strategies to increase openness

Based on the lessons identified above, there are a number of strategies that can help agencies increase their openness to experimentation:

- **Identify and mitigate internal concerns:** An aversion to RCTs often stems from a lack of familiarity with the method. For example, a typical concern was having to deny support to a control group. INNOSUP-06-2018 agencies facing this concern learned that trials can be designed without a 'control group' to compare different forms of support.

- **Showcase the impact of experiments:** Presenting examples from other agencies can help create an internal appetite for experimentation. Even small, low risk experiments can increase appreciation for the approach. Some agencies in INNOSUP-06-2018 told us they built their organisation's openness with small rapid fire experiments that provided 'quick wins'.

- **Identify an 'experimentation champion':** As mentioned, for some organisations an internal champion was central to developing an experimental culture. With support, such as that provided by IGL, champions can motivate colleagues to develop new ideas and get comfortable with learning from failure.

- **Advocate for legal and institutional changes**: Politicised administrations or with strong legalistic features can hamper attempts to introduce randomised experiments. Frameworks that allow for more flexibility and results-oriented strategies may foster more pro-innovation attitudes.

Capabilities to experiment

An agency will require access to the specific skills and resources to conduct experiments. Capabilities to undertake quantitative evaluation (e.g. knowledge of econometrics and data analysis) but to also embed research plans as they design, implement and monitor programmes. With experiments, evaluators need to be involved at all stages of policy development.

Capabilities can be sourced internally or externally. Some agencies that participated in the INNOSUP-06-2018 programme did not start with in-house expertise to run experiments, and found it invaluable to work with IGL and other evaluation partners to fill expertise gaps.

Some of the key capabilities required are:

- **Data infrastructure and availability:** Good data is crucial. Most agencies have monitoring and evaluation (M&E) systems in place, but too often they only capture process outcomes (e.g. number of businesses supported) or impact measures that are limited to self-assessed outcomes (e.g. a business owner's perception of the programme's effectiveness). The most developed M&E systems would cover objective outcomes (e.g. actual levels of investment) and use a range of data sources (e.g. matching to business records and novel sources such as public updates from company websites).

- **Research and evaluation capabilities:** Agencies with strong research and evaluation capabilities already routinely rely on research to improve programme design, and avoid making decisions purely based on intuition or political incentives. These capabilities may not be internal, but the agencies will be aware of what is required, able to access the external support and able to apply it to achieve a valuable outcome.

- **RCT expertise:** Some agencies had access to an internal research team that had experience with randomised experiments. Others, typically smaller, do not have capacity to run experiments in-house but could commission them. Even when the process is outsourced, familiarity with RCTs is still an asset to determine a realistic timeframe and allocate sufficient resources.

Strategies to build the capabilities for experimentation

Based on the feedback we received and our observations, there are a number of strategies that policy makers can use to increase the capabilities described above:

- **Invest in the necessary data infrastructure and research skills:** Agencies can rely on external support for running trials, but will need to be able to direct research to answer questions that will be of value to their decisions if they are to reap the full benefits of experimentation.

- **Find opportunities for peer learning:** As part of the INNOSUP-06-2018 programme, IGL organised regular workshops where agencies could learn from each other. Joining such a platform for peer learning can therefore be an important route to building internal experimentation capabilities.

- **Work with external partners to fill expertise gaps:** Where expertise gaps exist, agencies can work with external partners. Some agencies told us how beneficial it is to have a direct connection with University departments. Academics, such as those in IGL's research network, are often seeking opportunities to work with agencies, exchanging their knowledge and expertise for opportunities to test theories about innovation in the field.

- **Start small and learn by doing:** For an agency that is willing to experiment, but is still building its internal capabilities, starting with a small-scale, low-risk trial, such as rapid-fire messaging trials can be a useful way to gain expertise. Over time, the agency can become familiar with how randomisation can be applied and how experimental results can be reported.

Recommendations for running policy experiments

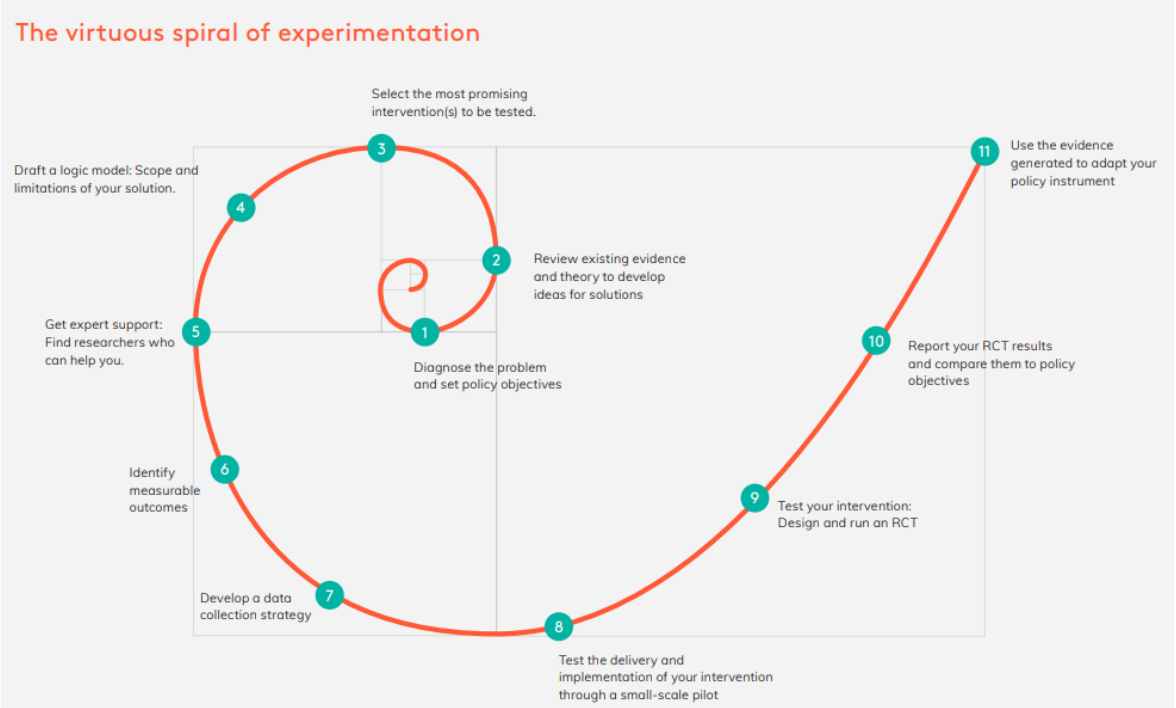
Even with the necessary willingness and infrastructure, running RCTs can be challenging as agencies need both the intervention and evaluation to work in tandem. Achieving this requires significant planning.

The need to recognise and account for this upfront demand was one of the key learning points for the INNOSUP-06-2018 projects. Public officials told us that they missed a clear pathway that they could follow to successfully set up their first experiment, which could potentially affect their openness and capabilities to experiment in the future. Sometimes, they acknowledged important steps that would have improved the quality of the experiment when it was too late.

In order to help innovation agencies avoid similar problems in the future, we present a spiral that seeks to describe the experimentation process, showing how to develop a policy experiment and avoid the risk of moving too soon.

As reality is often more complex, a project team may find themselves going back and forth along the spiral. This is an expected part of the innovation process - not all new ideas will work first time and each interaction can bring benefits, such as quality in the programme design and delivery.

Figure 6: The virtuous spiral of experimentation



1. Diagnose the problem and set policy objectives	The first step is to properly investigate the policy challenge or opportunity. Problem formulation can be supported by theory, qualitative evidence or design thinking. At this stage, it's crucial to form an initial understanding of the target group and policy goals (magnitude and nature of what you would hope to achieve through a policy intervention).
2. Review existing evidence and theory to develop ideas for solutions	Others may have tackled your policy challenge before or investigated the possible drivers. Review the literature for existing evidence, particularly from rigorous studies in different contexts. Other approaches can provide new insights to identify potential solutions, including design thinking, behavioural insights and participatory approaches.
3. Select most promising intervention(s) to be tested	As various potential solutions have been identified, those that are politically and financially feasible can be selected, using ex-ante cost-effectiveness assessments to narrow down to those intervention(s) with the highest potential for impact given the available time and resources.
4. Draft a logic model: Scope and limitations of your solution	Drafting a logic model will clarify the theory of change, with the causal pathways that you assume will connect actions to impact. It is essential that it openly states the limitations and assumptions of the model, which will help develop the experiment and be key to interpreting results.

5. Get expert support: Find researchers that can help you	Researchers and evaluation partners can help transform proposed solutions into feasible and testable interventions. Engaging with experimentation experts early on is advised to get a clear understanding of the costs and benefits of committing to the experiment and to develop a robust experimental protocol.
6. Identify measurable outcomes	The logic model will provide the basis for selecting measurable outcomes. Outcome indicators need to be comprehensive and should realistically capture the impact of an intervention within the timeframe of the experiment. This can be challenging given that it can take a long time to observe certain outcomes (e.g. a change in culture, business survival rates). Therefore, it is important to think about impact outcomes early on and specify realistic measures. Outcome measures will be tested and refined in the following steps, but should be finalised by the end of step 9.
7. Develop a data collection strategy	Before running an experiment, it is crucial to decide how the outcome data will be collected. Most commonly, surveys are used to collect data prior to ('baseline') and then again at least one after the intervention. If possible, it is advised to complement this with other data sources (e.g., administrative data such as business records, and qualitative data from interviews with participants). During this step, it is also important to develop strategies to avoid losing participants (attrition), which is a key challenge of running RCTs.
8. Test the delivery and implementation of your intervention through a small-scale pilot	It is recommended to start with a Proof of Concept pilot (PoC) to prove the feasibility of the intervention, identify any implementation challenges, assumptions made in the logic model and how outcomes will be measured. Depending on the results of the pilot, the intervention may need to be tweaked or completely reformulated.
9. Test your intervention: Design and run an RCT	Once the intervention is finalised and data collection systems are in place, it may be appropriate to run a randomised controlled trial (RCT) to evaluate the impact of the intervention. It's vital to set all experimental parameters in advance, such as how to randomise and the sample size required. A trial protocol records these parameters and provides the guide for the implementation of the RCT. For detailed guidance on setting up an RCT, see IGL's trial guide.
10. Report your RCT results and compare them to policy objectives	Once the data from the RCT are analysed, the results need to be reported in a transparent way so that it is clear whether the intervention had the desired impact and has met the policy objectives. To help others replicate and learn from your findings, reporting should go beyond impacts, considering whether the theory of change needs updating

	and lessons learned about recruiting and delivery. To get a broader understanding of the results it can be useful to take a mixed methods approach (e.g. include qualitative data).
11. Use the evidence generated to adapt your policy instrument	The results should be used to inform the design and delivery of current and future programmes. Ideally policy initiatives are scaled through a succession of trials to generate rigorous evidence of their effectiveness in different contexts. Organisations can build a culture of continuous improvement and learning by continuing to test improvements and evaluate impacts on interventions that are scaled.

6) Recommendations for future funding programmes for experiments

INNOSUP-06-2018 has clearly shown how a dedicated funding call can encourage innovation agencies to undertake policy experiments. By providing funding for both the intervention and evaluation in exchange for rigorous experimentation, the programme empowered internal champions in agencies to build support and find ways through the apparent barriers. It created a space for agencies to take risks with bold and innovative ideas, and subject the most promising to robust evaluation with results made accessible to wider policymakers regardless of their 'success'.

When it comes to the design of such calls, perhaps the main lesson that we can draw is the importance of helping innovation agencies understand the experimental methodology as they develop their proposals and not only when project parameters are largely fixed.

For most innovation agencies, INNOSUP-06-2018 was a rare opportunity to apply an experimental approach to policy development and learn 'hands on' about the use of RCTs. There are many factors for agencies to consider as they determine if, when and how to run an RCT. As outlined in the preceding section, this requires a change in the way most agencies undertake evaluation; the evaluation having to be planned in detail at the outset of the project, with a specific research question in mind and integrated into programme delivery.

IGL's initial project reviews and engagement with the teams led to a number of substantial changes in approach. Therefore, it's useful to consider that similar support is provided in future calls. However, agencies would also benefit from this additional engagement earlier in the process. For instance, it could be beneficial for potential participants to join more intensive workshops to test and refine their ideas before they submit their applications and the selection process 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, including the need for external evaluation partners.

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 they can be delivered and how a trial to evaluate their impact can be designed. When fundamental questions remain unanswered about the ability to deliver an intervention or measure outcomes, running an RCT is unlikely to be beneficial. Other approaches could more effectively and efficiently answer these initial questions. Funding calls can support this process by funding small scale pilots to ensure feasibility, perhaps as part of a staged approach where it is made possible to 'fail early and learn fast' inside a framework where ideas can progress to larger trials.

If the intention is to only support full impact evaluations then it is important to select projects where there is already confidence that delivery and outcome measurement are feasible. That does not necessarily mean that only perfectly designed trials should be funded. Trial designs can be strengthened 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 their experiments, it is also important to have clear expectations and requirements for each stage**.** For instance during early design the requirement to complete a trial protocol to a set standard and later, the parameters for analysis and reporting. We would also suggest that those managing a trial run first run pilots of the intervention and data collection before proceeding to the full trial.

The 'experimentation fund' approach, despite the challenges outlined above, has been shown to be a powerful tool to promote innovative and impactful evidence-based policymaking. However, there is also a need to develop agencies' openness and capabilities, maintain the incentives to scale ideas through an iterative process, and create rewards to share failures and to invest in rigorous evaluation that delivers the most benefit to others. The benefits from creating 'experimentation funds' are therefore likely to grow if they can become a long term feature of the policy funding landscape.

Annex A: Tools and resources

A1) Trial Protocol

A2) Statistical Analysis Plan

A1) IGL Trial Protocol Template

The following is a suggested template for trial protocols for those who are intending to run a full Randomised Controlled Trial.

This Trial Protocol is based on the SPIRIT statement[18], that sets out the items essential for study conduct, review, reporting, and interpretation of trials. The SPIRIT checklist includes scientific items that closely mirror the latest version of the CONSORT statement [19].

Each section contains suggestions of required content in both the boxes and footnotes[20]. Further guidance is available in IGL's guide to running RCTs or by contacting the IGL Team (innovationgrowthlab@nesta.org.uk).

The structure of the IGL template follows our approach for evaluating trial designs. Should issues arise in a section it is likely that these would need to be resolved by adapting plans in that section or earlier in the protocol. Therefore, we would recommend that you complete and discuss each section in turn rather than waiting until you have completed the whole template before seeking feedback.

1. INTRODUCTION

1.1 Complete project title	*Descriptive title identifying the study design, population and intervention.*
1.2 Trial registration	*Trial identifier and registry name. (It is now accepted best practice that all trials are pre-registered, we would suggest using AEA).*
1.3 Protocol version	*Date and version identifier* *Record any changes made to the trial design*
1.4 Roles and Responsibilities	*Names, affiliations, and roles of trial personnel*

2. MOTIVATION AND SETTING

2.1 Rationale	*Policy and research background and justification for undertaking the trial. For example, what evidence gap has been identified and what policy decisions are to be informed.*

[18] Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT).
[19] Consolidated Standards of Reporting Trials (CONSORT). Please visit this link for examples: http://www.consort-statement.org/examples/sample
[20] These can be deleted as you complete those sections.

2.2 Main Research Question[21]	*PICO Approach. Write down your research question that your RCT has been designed to answer, being clear about:* *Population - who are being compared* *Intervention/Control - what (and to whom) you are comparing* *Outcome – what do you want to see change*
2.3 Setting	*Where will this study be taking place?* *For example, will the study be confined to a specific geographic area or as part of an existing business support programme.*

3. PARTICIPANTS, INTERVENTIONS AND OUTCOMES

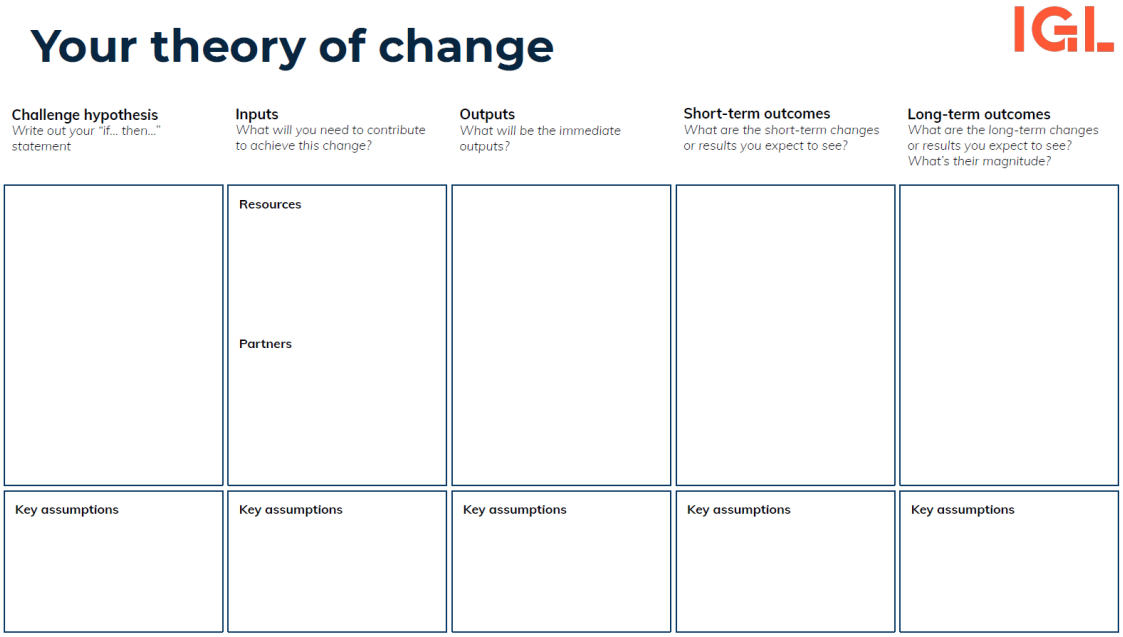
3.1 Participants	*Description of who is eligible and how they will be identified; description of exclusion criteria for participants if applicable.*
3.2 Interventions	*Details of the interventions for each group with sufficient detail to allow replication.*
3.3 Outcomes[22] [23]	*Clear definition of primary and secondary outcomes, 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, the proportion with score > 2), and time point of interest for analysis for each outcome.*

4. LOGIC MODEL

4.1 Logic Model

[21] Avoid setting too many primary questions to answer. Trials are typically powered to detect changes in a single primary outcome. Multiple primary outcomes require larger samples (or an analytical correction). Example: For SMEs (the population), does offering access to 30 hours of free business coaching alongside a grant (the intervention) lead to faster sales growth (the outcome) than offering the grant alone (the control)?
[22] If some of your measurement instruments (including composite scores) will be constructed, e.g. "attitude to new technology", please provide a description of how the outcome will be constructed from the main variables.
[23] As projects typically face sample and time constraints, we often recommend that projects identify proximate outcome measures that are key determinants of success - e.g. have SMEs advanced through different stages of adoption or delivered changes within their businesses that are expected to deliver positive impacts on productivity.

Setting out the underlying logic or theory of change and a set of assumptions about how an intervention works. [24]

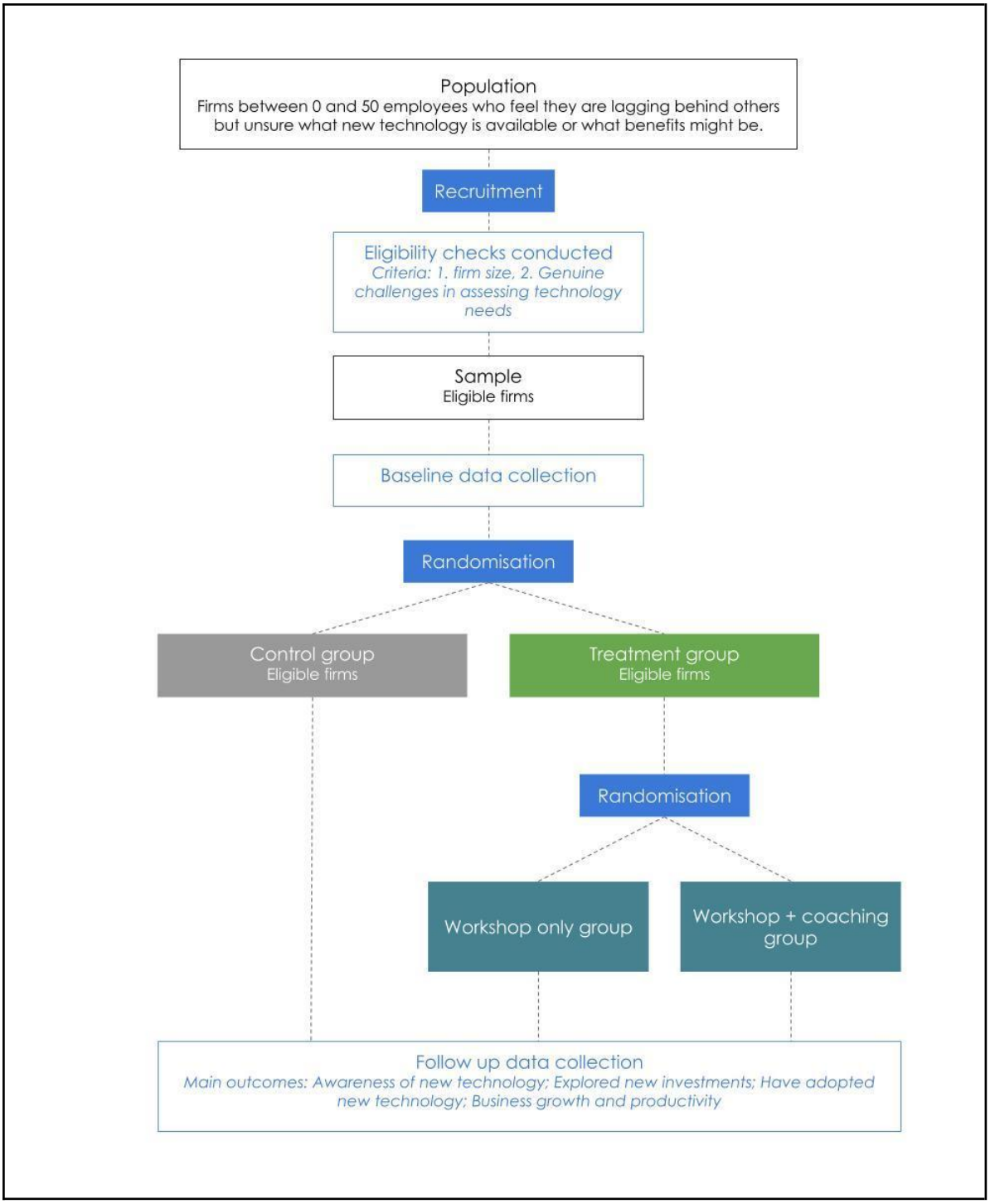


5. TRIAL DESIGN

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

5.2 Trial Diagram
Add a simple representation of the trial design. Below you can find an illustration of how this would look.

[24] A logic model will help to see the intended mechanism for change for each programme or policy and should also underpin what data needs to be collected for each evaluation. This can be helpful to identify the proximate outcome measures that can provide more timely measures than measures such as SME productivity. Logic models should also help to identify risks or contingency factors which may mean outcomes are not as envisaged.



6. RANDOMISATION AND ASSIGNMENT

6.1 Allocation Sequence	*Description of randomisation methods used to generate the allocation sequence* - *pure* - *stratified/blocked (please specify strata) - recommended* - *paired* - *cluster (please specify)*
6.2 Allocation Mechanism and Implementation	*How will you implement the allocation sequence - ie process by which assign participants to different arms of the trial.* *Who will be involved in this process.* *Will randomisation occur before or after baseline data collection and eligibility checks[25]?*

7. STATISTICAL ANALYSIS

7.1 Intended comparisons	*Please specify all the different comparisons to be carried out, e.g.* - *comparison of treatment groups along different outcome measures* - *comparison of different groups (e.g. treatment vs control, treatment 1 vs treatment 2 etc.)* - *comparison of different subgroups (e.g. male- vs female-owned businesses)*
7.2 Statistical methods	*Description of the statistical methods to be used to compare the groups on the primary and secondary outcome measures:* - *statistical test (e.g. t-test, chi2-test, linear regression with covariates etc.)[26]* - *treatment of standard errors*
7.3 Additional analysis	*Description of methods for any additional analysis (e.g. subgroup and adjusted analyses or mediation analysis).[27]*

[25] We almost always would recommend afterwards unless there is very good reason why this cannot be done.
[26] We typically encourage the use of two-sided tests for comparisons - if a one-tailed test is to be used, we ask that justification be provided.
[27] Please ensure that the analysis plan addresses all research objectives set out in the 'Objectives' sections above.

8. POWER AND SAMPLE SIZE

8.1 Sample Size[28]
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 below for examples of assumptions to consider.)*

Table 1: Potential assumptions relevant to sample size calculations [29]

8.2 Assumptions to consider	**Value/Details**
Criterion for statistical significance (probability level; typically 0.05)	
Power against alternative hypothesis (conventionally 80%)	
Allocation ratio, i.e. proportion of randomisation units assigned to treatment (e.g. 50% of the total sample assigned to treatment in a two-arm trial)	
Treatment compliance - participants switching treatment groups - participants in the treatment group deciding not to take up the offered program - participants dropping out of the experiment entirely (such that we no longer collect their data)	

[28] There are a number of factors involved in estimating the sample size, including:
Type and structure of the trial and research question; Effect size - 'Minimum Detectable Effect Size (MDES); The 'noise in the outcome variable' (i.e. unexplained variance).
Two approaches to estimation:
Fixed Sample: What is the smallest true impact that I can be confident of detecting given a sample of this size?
Target Impact: What sample size do I need to be confident of detecting an impact of this scale?
It is important to highlight that statistical significance is not the same as policy significance. What scale of impacts would inform your policy decision? For example, how much want to increase the proportion of SMEs who adopt technology x; % change in productivity or the ratio of net economic benefit to programme cost.
Ideally, want to align the Minimum Detectable Effect Size in the trial with policy significant outcome. So able to say with confidence whether impacts exceed or fall short of decision thresholds.

[29] Please use Table 1 to provide a summary of the key assumptions you have made, detailing and fully justifying your choices in the text below - where have the assumptions come from and why decided to use these (including why something is not relevant).

Number of individuals per randomisation unit (applicable to cluster randomised trials)	
Intra-class correlation (ICC) coefficient (rho) (applicable to cluster randomised trials)	
Proportion of variance in the outcome explained by covariates (R-squared) - if applicable	
Correction for multiple comparisons	
Software used for Minimum Detectable Effect Size calculations.	

Note: if the required sample size (or minimum detectable effect) is unrealistically large, i.e. the trial is unable to detect economically meaningful effects, we recommend reducing the number of comparisons (by dropping a treatment arm, or restricting subgroup analyses) and/or reducing the noise in the outcome measure (by using stratified randomization, more precise outcome measures or repeated measurements).

9. RECRUITMENT AND DATA COLLECTION

9.1 Recruitment	*How do you intend to recruit sufficient numbers to deliver the trial as designed and achieve your target sample size?*
9.2 Assessment of data collection	*Plans for assessment and collection of baseline, outcome and other trial data (incl. how and when - ie timeline for data collection and when data will be available).* *Have you considered any unintended effects of the intervention (these will be specific to each programme and should be thought through alongside the logic model/theory of change).*
9.3 Data collection instruments	*Description of data collection instruments (e.g. questionnaire, test, scale, rating, or tool) along with their reliability and validity, if known.*
9.4 Business retention plan	*Plans to maximise participation in data collection (e.g. survey response rates) including how data could be collected for businesses who drop out of programmes (if applicable).*

9.5 Cost-effectiveness analysis	*Are you intending to undertake an analysis of cost-effectiveness? If so, how will costs be estimated?*

10. IMPLEMENTATION AND PROCESS EVALUATION

10.1 Process evaluation and implementation: questions and purpose[30]	*Specify any implementation questions to be addressed by a process evaluation.* *How will the process evaluation complement the overall evaluation?*
10.2 Process evaluation: methods and data collection	*Description of methods used in the data collection (incl. why, how and when).* *This could include:* *- A summary of the methods you will use to assess programme fidelity so a 'non-compliance' analysis can be carried out.* *- The variable(s) used to estimate dosage and/ or compliance (eg how many workshops are attended), clarifying the level at which compliance is defined (e.g., individual participant, business, hub).* *- A summary of the methods you will use to assess usual practice at baseline and endpoint of the trial in treatment and control settings*
10.3 Wider Impact evaluation	*What other approaches are being used to assess and understand impacts, for example:* • Additional comparison groups or other quasi-experimental approaches for estimating impacts in addition to the main trial • *Externality benefits or disbenefits (eg economic displacement)* • *Wider social benefits - eg individual well-being or inclusion*

11. ETHICS

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

11.1 Ethical concerns	*Any ethical concerns, for example, could there be any harm caused to the businesses who receive the intervention, or to a comparison group who do not receive the intervention and how these will be mitigated. Please see the IGL trials toolkit pre-trial preparation section for information on ethical issues.* *If applicable, please explain the process for obtaining ethical approval, including timelines and responsible parties.*
11.2 Consent or assent for participation in the trial	*Describe the procedures for obtaining agreement to participate in the trial.*
11.3 Confidentiality	*Processes for ensuring data confidentiality - how will personal or otherwise identifiable information about potential and enrolled participants/businesses be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial?*
11.4 Data Protection	*Include a data protection statement relevant to the project.* *- If processing special categories of personal data, clearly describe the special data and the rationale for processing them with reference to the evaluation design.* *- Will you need to process special categories of personal data, provide a clear rationale for the legal bases selected for personal and special data, with reference to your organisational policies and the design of the specific evaluation project.*
11.5 Declaration of interest	*Any competing interests of evaluators should be declared.*

12. RISKS

Description of risks to the trial and how they might be addressed.

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 'on-treatment' 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 'on-treatment' analysis will be used. Process evaluation will monitor this.
Spillovers/ contamination	Likelihood: Low Impact: Moderate	Recruit firms operating in different markets who are not part of the same business networks and are unlikely to share information/resources with each other
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 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.

13. TIMELINE

Phase[31]	**Time period**
Phase 1: Trial design and preparation (trial protocol, survey design, etc.)	e.g. Sept. 2019 - Nov. 2019
Phase 2: Recruitment (engagement, baseline, randomisation, etc.)	e.g. Dec. 2019 - Jan.2020
Phase 3: Intervention Delivery (treatment period)	e.g. Feb. 2020 - Nov. 2020

[31] Although this is the most common time structure for trials, not all projects follow this clear path. Feel free to change the phases if necessary.

Phase 4: Data Collection and analysis (final follow-up survey, qualitative data, etc)	e.g. Decl. 2019 - Jan. 2021
Phase 5: Reporting (concluding analysis and evaluation report)	e.g. Feb. 2021

A2) IGL Statistical Analysis Plan Template for RCTs

This template is recommended by IGL for pre-specifying the details of the statistical analysis that will be used to assess the outcomes from randomised trials. Using a pre-specified statistical analysis plan adds greatly to the credibility of the findings of a trial, by demonstrating that the researcher has not engaged (even unconsciously) in specification search. The statistical analysis plan will also enable the evaluator or researcher to carry out key analysis rapidly once the outcome data becomes available, so that the key findings from the trial can be made available in a timely fashion.

The statistical analysis plan should be completed and registered online ***before the collection of outcome data takes place****. Preparing the statistical analysis plan provides an opportunity to revise the outcome measures that were defined in the trial protocol, based on learning about the measurement approaches from the baseline data and/or on changes in the project team's expectations of the outcomes that may be affected by the treatment(s). It is important to* ***review the outcome measures with the project implementation or delivery team*** *before completing the statistical analysis plan, so that any changes in expectations about the most appropriate outcome measures are reflected in this plan.*

Sections 4, 6 and 7 of this template include recommendations on approaches that are suitable for the majority of trials supported by IGL. These recommendations are in line with the guidance set out in IGL's Guide to Quantitative Analysis of RCT Data. However, since trials vary in their design and context, there may be good reasons for diverging from these recommendations in particular cases.

1. INTRODUCTION

1.1 Project title	*Descriptive title identifying the study design, population and intervention.*
1.2 Trial protocol	*Reference to version number and date of trial protocol (include a link if trial protocol is available online)*
1.3 Trial registration	*Link to trial registration (e.g. on https://www.socialscienceregistry.org/)*
1.4 Author(s) of statistical analysis plan	*Name and affiliation of the author(s) of this document*

2. DOCUMENT HISTORY

Version number	Date	Significant changes made

3. LOGIC MODEL

Has the project's logic model (setting out the underlying logic or theory of change and a set of assumptions about how an intervention works) changed since the trial protocol was completed? If yes, insert an updated version of the logic model and a brief description of the changes below.	*Yes/No*

4. PRELIMINARY CHECKS

Describe the checks that will be carried out before beginning data analysis. This will normally include a check that the treatment and control groups are balanced in their baseline (pre-intervention) characteristics, as a confirmation that the randomisation worked as expected and that there has not been significant attrition bias.

In most cases it is also useful to revise the power calculations set out in the trial protocol, to establish the minimum detectable effect size that can be estimated from the data available. If the minimum detectable effect size is larger than the minimum policy-relevant effect size, the findings of the trial may not be useful for informing future policy decisions: if so, a change in evaluation approach may be required.

	Default approach (IGL recommendation)	**Approach to be used**
Balance checks	*Produce a table showing the means of the baseline characteristics in each of the treatment and control groups.* *Carry out an F-test for joint significance of these characteristics in predicting treatment status. (More information here.)* *Carry out the steps above twice: once for the sample as originally randomised, and once for the sample as analysed.*	
Power calculations	*Revise the power calculations set out in the trial protocol, calculating the ex-post minimum detectable effect size with the sample available for analysis and with estimates of the standard deviations and (if relevant) intra-cluster correlations from the baseline data or from the control group in the final dataset. (More information here.)*	

5. CONSTRUCTION OF KEY VARIABLES

5.1 OUTCOME MEASURES

For each of the primary and secondary outcome measures, describe exactly how the measures will be constructed from the raw data. Enough detail should be included to allow your analysis to be replicated exactly. Annexing a file with the code that will be used to do this in your statistical software is ideal.

In the right-hand column, note any changes in the outcome measures that have been made since the trial protocol was finalised. This may include changes in the definition of the outcome measures or in whether each are to be considered as primary or secondary measures.

Primary or secondary outcome?	**Description of variable**	**Detailed definition** (referring to question numbers from survey instruments, if applicable)	**Any significant changes made since the trial protocol**

5.2 CONTROL VARIABLES

Describe the construction of each of the variables that will be used as control variables/covariates in your main analysis, if any.

Description of variable	**Detailed definition** (referring to question numbers from survey instruments, if applicable)

6. DATA CLEANING

Describe any steps that you intend to take to prepare the data for analysis, including whether any observations will be excluded from the analysis and how you will deal with missing data.

	Default approach (IGL recommendation)	**Primary approach to be used**	**Any alternative approaches to be used as robustness checks**
Handling of missing data in outcome measures	*Either exclude observations with missing values from the analysis (if there is limited missing data or evidence that it is missing at random) or calculate Manski bounds (if the outcome measure is binary or discrete and attrition is low).*		
Handling of missing data in covariates	*If less than 10% of observations have missing data, replace with the unconditional mean of the variable in the non-missing observations. Otherwise, replace the missing values with zero and create an additional variable indicating missingness, to be included as an additional covariate.*		
Criteria to be used to exclude observations from the analysis	*None*		

Any additional data cleaning	*None*		

7. MAIN ANALYSIS

Describe in detail how you will carry out the main analysis of outcomes in your trial.

The information below should apply to the analysis of both primary and secondary outcome measures. If a different approach is being used for the analysis of secondary outcomes, then this should be noted.

	Default approach (IGL recommendation)	**Primary approach to be used**	**Any alternative approaches to be used as robustness checks**
Type of treatment effect to be estimated	*Intention to treat (ITT)*		
Treatment groups to be compared	*If the trial has two arms: treatment group against control group.* *If the trial has more than two arms, specify which comparisons you expect to have sufficient statistical power for, and adjust for multiple comparisons in your inference (see below).*		
Type of statistical test	*First step: unadjusted t-test (for continuous variables) or chi-squared test (for binary variables)* *Second step: estimate linear regression/linear probability model using ordinary least squares (for continuous or binary variables)*		
Covariates	*First step: no covariate adjustment*		

	Second step: adjust for (i) stratification variables, (ii) baseline values of outcome variables, (iii) any other variables that are strongly predictive of the outcome in the baseline data.		
Weighting of observations	*Weight observations equally unless there is a reason for an alternative weighting.*		
Accounting for clustering in sampling or random-isation	*(Applies only if using a clustered design.) If there are at least 50 clusters, calculate cluster-robust standard errors. If there are fewer than 50 clusters, calculate randomisation inference-based standard errors, or use a cluster-aggregated approach.*		
Subgroup analysis	*None, or (if statistical power allows) only carry out subgroup analysis among groups that were used for stratification.*		
Correction for multiple comparisons	*(Applies if there is more than one primary outcome measure, or more than two trial arms, or if any subgroup analysis is being carried out.) Calculate the family-wise error rate, using Bonferroni correction or an alternative method.*		

Statistics to be reported	*Point estimates, 95% confidence intervals and continuous p-values*		

8. SUPPLEMENTARY ANALYSIS

Describe any additional analysis that you are planning to carry out with the trial data. This may include:

- Estimation of alternative types of treatment effect (e.g. estimate of the local average treatment effect, LATE, among those who complied with the treatment)
- Estimation of treatment effects on additional outcome measures
- Estimation of treatment effects among additional subgroups

Evaluators are free to conduct any additional exploratory analysis once the data is available. However, specifying in advance the analysis that will be carried out adds credibility to the findings, by reducing the potential for specification search.

Type of analysis	**Details**

Annex B: Project stories - The experiences of beneficiaries

IGL will be producing a number of blogs that capture the experiences of INNOSUP-06-2018 beneficiaries, giving other agencies who want to follow their approaches some of the inside stories of a team's motivations, challenges and learning from their experiments.

These project stories will be published as blogs on the IGL website. The first, which covers the experiences of the KEPA team who led the DCS-iSME project, is taken from the site, and reproduced below. The next, coming later in 2021, will look at the challenges faced by the RCT4MANU team as their trial hit the field during a turbulent economic period.

KEPA: A tale of piloting, exploring and scaling

In 2018, the European Commission introduced a new EU Horizon 2020 programme - INNOSUP-06-2018 - to encourage innovation agencies across Europe to experiment with their policy programmes. Here at the Innovation Growth Lab, we've been supporting both the EU and innovation agencies to succeed. This piece explores the journey of Greece's Business and Cultural Development Centre (KEPA) team, who are currently partaking in the INNOSUP programme.

In the words of a member of the team working on the EU funded project *'Design Customised Support for Innovative SMEs'*, it has the potential to demonstrate the famous adage: teach a man to fish and you feed him for a lifetime. The concept of design thinking - a human-centred approach to innovation - is a way of granting businesses the opportunity to intrinsically shift their approach and consistently yield greater long-term positive outcomes.

As the INNOSUP-06-2018 programme began, KEPA was launching the Hellenic Design Centre in Thessaloniki as part of wider efforts to promote the use of design thinking in both the public and private sector.

Design thinking allows organisations to approach problems in a human-centric way, researching and defining their users' needs, challenging assumptions and creating innovative solutions as a result. KEPA hopes to grow the use of this methodology within Greece, in order to enhance the competitiveness of Greek SMEs with the production of innovative products and the creation of high-quality services.

An earlier H2020 project, Design Shots, allowed the team to explore a light-touch design diagnostic for SMEs. Later, the INNOSUP-06-2018 programme afforded KEPA with an exciting opportunity to create a more substantial policy intervention that would target an apparent gap in the support available to SMEs after they had drawn down innovation funding. There is a dedicated section to the details of their trial on the IGL website.

Running the pilot as an experiment to demonstrate the effectiveness of design thinking was an idea with strong buy-in from leading government figures in Greece, as the majority of their information on the impact of design thinking was derived from other countries and large corporations. For Greek officials, there was a demand for more evidence which was based on the specifics of their environment, factoring in the smaller nature of some Greek organisations and their idiosyncrasies.

Though KEPA had previously been unfamiliar with experimentation in general, and typically planned evaluations once programmes were underway, they now aimed to test the impacts of design thinking by using a small-scale pilot with some participants randomly selected to receive the full 'Design Customised Support' programme. In doing so, the team's objective was to build their own capacity to run experiments and help embed this approach across innovation policy.

The unforeseen impacts

A number of challenges were encountered as KEPA began refining the details while developing their planned experiment. As an entirely new intervention and with support to be tailored to business needs, it proved difficult to know in advance how best to define outcomes measures and collect data in a way that would enable statistical comparisons across the two arms of the trial. With ten SMEs able to participate in the actual pilot, however, the achievable sample would always have been too small to provide reliable estimates of the programme's impacts. The team therefore had to adapt plans, shifting the focus to more qualitative assessments of how SMEs would benefit from the programme and creating the basis for future quantitative evaluation.

Initially when the call opened, optimism was high within the team; there was a significant level of SME interest, with numerous organisations seeking more information. However, soon after launching this phase of recruitment, KEPA's experience was put to the test by the unprecedented spread of the COVID-19 pandemic within Europe. As a result, it was unclear not just as to when was best to resume activities, but whether proceeding at all would be viable.

The call had to be relaunched, and this time around, interest was - as expected - not as high. KEPA's offer was no longer as appealing to companies dealing with the devastating impacts of COVID-19, who were primarily concerned with staying afloat.

They had hoped to try and capture information from a much larger sample of interested SMEs and then select ten core participants from within this group to receive access to their 'Design Customised Support' programme. The small sample size and unusual timing of the call limited the team's ability to be selective with who entered the pilot and to draw insights on wider demand. There was also a need to overhaul the approach to delivering the intervention - moving to an entirely digital route.

Outcomes and next steps

Through running the experiment, however, KEPA has learned some valuable lessons in both what works and what does not when it comes to policy experimentation.

More broadly, they attribute part of their learning journey to their participation in the activities organised as part of INNOSUP-06-2018 by IGL. 'I have to say that the peer learning sessions were very helpful', one team member notes, 'especially being guided through the difficulties of implementing the project'. Learning from the experience of others undertaking experiments similar to theirs was incredibly beneficial for KEPA in knowing how best to approach their project.

Though KEPA did not get to explore their hypothesis to its fullest extent given the wider context, the team can now better recognise the potential of experimentation. The full benefits of running an RCT would come at the next stage of the intervention's development - at a larger scale and with proven measures. However, with hindsight, the team believes that their pilot could have yielded more valuable evidence if they were able to be more discerning about who took part, selecting SMEs of a similar size, sector and particularly ones facing similar problems.

KEPA has navigated and overcome the challenges faced, and the fact that this was a small-scale pilot has provided them space to develop their approach and learn not only about taking an experimental approach, but also how businesses would use and benefit from the Design support. The pilot granted KEPA direct engagement with the businesses and the ability to observe benefits of the more tailored support, gaining greater insight and understanding of the underlying mechanisms and how one could create more quantitative outcome measures to assess outcomes from a larger project.

With the feedback and support of IGL in designing a survey, next up for the team is the stage in which they organise interviews with SMEs involved.

The key learning KEPA hopes to take forward for future pilots is to avoid vague parameters and unclear indicators, and instead to be more specific with what they test, using more scientific methods to draw significant conclusions and robust information. Running a pilot before proceeding to the full trial has been game-changing for KEPA, as this preparation allows for more time-efficient, well-prepared activities later down the line; they have a better sense of best practice and how to avoid potential pitfalls. For the team, the unforeseen barriers to experimental success have only served as further proof of what is potentially possible: this is just the beginning of their experimentation journey.

Publications Office

doi:10.2826/43979