technopolis group

WHAT METHODS WORK FOR EVALUATING THE IMPACT OF PUBLIC INVESTMENTS IN RD&I

A report prepared for the Department of Science, Innovation and Technology (DSIT)

May 2024

Authors: Cristina Rosemberg, Claudia Obando, Thoraya El-Rayyes

With contributions from: Nadya Mihaylova, Rebecca Babb, Juan Pablo Centeno, Anete Vingre, and Neil Brown

Acknowledgements: *The study team would like to thank the DSIT team for its contributions, in particular Madhavi Cherian and Will Steele. Our thanks also go to the Study Steering Group whose insights have significantly enriched this work.*

Any enquiries regarding this publication should be sent to us at: correspondence@dsit.gov.uk

Contents

Executive Summary

The effective implementation of **Research, Development and Innovation (RD&I) public interventions is key to delivering on the UK government's Innovation Strategy and to addressing the main societal challenges of our generation**, from Net Zero to healthy living, to sustained (and sustainable) economic growth. This comes with an increasing need for robust evaluation methods to track progress and determine the impact of those investments.

However, evaluating the impact of RD&I public interventions is extremely challenging. Evaluators are confronted with the traditional evaluation challenges, i.e. how to robustly establish if challenges observed – if any – can be safely attributed to the intervention and to what extent. Additionally, they face challenges associated with the nature of the intervention and the expected outcomes, including the time that it takes for outcomes and impacts to materialise (from a few years to decades), the different public (and private) investments that are likely to play a role in those results, and the fact that the selection of participants is, in many cases, based on a pre-set criteria that aims at funding excellence, creating selection bias problems, to name a few.

Many of those challenges, and its potential solutions, are not adequately covered in the standard textbooks and monitoring and evaluation guidelines. This study aims to cover this gap in the knowledge. It offers a view of "real-world" examples of impact evaluations of RD&I interventions nationally and internationally. This includes compiling and rapidly assessing 107 reports and developing more in-depth case studies for 20+ of them to delve deeper into the methodological approaches and strategies used. To further support the work of commissioners and evaluators, the core of the report is organised by the main challenges faced in the evaluation of RD&I interventions and the main methodological approaches and strategies identified in the literature (as shown in the graph below).

Lagged effects of research investments

- **Theory-based Evaluation Methods (TBE)**
- **Lead indicators (early signals of future impact)**
- **Economic forecasting**
- **Revealed preference techniques**
- **Assessment based on qualitative data / perceptions / self-assessments**

Intangible benefits	• **Citation analysis using bibliometrics (or patents) data** • **Mobility assessment using bibliometric data** • **Social Network Analysis (SNA)**
Low observability of research impacts and knowledge flows	• **Measuring / capturing development of intellectual property (IP)** • **Assessing changes in companies' valuations** • **Social Network Analysis (SNA)** • **Assessment based on qualitative data / perceptions / self-assessments**
Skewness of impact	• **Assessment based on qualitative data / perceptions / self-assessments** • **In-depth case studies using purposive outlier sampling**
Additionality / attribution, and contribution	• **Quasi-experimental design (QED) methods** • **Theory-based Evaluation Methods (TBE)** • **Analysis of several "generic" data collection methods (surveys, interviews)**

The rapid assessment of 107 cases and further development of 20+ cases identified in this study revealed some initial findings with respect to UK and international RD&I evaluation practice:

- **There is significant interest from countries to understand how to measure the impact of RD&I public interventions.** This is particularly the case in the US, EU, France, Germany, and the UK where more impact evaluations were found.
- **Most of the evaluations tend to be programme specific.** A few exceptions are evaluations looking at the effects of a number of different supports and interventions from a single agency or organisation over a period of time (e.g. Fraunhofer).
- **Most impact evaluations are carried out ex-post** (just after the end of the intervention). Although some medium-term evaluations collect data and evidence to try to capture early outcomes and identify impact pathways, particularly in the UK. Multi-phased evaluations are less common but useful to set up appropriate baselines, inform delivery and incorporate lessons learnt into future evaluation rounds.
- **The UK is more ambitious in its attempts to capture final impact.** The majority of the evaluations tend to measure what in the UK would be considered outcomes. Measuring impact is less common due to timing, complexity, and data availability. Also, the challenge of lagged effects makes it difficult to capture robust evidence of impact.
- **Measuring the impact of RD&I is challenging and there is no methodological silver bullet.** Evaluators are confronted not only with the traditional evaluation challenges, i.e. how best to establish what would have happened in the absence of the intervention, but also with multiple challenges associated with the nature of RD&I interventions as described in this report. Evaluators tend to deal with these challenges by grounding the evaluation in a Theory of Change (ToC) and combining qualitative and quantitative methods.

- **Having good M&E data collection systems enables robust assessments and the estimation of counterfactuals, as well as reducing the time and cost of data gathering during evaluations.** The existence of good contact information on participants allows evaluators to link their information with pre-existing secondary sources. Data collected via secondary sources allows the building of counterfactuals and accessing time-series (which in turn improved the implementation of those counterfactuals). A simple addition to those systems, like including unique identifiers (VAT for companies and ORCID for researchers) could improve greatly the efficiency of data linking process (resources required and accuracy).
- Furthermore, resources such as Gateway to Research and Researchfish (used by UKRI) are also great resources to inform evaluations.

The national and international practice shows that a robust RD&I evaluation must include the following elements to address its most important challenges:

Maintain a sense of proportionality, with a design that takes into account the policies being evaluated, the evaluation questions being asked, and the resources available

A ToC at its core, to set up expected impact pathways, guide the methodological approach, and inform the conclusions and findings

A combination of different quantitative and qualitative methods, and triangulation of evidence across a variety of sources and methods

The use of counterfactual and benchmarks (either quantitative and qualitative) to tease out what the intervention is delivering in excess of what would have happened anyway

A clear description of assumptions and statistical tests for quantative methods including for quasi-experimental approaches, and sensitivity analysis

A clear discussion of limitations and caveats of methods employ, and frameworks to asses the strength of the evidence

Different reports for different audiences, with for instance, short executive summaries catering for policy and decision makers, extended main reports for programme managers, and technical annexes for policy analysts and other evaluators

1 Introduction

1.1 The purpose of this study

The effective implementation of **Research, Development and Innovation (RD&I) public interventions** is key to delivering on the UK government's Innovation Strategy and to addressing the main societal challenges of our generation, from Net Zero to healthy living, to sustained (and sustainable) economic growth. This comes with an increasing need for robust evaluation methods to track progress and determine the impact of those investments.

Evaluating the impact of RD&I public interventions is extremely challenging. Evaluators are confronted with the traditional evaluation challenges, i.e. how to robustly establish if challenges observed – if any – can be safely attributed to the intervention and to what extent. Additionally, they face challenges associated with the nature of the intervention and the expected outcomes, including the time that it takes for outcomes and impacts to materialise (from a few years to decades), the different public (and private) investments that are likely to play a role in those results, and the fact that the selection of participants is, in many cases, based on pre-set criteria that aim at funding excellence, creating selection bias problems, among other things.

Many of those challenges, and their potential solutions, are not adequately covered in the standard textbooks and guidelines.

In this context, this study – commissioned from Technopolis by the Department for Science, Innovation, and Technology (DSIT) – provides **an overview of national and international best practice in impact evaluation methods used across different types of RD&I public interventions.** The study has been guided by three key objectives:

To identify **real-world impact evaluations of RD&I programmes** in advanced economies covering a wide range of funding bodies, types of grants/interventions and types of RD&I.

To identify **which methodologies are used to address** and overcome challenges encountered when conducting RD&I evaluations.

To **synthesise good and feasible impact evaluation practices**/techniques to inform future evaluation design.

The study intends to showcase good examples on how best to address **key methodological challenges** when evaluating RD&I public interventions, in the context of work commissioned by RD&I public funders, which tend to have the following characteristics:

- Addressing several evaluation questions to cover the needs of policy makers and relevant stakeholders (including sometimes process-related questions).
- Delivering in a relatively short period of time (e.g. 1 year for a reporting stage, rather than 3–5 years).

- Dealing with data gaps to address the evaluation questions to the best of their abilities and include methodological steps to deal with incomplete information.
- Using mixed or multi-methods to ensure triangulation of evidence.

As such, the evaluations and reports identified and further characterised in this study have been selected because they are expected to be more in line with the types of work that DSIT and its executive agencies will seek to undertake and commission.

In line with this sharp focus on "real world" examples, we have focused our search away from the academic literature, given that, in contrast with commissioned work, academic papers (i) tend to concentrate on a more limited set of questions (driven mostly by researchers rather than policymakers), (ii) tend to have more time to develop the research, and (iii) could decide to conclude there is limited information to make a final judgement (rather than arriving to a conclusion based on the – limited – information at hand). Having said so, academic research, by nature, is concerned with pushing the frontier of what is possible and offers good insights into emerging developments that are being (or may be more) widely adopted by evaluators in the coming years. In Section 3.2, we provide a reflection on two relatively new methodological developments which may become more prevalent in real world evaluations in the coming years.

1.2 Methodology in a nutshell

The study has been conducted in **two phases.**

Phase 1

Focused on **identifying and selecting the relevant impact evaluation reports of real-world national and international evaluations,** following the 4 steps briefly described in Figure 1 below. The search focused on capturing examples across a diversity of:

- **RD&I stages and interventions** – making sure a variety of interventions and stages of support relevant to DSIT and its executive agencies are well covered.
- **Methods** – looking to cover the majority of methods listed in the 2020 HMT Magenta Book (Chapter 3 – Evaluation methods, pp. 40–53). These include both quantitative and qualitative approaches to impact evaluation.
- **Geographies** – considering evaluations funded in the UK and in other advanced economies (including the US, France, Germany, Israel, Finland and Canada).

We identified 107 cases. The main characteristics of these examples are listed in Figure 2.

Figure 1: Steps for identifying national and international real world examples

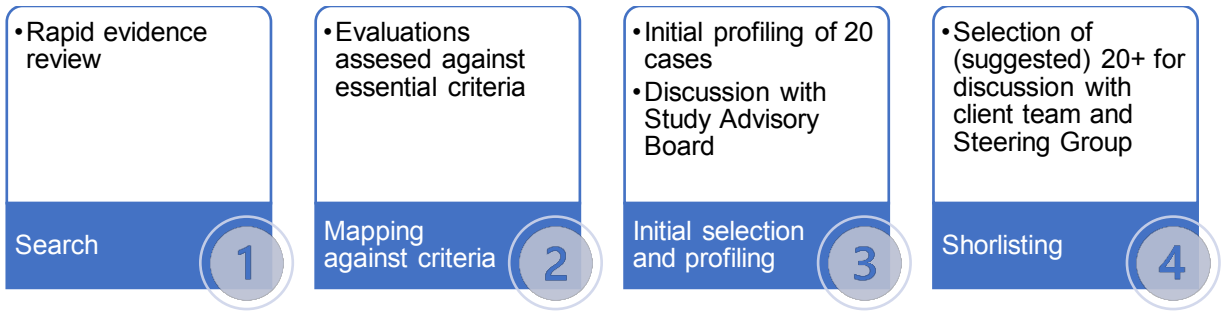


Figure 2: Main characteristics of long and short list of cases

	Long list (107 cases)	Selected cases (20+ cases)
	7+ types of interventions (from research & innovation grants to research infrastructures and fellowships)	**5+ types of interventions** (from fellowships and research infrastructure to networking activities and research grants)
	9+ types of research involved (from basic and applied research to collaborative R&D and challenge driven/mission oriented)	**8+ types of research involved** (basic and applied research to collaborative R&D and, mission oriented/challenge-driven)
	12 countries/regions (including the UK, the US, Israel, Korea, France, Germany and the EU).	**8 countries/regions** (including the UK, the US, France, Finland, Germany, Australia, the EU and Ireland)

Phase 2

Focused on **further expanding on 20+ selected cases**. Developing case studies of a selection of impact evaluation reports has provided the opportunity to delve deeper into the challenges faced by evaluators and the methodological decisions and strategies they have undertaken.

They have served as the basis for this report, and are presented in full as a separate Annex report.

The cases have been developed based on an in-depth review of the reports that are publicly available, further complemented by 7 interviews with funders and evaluators (covering 8 cases) to bring further insights into the challenges faced by the evaluation.

Please see 0 for further information on the methodology.

1.3 This report and how to navigate it

The central chapter of this report (Section 0) has been organised according to the **five main challenges faced by impact evaluations of (RD&I) public interventions** (see diagram below). For each of them (and its corresponding sub-section), we provide:

- A description of the challenge
- A brief description of the main methodological approaches and strategies identified across the selected cases to address that challenge
- Two cases to exemplify how evaluations have implemented those strategies in practice. These cases include a brief description of the programme, main evaluation questions and methods. No methodology is exempt from potential limitations and caveats (mostly when implemented in the “real world” conditions we have described above) and those are also presented in the cases, along with a discussion on potential transferability to the UK context.

Additionality, attribution, and contribution represent one of the most important challenges for any evaluation (not just RD&I interventions). However, the corresponding sub-section is presented at the end of the section as, in addition to the 3 main components described above, we also present a cross-cutting analysis from the 20+ cases as to how evaluations establish counterfactuals, their potential uses, and their main limitations.

Most evaluations faced a combination of these key challenges, but for expository purposes each sub-section and the corresponding examples focus on one specific challenge. The key methodological approaches and strategies to address each challenge are also summarised in the diagram below.

Section 3 provides reflections based on the main patterns observed across the 107 cases (Section 3.1), as well as insights into emerging trends in evaluation (Section 3.2). Finally, we also provide some recommendations on standards (Section 3.3).

 **Lagged effects of research investments**	• **Theory-based Evaluation Methods (TBE)** • **Lead indicators (early signals of future impact)** • **Economic forecasting** • **Revealed preference techniques** • **Assessment based on qualitative data / perceptions / self-assessments**
 **Intangible benefits**	• **Citation analysis using bibliometrics (or patents) data** • **Mobility assessment using bibliometric data** • **Social Network Analysis (SNA)**
 **Low observability of research impacts and knowledge flows**	• **Measuring / capturing development of intellectual property (IP)** • **Assessing changes in companies' valuations** • **Social Network Analysis (SNA)** • **Assessment based on qualitative data / perceptions / self-assessments**
 **Skewness of impact**	• **Assessment based on qualitative data / perceptions / self-assessments** • **In-depth case studies using purposive outlier sampling**
 **Additionality / attribution, and contribution**	• **Quasi-experimental design (QED) methods** • **Theory-based Evaluation Methods (TBE)** • **Analysis of several "generic" data collection methods (surveys, interviews)**

It is important to note that this report intends to provide concrete examples on how challenges can be overcome based on a selected number of cases. There are several guidelines and reports that provide a detailed list of methods and the theoretical and practical advantages of each of them, including the HMT Magenta Book (2022)[1] or the Overview of Evaluation Methods for RD&I programmes (Ruegg, 2011).[2] This report does not intend to reproduce this prior work but to build upon it by focusing on the identification and further development of national and international real-world evaluations from which the UK and DSIT can learn.

Nevertheless, Box 1 provides an overview of relevant guidance material that could be of use to inform the choice of methodologies.

Box 1: Relevant M&E guidelines

The following are a set of high-quality resources that provide comprehensive guidance on the design and implementation of evaluations for public policies, programmes, and projects. These resources cater for a wide range of evaluation needs and contexts.

[1] https://www.gov.uk/government/publications/the-magenta-book
[2]https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=bf6be32ec7c97e885140cfec9b6d1505c43be292

- Magenta Book. (2020) HM Treasury's Magenta Book provides guidance on how to evaluate the efficiency, effectiveness and impacts of governmental interventions. It is designed to guide the development of transparent, objective, evidence-based evaluations to inform governmental decision-making. It covers the following key topics: what evaluation is; why, how and when to evaluate; designing an evaluation; evaluation methods; data collection for evaluation; managing an evaluation; and the use and dissemination of evaluation findings.

- Green Book. (2022). HM Treasury's Green Book provides guidance on how to appraise public policies, programmes and projects, including guidance on the design and use of evaluation. It is designed to be used alongside the Magenta Book, which provides more detailed information on a range of evaluation methods. In contrast, the Green Book focuses on guidance related to valuing the costs and benefits of public policies, programmes and projects. This includes guidance on the valuation of monetisable, unmonetisable and non-market costs and benefits.

- CSIRO Impact Evaluation Guide. (2020) This guide presents the impact evaluation framework employed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia's national science agency. It covers the following topics: why evaluations are conducted; evaluation design; evaluation methodologies; aggregation and comparability of impacts across programmes of work; and sensitivity analysis and reporting. The guide focuses on cost benefit analysis, which is CSIRO's primary methodology for research impact evaluation.

- SciencesPo Laboratory for Interdisciplinary Evaluation of Public Policies Handbook. (2023). This open-access handbook presents 24 qualitative, quantitative and mixed methods approaches to evaluation. The methods presented in the handbook are wide ranging and include experimental methods, quasi-experimental approaches, value for money analysis, theory-based evaluation methods and traditional qualitative research methods such as interviews, focus groups and case studies.

- UNDP Independent Evaluation Office Methodology Center. The UNDP's Evaluation Methodology Center provides a comprehensive overview of approaches to evaluation. It covers the following topics: evaluation methodologies; data collection approaches and methods; data analysis approaches and methods; and approaches to assessing cross-cutting themes in evaluations.

- US Department of Energy EERE R&D Programme Standard Impact Evaluation Method (2014). This handbook provides guidance for impact assessments of R&D programmes for the US Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE). It covers the following key topics: evaluation planning; assessing additionality; estimation of economic costs and benefits; estimation of environmental impacts; estimation of energy security impacts; estimation of knowledge impacts; calculating economic performance measures; and sensitivity analysis. (An updated version of this handbook is due to be published soon).

1.4 Further considerations

1.4.1 Considerations on robustness

In this report we present a variety of methodological approaches and strategies identified from national and international best practice. All methods presented were deemed as appropriate for addressing the challenge at hand, even if limitation and caveats remain. It is fair to say that the level of robustness varies across methods; however, it is difficult to assess this against a standard scale because of the differences in the scope and nature of the cases.

The Maryland Scientific Methods Scale[3] offers a potential way forward, but its assessment is based on the lack or inclusion of control groups. Evaluations of interventions without a control group are immediately allocated the lowest level in the scale (1 out of 5). This makes it less relevant in the context of impact evaluations of RD&I interventions since:

- Not having a clear control group is not uncommon in RD&I public interventions, especially those that support areas of high specialism or small markets (e.g. the upstream space sector or the emergent quantum economy) or pockets of excellence (e.g. most notable researchers in their field), where virtually all major and relevant players have been supported in one way or another via public interventions.
- Quasi-experimental methods are best applied when interventions are relatively simple (i.e. there is a clear direct connection between the inputs of the programme and the expected outcomes), and where the expected outcomes are easily measurable.

Also, as presented in this report, some methodological challenges need to be addressed via other types of quantitative or qualitative methods (beyond quasi-experimental methods), and ideally via the triangulation of evidence emerging from both.

Finally, another related consideration when reading the report is that the budgets of the individual evaluations are not disclosed. Any methodological decision needs to be judged in the context of its robustness, but also in relation to the time and resources available for its implementation. This is in line with the principle of proportionality advocated in the former UK Department of Business, Energy and Industrial Strategy (BEIS) Monitoring and Evaluation Framework (2020). The document states that impact evaluation design "should be appropriate to the design of the policies being evaluated, the evaluation questions being asked, and the resources available" (p. 47).[4] This is something important to take into account when considering implementing some of the more sophisticated methods identified here via case studies.

[3] https://whatworksgrowth.org/resource-library/the-maryland-scientific-methods-scale-sms/
[4] BEIS Monitoring and Evaluation Framework (2020) https://assets.publishing.service.gov.uk/media/5fe341ad8fa8f56af2a85fcf/beis-monitoring-evaluation-framework.pdf

1.4.2 Gaps in the evidence base

The extensive search of real-world examples in the context this study revealed two gaps in the case studies selected:

Theory-based evaluation (TBE) methods – Our search of real-world examples (long list) reveal that these methods are more commonly used in UK evaluations of RD&I interventions. We only found one case outside the UK that used TBE as described in the Magenta Book (e.g. Contribution Analysis, Process tracing, Outcome Harvesting). This shows that the UK is leading the way in the use of these methods, but limits the insight that can be gained internationally.

Spillover effects – Our search also shows that there are few real-world examples on how to deal with spillover effects at programme level. Most studies that do explore the issue of spillovers tend to use econometric methods to model the impact of public R&D investment (beyond a specific programme) on the economy.

In 2014, the Department of Business, Innovation and Skills (BIS) commissioned a study to identify examples of economic analysis of spillovers from programmes of technological innovation support.[5] The authors concluded that "quantifying the value of spillovers at a given point in time is notoriously problematic. Added to this, the literature provides little to no quantitative evidence on the linkages between the presence of factors in the innovation system/support programme and the scale of spillovers". The authors did, however, find some examples of *private return* on R&D investment.

One of the few recent examples includes Beckter et al., 2023, "Assessing innovation spillovers from publicly funded R&D and innovation support: Evidence from the UK". Technovation, Volume 128, 2023.

[5] BIS (2014). "AN ECONOMIC ANALYSIS OF SPILLOVERS FROM PROGRAMMES OF TECHNOLOGICAL INNOVATION SUPPORT" https://assets.publishing.service.gov.uk/media/5a7c1c5a40f0b645ba3c6c21/bis-14-653-economic-analysis-of-spillovers-from-programmes-of-technological-innovation-support.pdf

2 Methodological challenges and strategies to address them

2.1 Lagged effects of research investments

Challenge description

Research and innovation support often takes place over several years, and the subsequent impact can take a decade or more to realise. The timing of impacts depends on the type of research, complexity of the technology being developed and contextual factors. This challenge is more prominent if programmes support basic/fundamental research which may not have a known application at the start or end of the project, with unexpected benefits materialising years after the original investment, and with support of further efforts from other funders and organisations. This challenge is further exacerbated when relevant data sources also have a lag in the publication of data (e.g. business data from official sources like ONS has one-to-two-year lag).

Additionality, research and innovation is not linear. There is a high degree of uncertainty on the extent to which results initially expected will materialise and what they would look like, with unexpected or unforeseen applications emerging over time. Moreover, positive outcomes are likely to be skewed towards a minority of the final investments (see Skewedness of impact, Section 2.3).

The larger the timespan between the initial research and the medium, long-term impacts, the more difficult it is to trace back to the original investments (see Low observability of impacts, Section 2.2) and attribute the final results to the intervention under evaluation (see Additionality, attribution and contribution, Section 2.5).

To name some examples, Planck and Einstein's work on wave-particle dualism and light photons formed the basis of lasers and digital cameras, but these applications were realised with a time lag of many decades. Similarly, technologies developed for particle physics research at CERN have since found applications in other areas. For instance, accelerators developed for high-energy physics have been adapted for cancer treatment through techniques like proton therapy, with research developments that can be traced back 20 years (with multiple public and private investments along the way).

Methodological strategies to address challenge

- **Theory-Based Evaluation Methods (TBE):** TBE methods offer a structured way of understanding an intervention. They involve drawing hypotheses as to how and why it works and then testing those hypotheses, using qualitative and quantitative data. They are typically used for complex interventions and/or when a control is not available to draw a counterfactual scenario. TBE methods include the development of a Theory of Change (ToC) for the intervention. The ToC identifies the (potential) pathways of impact by articulating the expected chain of effects from investments to outputs, outcomes and impacts in the short, medium and long term. It also highlights the risks and assumptions that underpin the achievement of those results. It tackles the challenge of lagged effects in so far as it allows a programme to be evaluated with respect to what would be expected given the timelines (from start to the point that the evaluation is conducted) and consequently assess if the intervention is "en route" to achieve its intended impacts. The ToC also allows the identification of lead indicators (See Table 1 below and Case 19 in Annex for examples).

- **Lead indicators (early signals of future impact):** In line with the use of TBE and TOCs, lead indicators offer a way to identify early signals of future effects, based on progress made in terms of short and medium-term outcomes. The main advantage of this approach is that it can offer the opportunity to quantify impact at an early stage of an intervention. Its main limitation is that it does not directly tackle the fact that RD&I is not linear, but this limitation is common to all methods not involving "historical tracing" of results (i.e. focused on looking 10–20 years back) (See Cases 1 and 16 in Annex for examples).

- **Economic forecasting**: This strategy entails using economic models to derive future effects of R&D investments. It tends to follow a "production function" approach to model how inputs (R&D investments) would be expected to translate into economic results (such as turnover or productivity). This method can also incorporate diffusion patterns of associates to specific technologies, using life cycle models. As above, the main advantage is that it can offer the opportunity to quantify impact at an early stage of an intervention. The main limitation is that any forecasting model will need to rely on strong assumptions (See Table 2 below and Cases 15 and 17 in separate Annex for examples)

- **Revealed preference techniques:** This approach is used to infer the value attributed by users/beneficiaries/stakeholders to a benefit, which does not have a price in the market and consequently cannot be directly monetised. They are also used to assess the impact of investments whose effects are expected to materialise in many years to come by asking individuals to make an assessment today of how much they will be willing to pay (or to accept) in perpetuity to maintain those investments. This strategy is typically used outside the RD&I policy environment to assess the value of natural[6]

[6] https://assets.publishing.service.gov.uk/media/6017e8378fa8f53fc01c78d4/ncc-natural-capital-valuation.pdf

or cultural and heritage capital.[7] In the RD&I policy environment it has been typically used to measure the impact of research infrastructures.[8]

- **Assessment based on qualitative data/perceptions/self-estimations:** This approach entails gathering evidence about the future effects of R&D investments, via interviews with experts in the field, programme managers & directors, and end-users (of the research). These are of course highly subjective and need to be appropriately weighted depending on the knowledge and seniority of the informant, as well as triangulated with other sources of evidence. Forecasting based on estimations made by participants could also add another layer of information, but it is likely to have a high degree of positive bias. To increase the robustness of this data, evaluators tend to analyse it in the context of a TBE approach, and to focus on outcomes that cannot be easily quantified but that are still important to capture according to the ToC of the intervention. (See Cases 4 and 23 in separate Annex for examples).

Multi-phased evaluations could also help with this challenge area. This is not a technique per se but relates to the overall set-up of the evaluation. Multi-phased evaluations could serve to identify early signals of impact at the interim stage (lead indicators) to (i) provide timely evidence to (immediate) decisions on the future of the investments, and then (ii) collect further evidence on progress towards impacts 1–2 years after the programme (projects) have ended. This approach does not fully solve the issue of lagged effects, but it does strengthen the evidence base in the short and medium term.

Finally, **historical (backward) tracing** is also a relevant approach to showcase how an innovation breakthrough can be traced back to an original investment (e.g. cancer treatment using proton therapy) and further exemplify how final effects can take decades to materialise. The 2012 evaluation of long-term impact of the EU Framework Programmes offers an interesting approach.[9]

[7] https://historicengland.org.uk/research/current/social-and-economic-research/culture-and-heritage-capital/
[8] Florio M, Giffoni F. A contingent valuation experiment about future particle accelerators at CERN. PLoS One. 2020 Mar 11;15(3):e0229885. doi: 10.1371/journal.pone.0229885. PMID: 32160265; PMCID: PMC7065825.
[9] Arnold, Erik (2012). "Understanding long-term impacts of R&D funding: The EU framework programme". doi:10.1093/reseval/rvs025

Table 1 Challenge: Lagged effects of research investments (example 1)

Case	Evaluation of the Energy Entrepreneurs Fund EEF UK
The programme	EEF supports SMEs and early-stage innovators to develop innovative low-carbon technologies and processes to leverage private sector funding for pre-commercial operation and help firms to progress across TRLs to produce technologies ready for the market within the 4 years after completion of the grant.
The challenge	Key economic and environmental impacts are expected to emerge beyond the life of the evaluation (mostly due to the early stage of most technologies supported by the programme). Additionally, expected impacts will require other changes in the energy system, such as new regulations, standards, and user and consumer practices to fully realise the benefits of some technologies. The study (and methodology) also addresses the challenge of establishing an appropriate counterfactual.
Approach	To overcome the lagged effects of the R&D activities, the evaluation adopted a mixed method approach combining Qualitative Comparative Analysis (a type of TBE method) with quasi-experimental and logistic regression analysis to capture evidence across six main impact domains: regulation and availability of follow-on funding, R&D spending, technological development, commercial readiness, follow-on investment, commercialisation and adoption, economic benefits, environmental impacts and spillovers. • **Qualitative Comparative Analysis (QCA):** The QCA focuses on the most advanced projects to demonstrate the type of results that could be expected from the wider portfolio. The case studies were selected using a purposive sampling approach, so by design the analysis focuses on projects showcasing the highest commercial outcomes (across the seven funding rounds and technological areas). • **Quasi-experimental design (QED) method:** The QED was implemented to measure impact on R&D activity and spending, leverage of follow-on funding, employment, turnover and GVA, valuations and the effectiveness of incubation services (comparing to grants awarded without it, e.g. Innovate UK). Unsuccessful applicants acted as a control group, with the adjustments made to ensure no systematic differences between successful and declined applicants (comparable assessments scores, similar baseline characteristics and removing outliners, e.g. firms with > 250 employees). Note that the economic analysis of the programme

	utilises “Company valuation” of the participating firms as a lead indicator of the expected future profits of the company.
Limitation & caveats	A limitation of using company valuation as a proxy is that it assumes that markets operate efficiently, but the logic of the EEF intervention assumes they do not. The approach of focusing on developing case studies for projects with highest commercial outcomes is sensible but may limit the ability to provide conclusions on the entire portfolio (given the likely skewedness of impacts). The environmental analysis is made based on 11 projects that have trialled technologies in labs and have provided an ex-ante assessment of their CO2 reductions. These estimates were elaborated by the project team (beneficiaries) and may overestimate their effects (due to positive bias, or effects decreasing over time). The evaluation addresses this limitation by presenting estimates of low and high effects.
Transferability	This evaluation is done in the UK context, using monitoring and programme management data, alongside primary sources of information and some secondary data for economic and environmental impact assessment. Monitoring and management information are instrumental for the evaluation, particularly the Commercial Progress Reports (CPRs) provided by grant holders by the end of their project. This methodology can be replicated for evaluating similar programmes supporting experimental development and industrial research provided that monitoring data as well as CPRs are available from the start of programme.

Source: Department for Business, Energy and Industrial Strategy (BEIS), 2023 (https://assets.publishing.service.gov.uk/media/63f78f58d3bf7f25f2d087ea/evaluation_of_the_energy_entrepreneurs_fund.pdf).

Table 2 Challenge: Lagged effects of research investments (example 2)

Case	World Class Ecosystems in the Finnish Economy
The programme	The programme supports the creation and consolidation of 33 business ecosystems through “orchestration” funds (funds supporting innovation cooperation and joint activities) and capital loans, across a variety of sectors (including Energy, Health, Manufacturing, Mobility & logistics, ICT). Orchestration funds provide up to 50% of the costs of collaborative innovation projects between private companies, associations, and foundations (and in exceptional cases research institutes and public bodies). Loans between €2m and €10m are offered to anchor (platform) companies to facilitate ecosystem development.
The challenge	Lagged effects at company level and at the level of the ecosystems as impacts are expected to materialise in years to come, and beyond the life of the evaluation.
Approach	The approach used in this impact study consisted of a balanced mix of qualitative and quantitative methodologies, and economic modelling was then used to outline the main economic results of the ecosystems. The evaluation focuses on 13 ecosystems operating across three different phases: 1. Exploration/emergence, 2. Birth/start up (Experimentation), 3. Growth/Expansion, with 10 of them in phase 2 and 3. It determines that most ecosystems (16 out of 33) are in the experimentation phase, characterised by a variety of competing initiatives and several start-ups and spinoffs (from research organisations or corporations), and a focus on finding a solution-market-fit through piloting and demonstration. This framework allows the exploration of relevant economic indicators, as well as explanation of results. The evaluation includes economic modelling (turnover, employment, added value and exports) for businesses included in 13 ecosystems, using results from an early period of intervention (2016–2018) to extrapolate results over the next 10 years. Since a control group is not available evaluators model two scenarios, one with and one without public investment. To address uncertainty in the achievement of results, for each scenario evaluators provide three estimates, for low, medium and high growth. Other qualitative and quantitative methods were used to provide evidence and make assumptions on the degree of input and output additionality. These methods sought to address two main challenges, lagged effects

	and additionality of the ecosystem policy. The evaluation assessed input, and behavioural additionality, as well as impact on the economy and society.
Limitation & caveats	The main limitations of the evaluation are: • The timeframe of the evaluation, two years after the launch of the programme, posed significant challenges in terms of data availability and evaluation impact. In fact, many ecosystems started in 2018, and data available covered the period 2016–2018. • The unstable context that dominated the period of analysis may reduce its strength.[10]
Transferability	Implementing a similar approach requires access to company data, ideally for the period before and after the evaluation (e.g. BERD dataset from the Office for National Statistics). Monitoring and evaluation data tracing interactions and collaborative work of members of ecosystems would be desirable (not significantly used for this case) to strengthen the results of the evaluation.

Source: Technopolis and 4Front, 2022 (https://www.businessfinland.fi/4ad697/globalassets/julkaisut/World-class-Ecosystems-in-the-Finnish-Economy-2-2021.pdf)

[10] The conditions under which the ecosystem policy was implemented were quite turbulent given the Tekes and Finpro merger in 2018 and the COVID-19 crisis from Spring 2020 onwards.

2.2 Low observability of research impacts and knowledge flows

Challenge description

Knowledge creation and diffusion is one of the key outputs expected to emerge from publicly funded research and innovation activities. Publications can be documented as part of routine monitoring, however the outcomes and impacts expected to emerge from this new or enhanced understanding and knowledge are more difficult to observe and capture.

This includes improved skills and capabilities (scientific, technological, strategic/managerial), improved careers and mobility, changed behaviour/perceptions (across different actors), improved solutions (and further effects on productivity and competitiveness), and improved policy design and implementation, among others. Even the best monitoring systems may struggle to capture this information, partly because researchers will tend to have limited visibility as to how their publications have enabled economic and societal impacts. Systems like Researchfish provide valuable data and information for evaluators on those same outcomes but they do not fully solve the problem of "low observability".

Additionally, knowledge can move with people to different organisations, countries, and sectors creating spillover benefits outside those intended by the programme or intervention. Scientific mobility, for instance, plays an important role in knowledge diffusion and exchange. On the flip side, the movement of researchers or team members to different organisations may also lead to loss of data that could have been used for evaluation.

This challenge is also connected to the challenge around intangible benefits which is further discussed in Section 2.3.

Methodological strategies to address challenge

- **Citation analysis using bibliometric (or patent) data.** This strategy entails tracing the publications emerging as a direct result of (funded by) an intervention and measuring the citations of those publications in a variety of outlets including:
    - other publications,
    - patents (to measure their influence on innovation activities),
    - policy documents (to measure their influence on policymaking), or

- on platforms available to the society at large, such as news or Wikipedia entries (to measure their influence on information available to the general public).

In this way, evaluators can tackle the challenge of tracing knowledge flows over time, beyond what can be reasonably captured via monitoring systems (at least when it comes to codified knowledge). This strategy can be complemented with comparisons with relevant benchmarks to further assess additionality. This may include comparisons with uptake of publications funded via other interventions (for the same funder or at national level), and /or comparisons with the other publications of the same researchers involved in the intervention, with researchers acting as their own control group. This approach is more relevant for programmes where publications are expected (including pre-prints, peer-review articles, books, and conference proceedings).

The strategy needs to account for the time lags involved in the production of publications (time between awards being provided, research conducted, and publications submitted and approved), as well as the different citations patterns across disciplines (See Table 3 below and Case 24 in separate Annex for examples).

Similar analysis can be conducted using citations within patents, to trace prior art (i.e. other patents) that influenced the development of a new patent.

- **Mobility assessment using bibliometric data.** Similar to the strategy described above, this approach also uses bibliometric data, in this case to track researchers' institutional affiliation, and test how it changes over time. This strategy allows one to track changes across geographies (regions, countries), providing a measure of mobility, providing another avenue to measure and tackle the challenge of knowledge flows. It also provides estimates of brain drain/circulation. Links to the interventions would require establishing benchmarks or a counterfactual scenario. (See Cases 6 and 24 in separate Annex for examples).

- **Social Network Analysis (SNA):** This strategy entails quantifying and characterising networks, and the extent to which they change as a consequence of the intervention. This approach can be applied to the analysis of collaborations and interactions that take place beyond codified knowledge (i.e. beyond collaboration in publications) to include other forms of interactions. This strategy deals with the challenge of measuring knowledge flows indirectly, by focusing on the spaces in which that knowledge flow occurs (the networks) (See Table 4 below and Case 6 in separate Annex for examples).

Capturing evidence on other expected outcomes flowing from knowledge generation requires primary data collection (such as **surveys, interviews, and case studies**).

Table 3 Challenge: Low observability of research impacts (example 1)

Case	The Strategic Priorities Fund
The programme	The Strategic Priorities Fund (SPF) is an £831m UK Research and Innovation (UKRI) Fund with three high level objectives: • To drive an increase in high quality multi- and inter-disciplinary R&I (MIDRI) • To ensure UKRI's investments link up effectively with cross-departmental R&I priorities • To respond to strategic priorities or opportunities
The challenge	Measuring the degree of multidisciplinarity of research outputs and measuring the extent to which this is in excess of what is already produced via other sources of funding, and tracing knowledge flows and the uptake of results beyond academia (in particular, among policymakers). The study (and methodology) also addresses the challenge of establishing an appropriate counterfactual.

Approach	The evaluation measured the degree of MIDRI in publications, based on different indicators that captured:
	1. the diversity of co-authors' disciplinary background and aimed to capture the collaborative aspect in cross-disciplinary research (a proxy for multidisciplinarity), and
	2. the diversity of disciplines within the citations included in those papers (i.e. the reference list of papers). These aimed to capture the knowledge integration dimension in cross-disciplinary research (a proxy for interdisciplinarity).
	The evaluation followed a mixed methods approach that included:
	• **Tracing knowledge flows via citations**. All the MIDRI indicators were normalised by subfield, year, and document type, using the world level as a reference. SPF papers were also assessed in relation to:
	  ○ MIDRI publications from SPF researchers prior to SPF and in parallel to SFP (to provide a counterfactual scenario)
	  ○ All UK publications funded by UKRI and by any other funder (to provide a further benchmark)

	The evaluation also analysed the degree of citation of publications emerging from SPF programmes within policy-related literature. • **Case studies**. Qualitative evidence collected via case studies was used to show concrete examples on how research outputs and insights emerging from SPF programmes were helping to inform policy decisions.
Limitation & caveats	Wider analysis shows that only around 30% of citations are accrued in the 2 years following the publication of a paper. Since SPF started in 2018 and the analysis was conducted in 2021, this means that the results only provide an early signal of impact.
Transferability	This approach has been applied in the UK context. Implementing it in the context of other interventions requires: • Tracking publications linked to the programme via monitoring systems such as Researchfish. • Identifying SPF (and non-SPF) researchers, which can also be obtained from monitoring systems but require a great deal of data cleaning to match with bibliometric data. • Access to proprietary data sources that link grant, publication, and policy document data. Additionally, it is worth noting that MIDRI is a multifaceted topic that materialises beyond what could be captured in bibliometric data (e.g. interactions among member of different disciplines beyond authorship or references), which calls the use of other techniques to complement results from bibliometric analysis (e.g. surveys, interviews, and case studies).

Source: Technopolis, 2023 (https://www.ukri.org/publications/strategic-priorities-fund-spf-interim-impact-evaluation-jun-2023/)

Table 4 Challenge: Low observability of research impacts (example 2)

Case	Final Impact Assessment Study for the COST association
The programme	Established in 1971, COST is an intergovernmental programme dedicated to promoting research networks among researchers from COST member countries as well as from partner countries known as Near Neighbour Countries (NNC) and International Partner Countries (IPC). It has three strategic priorities: 1. Promoting and spreading excellence, 2. Fostering interdisciplinary research for breakthrough science, and 3. Empowering and retaining Young Researchers. The main instrument of COST is the funding of COST Actions, which are networks of researchers and innovators.
The challenge	To capture scientific collaboration and knowledge generation and flow emerging from informal interactions such as meetings, training schools, short-term scientific missions (STSMs), and conference grants. (i.e. beyond what is captured in publications). The study (and methodology) also addresses the challenge of establishing an appropriate counterfactual to establish the added value of COST.
Approach	The methodological approach includes two main elements: A **social network analysis** at the participant and regional level, aiming at: (a) providing a descriptive analysis of the networks underlying the COST Actions and characterising the structure of these networks, and (b) comparing networks constituted by the COST Actions with "default" networks in science and technology (to provide a counterfactual scenario). The analysis focused on two aspects: • The links between participating researchers to COST activities and the characteristics of those links: interdisciplinarity, early vs. advanced career stage (professional background), gender, geographical localisation, etc. • The extent to which the networks generated as a result of the COST activities compare to (public-funded) scientific collaboration networks and what the value-added of COST is regarding the original features of those COST networks. A **bibliometric and textual analysis** to create an overview of structured and unstructured outputs of the COST Actions, focusing on the scientific and societal breakthroughs supported by the programme. As economic/societal breakthroughs are wider than the impact of COST

	Actions on patents, only one societal/economic breakthrough was selected using the bibliometric data. The other societal/economic breakthrough was selected using natural language processing. Case studies are presented alongside the bibliometric analysis to help in identifying significant breakthroughs among COST Actions.
Limitation & caveats	A key limitation of the evaluation is that societal impacts are mainly captured via proxies, using innovation and scientific indicators (citation, patents, publications) that do not necessarily translate into further societal impact. Furthermore, the uptake of the breakthroughs by industry is supported by citation data, which only provides a partial view as to how innovations are diffused among firms.
Transferability	The methods employed in this study can be applied to the context of the UK, provided the evaluation has the time and resources to collect primary data on those networks or that monitoring systems allow to construct them based on secondary data. Like in the case of COST, multiple R&D programmes and instruments could have supported the pre-existing relationships among the networks. However, the focus on the characteristics of the network supported by a specific programme, and the extent to which those characteristics are different from “business as usual” could offer valuable insights into the effectiveness of programmes aimed at widening participation.

Source: Technopolis, 2021 (https://www.cost.eu/uploads/2022/01/3560_COST_FINAL_IA_Final_report_August2021.pdf)

2.3 Intangible benefits

Challenge description

RD&I programmes are expected to generate intangible benefits and assets that are difficult to measure or express in quantitative or monetary terms.

RD&I interventions would usually expect to support increased skills, as well as IP and know-how. Improved organisational designs and processes may also be expected, depending on the focus of the intervention. Increased reputation, in turn, could emerge from engaging with funders when they also act as direct clients with strong technical specifications (public procurement of innovation), and by de-risking initial R&D activities such that researchers and/or companies can demonstrate to private funders they are capable to progress complex, uncertain projects. Strengthened networks are expected to emerge from those programmes that support collaborative RD&I and/or engagement and dissemination activities.

Quantifying increased reputation and strengthened networks is particularly challenging (and the COST case in Section 2.2 above provides a useful example).

Methodological strategies to address challenge

- **Measuring/capturing development of intellectual property (IP):** The strategy entails identifying IP activity emerging directly from (or enabled by) the intervention, usually via monitoring data & systems, and/or via survey. Similarly to the "Citation analysis" approach (see Section 2.2), this can then be used to trace how IP supported by the programme has influenced further IP activity (to measure spillover effects). This analysis could be complemented with "Patent valuation" which relies on modelling to assign a monetary value to a patent portfolio to provide further assessment of intangible benefits. (See Table 6 below and Case 1 in separate Annex for an example).

- **Assessing changes in companies' valuation:** This strategy is used to provide lead indicators on the companies' future prospects (see "Lagged effects of RD&I investments", Section 2.1) and it also offers approximation of the intangible value of a company, and how this changes over time (and as a result of an intervention). (See Case 18 and 19 in Annex for examples). Another measure (albeit not covered in the case studies analysed) would be to assess the difference between the (physical) assets of the company (fixed assets, investment, profits, and other indicators included in financial reports) with the companies' valuation.

- **Social Network Analysis (SNA):** Quantifying network effects of collaborative RD&I programmes is another way to identify and measure intangible benefits of R&D public

investments. This entails assessing the "degree of centrality" and "betweenness centrality" of organisations within a network to test the extent to which this position changes after the intervention. When complemented with quasi-experimental methods (e.g. difference-in-difference) it can offer a quantitative measure of the impact of the intervention in this area of analysis. (See Table 5 below and Case 14).

- **Assessment based on qualitative data/perceptions/self-estimations:** This approach entails gathering evidence and perceptions on expected intangible benefits from an intervention, e.g. the extent to which programme participants believe that their reputation has been strengthened as a consequence of an intervention, how that has expressed itself in tangible benefits (such as faster career development, high-profile contracts). These are of course highly subjective and need to be appropriately weighted depending on the knowledge and seniority of the informant, as well as triangulated with other sources of evidence such as case studies, drawing from the impacts pathways as captured in the ToC. To increase the robustness of this data, evaluators tend to analyse it in the context of a TBE approach, and to focus on outcomes that cannot be easily quantified but that are still important to capture according to the ToC of the intervention. (See Cases 18, 21 and 23 in separate Annex for examples).

Table 5 Challenge: Intangible benefits (example 1)

Case	Assessing the collaboration and network additionality of innovation policies: a counterfactual approach to the French cluster policy
The programme	Established in 2005, the French government's Competitiveness Cluster Programme (CCP) aims to boost the competitiveness of the French economy by supporting collaborative innovation, with a focus on networking, knowledge exchange and collaborative R&D projects between cluster members.
The challenge	Measuring and quantifying network effects and intangible benefits. This challenge is mainly addressed by attempting to measure changes in the networks linked to the programme.
Approach	A quasi-experimental method combining propensity score matching with difference-in-difference (DiD) is used to assess the impact of the CCP based on the patenting activity of cluster members. The analysis estimates the effect of CCP membership on four outcome indicators: • The "co-invention rate": the collaborative share of a business' total innovations, defined as the number of co-invented patent applications associated with an individual business.

	• The regional co-invention rate: the number of co-invented patent applications associated with an individual business, where at least one of the co-inventors is in the same region as the business. • The size of a business innovation network: the number of direct collaboration partners that an individual business has. • The centrality of a business in the overall innovation network. This employs a concept from social network analysis called "betweenness centrality" which measures an entity's centrality in an overall network. The analysis includes control variables for whether a business: 1) has a history of collaborative patenting activity in the pre-intervention period, and 2) is a member of the EU's Framework Programme, a key policy instrument to support medium- to large-sized collaborative research projects in Europe.
Limitation & caveats	Patenting activity is more common in certain economic sectors, such as the energy sector and industrial manufacturing. Research has shown that businesses are less likely to apply for patents related to their process innovations as compared to their product innovations. Using only firms engaged in patenting activity, 116 enterprises out of over 10,000 participating firms in the programme, makes it highly unlikely to be a representative sample of CCP members. The DiD estimator used divides the sample into a single pre-treatment and post-treatment period, when businesses joined the programme on an ongoing basis. A preferable approach would be to use a staggered DiD estimator, which accounts for the fact that businesses are exposed to the treatment at different time periods. As a good practice, the study could have presented a formal test of the parallel trends assumption to help the reader to formulate a judgement on the validity of DiD research designs.
Transferability	The methodology is transferable to the evaluation of large-scale collaborative RD&I programmes expected to increase collaboration on patentable innovation, and non-patentable innovation. Note that the construction of the required dataset is labour-intensive. For non-patent innovation, this would require a sophisticated survey instrument to measure outcomes of RD&I collaboration, which can require considerable financial resources. The "betweenness centrality" measure could be applied to bibliometric data on the authorship of academic research papers to evaluate the effect

	of collaborative research programmes on the network positionality of participating.

Source: N'Ghauran, K. and Autant-Bernard, C., 2020 (https://academic.oup.com/icc/article-abstract/30/6/1403/6278428?redirectedFrom=fulltext)

Table 6 Challenge: Intangible benefits (example 2)

Case	The Influence of Patents in Twenty R&D Portfolios Funded by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy
The programme	The Energy Efficiency and Renewable Energy EERE is a $3bn R&D organisation within the Department of Energy in the United States. It funds R&D, technology demonstrations and deployment through ten offices in broad thematic areas such as advanced manufacturing, bioenergy, building technologies, geothermal technologies, hydrogen and fuel cell technologies, solar energy, vehicle technologies, waterpower and wind energy.
The challenge	Measuring intangible benefits of R&D investments (in this case intellectual property). This challenge is mainly addressed by using patents and assessing how they enable subsequent technological developments and innovations.
Approach	The EERE carried out evaluations for each of its offices to determine how the funding allocated to research portfolios across offices has advanced innovation through patenting activity. This report is a synthesis evaluation looking at 20 research portfolios. It provides a high-level assessment of the contribution of funding research portfolios between 1976–2018, underpinned by a patent citation tracing exercise. Patent citation analysis focuses on the links between generations of patents that are made by these prior art references (patents). The evaluation used **forward tracing and backward tracing:** • Forward tracing identified all patents in each EERE-funded portfolio to evaluate the influence of these patents on later generations of technology. This tracing was not restricted to subsequent patents from the technology associated with each portfolio, since the influence of a body of research may extend beyond its immediate technology. • Backward tracing took a particular technology, product, or industry, and traced it back to identify the earlier technologies upon which it

	was built. The leading organisations (beneficiary firms) in each technology (in terms of patent portfolio size) were identified, and tracing was carried out backwards in time from the patents owned by these organisations. This made it possible to determine the extent to which innovations associated with these leading organisations build on earlier EERE-funded research.
Limitation & caveats	The patent tracing analysis requires a specialised skill set to set up the correct filters to identify the patents to be traced. Without them, it is likely to miss out on important data, which can generate misleading evaluation results. Verifying EERE funding in those selected patents presented significant challenges for the evaluation lead at EERE. The monitoring and evaluation data collection system of national labs, recipients of around 50% of the EERE funding, do not distinguish sources of funding and track their data, and therefore do not report to EERE the final project outputs. This is a labour-intensive task to ensure the data set is properly built. The responsibility for adding prior art references differs across patent systems. According to the evaluation authors, in the U.S. patent system, patent applicants must reference (or cite) all prior art of which they are aware that may affect the patentability of their invention. Patent examiners may then reference additional prior art. In contrast, for patents filed at the EPO and WIPO, the examiner adds prior art references. The number of prior art references on EPO and WIPO patents thus tends to be much lower than the number on U.S. patents. Sectors differ in their patenting activity and strategies.
Transferability	The methodology has a good potential to be transferred to the UK context with some limitations regarding prior art references in WIPO and the EPO. The overall approach can be used similarly to assess a portfolio of programmes or for single interventions.

Source: Office of Energy Efficiency and Renewable Energy, 2022 (https://www.energy.gov/sites/default/files/2022-03/The%20Influence%20of%20Patents%20in%20Twenty%20R%26D%20Portfolios%20Funded%20by%20the%20U.S.%20Department%20of%20Energy%27s%20Office%20of%20Energy%20Efficiency%20and%20Renewable%20Energy.pdf)

2.4 Skewedness of impacts

Challenge description

The impacts of innovation tend to be highly skewed towards a small number of very successful projects with a long tail of low or no-impact projects. This presents a significant challenge for evaluating public investments in RD&I, as traditional evaluation approaches, in particular quantitative methods, are not well-suited to capture such impacts. There are two principal reasons for this:

- Focus on average outcomes: Traditional (quantitative) evaluation methods often estimate programme impact based on an average outcome (e.g. based on average values of programme monitoring data) or average treatment effect (estimated using a quasi-experimental or experimental research design). These approaches can obscure the transformative impact of a small number of highly successful projects.

- Short-term focus: Traditional evaluation methods frequently focus on short-term outputs and outcomes, such as the number of patents filed or immediate commercial successes. This short-term lens is inadequate for capturing the long-term, transformative impacts of highly innovative projects, which may take years or even decades to fully materialise (i.e. this challenge if further compounded by expected lagged effects of RD&I investments, see Section 2.1).

Methodological strategies to address challenge

- **Assessment based on qualitative data/perceptions/self-estimations.** Impact profiling involves collecting data across the entire population of programme participants to examine the range and variance of impacts, identifying outliers and exploring their significance.

  This exercise provides a basis for evaluators to develop a strategy for capturing skewed impacts where they exist. Such a strategy may involve focusing on the analysis of highly successful cases (see purposive outlier sampling, Table 8). Alternatively, it may involve creating aggregate measures of the total impact from all cases, rather than relying on evaluating average outcomes. (See Table 7 below and Cases 18 and 22 in separate Annex for examples).

- **In-depth case studies using purposive outlier sampling.** Purposive outlier sampling involves systematically identifying and studying cases that are unusual in some way, such as outstanding successes or notable failures. This approach allows evaluators to capture skewed impacts in those public investments in RD&I characterised by a small number of very successful projects with a long tail of low or no-impact projects. (See Table 8 below and Cases 18 and 22 for examples).

An outlier sample can be analysed using several approaches. This includes the development of in-depth qualitative case studies, which may utilise theory-based evaluation approaches to examine issues of additionality, attribution and contribution. The development of in-depth case studies also provides evaluators with an opportunity to analyse success factors for such cases, with a view to informing recommendations for future relevant public interventions.

Economic analysis or case study analysis can also be applied to outlier samples. In such cases, evaluators may conduct an in-depth analysis of the economic benefits accruing from highly successful cases. These benefits can then be aggregated to arrive at a conservative estimate of the overall economic benefits of a public investment in RD&I.

Table 7 Challenge: Skewedness of impacts (example 1)

Case	Cooperative Research Centres (CRC) Programme Impact Evaluation
The programme	The CRC Programme was established in 1990 to promote collaboration between industry and the research sector in Australia. A flagship initiative of the Australian government, it is of high strategic importance to the government, involves significant funding and has a high public profile. Over the period 1990–2020, the Australian government invested around AU$5.1bn in the programme to support 230 CRCs.
The challenge	Capturing skewed impacts. Given the nature of the research supported, it was expected that not all projects will be successful in generating economic outcomes. The study (and methodology) also addresses the challenge of establishing an appropriate counterfactual.
Approach	The evaluation team constructed a dataset designed to cover all CRC projects in receipt of funding over the period 2012–2020. A comprehensive survey of all CRC projects active during this period was fielded to identify the economic impacts of each project. This included information on cost savings, contract income, increased sales/revenue, value of patents sold and value of spin-off companies. To validate the survey findings and fill remaining data gaps, the evaluation team supplemented the survey data with information from secondary sources, including company reports and CRC Programme Exit Reports. Particular attention was given to verifying claimed high-value impacts.

	The evaluation team obtained information on the economic impacts of 77 percent of CRC projects active during 2012–2020. This dataset was used to compute a dynamic, global computable general equilibrium (CGE) model that quantified the total impact of all CRC project economic impacts on Australia's economy. The CGE model was used to compare the actual performance of the Australian economy (i.e. the scenario in which the CRC Programme was implemented) to a counterfactual scenario in which CRC Programme funding was allocated across other government expenditures, potentially having positive impacts elsewhere.
Limitation & caveats	CGE models are based on several high-level assumptions regarding market behaviours, agent decision-making, and economic structures. Data for some CRC projects was not available, while data for some others was incomplete. As a result, the findings of the CGE model should be interpreted as a conservative estimate of the CRC Programme's economic impacts on Australia's economy.
Transferability	The principal challenge to implementing this approach is the construction of a comprehensive dataset on project-level impacts. This data collection process is a labour-intensive endeavour that involves surveying programme beneficiaries, compiling data from programme reports and, in some cases, correspondence with beneficiaries to collect supplementary data. The feasibility of compiling such a dataset for a particular programme should be carefully considered before employing this evaluation approach. In particular, this approach is unlikely to be well suited to the evaluation of programmes with a large number of projects/beneficiaries.

Source: ACIL Allen, 2021

Table 8 Challenge: Skewedness of impacts (example 2)

Case	Returns on Research Funded under the National Institute for Health and Care Research (NIHR) Health Technology Assessment (HTA)
The programme	The NIHR's HTA Programme was established in 1993 and is the largest research programme dedicated to the UK's National Health Service (NHS). It funds research on the clinical effectiveness, cost-effectiveness and broader impact of healthcare treatments. The purpose of the programme is to ensure that high quality research evidence is made

	available to policymakers, practitioners and patients in a timely and efficient manner.
The challenge	Capturing skewed impacts. The risky and uncertain nature of research related to health treatments means that economic and policy impacts are expected to emerge in a sub-set of supported projects/initiatives. Also, in this case, the study was also facing a challenge of lagged effects.
Approach	The study involved two components: an economic analysis and a case study approach. • **Economic analysis:** The economic analysis aimed to quantify the potential (rather than actual) economic benefits of HTA funded research. The evaluation team purposively sampled ten HTA studies judged likely to have the largest potential economic benefits. The purposive sampling strategy employed an iterative process, combining the expertise of programme staff with objective case selection criteria. In the first stage of sampling, programme staff identified an initial list of potentially high impact HTA studies. However, only three met the inclusion criteria developed by the evaluation team to identify high impact cases. In the second stage, programme staff provided a further list of potentially high impact studies and the evaluation team additionally scanned recent volumes of the NIHR's HTA Journal to identify further studies. The evaluation team calculated the potential net benefit of the interventions if the selected HTA studies were implemented in the NHS. This was compared to total programme spending. • **Case studies:** The purpose of the case study component was to determine the programme's impact on policy and practice. The case studies were selected in consultation with programme staff, drawing on their knowledge of the HTA portfolio to identify studies likely to have an impact on policy and practice. These case studies were separate from those in the economic analysis, although there was some overlap. For each case study, data was collected based on document review and interviews of HTA study researchers.
Limitation & caveats	This evaluation methodology does not address the issue of programme additionality/deadweight (i.e. whether these studies were likely to have been undertaken in the absence of the programme).

	The studies included in the economic analysis were partly selected based on the feasibility of monetising their impact. Thus, studies with high potential impact may have been excluded from the economic analysis due to methodological challenges, rather than for substantive reasons.
Transferability	The purposive outlier sampling strategy employed in this evaluation is appropriate for programmes expected to have highly skewed impacts. Importantly, the resulting sample of cases can be analysed using quantitative and/or qualitative methods, as appropriate for the specific objectives of a given RD&I evaluation. This approach is particularly effective for conducting value for money analysis when programme impacts are skewed and the evaluation faces resource and/or data limitations. It allows evaluators to arrive at a credible conservative estimate of overall value for money with considerably fewer resources than are required to collect data for all programme initiatives/ projects.

Source: Guthrie, Hafner, Bienkowska-Gibbs, & Wooding, 2015

2.5 Additionality, attribution, and contribution

Challenge description

Additionality. A classic challenge in evaluation is to estimate the effect of a programme in excess of, or addition to, what would have happened anyway if an investment had not been made. That is, what difference did the investment make? An important consideration in relation to UK publicly funded research and innovation is the extent to which other public funders, research organisations or private sector organisations would have undertaken the research anyway in the absence of investment for a given organisation.

Attribution. A related classic challenge is the difficulty of attributing impacts to specific interventions, as there may be many confounding factors and external influences in a complex innovation system that contribute to an observed impact. In the context of evaluating public investments in RD&I, this challenge is especially pronounced in contexts where multiple RD&I public interventions are occurring simultaneously, or an intervention is implemented during a period of significant economic and/or technological change.

Contribution. The contribution challenge in policy impact evaluation focuses on identifying and understanding the role of a specific intervention within the broader context of influencing factors and interventions. This challenge acknowledges that outcomes often result from multiple factors, and isolating the contribution of a single policy or programme can be complex.

Methodological strategies to address challenge

- **Quasi-experimental design (QED) methods.** Quasi-experimental counterfactual analysis is a powerful tool for evaluators to establish programme additionality, particularly when randomised controlled trials are not feasible. By mimicking experimental conditions, quasi-experimental designs such as propensity score matching, or difference-in-differences help to isolate the effect of the programme from other external factors. (See Table 9 below). An overview of the quasi-experimental approaches most commonly used in policy evaluation, along with their uses cases, limitations and common implementation challenges is provided below in Section 2.5.1.
- **Theory-based Evaluation Methods (TBE).** As explained above, TBE methods offer a structured way of understanding an intervention and involve drawing hypotheses as to how and why it works and then testing those hypotheses, using qualitative and quantitative data. They are typically used for complex interventions and/or when a control is not available to draw a counterfactual scenario. They enable evaluators to leverage a range of qualitative and quantitative data to trace the direct and indirect

effects of a programme, distinguishing between outcomes that are genuinely attributable to the programme and those that might have occurred regardless.

These methods are particularly effective in addressing the attribution and contribution challenges in policy impact evaluation. They address these challenges by analysing the various causal mechanisms and interactions at play in a given intervention, enabling a nuanced assessment of the intervention's contribution to the outcomes observed. (See Table 10 below and Cases 5, 7 and 21 in separate Annex for examples).

- **Analysis of several "generic" data collection methods (surveys, interviews).** In some instances, resource and/or data limitations may preclude evaluators from utilising counterfactual analysis or theory-based evaluation methods to examine additionality, attribution and contribution. In such cases, an alternative approach is to report descriptive findings from a combination from several "generic" data collection methods, such as interviews and surveys. This approach can provide insights from programme beneficiaries and other programme stakeholders into the perceived additionality, attribution and contribution of a programme; however, it may be less reliable than the other approaches as it is often more susceptible to bias and misreporting.

  This approach can also be used to complement counterfactual analysis or theory-based evaluation methods, serving to validate findings obtained using the latter methods and strengthen the evidence base for evaluation findings. Combining these approaches is also useful in cases where counterfactual analysis and/or theory-based evaluation methods prove insufficient to answer all evaluation questions. In such cases, descriptive findings from "generic" data collection methods can be used to fill gaps in the evaluation methodology (See Cases 3, 4, 5, 15, and 20 in separate Annex for examples).

The similarities and complementarities of various concurrent RD&I investments (public and private) often make it difficult to disentangle their effects in order to address the three related evaluation challenges of additionality, attribution and contribution. Consequently, evaluation design should consider performing multi-programme evaluations where programme objectives are related, employing methods such as contribution analysis that allow for a deeper understanding of each programme's causal mechanisms (Stern et al., 2012).

Table 9 Challenge: Additionality, attribution, and contribution (example 1)

Case	Evaluation of the Enterprise Ireland Research, Development and Innovation Programme
The programme	The Enterprise Ireland (EI) RD&I Programme aims to support Irish companies of all sizes to engage in innovation and research activities. It is a comprehensive programme offering a range of direct financial supports and various indirect supports.
The challenge	Assessing additionality and attribution as observed effects are expected to emerge from a variety of factors.
Approach	The evaluation team constructed a dataset of Irish Enterprise Development Agency client firms. The dataset contained information on firm characteristics, economic performance, and receipt of EI RD&I Programme funding (or lack thereof). The dataset was constructed by linking data from the Irish government's Annual Business Survey of Economic Impact (ABSEI) with EI administrative data. Propensity score matching was used to compare beneficiaries of EI RD&I Programme funding to a comparator group of firms with similar characteristics that had not accessed this funding. The econometric analysis benefits from the availability of rich, longitudinal firm-level data on RD&I outcomes and economic performance captured through the ABSEI. It considered the impact of EI RD&I Programme funding on turnover, employment productivity and exports. Additionally, an online survey of EI R&D Programme participants was conducted. This survey measured self-reported impacts and benefits of EI RD&I Programme funding, as well as firms' self-reported attribution of improved R&D performance to programme support. Triangulating the results of the econometric analyses with the online survey data served as a validity check for the econometric results. The survey data also provided information on firms for which econometric analysis was not possible due to data limitations. This includes micro-enterprises, which are not included in the ABSEI. Interviews with EI RD&I Programme beneficiaries and programme staff were used to provide insights into the causal mechanisms linking EI RD&I Programme funding to the various impacts observed and to explore impacts which could not be studied quantitatively due to data limitations.

Limitation & caveats	The online survey of programme beneficiaries had relatively low response rates. This may have had implications for the representativeness of these surveys. Due to data availability issues, the econometric analysis focused on the impacts of the programme's direct supports only and did not include indirect supports (investments made via other agencies, such has the European Space Agency).
Transferability	Implementation of this approach requires: • Availability of rich, longitudinal firm-level data on RD&I outcomes and economic performance for both programme beneficiaries and plausible comparator firms. • Large sample size of programme beneficiaries and comparators. This helps to ensure that it is possible to match programme beneficiaries with suitable comparators and that the statistical analysis yields reliable estimates. • Sufficient resources for the collection of original survey and interview data.

Source: Technopolis Group & Economic and Social Research Institute, 2020 (https://enterprise.gov.ie/en/publications/evaluation-of-enterprise-ireland-research-development-and-innovation-programme.html)

Table 10 Challenge: Additionality, attribution, and contribution (example 2)

Case	Evaluation of investments in Research and Technological Development (RTD) infrastructures and activities supported by the European Regional Development Funds (ERDF) in the period 2007–2013
The programme	Over the period 2007–2013, €17bn in ERDF funding was invested in projects to support RTD activities through 215 operational programmes (OPs). These activities included support for: scientific R&D activities, collaborative research activities, the internationalisation of research activities, development of researchers and other personnel involved in R&D activities, technology-transfer activities and the valorisation of research results.
The challenge	Assessing additionality, attribution, and contribution. Due to a lack of non-treated EU regions, quasi-experimental methods such as matching or difference-in-differences could not be used.

Approach	The evaluation employed Contribution Analysis (CA), a theory-based evaluation method that focuses on determining causal relationships between specific interventions and observed changes. The CA draws evidence from case studies, and quantitative methods were employed in the CA. • **Case studies**: In the case study component, the principles of CA were used to guide the collection and processing of evidence into seven case studies at Member-State level addressing nine OPs, including in-depth analyses of three selected policy instruments per case study. The programme's outputs, outcomes and impacts were analysed to understand what happened (i.e. the direct effects of ERDF support). Cross-case analysis was then conducted to understand how policy instruments work collectively as part of a broader "causal package". This involved analysing the underlying assumptions, pre-conditions and supporting factors influencing the achievement of causal packages to explain why and how the observed effects had occurred. The case study analysis was informed by over 200 interviews conducted with stakeholders and project beneficiaries, analysis of programme monitoring data, document review and literature review. • **Quantitative methods**: The quantitative component of the CA involved conducting multivariate regression analysis to test hypotheses regarding how ERDF support contributed to regional economic and innovation outcomes (e.g. patenting activity, private sector R&D expenditure). These hypotheses were developed based on preliminary findings from the CA. The regression analysis tested the presence of correlations between ERDF support and the abovementioned outcomes. However, it cannot provide insights into the causality of these relationships. Rather, this analysis is employed as a complement to the case study analysis.
Limitation & caveats	The evaluation relied heavily on data collected via interviews but mitigated this limitation by triangulating the analysis with programme monitoring data, secondary sources and complementary quantitative analysis.
Transferability	Implementation of this approach requires: • Availability of comprehensive programme monitoring data.

	• Sufficient resources for a comprehensive programme of interviews with programme beneficiaries and stakeholders. • Sufficiently large number of observed entities (e.g. individuals, firms, regions) to conduct a reliable multivariate regression analysis.

Source: European Commission, 2021 (https://ec.europa.eu/regional_policy/en/information/publications/evaluations/2021/evaluation-of-investments-in-research-and-technological-development-rtd-infrastructures-and-activities-supported-by-the-european-regional-development-funds-erdf-in-the-period-2007-2013)

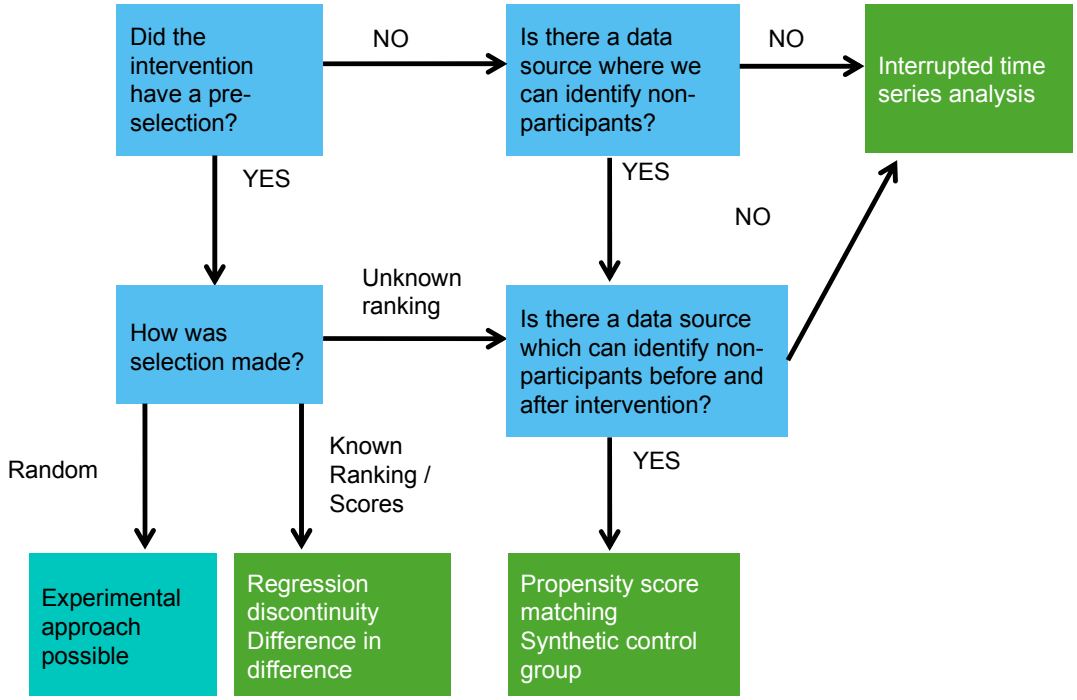
2.5.1 A note on counterfactuals

When evaluating the impacts of publicly funded programmes, it is good practice to establish a counterfactual scenario in which the programme did not take place to attempt to account for deadweight (what might have happened anyway) and additionality (how much has the programme contributed after discounting for deadweight).

A counterfactual scenario cannot be directly observed (one cannot observe the outcomes for the same participating company or entrepreneur with and without the programme). Consequently, evaluators must rely on drawing comparisons with control/comparator groups, making sure that they are as similar as possible to the participants before the programme took place. A further challenge emerges if those that win a place in a programme are more dynamic and have the most to gain from it, creating a selection bias “problem”. Additional statistical techniques are required to minimise this problem.

The choice of counterfactual. The choice of a counterfactual can be influenced (or restricted) by a range of factors, and it is useful to think about it as decision tree (see Figure 3). In particular, there are two key factors that drive that decision tree: **programme design** and the **availability of data** (on the treatment and control groups).

Figure 3: Counterfactual analysis – decision tree



Source: Technopolis. Developed in the context of the BEIS M&E training delivered in partnership with UCL.

The first important question to ask is whether or not the intervention had a selection process and, if so, how that was conducted. Almost all programmes will have a selection process, while interventions such as regulation change or new open access research infrastructure would not.

The funders of the programme could have decided to randomise the assignment of applicants/participants to either the intervention or control group. This allows for the implementation of a Randomised Control Trials (RCT) approach (provided data on treatment and control groups is available for before and after the intervention, either via primary data collection or secondary sources). See Box 2 for further reflections on RCT.

However, **random selection is often not possible or desirable in the context of R&I interventions**. In fact, it is common for funders of public R&I programmes to allocate funding competitively rather than randomly because the greatest social impact is likely to be derived from selectively supporting the strongest ideas and actors. In fact, we did not find any examples of RCTs in our search of real-world examples (although some are presented in the box below).

When participant selection is done through a selection process (i.e. not at random), it is common to use unsuccessful applicants as a counterfactual group. This ensures similarities between the programme beneficiaries and the control group in terms of the reasons for applying and the characteristics that are likely to affect the programme outcomes. There were various examples of this approach in the case studies analysed and this is further documented in Table 11 (at the end of this sub-chapter). Moreover, in a recent study we conducted for

Innovate UK,[11] we analysed the methodologies (and findings) used across 16 evaluations the organisation had commissioned. We found that 10 used quasi-experimental methods, and all relied on unsuccessful applicants as the control group.

There are limitations around the use of unsuccessful applicants as a control group. One could argue their characteristics are negatively correlated with the expected outcomes, i.e. that the unsuccessful applicants are less likely to perform well in the future, and observed differences between control and treatment after the intervention may be due to those characteristics, rather than the effects of the programme. To mitigate this limitation, evaluators tend to:

- Implement matching techniques to find the most appropriate control groups among unsuccessful applicants, noting that this will only account for observable characteristics. This is only possible if the sample of unsuccessful applicants is big enough (which is not always the case).
- Apply a difference in difference approach after matching to further reduce bias (and account for unobservable characteristics that do not change over time).
- Apply regression discontinuity design to increase comparability across groups to further reduce bias.
- Test different controls groups to help to increase the robustness of the results, such as completing and comparing the results emerging from using unsuccessful applicants as control groups, with results emerging from using drawing control groups from general population (via matching techniques).

These approaches are documented in Table 11. Note that in line with what is considered best practice, the majority of evaluations included in the case studies used mixed methods approaches (given not all impact types/indicators of interest can be subject to a "counterfactual treatment") and a Theory of Change to guide the evaluation.

When there are no clear comparison groups to rely on, there are at least two alternative methods that can be employed:

- Pipeline design, where later cohorts are used as comparators for earlier cohorts. This is only possible if the cohorts were similar at the various stages of the programme.
- An interrupted time series approach, to track the changes in treatment group over time and detect if participation in the programme lead to a marked step change in performance.

Another option is to use "intensity analysis". This is not a counterfactual option per involves comparing the difference between "low" involvement and "medium" or "high" involvement. This may be relevant for interventions that combine free access to knowledge and information ("low" involvement) with more dedicated and targeted support ("medium/high" involvement).

[11] Technopolis (2023). "A review of Innovate UK evaluation practice". Internal report.

These methods are less robust than the quasi-experimental approaches listed above but do offer a viable option when control groups are not available.

Unit of analysis and impact types. In our experience evaluating public RD&I programmes, and as showcased in the examples identified in this study, counterfactual methods are more commonly implemented where companies are the major beneficiaries and the primary **unit of analysis**. Out of 107 cases that we included in our repository, 31 used quasi-experimental methods. Of those 31, 74% had companies as the unit of analysis.

The focus on companies as the unit of analysis is due to several reasons:

- The key indicators for economic impact are measured routinely (e.g. investment, income, exports, employment) and can be found in secondary data sources (which makes it viable to run analyses and comparisons when time and resources limit primary research).
- The indicators of interest can be used to derive conclusions in terms of the economic impact of the programme (on growth, employment, productivity, etc.).
- A larger number of participants in a programme means that statistical analyses can be conducted where this may be more difficult with smaller populations (although synthetic control groups could help to mitigate this data requirement). It is easier to find programmes that have supported 100+ companies than 100+ researchers or 100+ research organisations.
- Comparing the performance of businesses tends to be more readily accepted by RD&I funders, in comparison with comparing the performance of researchers, in part because the former are easier to quantify (as indicated in point 1, above) and in part because of the complexities in defining and measuring performance of the academic community. The Research Excellence Framework is currently the standard in the UK for defining research excellence, and costs £471m to run[12] (albeit covering 150 institutions and all disciplines).

The case studies identified in this study show that there are ways forward in terms of using researchers as units of analysis, for example in the evaluation of the Strategic Priorities Fund (Table 3) or Final Impact Assessment Study for the COST association (see Table 4).

There are also several examples of having individuals as units of analysis when implementing behavioural economics approaches, which also tend to randomise. However, those are usually restricted to specific types of public interventions (e.g. nudging researchers to change citation behaviour around data), which may be relevant for community-wide public interventions related to public engagement or knowledge exchange or open science.

Box 2 presents a note on Randomised Control Trials (RCT), while the table below provides an overview on how the counterfactual has been built across the 20+ cases included in this study.

[12] Technopolis (2023). REF2021 Cost Evaluation.
https://repository.jisc.ac.uk/9184/1/REF_2021_cost_evaluation_final_report.pdf

Box 2: A note on Randomised Control Trials (RCT)

RCTs are considered to be the gold standard of evaluation as they are the most effective way of addressing the problems of bias described above. Endogeneity/selection bias emerges when those who seek to be part of the intervention or have been selected to be part of it have the most to gain from the intervention. Randomising the selection of participants helps to address selection bias in so far as this selection is not based on those characteristics that can also influence the expected outcomes (higher productivity, more impactful research).

Randomising is not common in RD&I programmes since such interventions tend to target those likely to be more successful, with strong processes in place to fund excellence. As such, randomisation, if it happens, needs to be part of the programme design (rather than the evaluation design). None of the examples identified in this study (either in the long or the short list) use RCT, indicating that, in all those cases, public funders considered that it was not appropriate to randomise their interventions. There are however a few examples of RCTs in the innovation field, mostly related to innovation vouchers, including for example "Creative credits".[13]

That is not to say that RCT should not be considered by funders such as DSIT, as they can serve as a powerful approach to support experimentation at programme design stage. The Innovation Growth Lab guidelines offer a valuable tool for understanding when and how to implement RCTs.[14] In recent years, there has also been a push from Central Government to promote experimentation. A report from the What Works Network and Economic and Social Research Council (ESRC), "Rise of Experimental Government", documents cross-departmental efforts in that direction and describes 52 projects supporting randomisation (note that only three examples are focused on Business Support and none relate to research and innovation more specifically).[15] More recently, the Evaluation Task Force, a Cabinet Office-HM Treasury unit, was set up in 2020 to tackle the main barriers to robust evaluations in government and foster a culture of evaluation and experimentation. In the RD&I space more specifically, DSIT/UKRI have set up a Metascience Unit, to fund experiments on what works in terms of supporting research and innovation. There are also recent examples of randomisation used in relevant aspects of RD&I such as for the peer review process.[16]

[13] What works growth (2018). How to evaluate innovation: UK "creative credits" programme (randomised control trial). https://whatworksgrowth.org/resource-library/how-to-evaluate-innovation-uk-creative-credits-programme-randomised-control-trial/

[14] Innovation Growth Lab (2016). Running randomised controlled trials in innovation, entrepreneurship and growth: an introductory guide. https://www.innovationgrowthlab.org/sites/default/files/field/image/a_guide_to_rcts_-_igl_09aKzWa.pdf

[15] The Rise of Experimental Government: Cross-Government Trial Advice Panel Update Report (2018). https://assets.publishing.service.gov.uk/media/5bdc3a1be5274a6e355faa11/RiseExperimentalGovernment_Cross-GovTrialAdvicePanelUpdateReport.pdf

[16] Technopolis (2023) Review of Peer Review. https://www.ukri.org/wp-content/uploads/2023/07/UKRI-060723-Review-of-peer-review-Final-report-revs-v2.pdf

Table 11 Quasi-experimental design as identified in case studies

Approach	Applicability	Challenges and limitations
Difference-in-differences (DiD)	1. **Comparable treatment and control groups**: DiD is most suitable for scenarios where there are groups that have not been exposed to an intervention but are otherwise similar to the treatment group. This similarity allows for a more accurate estimation of the counterfactual. A fundamental assumption underpinning the DiD methodology is the parallel trends assumption. This assumption posits that, in the absence of the intervention, the treatment and control groups would have followed similar trends over time in the outcome of interest. Ensuring or testing for the parallel trends assumption often involves analysing pre-intervention trends to confirm that they are indeed parallel. 2. **Sufficient sample size**: DiD is most suitable for evaluating the impact of programmes with a medium to large number of beneficiaries and potential control units. This is crucial for ensuring the statistical power of the analysis, in order to draw reliable conclusions.	• Testing the parallel trends assumption requires time-series data for the period before the intervention for the outcomes of both the treatment and control groups. If the parallel trends assumption is not met, it may be possible to proceed with the analysis by combining the DiD method with a matching method.
Dynamic panel data models	1. **Panel data:** Where panel (longitudinal) data is available, it may be possible to use dynamic panel data models for counterfactual analysis. These are panel data models that employ instrumental variable techniques, such as the Generalised Method of Moments and the Arellano-Bond estimator.	• This method requires panel data for a large number of entities (individuals, firms, etc.).

	2. **Sufficient sample size:** Statistical analysis using panel data models is most suitable for evaluating the impact of programmes with a medium to large number of beneficiaries. This is crucial for ensuring the statistical power of the analysis, to draw reliable conclusions.	
Matching	1. **Programmes with comparable groups**: Matching is effective for evaluating programmes where there are clear, identifiable groups that can be compared. These are situations where there are both beneficiaries and a plausible control group, such as unsuccessful applicants or comparable entities not participating in the programme. 2. **Sufficient sample size:** Matching is most suitable for programmes with a medium to large number of beneficiaries and potential control units. In such cases, the method can effectively identify comparable units within the dataset. Moreover, a sufficient sample size is crucial for ensuring the statistical power of the analysis to draw reliable conclusions.	• This method requires data on relevant characteristics of both the treatment and control groups. This is essential for creating matched pairs or groups that are similar in key relevant characteristics, except for the treatment. The comparison is less plausible with a small number of matching variables. However, there may be computational challenges when implementing the analysis with a large number of matching variables. • Results of the analysis may be sensitive to the specific matching method used. This is due to differences in how various matching methods identify and pair treatment and control units based on their characteristics. Given this sensitivity, it is good practice to test the robustness of the findings. This may involve applying multiple matching methods to the same data

		and comparing the results for consistency across various methods. • Also, this method only accounts for observable characteristics.
Regression discontinuity design (RDD)	1. **Assignment based on a cut-off point:** RDD is used when individuals, regions, or entities are assigned to a treatment based on whether they fall above or below a certain threshold or cut-off point, such as test scores, income levels, or age. It is particularly useful for evaluating policies or programmes that have clear eligibility rules. For example, evaluating the impact of a programme where applicants are selected based on application scores and there is a pool of unsuccessful applicants that scored similarly to successful applicants. 2. **Sufficient sample size:** A fundamental requirement for the successful application of RDD is the existence of a sufficient number of entities (individuals, firms, etc.) both above and below the treatment cut-off point. This is crucial for ensuring the statistical power of the RDD analysis in order to draw reliable conclusions.	• This method requires data on where treated and potential control units fall in relation to the treatment cut-off point. In RD&I evaluations, this commonly involves analysing oversubscribed RD&I programmes where inclusion in the programme was determined by application scores that are available to the evaluator. • The focus on those just above/ under the threshold means findings may not apply to intervention as whole. • There is a need to ensure that the treatment and comparison groups are similar and that the threshold is not likely to affect outcomes.
Synthetic control method (SCM)	1. **Unique interventions in small treatment groups**: SCM is most suitable when the intervention or programme being evaluated is unique to a specific group or region. This method is often applied in cases with a small number of units receiving the treatment, making traditional quasi-experimental methods challenging.	• This method requires extensive data on both the treated unit and potential control units for the period before the intervention. This includes data on variables that are predictive of the outcome of interest.

	2. **Lack of a clear control group**: SCM constructs a synthetic control group by combining data from multiple untreated units to create a counterfactual scenario that closely matches the treated unit's characteristics before the intervention. 3. **Case studies:** The evaluation's focus is on understanding the impact of an intervention on a specific entity, rather than generalising findings across a broader population.	• When SCM is applied to regions or countries, the "attribution/contribution" story can become more tenuous and requires careful validation of the study's assumptions.

3 Reflections and recommendations

3.1 Overall reflections from assessing the national and international evidence

The rapid assessment of 107 cases and further development of 20+ cases identified in this study revealed some initial findings with respect to UK and international RD&I evaluation practice:

There is **significant interest from countries to understand how to measure the impact of RD&I public interventions**. This is particularly the case in the US, EU, France, Germany, and the UK where more impact evaluations were found. The fact that those reports were publicly available, in many cases in English, indicates that funders in these countries are open to sharing the results of their evaluations with the wider national and international public.[17]

RD&I funders in advanced economies also undertake impact evaluations of their main programmes regularly (e.g. CIFRE, ZIM, COST). This gives them the opportunity to incorporate lessons learnt, for both future programmes and evaluation, design, and implementation.

Also, most of the evaluations tend to be programme specific. A few exceptions are evaluations looking at the effects of a number of different supports and interventions from a single agency or organisation over a period of time (e.g. Fraunhofer).

Most impact evaluations are carried out ex-post (just after the end of the intervention). Although some medium-term evaluations collect data and evidence to try to capture early outcomes and identify impact pathways, particularly in the UK. One concrete example of a developmental evaluation is the RWJF Change Leadership Initiative in the US, where the programme designers and administrators work with the evaluation team to learn what works within the programme and translate that into ongoing improvements of the initiative.

Multi-phased evaluations are less common but useful to set up appropriate baselines, inform delivery and incorporate lessons learnt into future evaluation rounds.

Ex-post evaluations that take place 2+ years after the programme ended benefit from the fact that more time has passed for benefits to materialise. However, they have to grapple with the fact that participants may be more reluctant to take part in primary data collection exercises; and that it becomes more difficult to draw strong conclusions from secondary data sources when using quasi-experimental design, since many external factors may be at play in

[17] A notable exception has been South Korea. This is even though they have an advanced and complex evaluation system, including a specialised evaluation agency, carrying out regular assessments (yearly), using summative and developmental evaluation approaches (as reported in their Annual STI report). The evaluations however were all in Korean, which made it more difficult to engage with the content (even using automated translations).

explaining observed results. The triangulation of evidence from different sources may help to alleviate this problem.

The UK is more ambitious in its attempts to capture final impact. The majority of the evaluations tend to measure what in the UK would be considered outcomes (i.e. benefits to direct participants in a programme or intervention rather than to society at large). Measuring impact is less common due to timing, complexity, and data availability. Also, the challenge of lagged effects makes it difficult to capture robust evidence of impact.

Evaluators attempting to measure impact as part of their brief have to be content with providing examples of early impact or make extrapolations (qualitative or quantitative) based on early signals of outcomes.

Measuring the impact of RD&I is challenging and there is no methodological silver bullet. Evaluators are confronted not only with the traditional evaluation challenges, i.e. how best to establish what would have happened in the absence of the intervention, but also with multiple challenges associated with the nature of RD&I interventions as described in this report.

Evaluators tend to deal with these challenges by grounding the evaluation in a Theory of Change (ToC) or Logic Model (LM) to tease out the expected casual pathways that "theoretically" explain why the intervention is expected to facilitate the progression from outputs to outcomes and impacts. ToCs and LMs are then used to drive the methodological approach, with a combination of methods, both qualitative and quantitative being used to address the diversity of expected results. Even studies that make use of quasi-experimental approaches, complement these with extensive use of qualitative data which is fundamental to (i) explain not only what has been achieved but why and how, and (ii) capture outcomes that are not easily measurable but highly relevant to the objectives of the intervention (e.g. changed national and international reputation, increased confidence in investing or adopting new technologies, increased ability to work among multi-disciplinary teams).

We also found that the UK leads the charge in the use of Theory-Based Evaluation (TBE) approaches (as defined in the Magenta Book, e.g. Process Tracing and Contribution Analysis) for RD&I evaluations. When qualitative approaches are used internationally, they tend to be in the form of interviews and case studies.

Having good M&E data collection systems enables robust assessments and the estimation of counterfactuals, as well as reducing the time and cost of data gathering during evaluations. The existence of good contact information on participants allows evaluators to link their information with pre-existing secondary sources, from bibliometric data (e.g. Scopus, Web of Science, OpenAlex), to datasets on patents (e.g. EPO PASTAT, Orbis IP), investment (e.g. Crunchbase, Pitchbook, Dealroom), and business performance (e.g. Moody's FAME, ONS Inter-Departmental Business Register). Data collected via secondary sources allows the building of counterfactuals and accessing time-series data (which in turn improved the implementation of those counterfactuals).

Simple changes in the way the data is stored – especially the use of unique identifiers (e.g. VAT for companies, ORCID for researchers) – can help greatly reduce the resources needed for data linking.

Resources such as Gateway to Research and Researchfish (used by UKRI) are also great resources to inform evaluations, in particular to (i) report on outputs (minimising the data that needs to be captured via primary data collection), and (ii) draw benchmarks across different types of interventions (compare investments and outcomes across a variety of interventions e.g. fellowship programmes and grants over the past 10 years, grants and programme supporting international RD&I collaboration).

3.2 New and emerging methods

This sub-section presents notable recent advancements in evaluation methods that have introduced new possibilities for more rigorous and versatile evaluations in various contexts. While these methods have yet to be widely applied in real-world evaluations, their characteristics suggest they that hold considerable potential for improving evaluation practice.

Synthetic control method (SCM). SCM is a relatively recent methodological development which has quickly emerged as a key approach to quantitative impact evaluation. The main advantage of SCM over traditional quasi-experimental methods is that this technique is suitable for small samples, and consequently, to explore the effect of an intervention on aggregate units such as countries, regions, or cities. The seminal paper that proposed this technique was published just 12 years ago, in 2012 (to study the effect of the 1990 German reunification on West Germany's GDP per capita).[18] More recently several papers used this technique to explore the effect of Brexit on outward investment by UK firms[19] and employment,[20] among others.

In recent years, several notable innovations in SCM have further increased the versatility and reliability of this method. A key recent innovation in this field is the development of the synthetic difference-in-differences method.[21] This combines the advantages of difference-in-differences (DiD) and SCM approaches, offering a hybrid method that is effective across a wide range of evaluation contexts. The utilisation of "pure" SCM approaches requires extensive data over time to be available for both the treated unit and potential control units during the pre-intervention period. In real-world evaluation settings, these data requirements are highly restrictive. The utilisation of DiD, on the other hand, rests on an assumption of similar pre-intervention trends, a condition that is often challenging to meet. This hybrid method draws on

[18] Abadie, Alberto & Diamond, Alexis & Hainmueller, Jens. (2012). Comparative Politics and the Synthetic Control Method. American Journal of Political Science. 59. 10.2139/ssrn.1950298.
[19] Holger Breinlich, Elsa Leromain, Dennis Novy, Thomas Sampson (2020). Voting with their money: Brexit and outward investment by UK firms, European Economic Review, Volume 124, 2020. https://doi.org/10.1016/j.euroecorev.2020.103400.
[20] Elissaios Papyrakis, Lorenzo Pellegrini & Luca Tasciotti (2023) Impacts of the Brexit referendum on UK employment: a synthetic control method approach, Applied Economics Letters, 30:11, 1407–1410, DOI: 10.1080/13504851.2022.2056567
[21] Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, *111*(12), 4088–4118.

the strengths of both approaches and is applicable across a wider range of evaluation scenarios than traditional DiD or SCM methods.

Recent advancements in machine learning have also increased the relevance of SCM to evaluation. To give an illustrative example, the Synthetic Control Using Lasso method provides a flexible, data-driven way to construct high-quality synthetic control groups.[22] This approach allows for the development of synthetic control groups based on a larger number of control units than conventional approaches, making it particularly suitable for "big data" settings. Additionally, this approach automates the SCM model selection process.

These recent advances in SCM allow for more accurate, reliable, and nuanced counterfactual analysis across a wider range of evaluation scenarios. As the method continues to evolve, it is likely to play an increasingly important role in quasi-experimental approaches to evaluation.

Further developments in theory-based evaluation methods. Bayesian process tracing is a recent methodological development that aims to introduce greater transparency and rigour to process tracing methods.[23] This method involves making all process tracing judgements explicit, including by identifying the strength of each piece of evidence used in the analysis on a four-category scale. Comprehensive, formal Bayesian process tracing analyses are onerous and therefore not recommended for conducting evaluations; however, applying the principles of Bayesian process tracing to the most important pieces of information used in a Contribution Analysis or other theory-based method can serve to strengthen the empirical basis of the analysis.[24] As a relatively new methodological development, Bayesian process tracing has yet to be applied widely in evaluation although it was recently added to the UN's methodological toolkit for evaluation.[25]

3.3 Recommendations on good practice

National and international practice shows that a robust RD&I evaluation must include the following to address its most important challenges:

[22] Hollingsworth, A., & Wing, C. (2020). Tactics for design and inference in synthetic control studies: An applied example using high-dimensional data. *Available at SSRN 3592088*.
[23] Fairfield, T., & Charman, A. E. (2017). Explicit Bayesian analysis for process tracing: Guidelines, opportunities, and caveats. *Political Analysis*, *25*(3), 363-380.
[24] Bennett, A. (2023). "Causal Inference and Policy Evaluation from Case Studies Using Bayesian Process Tracing." In *Causality in Policy Studies: a Pluralist Toolbox* (pp. 187-215). Springer International Publishing.
[25] https://erc.undp.org/methods-center/methods/data-analysis-approaches-methods/process-tracing

Maintain a sense of proportionality, with a design that takes into account the policies being evaluated, the evaluation questions being asked, and the resources available

A ToC at its core, to set up expected impact pathways, guide the methodological approach, and inform the conclusions and findings

A combination of different quantitative and qualitative methods, and triangulation of evidence across a variety of sources and methods

The use of counterfactual and benchmarks (either quantitative and qualitative) to tease out what the intervention is delivering in excess of what would have happened anyway

A clear description of assumptions and statistical tests for quantative methods including for quasi-experimental approaches, and sensitivity analysis

A clear discussion of limitations and caveats of methods employ, and frameworks to asses the strength of the evidence

Different reports for different audiences, with for instance, short executive summaries catering for policy and decision makers, extended main reports for programme managers, and technical annexes for policy analysts and other evaluators

Appendix A Methodology

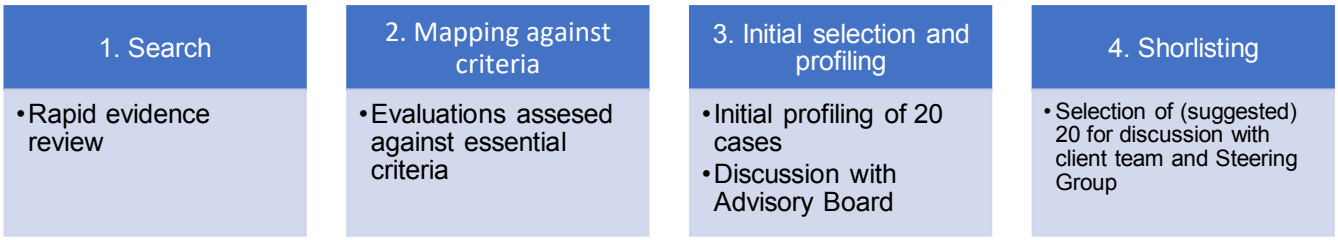
As briefly explained above, the study is being conducted in two Phases:

- **Phase 1 (May – August 2023):** Focused on identifying a long list of published evaluations of RD&I programmes in advanced economies (US, France, Germany, Israel, Finland, Canada, Australia and the UK) and looked at the effectiveness of the methodologies used to assess the impact of those interventions. Figure 1 summarises the methodological approach of the study.
- **Phase 2 (August 2023 – January 2024)**: Will focus on case study development and synthesis of findings. The main deliverable is a final report containing a compendium of 20 cases, and a synthesis to highlight what the UK can learn from these evaluations.

We will also produce a searchable repository which has been built using PowerBi and included as part of the Phase 1 deliverables and will be further updated with any relevant report/evaluation we may encounter before the end of the study.

Figure 4 summarises the steps followed in Phase 1, which is the focus of this report, with each step described in more detail below.

Figure 4: Phase 1 steps



Step 1: Search and initial classification. In this first step we conducted a rapid evidence review to identify a long list of real-world evaluations. Following the study specification, we searched for **four criteria: type of R&D, types of RD&I intervention, methodology and funding institutions**. Table 12 below showcases the categories identified for each criterion in our iterative search.

Table 12 Essential search criteria

Criteria	Including	Criteria	Including
Type of R&D	• Basic • Applied • Experimental development	**Methodology**	• Qualitative • Quantitative (RCT, econometrics) • *Following the Magenta Book guidance* (see below)
Types of RD&I intervention*	• Fellowships • Knowledge Transfer Partnerships • Research and innovation grants • Direct assistance • Networking activities • PhD grants • Research infrastructure • Others	**Institution**	• Universities • Government • Government supported institutions (e.g. Catapults in the UK, National Science Foundation in the US)

* As agreed with DSIT, fiscal incentives such as R&D tax credits were excluded from the search.

Table 13 Guidance on evaluation methods for search of potential cases

Type of evaluation approach	Method	Type of evaluation approach	Method
Experimental and quasi-experimental methods	Randomised Controlled Trials Interrupted time series analysis	**Methods for value for money evaluation**	Cost-benefit analysis Cost effectiveness analysis

	Difference-in-differences Regression Discontinuity Design Propensity Score Matching Synthetic control methods Instrumental variables	**Methods for synthesising evidence**	Systematic reviews Rapid evidence assessment Meta-analysis Meta-ethnography Realist synthesis
Theory-based methods	Realist evaluation Contribution analysis Process tracing Bayesian updating Contribution tracing Qualitative Comparative Analysis Outcome harvesting Most-significant change	**Generic research methods**	Interviews and focus groups Case studies Surveys and polling Output or performance monitoring Qualitative observation studies Consultative/deliberative methods

Source: https://www.gov.uk/government/publications/the-magenta-book

Finally, **other inclusion/exclusion criteria included publication year.** We focused mainly on evaluations published in the past 10 years to capture progress made in the implementation of relevant approaches to evaluate RD&I programmes in the past decade (including on implementing quantitative and qualitative approaches to address the counterfactual scenario in a more systematic way).

We relied on **four main sources**:

- Advanced economies funders' websites (as listed in Table 14) and intergovernmental organisations such as OECD.
- Technopolis' internal repository of evaluation reports. In practice, we asked our more senior colleagues in the Technopolis Group Science, Technology and Innovation (STI) community of practice to identify relevant evaluations and reports.

- Examples/evaluations identified in prior relevant exercises include the "What Works Growth Centre",[26] Evidence Review 9 Innovation: grants, loans and subsidies[27] and Nesta's Compendium of Evidence on the Effectiveness of Innovation Policy Intervention[28] which have carried out inventories and analysis of impact evaluation methods.
- Ad hoc searches in Google Scholar and Google.

Table 14 Funders' websites searched for evaluation report

Funder	Website	Country / region
National Science Foundation	https://www.nsf.gov/	US
Department of Energy DOE	https://www.energy.gov/	US
Department of Health and Human Services HHS	https://www.hhs.gov/	US
National Institutes of Health	https://www.nih.gov/	US
Department of Defense	https://www.defense.gov/	US
Advanced Research Projects Agency-Energy	https://arpa-e.energy.gov/	US
European Commission	https://commission.europa.eu/research-and-innovation_en	European Union
Research Council of Norway (RCN)	https://www.forskningsradet.no/en/	Norway
Agence Nationale de la Recherche (ANR)	https://anr.fr/en/	France
German Research Foundation,	https://www.dfg.de/en/index.jsp	Germany

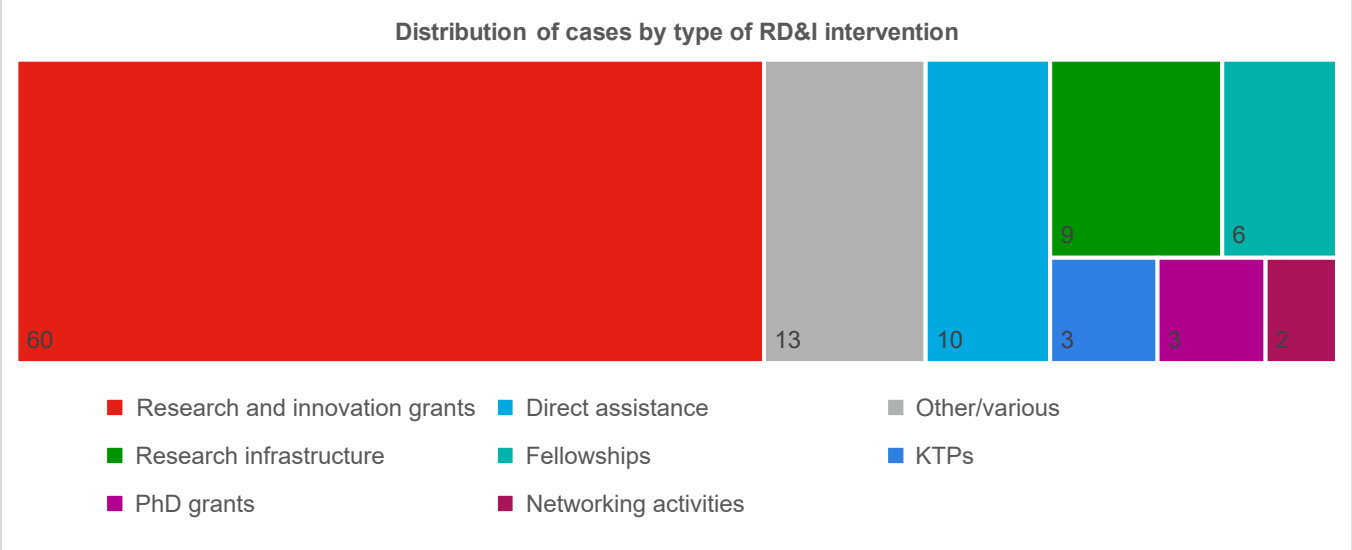
[26] https://whatworksgrowth.org/
[27] https://whatworksgrowth.org/wp-content/uploads/15-10-20-Innovation-Grants-Loans-Subsidies-Report.pdf
[28] https://media.nesta.org.uk/documents/impacts_of_innovation_policy_synthesis_and_conclusion_final.pdf and https://www.nesta.org.uk/report/the-impact-and-effectiveness-of-policies-to-support-collaboration-for-rd-and-innovation/

Israel Science Foundation	https://www.isf.org.il/#/	Israel
Israel Innovation Authority	https://innovationisrael.org.il/en/	Israel
Korea Institute of STI evaluation and planning	https://www.kistep.re.kr/eng/	South Korea

A total of 107 cases were identified following this approach. This was an iterative process whereby further efforts were made after 3–4 touching points to make sure that the cases identified offered a good spread across the categories showcased in Table 13 above. The cases form the repository of the study and are presented in a separate annex.

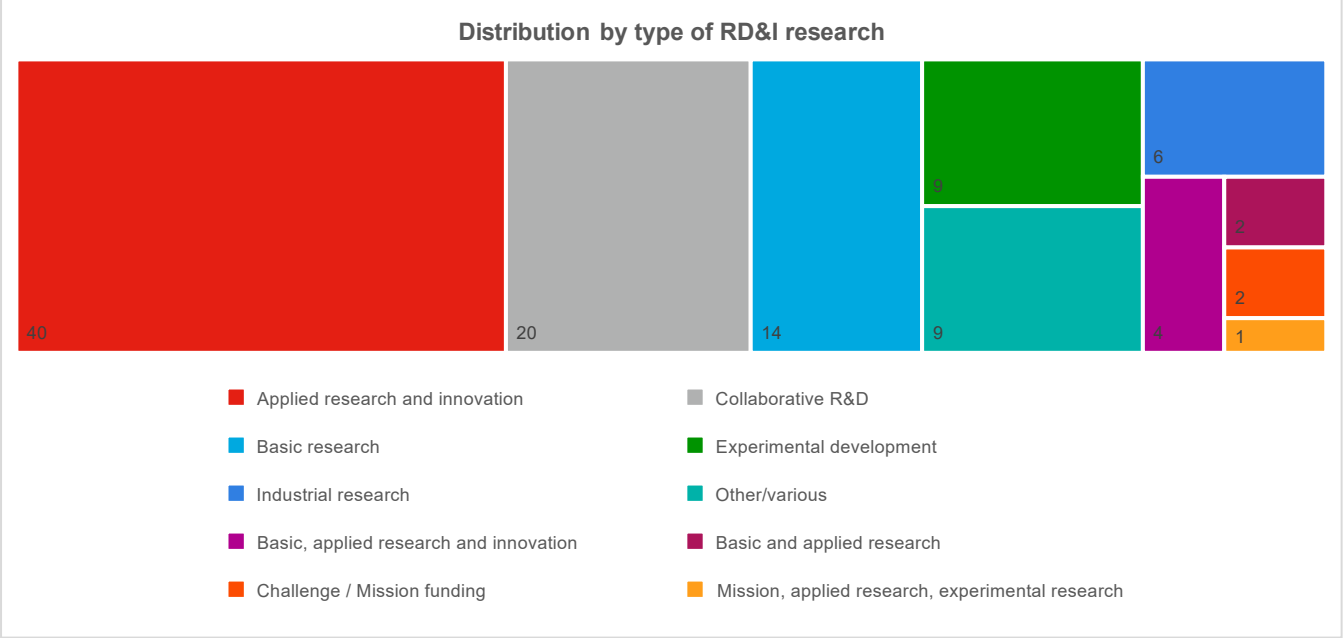
The spread of RD&I interventions is shown in Figure 5. Research and innovation grants represent 57% of the repository, followed by other interventions with 12% of the cases which combine several instruments, such as the 7th Framework Programme and the Research and Technological Development (RTD) infrastructures and activities supported by the European Regional Development Funds (ERDF). Direct assistance and Research Infrastructure account for 9% and 8% respectively. Fellowships, Knowledge Transfer Partnerships KTP, PhD grants and Networking activities are also included in the repository, although with fewer cases.

Figure 5: Distribution of RD&I programmes



The distribution of cases by type of RD&I research is shown in Figure 6. “Applied research” and “Innovation” ranks first with 37% of the evaluations identified, followed by collaborative RD&I with 19%, and 13% related to basic research. Other types of interventions (involving several types of research being supported by a single intervention) and experimental development represent 8% each, while mission and challenge-driven R&D have a lower share but are still represented in the repository, ensuring a balanced distribution of cases across different types of RD&I research from which to choose the representative case studies.

Figure 6: Distribution of cases by type of RD&I research



The spread of cases across countries is presented in Figure 7. Despite a skewed distribution of cases towards the UK, we have tried to ensure advanced economies selected are all well represented in the long list of evaluations.

Figure 7: Geographical distribution

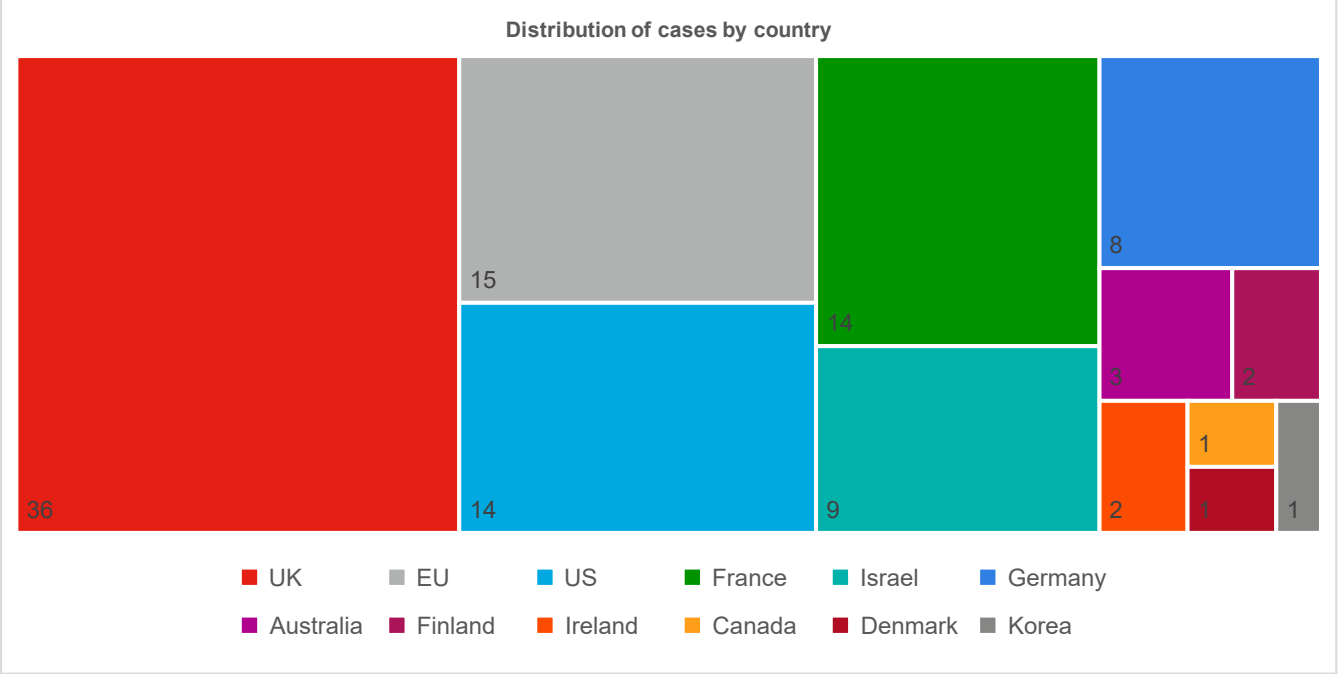


Table 15 summarises the distribution of evaluation methods used in the long list of reports. Generic methods (case studies, surveys, focus groups, output and performance monitoring) are the most used methods, alongside Value for Money evaluations. As mentioned earlier, experimental and quasi-experimental methods have become more used in recent years, and this is confirmed by the number of cases found and included in the repository (25).

Table 15 Distribution of methods

	Generic methods	Experimental and quasi-experimental methods	Value for money	Theory-based evaluation methods	Total
Generic methods	**52**		**1**		**53**
Experimental and quasi-experimental methods	**3**	**25**		**1**	**29**
Value for money	**6**		**8**		**14**
Theory-based evaluation methods	**1**		**1**	**8**	**10**

Step 2: Mapping against criteria. Once an evaluation was identified, it was profiled against the full set of essential criteria and additional criteria, including the objective of the evaluation,

data collection methods, outcomes and impacts evaluated, and challenges in measuring outcomes and impacts.

At this point, and to guide Step 3, we also assessed the extent to which each evaluation:

- Provided a detailed explanation of the methodological approach, and its relevance in terms of measuring the outcomes and impacts expected from the project.
- Used various data sources or methods to triangulate evidence.
- Offered a description of the limitations of the approach.

If an evaluation fulfilled all three aspects it was tagged as “High” (relevance for case study development), if it only fulfilled two it was tagged as “Medium” and “Low” if it fulfilled only one of these aspects.

Step 3: Initial selection and profiling. Based on the assessment done in Step 2, the core team selected 24 cases and developed individual fiches. They were put to the Advisory Panel for consideration and (alongside the full repository) for individual review and collective discussion.

Alongside the assessment of relevance, we also tried to strike a balance across the different types of RD&I interventions and evaluation methods (both quantitative and qualitative methods) based on the Magenta Book. Thus, we obtained a good distribution of evaluations across the criteria provided by DSIT.

Step 4: Shortlisting. As part of the shortlisting process, we formed an Advisory Panel[29] to bring expertise from various backgrounds in the STI field, to complement the experience of the core study team and support the selection of the 20 cases to be recommended to DSIT.

The Advisory Panel met on the 4th of July to analyse the composition and balance of the proposed cases, assess their quality in meeting the selection criteria, opportunities to draw lessons from them and their relevance for and transferability to the UK. A total of 24 cases were discussed.

From the discussion and systematic assessment, 15 cases were confirmed, panellists had divided opinions about 7, and 2 evaluations were excluded and replaced with alternative cases from the repository. The 20 cases shortlisted in the report incorporate the comments from the Panel, and the shortlist was adjusted accordingly. During the preparation of this report a further 3 examples were included from this list to address evidence gaps (i.e. examples on how to address key challenges).

[29] The Panel consists of experts with varied experiences and backgrounds in different areas of RD&I, among them, one of the main contributors to the Magenta Book, Professor Nigel Gilbert, Professor Stephen Roper, director of the Enterprise Research Centre (ERC) at the University of Warwick, Professor Erik Arnold, founder of Technopolis, and Fraser Macleod, Partner at Technopolis. The Panel is chaired by Cristina Rosemberg, Managing Partner at Technopolis and member of the UK Cabinet’s Evaluation and Trial Advice Panel (ETAP).

technopolis group

www.technopolis-group.com