Grantham Research Institute on Climate Change and the Environment

Centre for Climate Change Economics and Policy

Randomised controlled trials: Can they inform the development of green innovation policies in the UK?

Policy insight

Esin Serin, Nils Handler, Lewis Morey and Apurva Munjal

October 2022

Economic and Social Research Council

The Grantham Foundation for the Protection of the Environment

University of Leeds

The London School of Economics and Political Science

The Centre for Climate Change Economics and Policy (CCCEP) was established in 2008 to advance public and private action on climate change through rigorous, innovative research. The Centre is hosted jointly by the University of Leeds and the London School of Economics and Political Science. It is funded by the UK Economic and Social Research Council. **www.cccep.ac.uk**

The Grantham Research Institute on Climate Change and the Environment was established in 2008 at the London School of Economics and Political Science. The Institute brings together international expertise on economics, as well as finance, geography, the environment, international development and political economy to establish a world-leading centre for policy-relevant research, teaching and training in climate change and the environment. It is funded by the Grantham Foundation for the Protection of the Environment, which also funds the Grantham Institute – Climate Change and the Environment at Imperial College London. **www.lse.ac.uk/granthaminstitute**

About the authors

Esin Serin is a Policy Analyst at the Grantham Research Institute on Climate Change and the Environment.

Nils Handler is a visiting PhD research fellow at the Grantham Research Institute on Climate Change and the Environment and founder of the d\carb Future Economy Forum.

Lewis Morey is an MSc student in the Department of Geography and Environment at LSE.

Apurva Munjal is an MSc student in the Department of Geography and Environment at LSE.

Acknowledgements

The authors are grateful for review comments from Anomitro Chatterjee, Daire McCoy, Jacquelyn Pless, Anna Valero and Nigel Yau. Esin Serin acknowledges funding from the Grantham Foundation for the Protection of the Environment, and the UK Economic and Social Research Council (ESRC) through the Centre for Climate Change Economics and Policy (CCCEP). Nils Handler acknowledges doctoral funding from the Foundation of German Business. The views expressed in this report represent those of the authors and do not necessarily represent those of the reviewers, host institutions or funders. The authors have no relevant financial or non-financial interests to disclose.

Georgina Kyriacou copy-edited the report with production support from Natalie Pearson.

This report was first published in October 2022 by the Grantham Research Institute on Climate Change and the Environment and the Centre for Climate Change Economics and Policy.

Suggested citation: Serin E, Handler N, Morey L and Munjal A (2022) *Randomised controlled trials: can they inform the development of green innovation policies in the UK?* London: Grantham Research Institute on Climate Change and the Environment and Centre for Climate Change Economics and Policy, London School of Economics and Political Science.

Contents

Summary

This report explores if there is a case for the wider use of experimental methods, and randomised controlled trials (RCTs) in particular, as a source of evidence to inform green innovation policy development in the UK, with lessons applicable in other countries.

Why do we need green innovation policy?

Technology, alongside major behavioural change, is central to the global mission to move towards net zero emissions. Many low-carbon technologies required to deliver these ambitious targets already exist and they now need to be rolled out rapidly. But much progress is also needed to develop new technologies for reducing emissions where current technologies fall short. Innovations in technologies required to reach net zero also represent a significant growth opportunity. Direct policy support is needed for green innovation to happen at the pace and on the scale required to meet the world's climate targets. From targeted financial incentives to firms or individuals (e.g. R&D tax breaks and grants) to effective regulation (e.g. building codes), this is most effective when designed with the full path to technological commercialisation in mind. There must be sufficient 'push' on the supply side of innovation (research, development and demonstration [RD&D] of technologies) at the same time as creating 'pull' on the demand side of innovation (scaling up a market for these technologies).

Where might randomised controlled trials (RCTs) come in?

Rigorous policy evaluation can help understand which policies are most effective at driving green innovation and in turn ensure that public resources are spent most efficiently. A randomised controlled trial (RCT) is an experimental method of policy evaluation that enables understanding of the causal effects of policies with a high level of confidence. RCTs involve assigning subjects (individuals, households or firms) in a fully randomised manner to either a treatment group that receives the policy being studied or a control group that does not (thereby representing the 'counterfactual', i.e. what would have happened in the absence of the policy). Since the application of the policy is fully randomised, subjects in the control group should, in theory, be similar to those in the treatment group, so changes in their behaviour and outcomes when the policy is implemented can be associated with the policy itself rather than individual characteristics or wider trends or influences.

Another counterfactual-based method of policy evaluation is quasi-experiments. In quasi-experimental methods, the counterfactual is represented not by a control group that results from the random allocation of the policy but by a comparison group that can be identified, which is as similar as possible to the treatment group in terms of baseline (pre-policy) characteristics. These studies typically involve the use of control variables to ensure that the treatment and comparison groups are as similar as possible, which means the messages they produce may be less transparent and convincing for policymakers than those from an RCT, which allows a visible comparison of a treatment group with a control group.

Varying views on RCT-based evaluations

Some argue that RCTs represent a methodological 'gold standard', while others view them as normative, reductionist and insensitive to the importance of context. On the question of context, both RCTs and quasi-experimental methods often need to be complemented with other investigation methods to understand how any measured change comes about, or whether the same outcome would occur if the intervention were tried in another context or at a different scale. However, because RCTs tend to be more expensive and time-consuming and less politically feasible than most other methods of policy evaluation, their reliance on complementary investigations could lead to RCTs in particular being perceived as a 'high effort, low reward' method of evaluation by practitioners. One important way of

thinking about RCTs is to shift the focus away from 'evaluation' *per se* of existing policies to using RCTs to make innovation policy itself more innovative, exploring a wide range of policy ideas, testing them out on a small scale and learning which are likely to work better before scaling them up.

Using RCTs to close the evidence gap on what works

For many innovation policies there is limited evidence to indicate that they achieve their aims, let alone that they do so efficiently or more effectively than alternatives. RCTs can be an important tool for closing evidence gaps on what works, both for innovation generally and for green innovation specifically. Despite overall justifications for technology-specific policy support to drive green innovation, the shape that such support should take is not clear in many cases, and further research is needed on the role of different policy instruments when it comes to achieving specific technological objectives.

RCTs can help to investigate whether policies are achieving their intended objectives, in turn helping to ensure that government resources are spent most efficiently and effectively on policy designs that have proven potential to accelerate the transition to low-carbon technology. Because random allocation is not a feature that is possible to retrofit into policy design and implementation, only a proactive approach can unlock the full potential of RCTs as part of the policymakers' toolkit for evidence-building on green innovation.

Why are RCTs relevant for the UK's green innovation policy landscape?

The UK government has adopted innovation-led growth as a long-term ambition, having set out its largest ever R&D budget in 2022. The high-level commitments to green innovation in the *Net Zero Strategy* have been followed by concrete steps, including a £1 billion Net Zero Innovation Portfolio (NZIP) (later complemented with additional funding) and the Net Zero Research & Innovation Framework.

The UK has a long history of evidence-based policymaking, with strong roots in development aid in particular. The introduction of the Behavioural Insights Team (BIT) within the Cabinet Office in 2010 was a product of this foundation. More recently, the Department for Business, Energy and Industrial Strategy (BEIS) launched a large experimentation fund to test innovative ways to encourage SMEs to improve their productivity through new technology and practices. A range of organisations (such as the 'What Works' network, Nesta and J-PAL) act in the UK to promote the use of RCTs and other experimental methods in policy analysis. The Magenta Book, which provides formal guidance on policy evaluation, and the introduction of a Cabinet Office Evaluation Task Force in 2020, reflect the continued commitment from central government to evidence-based policymaking, including through the use of RCTs.

The many policy frameworks that have already been or are yet to be introduced to drive green innovation in the UK may create new opportunities for experimentation and the use of RCTs for building evidence on what actually works in this area.

Use of RCTs to date to inform green innovation policies: results from our review

From a comprehensive review of the global evidence base on RCTs we have identified just 29 trials that have already been conducted, are currently underway or are being planned which can inform the development of policies aimed to support *green* innovation. This limited number might imply a limitation of the evidence base generally, but there is also reason to believe that there are many additional RCTs that could be relevant for green innovation policy development but are not publicly registered or do not have their results published.

By assessing the attributes of the studies identified, we try to understand the circumstances that made an RCT the chosen method for building evidence in a green innovation policy area. Our review draws insights from the attributes of the RCTs identified, rather than their quality or policy impact. In other words, we only discuss 'if' RCTs have been used to build evidence on interventions relating to green innovation, rather than 'how' these studies have been used to inform policy, if at all. This is due to the scarcity of publicly available information on the latter aspect. One important exception is the energy labelling trial led by the erstwhile Department for Energy and Climate Change (DECC) in the UK, for which policy impact was reported directly. Findings from this RCT were fed directly into the European Commission's evaluation of the Energy Labelling Directive, and led to a call on UK retailers and consumer groups to consider changes in line with the lessons learned from the trial. Furthermore, while we focus specifically on RCTs relating to green innovation policies, there are other RCT studies that are about innovation generally which may generate important lessons for green innovation too.

Characteristics of the identified RCTs:

- Almost all have been or are being conducted for research purposes, aiming to inform policy development from the outside, by the academic community or members of non-profit, social-purpose organisations, rather than by the public sector itself.
- Study locations have a wide geographical range, although the US dominates, followed by the UK.
- Outcomes of interest for research differ somewhat by location – e.g. the primary focus being on climate benefits versus energy access.
- The number of RCTs that relate to interventions around the RD&D stage of green technologies (i.e. the 'supply side' of innovation) is especially limited (two in total).
- There are more examples of RCTs relating to interventions around the scale-up of green technologies (i.e. the 'demand side' of innovation).
- The untapped potential of RCTs for identifying what works might be especially large where it is most needed, given wide recognition that the need for government support is highest in the RD&D stage of the commercialisation pathway, which decreases as technologies progress through the pathway to reach scale in the market.
- Fewer than one-third of the RCTs that relate to the scale-up of green technologies are testing the effects of some kind of financial incentive. This might suggest that the benefits RCTs offer in terms of establishing accountability are even less exploited with regard to policy designs that involve immediately visible decisions about distributing government funding.
- A good number of studies appear to be following best practice, ensuring research credibility; another big group of studies have start dates before best practice was structured and formalised to this extent.
- Four studies contain signs of a potential publication bias, whereby RCTs that do not produce 'significant enough' results are not put forward for publication, and another six might be suffering from implementation issues. All insights should be shared in order to build a shared evidence base and inform future trials.

Why have RCTs not been used more widely to date for the development of green innovation policies in the UK?

We have identified eight broad areas and issues that may contain answers to this question:

1. Awareness, skills and institutional culture
2. Generalisability and transferability of lessons learnt
3. Cost

4. Time and timing
5. Experimental design constraints and risks
6. Complexity of the green innovation system
7. Fairness, equity and ethical considerations of random allocation
8. Political feasibility.

These are based on a review of the existing literature complemented with results from semi-structured interviews we held in summer 2022 with 10 policy professionals working specifically in green innovation policy or RCT design and implementation in the UK.

While our insights are indicative rather than conclusive, it is clear that all the possible answers across the eight areas named above are highly context-specific. This reiterates that a method of evaluation needs to be chosen that is appropriate to the design of the specific policies being evaluated, the evaluation questions being asked, and the resources available. Quasi-experimental methods can be used where it simply is not possible or desirable to use an RCT-based approach. Ultimately, **RCTs are one of the most rigorous but *not the only* rigorous method for policy evaluation**.

Barriers to conducting RCTs that relate to generalisability and transferability of lessons learnt, time and timing are not unique to RCTs but can affect any evaluation method for a green innovation policy. There are also concrete ways in which RCTs can be designed to deal with these issues, for instance using complementary investigations and embedding dynamic feedback loops aligned with key decision points, respectively. On the other hand, barriers to RCTs identified under the themes of awareness, skills and institutional culture; cost; fairness, equity and ethical considerations of random allocation; and political feasibility appear not to be unique to the green innovation policy landscape. These views appear to be rooted in the wider context around the use of RCTs in policy evaluation and evidence-based policymaking in the UK. These issues require further investigation involving the policy profession as a whole, which can build on the discussion we present here.

Experimental design constraints and risks, and the complexity of the green innovation system, appear to present some feasibility challenges unique to the use of RCTs to inform the development of green innovation policies. These are very likely to have implications for the applicability of RCTs as a method of building evidence on the efficacy of the UK's current and future portfolio of green innovation policies. Unlocking the full learning potential from RCTs to inform the development of green innovation policies would therefore require a concerted effort from policy professionals working specifically in the green innovation landscape.

Can RCTs have a greater role?

We have evaluated which of the UK's priority areas for innovation, as set out by the Net Zero Innovation Portfolio, may represent the greatest opportunities for RCT-based evidence-building. We focused purely on the two themes that contain challenges unique to the use of RCTs in the green innovation policy landscape: 1) experimental design constraints and risks, and 2) complexity of the green innovation system. We found that **energy storage and flexibility, homes and buildings, and hydrogen may present fertile grounds for RCT-based evidence building on innovation policies.** This is in line with two high-level observations regarding the feasibility of RCTs:

- Firstly, a large and homogeneous sample may be easy to find for policies operating at individual or household level, but not as easily achieved when it comes to innovation policies targeting what usually is a highly heterogenous pool of innovator firms. The pool of firms can be particularly small if the innovation area relates to large-scale technologies that also require the development of large-scale infrastructure such as carbon capture, usage and storage (CCUS) or nuclear energy.

- On the other hand, many green technologies are integrated to the energy system 'upstream' (e.g. offshore wind, CCUS) to decarbonise energy production or industrial processes, rather than being used directly by consumers. Policy interventions designed to create a 'pull' on the demand side for such technologies would likely need to be placed on other products and services enabled by the technology (e.g. green tariffs, low-carbon manufactured products), rather than on the technology itself. Policy interventions with such long, indirect causal chains of intended impact (in this case, technology uptake) are less feasible to test under an RCT design. It would be suitable to use theory-based evaluation methods in situations where the intervention is designed to make a change in a complex system consisting of diverse, interacting components leading to long, indirect causal chains between the policy intervention and intended impacts.

Conclusions

When complemented with other methods to reveal the contextual factors underpinning results, RCTs can contribute to an evidence base that can be used as a public good to accelerate net zero-aligned innovation and growth. However, other sources of rigorous evidence are available where RCTs are not appropriate or practicable. Ultimately, RCTs are one important enabler of evidence-based policymaking rather than an end in themselves; this is the lens through which the case for using RCTs to inform the development of green innovation policies in the UK (or elsewhere) should be evaluated.

Policy implications

- More evidence on the causal effects of policies designed to achieve specific technological objectives is needed to ensure public resources are used most efficiently in the drive to meet climate targets. RCTs can provide this.
- While RCT-based evidence on what works for driving green innovation is as yet limited, there is a lot to be gained from an open-minded approach to recognising and learning from transferable knowledge.
- Following best practice in research can ensure that knowledge from future green innovation RCTs can feed into a shared, credible evidence base that can be used as a public good to accelerate net zero-aligned innovation.
- An RCT-enabled experimental approach to policy development – identifying different policy alternatives with an open mind, testing promising ones at small scale and identifying those that work based on well-defined intermediate outcomes before scaling up – might be especially relevant for policies aiming to support the RD&D of green technologies as the potential of RCTs to generate evidence in this area appears largely unexploited.
- The heterogeneity and relatively small size of the pool of innovators targeted by many of the UK's innovation policies, as well as the long, indirect and complex relationships between many green technologies and their ultimate users, present challenges to implementing RCT-based evaluations for green innovation in the UK.
- Energy storage and flexibility, homes and buildings, and hydrogen may present fertile grounds for RCT-based evidence-building for the UK's innovation policies.
- RCTs are one of the most rigorous but not the *only* rigorous method for policy evaluation. Whichever method is chosen, it would be most useful if complemented with other investigations to extract information on 'why' and 'how' a policy works, to generate knowledge that can be used in a wider range of contexts.

1. Introduction

Purpose of this report

This policy insight explores whether there is a case for the wider use of experimental methods, and randomised controlled trials (RCTs) in particular, as a source of evidence to inform green innovation policy development in the UK. By green innovation policies we mean government-led initiatives that aim to foster the research, development, demonstration (RD&D) and scale-up of new technologies, business models and processes that can contribute to moving an economy towards net zero emissions.

While recognising that innovations in business models and processes are essential for the net zero transition, our focus in this report is limited to policies designed to spur green *technological* innovation as a starting point. We aim to contribute to the debate among policymakers, researchers and other stakeholders in the UK on how to better build and use evidence to inform spending decisions underpinning these policies while generating potentially valuable insights for similar debates around the world.

Outline

In Section 2 we set the scene on how RCTs can inform the development of green innovation policies. This includes a discussion of why green innovation policies are needed in the first place, an introduction to the different types of green innovation policies within our scope, and a description of where RCTs fit within the wider context of policy evaluation.

In Section 3 we compile examples from around the world of RCTs conducted to date, currently underway or being planned, which can inform the development of policies aimed to support green innovation. We review the attributes of these studies to understand their similarities and differences with an aim to reveal contexts in which RCTs have been more applicable than other kinds of study. The appendix contains further details.

This stock-taking exercise sets the stage for Section 4, where we draw on a review of the relevant literature and a set of interviews with UK policy practitioners to understand why RCTs have not to date been used more widely to inform the development of green innovation policies in the UK, whether they can play a greater role in this space going forward and, if so, where the greatest opportunities might lie.

Section 5 concludes and provides a set of high-level policy implications.

2. How can RCTs inform green innovation policymaking?

Green innovation as an environmental and economic priority

Technology, alongside major behavioural change, is central to the global mission to move towards net zero emissions. As of September 2022, around 140 countries had announced or were considering net zero targets, covering close to 90% of global emissions (Climate Action Tracker, 2022). Many low-carbon technologies needed to deliver these ambitious targets already exist and need to be rolled out rapidly. But much progress is also needed to develop new technologies for addressing emissions where current technologies fall short. In fact, almost half the emissions reductions required to reach global net zero by 2050 may need to come from technologies that are currently at the demonstration or prototype stage (IEA, 2021).

Increasing innovation in and diffusion of technologies required to reach net zero also represents a significant opportunity for growth and improved living standards (Curran et al., 2022; Stern et al., 2020; Rydge et al., 2018). These technologies have the potential to increase efficiency and save labour, thereby raising productivity both directly and through spillovers to the rest of an economy (Green Alliance and Nesta, 2022). We are already seeing evidence of increasing returns to scale in the discovery and production of green technologies (Ekins and Zenghelis, 2021), for example via the dramatic declines in the 2010–2019 period in the unit costs of solar energy (declining by 85%), wind energy (55%), and lithium-ion batteries (85%) (IPCC, 2022).

The need for green innovation policies around the world

Market failures

Despite the strong environmental and economic case for it, the incentive for firms and individuals to pursue green innovation is undermined by a number of market failures. These market failures affect innovation generally (Bloom et al., 2019), but two in particular are exacerbated in the specific context of green innovation: 1) knowledge spillovers and 2) environmental externalities.

Firstly, knowledge spillovers mean that an innovator is unlikely to be able to capture all the financial returns from investments in research and development (R&D), with the implication that these investments tend to be lower than the level that would be "socially optimal" (Stern and Valero, 2021). This is particularly relevant for green innovation as knowledge spillovers from green technologies are observed to be greater than those generated by their dirty counterparts (Dechezleprêtre et al., 2014). Secondly, in the absence of a robust carbon price, markets do not internalise the cost of the damage that greenhouse gas emissions inflict on others, creating negative environmental externalities. This reduces the incentive to invest in green alternatives.

Path dependence

Innovation systems are also subject to path dependence, whereby initial conditions and history matter for eventual outcomes. When a technology or fuel is prevalent in society, not only are significant amounts of money invested in its production but dependent technologies – for example, natural gas and gas cookers – also become widely adopted. Such technological 'lock-in' means that changing to a different technology or fuel (such as electricity) can be difficult, even if it is far cheaper and more effective. And in reality, many green technologies that will clearly be superior in the future are not cost-effective to begin with, giving rise to dynamic market failures (Martin et al., 2020). Path dependence may undermine not just the diffusion but also the innovation of green technologies in the first place. Acemoglu et al. (2012) demonstrated that when the network infrastructure and skills base are initially located in the dirty sector, the immediate returns to innovating there are higher, implying that green technologies may never overtake dirty technologies without government intervention.

Behavioural factors

There may also be behavioural anomalies and failures, ranging from biased beliefs to inattention, which impact the uptake of green technologies in the market. Such behavioural factors are becoming a commonly cited explanation for the 'energy efficiency gap', whereby consumers and firms fail to make seemingly positive 'net present value' energy-saving investments (Gillingham and Palmer, 2014). Therefore, technology adoption is often not simple or linear, but must change the wider physical and social systems that develop around certain key technologies (Seto et al., 2016).

These market failures, path dependencies and behavioural factors justify the need for coordinated policy action to switch innovation systems from a dirty to clean pathway to deliver emissions reductions at the pace required for global net zero (Aghion et al., 2014; Aghion et al., 2019; Stern and Valero, 2021; Ekins and Zenghelis, 2021). Many governments are already implementing such policies, with much capital being deployed in the process. As of October 2022, the policies database of the International Energy Agency's (IEA) identifies 525 policies in force, planned or announced around the world that support R&D and innovation for low-carbon technologies (IEA, 2022).

Types of green innovation policy

Carbon pricing can help re-orient R&D investments towards green technologies and away from dirty ones, alongside its role in directly reducing greenhouse gas emissions (Stern and Valero, 2021). However, given the range of complex factors affecting innovation systems, carbon pricing by itself may not be sufficient to drive the shift to a green innovation path at the pace and scale required to meet the world's climate targets (Carbon Pricing Leadership Coalition, 2019). If a zero-carbon transition is to be achieved, policy also needs to provide direct support for green innovation, for example in the form of R&D subsidies, or investments in appropriate infrastructure to accelerate the diffusion of green technologies, alongside effective regulation (Stern and Valero, 2021). **This report focuses on such direct forms of intervention to support green innovation under the label of 'green innovation policy'.**

Technologies face a long journey before they can reach widespread use and unlock investments in related infrastructure, skills and supply chains. The timeline between idea generation and commercialisation can be particularly long for net zero-enabling technologies in the energy sector, given many of these technologies are capital-intensive and the sector is heavily regulated. Indeed, technologies first must be researched, have prototypes developed, be demonstrated, be formulated into commercial products, be dispersed widely enough to replace enough of the old carbon-intensive technologies and, at each of these stages, secure sufficient funding from investors and public funds (Gallagher et al., 2012). Since unlocking such funding depends on the existence of an ultimate market for these technologies, green innovation policies need to be designed with the full path to commercialisation in mind. This means ensuring sufficient 'push' on the supply side of innovation (covering the RD&D of technologies) at the same time as creating 'pull' on the demand side of innovation (covering the scale-up and growth of technologies in the market).

Both firms and individuals can be the target of green innovation policies, operating on either the demand or the supply side of innovation. For example, on the supply side, R&D funding can be provided to firms conducting in-house RD&D as well as to individual researchers, commonly via universities. Similarly, on the demand side, feed-in tariffs incentivising the uptake of renewable energy generation often target both domestic and commercial users. While RD&D can also be carried out directly by governments, we are interested in policy interventions that can induce green innovation activities by private actors (that is, either firms or individuals). Indeed, the Climate Change Committee states that the increase in investment required to meet the UK's climate goals can, and should, be delivered largely by the private sector (CCC, 2020).

Table 2.1 summarises the types of green innovation policy interventions that are the subject of this report.

Table 2.1. Types of green innovation policy interventions

			Example policy interventions	
Type of policy intervention	**Definition**	**Target**	**Supply side of innovation (RD&D)**	**Demand side of innovation (scale-up and growth)**
Financial incentives	Financial instruments (non-debt or debt) that increase access to capital required for investment in RD&D or take-up of innovative technologies	Firms or individuals	R&D grants, R&D subsidies, R&D tax breaks, loans for R&D activities, prizes, competitions	Feed-in-tariffs, purchasing grants, VAT discounts, loans for take-up of low-carbon technology (e.g. UK Green Deal)
Regulation and obligations	Laws and regulations that affect innovation (positively or negatively)	Firms or individuals	Supplier obligations, sales bans, reductions in trade barriers, patent laws	Building codes, zero emission zones, environmental standards attached to government procurement
Information, networks and collaboration	Promoting information, networks, partnerships and relationships that can unlock innovation and/or create demand for innovative products	Firms or individuals	Managerial associations for production innovation, innovator workshops	Information campaigns, community workshops, product labels
Capacity-building	Interventions that improve the capacity of the environment surrounding the innovation system, including by enabling physical infrastructure and promoting skills, supply chain capabilities and institutional capacity	Firms or individuals	Programmes that provide technical assistance, training and mentorship to support RD&D activities, accelerators, incubator programmes, testbeds, investments in infrastructure	Training, reskilling and/or upskilling programmes that support the development of skills required for the production, installation and maintenance of new technology; investments in infrastructure

Source/note: Some of the intervention categories, definitions and examples used have been adapted from Sabet et al. (2017). Modifications to the original content have been made to make it relevant specifically for green innovation.

Policy evaluation and where RCTs fit

Policy evaluation is "the systematic assessment of a government policy's design, implementation and outcomes" (HM Treasury, 2020a). Well designed and executed evaluation provides an understanding of the actual economic, financial, social and environmental impacts of a policy, and/or provides an assessment of how it is/was implemented, why it did or did not deliver as expected, and whether it represents value for money (BEIS, 2020). This information can inform design choices to maximise the impact of an ongoing policy intervention, enable decisions about whether to continue, discontinue, replicate or scale up the policy intervention, and feed into future policy development decisions (ibid.).

One group of policy evaluations investigates questions of causality between a policy and its intended impacts by attempting to (re)construct a counterfactual. The counterfactual – i.e. what would have happened in the absence of a policy – is fundamentally unobservable (What Works Centre, 2014). This makes it highly challenging to confidently establish that a realised outcome is the result of a policy in question, rather than of other potential factors simultaneously in effect. Counterfactual-based evaluation methods, which include RCTs and quasi-experimental methods, address this issue by enabling estimation of the difference between the actual realised outcomes and the potential outcomes had the given policy not been implemented.

Table 2.2. Summary of the types of counterfactual-based policy evaluation

Type of policy evaluation	Examples	Description
Experimental	Randomised Controlled Trials (RCTs)	A study in which the receipt of policy intervention that is being tested is by explicit randomisation. This ensures that there are no significant differences between the 'treatment' group(s) (receiving the intervention) and 'control' group(s) (not receiving the intervention, thereby representing the counterfactual) on either observable (e.g. age) or unobservable (e.g. ability) characteristics.
Quasi-experimental	Difference-in-difference studies, regression discontinuity designs, matching methods, synthetic control methods	A study in which the counterfactual is represented not by a control group that results from the random allocation of the policy intervention but by a comparison group that can be identified, which is as similar as possible to the treatment group in terms of baseline (pre-intervention) characteristics. These studies typically involve the use of control variables to ensure that the treatment and comparison groups are as similar as possible on observable and unobservable characteristics.

Source/note: Magenta Book (HM Treasury, 2020a) and BEIS Monitoring and Evaluation Framework (BEIS, 2020). Relevant information and examples selected, summarised and complemented with further detail by the authors.

We can illustrate what an RCT versus a quasi-experimental evaluation in the green innovation policy landscape might look like with a few simple examples. An RCT investigating the efficacy of an R&D grant, for example, could be designed to randomly allocate these grants among the applicants of the scheme and statistically compare the innovative performance of the group of firms to which the grant is allocated with the group of firms to which the grant is not allocated. Outcomes of interest can include the green innovation output as well as wider trends such as changes in organisational culture and creation of networks with other innovating firms. Outcomes can be measured using self-reported information (e.g. using surveys) or observed data (e.g. patent applications) at one or more point(s) in time identified by the evaluators as appropriate to measure potential impacts.

A quasi-experimental evaluation, on the other hand, would exclude the random allocation element and instead be designed around how such an R&D grant scheme would normally operate. These schemes commonly use cut-offs on the grades or rankings produced by expert review panels assessing project proposals to determine the projects that ultimately receive a grant (Pless et al., 2020). These cut-offs are typically implicitly based on the total amount of money available (ibid.). This enables the use of a regression discontinuity design to statistically compare firms that fall just below the cut-off and do not receive the grant (i.e. the comparison group) with firms just above the cut-off who receive the grant (i.e. the treatment group) in order to infer the impact of the funding on innovative performance (ibid.). Furthermore, if cut-offs in firm size determine funding rates (which is the approach used by Innovate UK), comparing firms receiving grants that fall just below the firm size cut-off with those just above the cut-off could enable effects of the more generous funding rate to be interpreted as causal (ibid.).

Experimental and quasi-experimental methods (i.e. counterfactual-based approaches) have the primary aim of assessing the net impact of an intervention (HM Treasury, 2020a). But these methods may not in themselves produce insights about how any measured change comes about, or whether the same outcome would occur if the intervention were tried in another context or at a different scale (ibid.). In other words, they have limited external validity. External validity can be developed by complementing an experimental or quasi-experimental method with other investigation methods. These complementary methods exploring questions of 'how' and 'why' are typically qualitative and can include in-depth interviews, focus groups, case studies and observation (HM Treasury, 2020a).

Theory-based approaches provide an alternative to counterfactual-based approaches for policy evaluation, although the two approaches can be used in combination, too. Theory-based impact evaluations draw conclusions about an intervention's impact through rigorous testing of whether the causal chains thought to bring about change are supported by sufficiently strong evidence (usually gathered from multiple quantitative and/or qualitative sources) and that alternative explanations can be ruled out (HM Treasury, 2020a). This requires an explicit Theory of Change, which includes theories about alternative explanations for the outcomes. In other words, theory-based evaluations try to get inside the black box of what happens between inputs and outcomes, and how that is affected by wider contexts (ibid.). While these evaluations do not typically produce a single numeric estimate of the impact of an intervention, they can shed light on the extent of the impact and confirm whether it is in the desired direction (ibid.). Evaluators may need to rely fully on theory-based methods in cases where a counterfactual-based method may not be suitable, for example for the evaluation of complex interventions or simple interventions in complex environments (BEIS, 2020).

Randomised controlled trials as a highly rigorous method of policy evaluation

RCTs stand out among the different methods of policy evaluation by virtue of the unique strengths they offer. RCTs have the potential to provide strong internal validity, given allocation into the treatment or control group in an RCT design is by explicit randomisation.

This provides RCTs with a particularly credible counterfactual represented by the control group, whose only distinguishable difference from the treatment group should, in theory, be its receipt of the intervention being tested. As such, RCTs are particularly good at dealing with known and unknown biases, including a potential 'selection bias'. Selection bias occurs when recipients of a policy systematically differ from those that do not receive the policy. For example, firms that are already more innovative may be more likely to apply for and win grants. In such cases, estimates of the policy impact may be biased upwards if innovation outcomes are incorrectly attributed to the grant, rather than to the fact that these firms were more innovative to begin with.

While randomisation – where well-executed – inherently deals with a potential selection bias, quasi-experimental methods typically require additional measures. One common measure is modelling the probability (implicitly or explicitly) that a subject will be in the treatment rather than the comparison group, and then estimating the impact of the treatment, 'controlling' for any selection biases (Bakhshi et al., 2015). However, the messages coming out of an evaluation that relies on the use of such econometric approaches may be less transparent and convincing for policymakers than those from an RCT that allows visible comparison of a treatment group with a control group (ibid.).

Despite their strengths, RCTs have crucial limitations that mean they are not applicable in every context:

- *Cost and time:* Because of operational requirements inherent to their design, notably the random allocation of the policy intervention under investigation, RCTs tend to be more expensive and time consuming to implement than quasi-experimental methods (although not in all cases; for example, see Haynes et al., 2013).
- *Suitabality:* Not every policy intervention lends itself to random allocation. Even if it would be technically possible to randomise the intervention, RCTs might create risk for the subjects (particularly those in the control group withheld from receiving a potentially beneficial intervention) and consequently the policymaker responsible for the use of the evaluation method. (This is discussed in detail in following sections.)
- *External validitiy:* RCTs may have limited external validity, especially when conducted at small scale. For RCTs in the context of innovation systems, external validity problems may be exacerbated by the heterogeneity of small firms, the social and interactive nature of business and the importance (and diversity) of the contexts in which firms operate (Bakhshi et al., 2015).

  Limited external validity is actually not a problem unique to RCTs as it affects quasi-experimental methods too, as previously discussed. However, given RCTs generally require greater resources to implement than quasi-experimental evaluation methods, their limited external validity and the resulting need to use complementary investigations could lead to RCTs in particular being perceived as a 'high effort, low reward' method of evaluation by practitioners. Ideally, strong external validity can also be achieved by running multiple RCTs in different contexts of interest and then pooling the results using meta-analysis (for an example of this approach from the development literature, see Meager [2019]). This would likely increase resource requirements substantially, though.

A historical background on public policy RCTs

One particular RCT, conducted in the medical field in 2005, is famous for revealing that a steroid treatment commonly used at the time for head injury victims was in fact increasing the mortality rate of those receiving it (Edwards et al., 2015). Indeed, RCTs have long been applied in clinical research; only more recently have they become widely used in social sciences and public policy[1] as well, especially in the field of development economics (Cameron et al., 2016). This reflects a cultural change within the academic and professional community towards belief in such methods, commonly referred to as the 'credibility revolution' (Duflo et al., 2007; Athey and Imbens, 2017).

Varying views have emerged over time around the use of RCTs in public policy, with certain schools of thought putting more weight on their strengths over their limitations and vice versa. Vine et al. (2014) are among the proponents, arguing that RCTs represent a methodological 'gold standard' and should be used whenever possible. Contrasting views regard the push for RCT-based evaluations as being normative and reductionist, and insensitive to the importance of context and diversity in methodologies for social science research (Dalziel, 2018). Taking a more philosophical stance, Cooper (2018) argues that the focus should be on the specific knowledge demand, and that promoting RCTs for energy policy evaluation is "to force a square research design peg into a round 'policy epistemic' hole".

Several organisations and initiatives currently promote, fund and disseminate the results of RCTs as a tool to enable evidence-based policymaking. These include the Innovation Growth Lab and the What Works Network in the UK and the Abdul Latif Jameel Poverty Action Lab (J-PAL), the International Initiative for Impact Evaluation and the American Economic Association's RCT Registry in the US (Dalziel, 2018).

RCTs for informing the development of green innovation policies

Evidence to indicate whether many innovation policies achieve their aims, let alone that they do so efficiently or more effectively than the alternatives, is limited (Pless et al., 2020). The effectiveness of certain policies designed to drive innovation appears especially understudied – policies such as direct R&D grants, university incentives and mission-oriented policies – while there is relatively high quality and conclusive evidence on the effectiveness of tools including R&D tax credits and encouraging skilled immigration (Bloom et al., 2019). RCTs can be an important tool for closing evidence gaps on what works when it comes to fostering innovation (Pless et al., 2020; Bravo-Biosca, 2020; Carattini et al., 2020). Having such evidence is crucial for ensuring public resources are not wasted but are spent on policies that deliver intended outcomes. RCTs can also help governments establish and maintain accountability as they can produce transparent and communicable policy evidence that can be used to underpin spending decisions. The number of RCTs in the fields of innovation, entrepreneurship and business growth (in a wide sense, rather than specifically *green*) remains limited but has grown recently (Firpo and Phipps, 2019). The Innovation Growth Lab Database, which attempts to collect all RCTs in this field, currently counts 226 such experiments (IGL, 2022).

As with innovation generally, we are learning what works when it comes to fostering *green* innovation, but significant knowledge gaps remain. Evidence to date suggests that progress made in reducing the costs and increasing the deployment of key energy technologies – including solar PV, onshore and offshore wind, and lithium-ion batteries – was driven not by technology-neutral, generic R&D investments or carbon pricing but instead by innovation policy packages involving deliberate technology choices made by governments in many countries over the course of decades (Economics of Energy Innovation and System Transition [EEIST] Consortium, 2022). Despite overall justifications for targeted policy support for technologies, the shape that such support should take is not clear in many cases, and evidence from empirical studies can say more on the role of different policy instruments when it comes to achieving specific technological objectives (Popp, 2019).

[1] Among other fields, including the technology industry. For example, companies like Google and Amazon are known to be running tens of thousands of experiments a year (Breckon and Sutherland, 2020).

RCTs can help to investigate whether policies are achieving their intended objectives, in turn helping to ensure that government resources are spent as efficiently and effectively as possible, i.e. on policy designs that have proven potential to accelerate the transition to low-carbon technology.

Parallel to the urgent need to decarbonise their economies, many countries including the UK are currently experiencing the lasting effects of the COVID-19 pandemic on public budgets, which are tightening further with measures introduced to mitigate the ongoing global cost-of-living crisis. Squeezed public sector budgets can reduce the appetite for experimentation and give rise to the temptation to stick with conventional options as well as to focus on short-term rather than long-term goals (Rydge et al., 2018). However, it is important to consider the other side of this argument. The pressure on public sector budgets could also strengthen the case for building evidence for government spending decisions, including via experimental designs such as RCTs. Poorly designed policies that do not draw on robust evidence are likely to result in a waste of public money and may come with an opportunity cost of potentially more effective policies that could have been implemented instead.

One important way of thinking about RCTs shifts the focus away from 'evaluation' *per se* of existing policies to using RCTs to make innovation policy more experimental in itself (Bravo-Biosca, 2020). This refers to exploring a wide range of policy ideas, testing out the most promising at a small scale, learning which are likely to work better, and only then scaling them up (ibid.). Such an approach to testing new ideas is already used at scale within the technology industry. Changing the objective from answering 'is my policy working as intended?' to searching for 'what works' with an open mind might be especially relevant for maximising what RCTs have to offer to green innovation policymaking.

As previously discussed, RCTs offer unique strengths but they are not applicable in every context or to test every policy idea. Ultimately, there is no one-size-fits-all method of policy evaluation; the focus should be on choosing the most robust method of building evidence for the specific policy question, using available resources. Having an evaluation strategy upfront is best practice for any policy but most policy evaluation methods can be designed and finessed in parallel to policy design and implementation. What distinguishes RCTs from other methods of policy evaluation is that they need to be built into the implementation of the policy from the beginning, as random allocation is not a feature that it is possible to apply retrospectively. It is therefore of special importance to consider, sooner rather than later, whether RCTs can play a greater role for informing the development of green innovation policies in the UK over this critical decade for climate action and beyond. Only such a proactive approach can unlock the full potential of RCTs as part of the policymakers' toolkit to build evidence on what works for driving green innovation.

3. Taking stock: RCTs that can inform the development of green innovation policies

An assessment of the extent to which RCTs have been used to inform the development of green innovation policies to date can shed light on the role they can play in this area into the future.

This section provides a review of the attributes of RCT studies from around the world that have been conducted to date, that are currently underway or that are being planned, which can inform the development of policies aimed at supporting green innovation (with a particular focus on technological innovation). These studies have been identified by undertaking comprehensive searches within relevant websites, registries and databases[2] using keywords related to green innovation, and through conversations with experts in the field. By assessing the attributes of these studies, we have tried to understand the circumstances that have made an RCT the go-to method for building evidence in a green innovation policy area. As such, we are interested in the factors that have brought about the RCTs we have identified in the first place, rather than what these studies have found.

Our review is not limited to RCTs whose primary purpose is to evaluate already existing or proposed policies: it also includes RCTs implemented (or planned and designed but yet to be implemented) by researchers to test potential interventions and new ideas that could be adopted by the public or the private sector to spur green innovation. For this reason, our discussion in this section is framed under 'interventions' that can spur green innovation more widely, rather than '*policy* interventions' specifically. Furthermore, our review spans studies along the full path to commercialisation, from RCTs testing interventions that aim to encourage the RD&D to those aimed to foster the scale-up and growth of green technologies.[3] We do not filter the RCTs we identify by geographical location but we do keep a narrow focus on green technologies that are most relevant for the UK's net zero transition.

RCTs relating to interventions focusing on the RD&D stage of green technologies

RD&D represents the beginning stage of the innovation process. In Table 3.1 below, we present the key attributes of the only two RCTs we have identified in our review that relate to interventions relevant for the RD&D stage of green technologies. The second of these relates to a matching grant scheme that is not actually about environmental (or green) innovation exclusively as it targets innovation with high social impact as well, but the fact that the scheme targets these specific types of innovations (as opposed to innovation generally) justifies the inclusion of this study in our review.

---

[2] Scanned websites, registries and databases include: TRIMIS, US Office of Energy Efficiency and Renewable Energy list of policy evaluations, Nesta Innovation Growth Lab, American Economic Association RCT inventory, J-PAL, STIP Compass, IEA Policies Database, Innovate UK, Catapult Energy, Uchicago UrbanLabs, National Environmental Research Council, HAL Open Science and Google Scholar.

[3] It should be noted that we are interested in RCT studies testing the effectiveness of policies aimed to induce innovation of technologies rather than RCT studies testing the effectiveness of the technologies themselves.

Table 3.1. RCTs relating to the RD&D of green technologies

1. Innovator preferences and catastrophes: evidence from a climate change field experiment	
Researchers:	Jorge Guzman, Jean Oh, Ananya Sen, Jorge Guzman, Jean Oh, Ananya Sen
Researchers' primary affiliations:	University
Trial location:	United States
Research objective[1]:	To understand the ways in which innovators' interest to innovate changes in response to different information on the risks of climate change, and how this varies by certain individual attributes.
Type of intervention tested[2] and type of outcome data measured:	Information, networks and collaboration Revealed outcomes
Registration date: Stated trial period: Stated trial status:	2 May 2020 (pre-trial) 5 May-27 Oct 2020 In development
Post-trial information/ analysis published?	No[3]
2. Promoting high-impact entrepreneurship: an evaluation of a Mexican government matching grant scheme	
Researchers:	David Atkin, Alejandra Mendoza, Leonardo Iacovone, Eric Verhoogen
Researchers' primary affiliations:	University, World Bank
Trial location:	Mexico
Research objective[1]:	To investigate the overall impact of a government matching grant programme (targeting innovative SMEs with high environmental or social impact, or high growth potential) on firms' performance, and to explore how to more effectively pick high-growth firms.
Type of intervention tested[2] and type of outcome data measured:	Financial incentives Revealed outcomes
Registration date: Stated trial period: Stated trial status:	21 Dec 2016 (during trial) 1 Sept 2016-1 Dec 2018 In development
Post-trial information/ analysis published?	No

Notes: 1. Many studies presented in the report have broadly defined and/or multiple objectives. The 'research objective' presented here may be a selected segment of the overall objective of a given study which is found to be most relevant for the purposes of this review. It may not be representative of the full scope of the given study. 2. See Table 2.1. 3. All reasonable effort has been made to scan sources where this information could possibly be published. Nevertheless, omissions may have been made; any such errors are the authors' alone. This applies to the whole table.

RCTs relating to interventions focusing on the scale-up of green technologies

Taking a holistic approach to innovation systems (as previously explained), our review of RCTs expands beyond those that are about the RD&D stage of technological development (i.e. the 'supply side' of innovation) and covers those about the scale-up and growth of green technologies, too (i.e. the 'demand side' of innovation).

In our review, we are interested specifically in RCTs that deal with the uptake of and investment in green technologies, rather than those that investigate purely behavioural outcomes (for plenty of examples of the latter, see Nisa et al., 2019). For example, an RCT that looks at whether participants are reducing energy demand as a result of an information campaign *is not* within our scope, but an RCT looking at whether participants are adopting (or expressing willingness to adopt) a technology as a result of an information campaign *is* within our scope. Key attributes of studies we have identified in this category are summarised in Table A1 in the Appendix.

Insights from an assessment of the attributes of green innovation RCTs identified

Development statuses of the RCTs identified

Our comprehensive review of the international evidence base for the purposes of this report has found a limited number of RCT studies (29 in total) that relate to interventions designed to spur green technological innovation. It is important to note that there are other RCT studies out there that are about innovation generally, rather than green innovation specifically. General innovation interventions that have been subject to RCT-based evaluations (for example, see Kleine et al., 2022 and Bakhshi et al., 2015) may have spurred some level of green innovation, too. Important lessons can be learned from these RCTs and be applied to the specific context of green innovation. However, these fall outside the scope of our current report as we focus on the role of RCTs in the development of green innovation policies specifically.

A crucial limitation of our review is that it only captures RCTs that have been recorded on a public registry and/or have had their results published. While the limited number of identified studies might imply a limitation of the evidence base generally, a unique dataset on public funding for global climate change and energy research reveals that 9% of the 1,000 projects across 17 countries that have been compiled supported experiments and quasi-experiments (Sovacool et al., 2022). This might suggest that there are many RCTs that could be relevant for green innovation policy development that are not publicly registered or have not published their results. If this is the case, it would go against the growing move towards scientific best practice which requires RCTs to be registered prior to implementation, with the study design, intended analysis methods and expected outcomes declared (Cavanagh et al., 2021). Following this practice ensures credibility of research findings and mitigation of issues such as 'p-hacking'[4] and hypothesising after results are known. The American Economic Association's (AEA) registry for RCTs, launched in 2013, and the Open Science Framework (OSF), also started in 2013, are some of the main tools that researchers can use to register their studies, publish pre-analysis plans and improve the credibility of their research.

Within our review, five studies have been registered pre-trial, completed and had results published; one study has been registered pre-trial, recently completed but has not yet been published; and another five studies have been registered pre-trial but are still in

[4] P-hacking refers to misuse of data analysis to find patterns in data that can be presented as statistically significant when in fact there is no real underlying effect.

development. Subject to the latter six studies disseminating their results when available, these 11 studies can be said to be following best practice. A further group of 10 studies identified in our review have been completed and had their results published, but do not specify whether they were registered prior to implementation. Another two studies that have had their results published were registered on the AEA registry after their trial was completed. These 12 studies are perhaps not technically in line with latest best practice, but this is likely a natural result of their early start dates before best practice was structured and formalised through the use of common tools among researchers to this extent. Indeed, trial start dates of almost all 12 of these studies date back to before 2013 when neither the AEA registry nor the OSF had yet been set up.

Four studies we have identified were registered after or during the trial and are stated as 'completed' on the AEA registry but remain unpublished, despite two having trial completion dates as early as 2013. These studies might be reflective of a publication bias (also referred to as the 'file-drawer problem'), whereby RCTs that do not produce 'significant enough' results are not put forward for publication (Dellavigna and Linos, 2022). This problem creates significant inefficiencies as more investigators might then test the same issues and find similar results, resulting in efforts being duplicated rather than building a shared evidence base. Furthermore, there are six studies in our review stated to be 'in development' on the AEA registry that were supposed to have been completed already, according to self-reported trial timelines. These studies might be suffering from implementation issues (discussed further in the following section), which can include delays, participant drop-outs and insufficient data to make firm conclusions. It may not be possible to draw robust findings from such trials to submit for formal publication, but there is still a strong case for disseminating lessons from what has gone wrong in order to inform future trials (for examples of researchers sharing lessons from implementation challenges, see Campos et al. [2014] and Riom et al. [2022]). Where a pre-analysis plan is published, this can also enable other researchers to potentially reproduce the trial in a different context and hopefully achieve successful implementation.

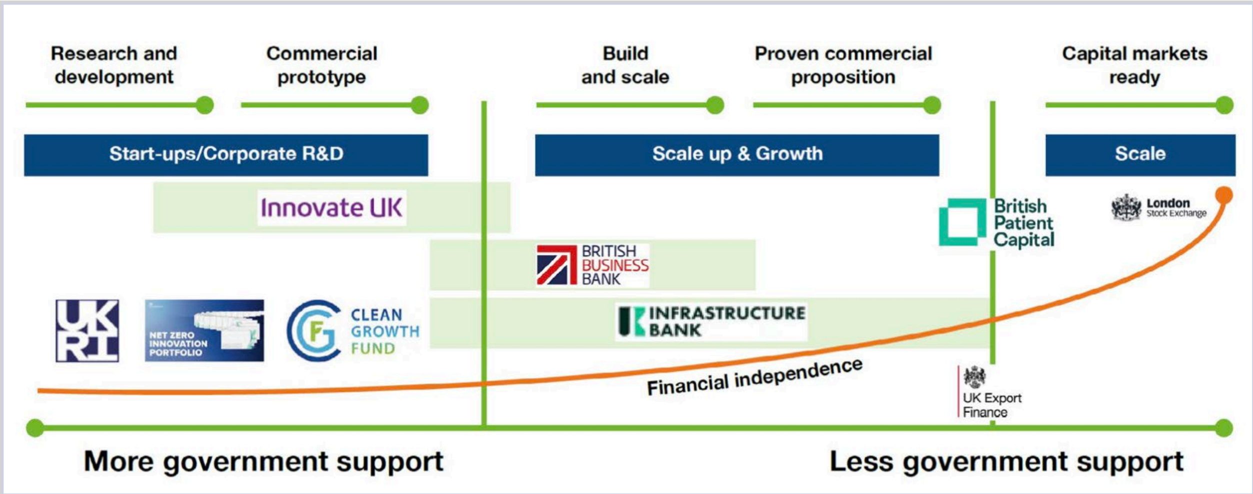
Trial start dates for a majority of the studies identified took place within the past decade or so, potentially reflecting the relatively recent rise of popularity of RCTs in the field of innovation, entrepreneurship and business growth. Interestingly, the number of studies is spread somewhat evenly through the 2010s with no apparent surge over time. This might suggest that the number of RCTs in the green innovation landscape is not increasing in line with growing climate commitments from around the world that only really gained momentum towards the end of the decade. Indeed, explicit labelling of firms or technologies as 'green' is still relatively new, and so are policies and private sector efforts explicitly targeting development in these areas, which in itself may partly explain the limited number of RCTs used in this context to date. Indicative evidence for this is seen in the RCT conducted by Kleine et al. (2022), who analysed the effect of the UK's nationwide innovation voucher scheme for small and medium-sized enterprises (SMEs). They found that it was not possible to analyse years of the scheme when application rounds were under specific categorisations such as energy and water. Specific categories on their own had relatively low levels of private sector innovation and growth up to that date, and therefore did not yield a large enough number of applicants to satisfy the RCT design.

Trial characteristics of the RCTs identified

A review of the trial characteristics of the RCTs we have identified produces some high-level insights. There is wide recognition that the need for government support is highest in the RD&D stage of the commercialisation pathway, but it decreases as technologies progress through the pathway to reach scale in the market (this is illustrated in Figure 3.1, which reflects the UK Government's thinking on this issue). Conversely, the number of RCTs identified in our review that relate to interventions focusing on the RD&D stage of green technologies is especially limited (two in total), while there are more examples of RCTs relating to interventions focusing on the scale-up of green technologies (27 in total). This may suggest that the unexploited potential of RCTs in generating evidence on what works might be especially large where it is most needed.

Indeed, when we look at technological innovation more broadly – which can hold important lessons for the development of green innovation policy specifically – the evidence to date on the effects of direct funding programmes on RD&D related outcomes is mixed (Pless et al., 2020). This is based on a review of 44 papers that use rigorous statistical methods (many of which are quasi-experimental) to evaluate impacts of such programmes. To illustrate, eight of these papers show positive impacts of direct innovation funding on patent outputs but another six show either mixed or zero impact. Bloom et al. (2019) also suggest that a knowledge gap might exist on the effectiveness of direct R&D grants, judging that the quality and conclusiveness of policy evidence in this area is 'medium'. A very small portion of the evidence base on the effects of public funding on innovation so far comes from RCT-based evaluations (see examples in: Kleine et al., 2022; and Bakhshi et al., 2015).

Figure 3.1. Public finance interventions across the different stages of commercialisation



Source: UK Net Zero Research and Innovation Framework (HM Government, 2021)

A majority (23 out of 29) of RCT studies identified are testing interventions relating to information, networks and collaboration. This is followed by nine studies, making up less than one-third of the total, testing some kind of financial incentive. This is unsurprising given the potential challenges and scrutiny that an investigator wanting to randomly allocate a financial instrument would face with regard to the fairness, equity and ethical implications of such practice, even if it is for research purposes (see Section 4 for more).[5] However, this may suggest that the benefits RCTs offer in terms of establishing accountability are even less exploited with regard to policy designs that involve immediately visible decisions about distributing government funding. Every policy comes at an operational cost but arguably, a policy with a visible 'price tag' (e.g. a government handing

[5] Note that this interpretation is underpinned by an oversimplified assumption that perceptions of fairness, equity and ethics will be dictated purely by the existence of the random allocation of a financial incentive within a given study without accounting for the value of the financial incentive and how that might compare to the cost underlying a non-financial intervention, which may serve a similar purpose. In reality, a non-financial intervention such as personalised information can be more expensive than small financial incentives. Furthermore, depending on sample size, the overall cost of an RCT that does not entail the random allocation of a financial incentive can far exceed that of one that does entail the random allocation of a financial incentive. However, since cost information of the RCTs we have identified in our review is not publicly available, our discussion is informed purely by the existence of the random allocation of a financial incentive within these studies.

out heat pump grants) is more likely to face public scrutiny on cost grounds than a government-run information campaign.

Our review has identified a single study testing an intervention relating to capacity-building and no studies testing an intervention that would be categorised as a regulation or obligation. Furthermore, in almost all the studies, the intervention is randomised among a sample of subjects that are individuals or households. These results are consistent with considerations we discuss in the next section that might explain the limited use of RCTs to date to inform green innovation policies. These include sample size requirements being easier to meet when working with individual or household subjects rather than firms, and discrete interventions such as information, grants and subsidies lending themselves better to randomisation than system-wide regulatory changes.

Study locations have a wide geographical range but the US clearly dominates as the location for 11 of the 29 examples identified. With only five studies, the UK holds the second highest number of RCTs identified. However, this information should be approached with caution as it might be reflective of RCT studies from the US and the UK naturally being more likely to appear in English-speaking journals and registries, despite many of these sources being global in scope. The outcomes of interest for research also appear to differ somewhat by location. For example, while the studies in the US and the UK focus primarily on the climate benefits of the technologies they relate to, studies in some other locations focus more on outcomes relating to improved energy access through the use of green technologies (see the studies in India and Argentina focusing on off-grid solar technologies). Furthermore, 21 of the identified studies measure revealed outcomes (e.g. technology sales, patents, productivity), while seven of them measure stated preferences (e.g. willingness to pay), and one study does not specify its measured variables.

Almost all the RCTs identified have been conducted by the academic community or members of non-profit, social-purpose organisations. In other words, they have been conducted not for policy evaluation *per se* but for research purposes, although ultimately still aiming to inform policy development. Very few of the RCTs in our review can be attributed to the private sector or government bodies. Notably, the only examples of direct government involvement are seen in the UK and the US. However, the limited number of examples might be a natural consequence of any RCTs conducted by the private sector or government being less likely to get published, given an expected level of confidentiality in such experiments. Indeed, many governments around the world have behavioural insight and experimental units running such studies without necessarily publishing results (OPSI, 2022). Similarly, many well-known companies like Amazon, Google, Uber and Deliveroo conduct experiments – some on a very large scale – but do not publicise the results or publish them in academic journals. Furthermore, there may be links that are not immediately apparent, such as academic researchers acting on behalf or in support of public or private bodies interested in conducting an RCT study (for a published example of an academic partnership with a private organisation, see Gosnell et al., 2020).

How have green innovation RCTs been used to inform policies?

Our review of the green innovation RCTs to date draws insights from the attributes of the RCTs identified, rather than their quality or policy impact. In other words, we only discuss 'if' RCTs have been used to build evidence on interventions relating to green innovation, rather than 'how' these studies have been used to inform policy, if at all. While we recognise the potential differences in the quality and policy relevance of the studies, our discussion is limited by the scarcity of publicly available information on these aspects (note that this is about information on the policy impact of trial results, rather than the publication of trial results in itself). While this information might exist but just not be publicly available, it is also possible that some studies have not effectively communicated or fed results into policy processes or have not had any notable impact on policy decisions, despite feeding into processes. As previously discussed, at least seven of the identified RCTs also appear to be incomplete and a few others have missing or out-of-date status information.

One study with directly reported information on policy impact is the energy labelling trial led by the erstwhile UK Department of Energy and Climate Change (DECC) (DECC, 2014b). Conducted in partnership with the department store retailer John Lewis and the Behavioural Insights Team (which was part of the Cabinet Office at the time), this RCT, which ran between September 2013 and June 2014, tested the effectiveness of providing information on electricity lifetime running costs (as opposed to only on a KWh per year basis) for moving people's purchasing behaviour towards buying appliances with lower energy consumption. The trial placed each John Lewis store in the study randomly in either an intervention group (where appliance labels included lifetime running costs in addition to kWh per year) or a control group (where appliance labels only reported KWh per year). There was a statistically significant decrease in the average energy consumption of the washer-dryers sold in the intervention stores over the course of the trial compared with the control stores, but no significant effect was observed with regard to other appliances in scope (washing machines and tumble dryers). Focus groups were conducted during the trial as well, identifying further ways in which the presented information could be made more meaningful to consumers (for instance, by presenting running costs on an annual rather than lifetime basis).

Overall, the findings supported the idea that small, low-cost changes to information provision can, in certain contexts, help to reduce energy demand. The DECC policy team contributed the findings to the European Commission's evaluation of the Energy Labelling Directive at the time. DECC also called on retailers and consumer groups to consider the lessons learned from the trial and consider making running costs more 'salient' to consumers, both online and in store. John Lewis decided to continue with the running cost approach to energy labelling (although detailing the annual rather than lifetime costs) and to explore extending it to other product categories as well as to online sales.

A notable scope for policy impact is also seen in the RCT conducted by Fowlie et al. (2018), who challenged the conventional wisdom that investments in energy efficiency bring positive private returns and generate environmental benefits and that people's reluctance to make these investments is thus ill-conceived. Their experimental evaluation of the largest residential energy efficiency programme in the US – the Weatherization Assistance Program (WAP) – conducted on a sample of approximately 30,000 households in Michigan, found that in fact the upfront investment costs entailed in the programme were about twice the actual energy savings. They suggested that the costs substantially outweigh the benefits, even when accounting for the broader societal benefits derived from emissions reductions, while recognising that the full set of policy goals may not be reflected in this kind of conventional cost-benefit test. Indeed, based on its own evaluation, the US government emphasises the jobs and economic growth induced through the programme rather than savings to households, suggesting that every $1.00 invested returns $2.78 in *non-energy* benefits (US DOE, 2021). Perhaps it is for this reason that the WAP is still in operation and, in fact, is set to see a tenfold increase in its budget, with a $3.2 billion influx as part of the $1.2 trillion US infrastructure bill signed into law in November 2021 (Protocol, 2022).

4. Using RCTs to inform the development of green innovation policies in the UK

Our review of RCTs that are relevant for green innovation policy development has shown that implementing RCTs in this field is possible but there are not yet many examples. This sets the stage for an assessment of whether there can be a greater role for RCTs for informing the development of UK green innovation policies into the future and, if so, how this could be achieved and where the greatest opportunities might lie.

Evidence-based policymaking and public policy RCTs in the UK

Within the UK government there is a long history of evidence-based policymaking – that is, using rigorous information to support policy decisions (Valters and Whitty, 2017). RCTs have emerged as a product of this culture with a wave of trials implemented in criminal justice policymaking in the late 1990s. This gave rise to some controversy about the compatibility of RCTs with policymaking (Cooper, 2018). Nevertheless, development aid became another hotspot for RCTs in the 2000s. These RCTs were used to provide well-researched justifications for development interventions (Valters and Whitty, 2017). Minouche Shafik, who served as the permanent secretary of the Department for International Development (DfID) between 2008 and 2011, was among the champions of a move towards the central coordination of evidence that could be used as a public good. Her views on the importance of evidence-building are evident from a speech she gave several years later, stating: "Scientific progress occurs as evidence is accumulated through better and better guesses – hence the importance of putting in place processes for challenge to avoid the tyranny of misguided ideologies" (Shafik, 2017).

In 2010, the Government created a Behavioural Insights Team (BIT) within the Cabinet Office. Not long after, the Cabinet Office released a report encouraging the increased use of RCTs across government (Haynes et al., 2012), while other parts of government also called for such methods of evidence building (HoC Health Committee, 2009; NAO, 2013). The BIT's intervention led to a small number of trials being implemented within the then Department of Energy and Climate Change (DECC, 2013, 2014a, 2014b), which provoked some debate within the department, perhaps echoing questions already raised in other areas of policy regarding the compatibility of RCTs with policymaking (Cooper, 2018). There were also a series of large-scale industrial policy experiments from the then Department for Business, Innovation & Skills (BIS),[6] including on Growth Vouchers (offering expert advice to small businesses), mentoring for high-growth firms, and university–industry cooperation (see Bakhshi et al., 2015; Phipps, 2017). Although we know views are varied, Bakhshi et al. (2015) argue that this reflected acceptance by (at least some) policymakers in the UK that randomisation permits more effective policy evaluation and testing, and in the longer run may help avoid poor (and expensive) policy decisions. More recently, the Department for Business, Energy and Industrial Strategy (BEIS) launched a large experimentation fund to test innovative ways of encouraging SMEs to adopt existing technologies and management practices to improve their productivity (BEIS, 2018).

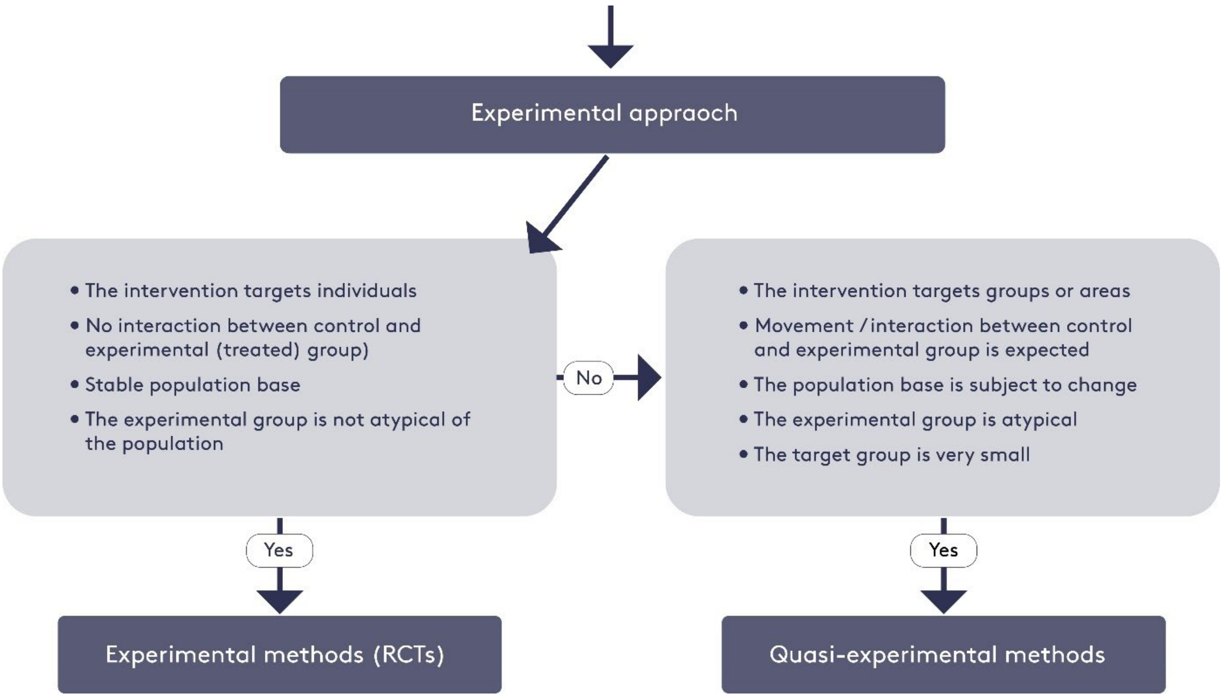
The Government's current guidance on policy appraisal and evaluation is contained in the Green Book (HM Treasury, 2022). This guidance requires monitoring and evaluation to be part of the development and planning of an intervention from the start, emphasising their importance for the "successful implementation and the responsible, transparent management of public resources". The Magenta Book, meanwhile, provides formal

[6] The Department of Energy and Climate Change (DECC) and the Department for Business, Innovation and Skills (BIS) were replaced by the Department for Business, Energy and Industrial Strategy (BEIS) in July 2016.

guidance specifically on policy evaluation (HM Treasury, 2020a). The latter does not serve as a rigid prescription of evaluation methods to be used in certain settings or for certain types of policies but rather allows the evaluator to judge which method is appropriate, based on an assessment of a set of practical criteria, perhaps in recognition of the context-specific nature of this choice.

The snapshot provided in Figure 4.1 below of a segment of a decision tree taken from the Magenta Book outlines the criteria that need to be considered when choosing between an RCT-based or quasi-experimental evaluation method, assuming an experimental approach to impact evaluation is taken (i.e. a 'counterfactual-based' approach, by the definition adopted in this report). While this kind of high-level guidance can provide a general idea about the criteria that need to be factored into the choice of policy evaluation method, especially for policy actors not directly involved in conducting evaluations, oversimplifying this choice into a set of binary criteria may be misleading. For example, contrary to what the decision tree suggests, certain RCT designs (e.g. cluster randomised trials) could work with interventions that target groups or areas too, not just individuals. Ultimately, as the Magenta Book itself states, an evaluator trying to choose the most suitable evaluation method must consider the full set of options available under the specific circumstances of the evaluation, including the type of question(s) to be answered, the intervention being investigated, the context in which it is being implemented and the information/data available (HM Treasury, 2020a).

Figure 4.1. Segment of decision-tree for selecting the most appropriate impact evaluation approach, from the Magenta Book



Source: Recreated by authors based on the Magenta Book (HM Treasury, 2020a)

Today, a range of organisations in the UK promote the use of RCTs and other rigorous methods in policy analysis. The introduction of a Cabinet Office Evaluation Task Force in 2020 is notable in this regard. They describe themselves as focused above all on impact and value-for-money evaluation, putting in place central evaluation teams within government departments and providing heightened scrutiny of evaluation practices within government. Other champions of rigorous evaluation and evidence gathering to support UK policy include the 'What Works' network, Nesta (which has now acquired the BIT) and its Innovation Growth Lab, J-PAL, as well as small organisations like The Behaviouralist. These organisations have been able to run, fund and help design an increasing number of RCTs and other evaluations in several policy areas.

The UK's green innovation policy landscape

Spending commitments for innovation

The UK has a strong history of pioneering and game-changing innovation in a variety of fields. However, despite the country's strong research system, its patenting intensity – a key measure of innovation output – currently lags other innovative countries: on average, patenting intensity (measured by patents per 1,000 employees) across France, Germany and the US is over twice that in the UK (Oliveira-Cunha et al., 2021). The Government has outlined innovation-led growth as a long-term strategic ambition for the UK (BEIS, 2021). In 2017 the Government committed to stepping up its direct support for innovation with an objective to invest 2.4% of GDP in R&D by 2027. In 2022 the Government set out its largest ever R&D budget, worth £39.8 billion for 2022–2025, to cement the UK's position as a science superpower and innovation nation.

Net Zero Strategy

How does innovation factor into the UK's transition towards a net zero economy? Since becoming the first major economy to legislate a net zero emissions target, the UK has set out its priority areas of action for getting there in its *Net Zero Strategy* (HM Government, 2021a).[7] The Strategy involves ambitious targets that entail rapidly scaling up a range of proven low-carbon technologies. These targets include: achieving a fully decarbonised electricity system by 2035; ensuring all new cars on the road are fully zero emissions capable by 2035; installing at least 600,000 heat pumps annually by 2028; bringing as many homes as possible to a minimum energy performance level of grade 'C' by 2035; and reducing industrial emissions by 63–76% (vs. 2019 levels) by 2035. Alongside commitments to known solutions, the Strategy puts innovation centre-stage for its role in delivering both emissions reductions and economic benefits including job creation. The UK currently has a technological advantage in innovation related to clean technologies overall, but faces strong international competition, with countries including Denmark, Norway, Austria, Germany and France patenting more intensively in this area than the UK (De Lyon et al., 2022).

Net Zero Innovation Portfolio

To accelerate the path for low-carbon technologies, systems and processes to market, the UK has introduced a £1 billion Net Zero Innovation Portfolio (NZIP), replacing the BEIS Energy Innovation Programme, which focuses on 10 priority areas for research and innovation (see Figure 4.2). Additional funding for the Portfolio was announced in the 2022 Spring Spending Review, bringing total funding for net zero innovation to £1.5 billion over the next three-year spending review period (up to 2024) (CCC, 2022). The overall funding is distributed into priority areas, drawing on the Government's new Net Zero Research and Innovation Framework determined in line with Climate Change Committee recommendations (CCC, 2022), and is then allocated to innovators typically on a competitive basis through a range of competitions run under the NZIP. The Net Zero Research and Innovation Framework focuses on objectives for the next five to 10 years, allowing for adjustments as the net zero transition evolves (HM Government, 2021b).

---

[7] In September 2022, the BEIS Secretary of State launched an independent review of the Government's approach to delivering its net zero target to ensure that delivering the target does not place an undue burden on businesses or consumers. The findings of the review, being led by Chris Skidmore MP, could lead to a new iteration of the Net Zero Strategy with potential updates to these priority areas.

The UK's growing commitment to spurring net zero-aligned innovation underpinned by recently introduced policy frameworks may create new opportunities for experimentation and the use of RCTs for building evidence on what actually works in this area.

Figure 4.2. Priority areas of the UK Government's Net Zero Innovation Portfolio

Source: Energy White Paper (HM Government, 2020)

Exploring the limited use of RCTs to inform green innovation policies in the UK: analysis from the literature review and interviews

Our review in the previous section has demonstrated that examples of the use of RCTs to inform the development of green innovation policies in the UK are very limited to date. The existing literature points to several possible explanations for this. The scarcity of existing examples is not exclusive to *green* innovation policy or to the UK: the application of RCTs is also limited in industrial and innovation policy more widely (Bravo-Biosca, 2020) and in other countries, as our review has shown. Our discussion of the reasons that may explain the limited use of RCTs in green innovation policymaking in the UK draws from this wider context.

We complement our review of the literature with results from semi-structured interviews we conducted with 10 policy professionals working specifically in green innovation policy or RCT design and implementation in the UK between June and August 2022.[8] Our interviewees belonged to governmental or non-governmental organisations with roles including managing million-pound budgets for green innovation, providing small packages of one-to-one support for entrepreneurs and offering help and advice on the implementation of RCTs.

We structure our discussion under eight themes that may contain answers to the question of why RCTs have not been used more widely to inform green innovation policies to date: (1) awareness, skills and institutional culture; (2) generalisability and transferability of lessons learnt; (3) cost; (4) time and timing; (5) experimental design constraints and risks; (6) complexity of the green innovation system; (7) fairness, equity and ethical considerations of random allocation and; (8) political feasibility.

As well as the literature we review for the purposes of this discussion not being specific to green innovation policy or the UK, our interview results (while being specific to this context) have substantial limitations as they rely on a very small and non-representative sample of

[8] See Morey (2022) for further methodology notes including data collection, data analysis and limitations.

green innovation policy professionals in the UK. Therefore, the insights we draw from our discussion are indicative rather than at all conclusive.

1. Awareness, skills and institutional culture

Insufficient awareness or skills, at either managerial or practitioner level, may inhibit the use of RCTs in informing green innovation policy development. Conducting an RCT typically requires teams with extensive skills in qualitative and quantitative research methods, the right combination of disciplinary approaches (e.g. economics, econometrics, sociology, psychology), knowledge of best practice, as well as experience in conducting such trials, disseminating findings and engaging with stakeholders (Masset et al., 2019). In 2016 Nesta's Innovation Growth Lab surveyed 170 academics, policy officials and practitioners from the public and third/charity sectors, asking what was, in their view, the single most important barrier to running RCTs and to using more evidence to inform their policies (Firpo, 2017). Many answers mentioned a lack of knowledge, including a lack of understanding of RCTs among senior officials and a lack of knowledge or skills to implement RCTs.

In some cases, an organisation may lack the institutional culture that necessitates or is conducive to making evaluation methods more rigorous. Firpo and Phipps (2019) observe that experimenting is not something that comes naturally to a lot of public organisations. Views on RCTs may also vary between different actors in the policy landscape. For example, within government, policy evaluators may be very willing and able to conduct RCTs but experience little demand from decision-makers for the use of rigorous evaluation methods. The aforementioned survey by Nesta highlighted barriers that relate to 'knowledge inertia', a term used by its researchers to categorise answers suggesting that decision-makers often believe their opinion is correct and do not demand better evidence of their programmes' impact (Firpo, 2017). On the other hand, it may be the evaluators that resist calls for such studies from decision-makers due to budgeting constraints.

Interview insights

There was scant evidence from our interviews that awareness or skills of staff responsible for evaluations are factors restricting or deterring the use of RCTs to inform the development of green innovation policies in the UK. Interviewees even outside of dedicated evaluation roles showed a fair level of awareness of RCTs and three of the interviewees had been directly involved in RCTs in the past. Furthermore, it was pointed out that even outside the internal capabilities of these organisations, there are third party organisations that can be contracted to run RCTs, and the hiring of such organisations is not unusual.

There was no sign of institutional cultures disregarding the value of rigorous evidence to inform decision-making. A motivational focus on the ultimate goal of net zero as well as genuine willingness to improve policies were apparent in the interviewees' responses. However, the general stance was that there was already adequate evidence available from existing monitoring and evaluations on many programmes to suggest that they were achieving their aims. These monitoring and evaluation processes were described in many cases as going 'above and beyond' the minimum requirements.

Our findings are inconclusive on the extent to which the use of RCTs in green innovation policymaking so far may have depended on factors relating to awareness, skills or institutional culture. However, there is good reason to believe that views on what constitutes 'adequate' evidence and the demand for such evidence vary among individuals and/or the different types of roles in the policy landscape. The demand for RCT-based evaluations may have been low if existing evidence, likely underpinned by quasi-experimental or theory-based methods, is viewed as adequate.

2. Generalisability and transferability of lessons learnt

There is a risk that an RCT will produce knowledge that is too narrow and local to be useful for policy – that is, that has limited external validity – especially where the experimental setting is so tightly controlled that it no longer accurately reflects the broader real-world environment (Frederiks et al., 2016; Masset et al., 2019). With such an RCT, questions are likely to remain about the contextual factors influencing the effects of the policy and how these effects could vary across different individuals or firms outside the scope of the specific study (Steventon et al., 2012; Ettelt et al., 2015; Dalziel, 2018). Especially as innovation outcomes are heavily influenced by external factors, an innovation policy observed to work in a particular experiment does not necessarily mean that it would work elsewhere or with different participants. On the other hand, if the study generates negative results, it may provide no other recommendation apart from confirming that a particular type of incentive does not work (Masset et al., 2019).

Interview insights

Some interviewees argued that the use of RCTs is not appropriate even in the absence of any practical barriers, principally because RCTs do not provide policy professionals with the kind of data they need to improve their programmes. This was linked to comments about RCTs having weak external validity, suggesting that RCTs only provide a measurement of the impact of a programme, while providing little information on how this impact was achieved or the ways in which the programme could improve.

The interviewees said they would rather see new evaluation regimes that collect a variety of rich qualitative and quantitative data and begin supplying insights to those implementing the policy so they could improve the programmes as quickly as possible. In general, there seemed to be an emphasis on the usefulness of talking at length with the beneficiaries of the innovation programmes and receiving detailed feedback on the support provided.

Critics' heavy focus on external validity suggests this may have been a limiting factor to the use of RCTs in green innovation policymaking to date. However, as previously discussed, limited external validity is not a problem unique to RCTs: other policy evaluation methods usually run into similar issues as well (Firpo and Phipps, 2019). Complementary investigations are commonly used to understand the mechanisms by which effects of the policy are produced and how the effects may vary in different contexts – in other words, to identify the 'why' and 'how' in addition to 'if' the policy is working as intended. Deaton and Cartwright (2018) go further by stating that the *only* way RCTs can play a role in building scientific knowledge and useful predictions is if they are used as part of a cumulative programme and combined with other methods, including conceptual and theoretical development.

BEIS already requires all experimental (and quasi-experimental) methodologies to be complemented with qualitative research to "learn why and how different groups were affected, to improve delivery and inform future policy design" (BEIS, 2020). An example of a mixed methods approach is demonstrated by Bakhshi et al. (2015), who combine an RCT with qualitative data collection through case studies consisting of semi-structured interviews, multi-organisational workshops and ethnographic observations. Such insights into causal processes underlying quantitative findings offer lessons that may be generalised and transferred to any subsequent or wider implementation of the policy.

3. Cost

RCTs have operational requirements that can be expensive to implement. These include scientifically sound sampling of subjects, standardised delivery of the intervention under a random allocation design, valid and reliable measurement of key constructs, and appropriate data collection (Frederiks et al., 2016). However, the costly aspect tends to be true of high-quality policy evaluations in general, not just RCTs (Dalziel, 2018). To illustrate, the head of the World Bank Development Impact Evaluation Division estimates the average cost of an impact evaluation in the field of international development at $500,000 (Nature, 2015). Despite large costs upfront, it could be argued that RCTs (or other forms of high-quality impact evaluations) are cost-effective in the long run because they will help improve policy design and inform policies in the future or elsewhere. However, Masset et al. (2019) argue that little empirical evidence is available in support of such a hypothesis.

Interview insights

Cost was one of the most mentioned problems associated with running an RCT, discussed by all interviewees. There were unilateral expressions of the preference that as much of the overall budget as possible be spent on implementation of the programme in question, with some reluctance observed when the prospect of raising the proportion of spending for evaluation was discussed. This thinking does not appear to arise from a general lack of funding for programmes or their evaluation; most parties agreed that it is not difficult to achieve funding (although smaller organisations acknowledged higher funding pressure generally). Instead, this preference seems to arise from the desire that as much of the money as possible be put into creating a positive impact. This suggests that cost is not so much the restrictive factor here but rather the bigger determinant is the extent to which RCT-based evaluations are viewed as being able to facilitate creating a positive impact.

While cost certainly appears to have contributed to decisions about using RCTs as a policy evaluation method in green innovation, this seems to be about making a judgement of the perceived value of RCTs rather than about a genuine lack of money available. If it could be established in advance that RCTs were worth the money as they would generate useful lessons for improving policies, more RCTs would probably be conducted. However, this is easier said than done; further work is needed to build empirical evidence on the cost-effectiveness of RCTs themselves.

4. Time and timing

RCTs require long time horizons to design and implement. Design steps of an RCT – which all need to be fulfilled prior to implementation – include formulating a research protocol consisting of the research hypotheses and a framework for population sampling, randomisation, data collection and analysis; having the research protocol peer reviewed, modified as required and registered; and having clearance from any collaborators and implementation partners. In Nesta's survey, 'limited resources and time constraints' was selected by over 70% of respondents as a barrier to conducting RCTs (Firpo, 2017).

When it comes to implementing the RCT, the time required for innovation- or productivity-related outcomes from projects targeted by an innovation policy to become measurable can be several years, especially where the objective is to measure revealed outcomes rather than stated preferences. This issue affects the evaluation of green innovation policy in general, irrespective of the chosen evaluation method. For example, recent work has shown that more than a decade is needed for the full effect of public funding for energy R&D to be realised (Pless et al., 2020). Not only is the time lag between public innovation support and ultimate innovation outcomes long, but it is also uncertain. Adopting a single impact period may therefore either miss or incorrectly estimate the true effect of these policies (Bakhshi et al., 2015). For example, Bakhshi et al. (2015) adopt a longitudinal data collection strategy in their 'RCT+' study of Creative Credits. They collect data in four stages spanning around two years from project start and find important insights from later stages that would have been lost had it not been for the longitudinal design.

What is arguably more relevant than simply looking at the duration of an RCT is how that duration aligns with related policy or financial cycles and whether the evidence generated from the RCT will be available in time to inform decisions. Evaluation results that come too late in the decision process may be redundant for the decisions at hand and irrelevant unless they are meant to inform decisions about other projects or to generate knowledge that can be used in the future (Masset et al., 2019). In Nesta's survey, recurrent answers around the main barriers to RCTs included the short nature of policy cycles (Firpo, 2017).

Interview insights

The time it takes to run RCTs was a very commonly-raised issue in interviews. These trials tend to take longer to complete than other forms of evaluation, which means that the information provided by the evaluation takes longer to be fed into policy decisions, slowing potential improvements in successive programmes. This was discussed alongside the fact that public funding is linked to policy cycles, and once a cycle is in flight, significant changes must wait until the next cycle is designed and released to see learning implemented in policy. Alternative evaluation techniques were mentioned that can give programme feedback while they are still in flight, in turn enabling faster learning to take place.

Interviewees also raised that RCTs are best built into the design of a programme from the beginning and not possible to retrofit because of requirements such as randomisation, which needs to be conducted prior to programme implementation. This limits the applicability of RCTs for many of the UK's programmes designed to support net zero innovation, which are already underway.

Time and timing considerations may have posed a real limit on the applicability of RCTs to inform green innovation policy development to date. However, the time lag between when information from an RCT starts emerging and the point at which it can be fed into policy decisions is not inevitable. Intermediate outcomes can be defined within the RCT to test if change is happening in the intended direction. For example, if the ultimate outcome of interest is the generation of a green patent, an intermediate outcome could be defined as 'plans to conduct R&D in green technology signed off by senior management', or 'actual spending on R&D for green technology' by targeted firms. That being said, since the development of new energy technologies can take decades, even patents may be more appropriate as intermediate outcomes (Pless et al., 2020).

Pritchett et al. (2013) outline how RCT-based evaluations can be designed specifically to improve projects while in operation by embedding dynamic feedback loops aligned with key decision points. This approach extends the principles of RCTs inside the policy implementation process, starting by articulating the different available policy alternatives and then simultaneously trying out those identified as being promising and adapting the policy sequentially, drawing on a measurement and assessment of intermediate outcomes. At the core of the approach is the argument that variations within policy design can serve as their own counterfactual and allow for robust evaluation. This kind of experimental stance was illustrated by the BEIS Growth Vouchers Programme, where policymakers approached the whole programme as a policy experiment, rather than starting with a single policy design (Bravo-Biosca, 2020). The trial evaluated alternative delivery strategies, including messaging to attract applicants and different diagnostic tools to guide applicants' support decisions, in addition to the effectiveness of the voucher itself (ibid.).

5. Experimental design constraints and risks

RCTs require samples that are large and homogenous enough to be able to credibly detect causal effects. This relates to statistical power, which is the probability that a statistical test will find a statistically significant difference between the treatment and control groups when such a difference actually exists. Statistical power increases the larger the sample size and the smaller the variability among subjects in the sample (Norton and Strube, 2001). While a large and homogeneous sample may be easier to find when the subjects are individuals or households, it is more difficult to achieve when working with firms, especially when thinking about those that are likely to be targeted by UK innovation policies.

Firstly, the number of firms within the scope of an innovation programme may be relatively small to begin with. This relates to many green technologies, including in energy generation or industrial decarbonisation, which require large-scale infrastructure development being likely to be driven by a relatively small number of large firms. Secondly, take-up of the programme may be too small to yield an ultimate sample size that would be sufficient for an RCT, even if the number of firms targeted by the programme is large enough. This issue could be somewhat mitigated by targeting a very large pool of firms, so that the resulting sample will still contain enough firms that take up the programme, but this approach can be very expensive (McKenzie, 2010). Thirdly, innovators operating in the UK net zero landscape are highly heterogeneous, varying in size, activities and capabilities. Such heterogeneity in the experimental sample can be especially detrimental to statistical power if one of the objectives of the study is to investigate how different types of firm respond to a certain programme, as the number of firms of any given type may be especially small (ibid.)

The UK's public funding for R&D is allocated primarily through competitions with specific, narrowly defined scopes, which are usually run in multiple phases. The NZIP provides the latest major example of this approach. For illustration, only 11 projects have been awarded funding in the first phase of the first stream of the Heat Pump Ready Programme under the NZIP, which clearly cannot provide a meaningful size, even for a single trial arm, let alone an entire sample to randomise within. The fact that only applicants who have been successful in the first phase can apply for the second phase means that any future phases of such competitions will be even less likely to meet the sample size requirements of any prospective RCT.

Even with a well-designed experiment, there is a risk that implementation will not go as planned. There can be many threats to the ultimate robustness of the experimental sample, and consequently of the data drawn from it, due to issues with recruitment or retention of the subjects, the delivery of the intervention, or possibilities of contamination.[9] Researchers and evaluators may steer away from conducting RCTs given these risks, even if the payoffs from a successful trial and from drawing causal conclusions is attractive.

Interview insights

There was much reference in the interviews to the huge variety of innovations that receive funding from UK organisations: from the very small to the very large, from the circular economy to renewables. This level of uniqueness of participants means that it is very difficult to point to an average innovator across whole innovation portfolios, and if RCTs are to be implemented the comparison can be made only across comparable

[9] Contamination refers to when members of the 'control' group inadvertently receive the treatment or are somehow exposed to the intervention.

companies, most likely within the same programme (and as the next point shows, this is often insufficient for the creation of a large enough sample size for an RCT).

Sample sizes that were too small was raised as a significant problem for many of the programmes at all levels of funding; the largest projects, such as new energy plant technologies, involve funding only one beneficiary to build a demonstrator or final product. Other forms of innovation programmes involve as many as 20 directly comparable companies, selected initially after a competition, but this number tends to be reduced over consecutive phases as the best projects are prioritised. Many of the UK net zero portfolios of various organisations largely consist of such programmes, with the reason often being that there are not many innovators to choose from in any given area. This means that there are simply not sufficient sample sizes available for many of these existing programmes to be adapted into RCTs.

Interviewees also highlighted the difficulty of retaining subjects over the duration of the experiment. Participants in the programmes are generally found to be willing to provide internal data and performance figures when they are offered money or support in return, and will even do so for a period after they stop receiving support. However, several interviewees suggested that innovators might not be willing to provide outcome data across the time spans required to run an RCT, and also pointed out that those in a control group who receive no intervention will have little incentive to supply data.

Both the existing literature and our own interview findings suggest that there are real design constraints and risks that may be inhibiting the use of RCTs to inform green innovation policies in the UK. Nevertheless, not every RCT-based evaluation of green innovation policy will suffer equally from these issues and the specific circumstances should dictate the ultimate choice of an evaluation method. For example, Bloom et al. (2013) were able to conduct an RCT with a very small sample of only 17 textile firms in India to test the effect of free consulting about management practices on firm performance. The advantage they had was that these firms were all in the same sector, delivering a sufficiently homogeneous sample to work with, and there was rich data available on weekly production, making it possible to measure impacts of the intervention.

6. Complexity of the green innovation system

Complex systems are made up of and emerge from many diverse, interacting components (actions, layers, organisations, government departments), and non-linear and non-proportional interactions between those components (HM Treasury, 2020b). Green innovation occurs across such complex systems of energy, transport, food and the built environment, requiring changes in technology, market structures and infrastructure. Driving this kind of system-wide change is generally not feasible for a single policy to achieve and so instead requires a combination of policies (EEIST, 2022). These policies inevitably interact with each other and can even be mutually reinforcing (ibid.). RCTs become less feasible under these circumstances, i.e. if the intervention they are concerned with is being applied to a complex and emergent system and/or is difficult to disentangle from other programme interventions (HM Treasury, 2020a).

Some of the policies in the green innovation policy mix are discrete, delivered directly to the innovators or users of technology (e.g. grants, subsidies or information treatments), but some are operationalised at the market-level, changing relative prices or conditions overall in favour of lower-carbon alternatives (e.g. reforms in the regulatory environment or an economy-wide carbon tax). Market-level interventions do not lend themselves to randomisation under an RCT design in the same way that offering discrete interventions to individuals or firms does (McKenzie, 2010; Campos et al., 2014; Vine et al., 2014). It is rarely possible to change the ‘system’ (or the market) for some subset of the population (Chater and Loewenstein, 2022), which makes it highly challenging to construct a counterfactual.

As a notable exception, Atkin et al. (2017) demonstrate how randomising market conditions for a subset of the population can be possible. In their RCT design, by generating exogenous variation in the access to foreign markets for rug producers in Egypt, they are able to test the impact of exporting on firm performance.

Even with a discrete policy intervention, the nature of the relationship between the technology and the ultimate users of the technology matters for the feasibility of testing a given policy intervention under an RCT design, especially on the demand side. Some of the technologies in the green innovation system are designed to be used by consumers (individuals or firms) directly, enabling the consumption of green energy (e.g. electric vehicles [EVs], heat pumps, hydrogen-ready boilers), while some other technologies are designed to decarbonise the production of energy or industrial processes 'upstream' (e.g. nuclear power, carbon capture, usage and storage [CCUS]). The latter set of technologies reduce the emissions embedded in the production of other products and services which are then used directly by consumers (e.g. low-carbon electricity, low-carbon manufactured goods), although they themselves do not have a direct consumer interface. Demand-side policy interventions designed to drive the uptake of consumer-facing technologies can be placed directly on these technologies (e.g. EV grants). But demand-side efforts to drive the uptake of technologies that integrate into the energy system upstream are likely to require policy interventions to increase the demand for the consumer-facing products and services they enable, instead of interventions on the technologies themselves. Such long, highly interconnected, indirect causal chains between interventions (placed on the demand-side) and impacts (occurring upstream) can make causality hard to prove, because of the difficulties it creates for standardising the intervention or isolating a control group (HM Treasury, 2020b).

Furthermore, recent research suggests that policy to drive rapid innovation and growth in green technologies should be adaptive, having the ability to change as markets change and more information becomes available (EEIST Consortium, 2022). Given an effectively unlimited range of different pathways that a transitioning innovation system could take, it is hard (if not impossible) to identify a single 'optimal' policy choice ex-ante. To remain effective and durable, policy must be able to adapt to relevant macroeconomic, demographic, social and geopolitical trends and developments, along with changing dynamics (including those shaped by the policy itself) in technology, infrastructures, markets, preferences and politics. An RCT-based evaluation becomes less feasible if the policy intervention it is concerned with adapts over time (HM Treasury, 2020a), although there are ways to embed principles of adaptiveness in RCTs, for instance by adopting the approach from Pritchett et al. (2013), set out in discussions above.

Interview insights

Several interviewees noted challenges relating to causal complexities in innovation systems. They suggested that there is a lack of formal theory that describes how innovation and innovation support happen, with a lack of detailed definition or modelling of the mechanisms involved and how they interact with different contexts. The result is that it is difficult to test the mechanisms behind innovation policies to confirm they have impact in reality. While testing the impact of policies on generating innovation outcomes is immensely complex to begin with due to dynamic interactions between markets, policies and physical processes underpinning the innovation system, an even bigger challenge specific to *green* innovation policy is how to pin down the impact of a given innovation on emissions reduction.

Overall, the complexity of the green innovation system may pose a real limit to the extent of policy interventions in this landscape that can be feasibly tested under an RCT. Indeed, the Magenta Book (2020) suggests it would be suited to turn to theory-based evaluation methods in situations where there is a complicated policy landscape with a combination of interventions, the intervention is designed to make a change in a complex system, the intervention is changing over time or where outcomes are emergent and cannot be predicted at the outset.

7. Fairness, equity and ethical considerations of random allocation

By definition, RCTs require the random allocation of a policy intervention. While for learning purposes this would be limited to the scope of the study, withholding potentially effective interventions from certain participants in pursuit of scientific rigour may still give rise to fairness and equity concerns (Frederiks et al., 2016). Policymakers may oppose the random allocation of innovation support on the basis that it would undermine expertise developed over many decades that informs funding decisions (Pless et al., 2020) and result in an inefficient use of public investment (Bakhshi et al., 2015). For example, Campos et al. (2014) report governments' unwillingness to randomly select recipients of grants as one of the main reasons why their attempts to conduct experimental evaluations of matching grant programmes in Africa have failed. Even if policymakers themselves were open to random allocation of financial resources, this is unlikely to be politically feasible if business leaders use their financial positions and relationships with decision-makers as leverage to campaign against the random allocation of a resource on which they may have come to depend (Dalziel, 2018).

There are contrary opinions that suggest random allocation of a policy intervention or support could actually be (or be perceived as) the fairest approach, at least in certain circumstances (Dalziel, 2018). These circumstances could include when demand for a programme exceeds available resources or when the programme targets early start-ups, which might be difficult to screen due to a lack of track record. The resource-intensive nature of screening applications for a programme may also create an economic rationale for random allocation, especially if the support being offered to each firm is relatively small, as is typically the case in innovation voucher schemes (Bakhshi et al., 2015). One further view is that the alternative to random assignment – that is, allocating grants through a peer review system – may in fact be more biased or have less legitimacy than random assignment (Roumbanis, 2019; Mayo et al., 2006; Day, 2015; Lee, 2015; Pier et al., 2018). Some science and innovation funders are in fact already awarding grants (at least partly) randomly, using lottery systems after an initial screening to ensure the participants meet a minimum standard (Adam, 2019). An example is the YouWin programme run by the Nigerian government, which awarded grants at random to business owners as an assurance against corruption – and to enable the rigorous, RCT-based evaluation of the programme (Kopf, 2015).

Regardless of the different views, there may be regulation and legislation against research practices underpinned by the random allocation of a treatment. It may be that regulation requires all similarly situated customers to be treated (simultaneously) in the same manner, and that nobody can be worse off as a result of exposure to an experimental treatment (Vine et al., 2014; Hahn et al., 2017). For example, the proceedings of the Retail Market Review by Ofgem in 2010 made it difficult for energy suppliers to randomly assign offers of discounts or rewards for behavioural change to some customers but not to others (ibid.). Government officials themselves may also be reluctant to 'control' decision-making by individuals in the market that are supposed to be outside the control of government (Vine et al., 2018).

Views about fairness, equity and ethical implications of random allocation appear highly varied, with no clarity on how these factors may have affected the use of RCTs for informing green innovation policies to date. It is worth emphasising that the severity of concerns around fairness, equity and ethics highly depends on the experimental setting, nature and the aspect of the intervention being randomised and the design of the particular RCT. Such concerns could be minimal for an RCT that randomises aspects of the delivery of policy support rather than the policy support itself. These RCTs can generate lessons about the most effective way of delivering a certain type of policy support without blocking any subject from accessing the policy support. 'Encouragement designs', for example, make the policy intervention simultaneously accessible to both the treatment and control group but actively encourage the treatment group to take up the intervention (for instance, using information and advertising).

Interview insights

Our interviewees did not raise fairness, equity or ethical concerns as directly discouraging the use of RCTs to inform green innovation policies. This may be explained by innovation policy conventionally being understood as R&D policies targeting firms, and a view that randomly allocating public support among firms arguably is more ethically acceptable than adopting a similar allocation approach to the distribution of support among financially disadvantaged individuals (Bakhshi et al., 2015). Alternatively, this may be a result of our interviewees representing a segment of the policy landscape that has either direct experience or strong familiarity with RCTs. Certainly in Nesta's survey, respondents whose organisation had run an RCT were not as likely to mention ethical concerns or public reactions as a barrier (Firpo, 2017).

Even where the objective is to test the impact of the policy support itself, RCTs can be designed in ways that mitigate concerns about fairness, equity or ethics. For example, a phase-in design applies the policy intervention to a randomly assigned treatment group some time (months or years) before applying it to the control group. This provides an interim period of data collection, which can be used to identify causal effects of the intervention without actually preventing the control group from receiving the intervention. This design can be especially relevant if there are constraints on budget that prevent roll-out of the intervention to the whole population in one go. Furthermore, random allocation of interventions that are not about financial support may be less sensitive to fairness or equity concerns. For example, by randomly assigning a requirement to collaborate with specific types of firms, national laboratories or universities among the recipients of a research grant it could be investigated how collaborations affect technology commercialisation (Pless et al., 2020).

8. Political feasibility

Officials within government may be generally opposed to the use of rigorous evaluation methods, including but not limited to RCTs. The reward for showing a programme is successful may typically be less for government officials than the consequences (or perceived consequences) of proving that the programme has not worked, leading to a reluctance to carry out an evaluation (Campos et al., 2014). In fact, in Nesta's survey, the largest group of answers among policymakers and public sector practitioners regarding barriers to RCTs revolved around concerns with public reactions and fear of failure (24% of all answers) (Firpo, 2017). If an RCT demonstrated the failure of a policy programme, there would be a natural expectation for policymakers to adopt a different course of action. However, policymakers, perhaps especially if they are elected politicians, do not want to be accused of an error of judgement or be cornered into making a 'U-turn' on a policy decision, as this is often seen as a sign of weakness, indecisiveness or incompetence (Ettelt et al., 2015).

The factors identified above may not stop RCTs from being conducted entirely, but they might mean that the evidence generated from an RCT will not be used as intended to inform policies. For example, Ettelt et al. (2015) analysed three examples of policy experiments (two of which involved large-scale RCTs) in national health and social care policy in England, and concluded that these policy experiments were mostly seen as a strategy for demonstrating the effectiveness of a previously chosen path rather than to answer open questions about whether or not a policy was likely to work. The policy experiments were used strategically early in the policy process to establish a narrative of 'evidence-informed policy' in support of existing policy decisions, but there was a lack of genuine engagement with the findings (ibid.).

Interview insights

There was little evidence from the interviews to suggest a fear of failure or subsequent negative consequences were preventing evaluation from happening. There was some acknowledgement that lower performance figures could translate into less funding being supplied, and again the smaller organisations recognised more pressure here. However, this was largely overshadowed by a unilateral strong enthusiasm for the usefulness of constructive feedback provided by evaluations. The possibility of bad results and learning from these mistakes, therefore, seems to be accepted as a natural aspect of policy evaluation.

Interviewees commonly said they might implement an RCT in order to communicate the success of their programmes more effectively. This, however, was not the result of external pressure to present a certain image to the scrutinising media, but was due to the incentives produced by internal funding processes. Several interviewees emphasised that while RCTs could not supply them with the type of information they find most useful (rich and detailed qualitative information, giving insights into 'why' certain things happened), voluntarily supplying rigorously generated, RCT-based data on the impact of their programmes could help them to secure more funding in the future. This was not an indication of strong top-down pressure to prove the impact of programmes or otherwise lose funding: the continued survival of programmes could in fact be secured with quite simple performance indicators. Rather, RCTs were seen as something that could maximise the impressiveness of a programme to secure more funding for the next cycle. Interviewees from smaller organisations who received money from the Government did mention more pressure to demonstrate the success of programmes, but this was not pressure for rigorous evaluation.

While a fear of failure does not appear to be a big concern for the interviewed policy practitioners, the literature suggests that such a fear may still be present at the top levels of government. This may be reducing the top-down pressure to implement rigorous methods of policy evaluation such as RCTs. Ultimately, RCTs require openness from both evaluators and top-level officials to accept that some policies or programmes may not be having the impact they were thought to be having. In fact, insights gathered from impact evaluations across a range of policy areas suggest a general rule that "80% of things don't work" (White, 2019). Communicating policy impact effectively and convincingly, either to the public or for the purposes of internal funding processes, is a legitimate motivation to conduct RCTs. But any upfront assumption that an RCT can be used to communicate success and maximise the impression of impressiveness would imply a lack of interest in the actual results and the consequent learning opportunity, including from failures (Heiskenan et al., 2022).

Could RCTs be used more widely to inform the development of green innovation policies in the UK?

Across the eight themes discussed above, all the factors we have identified that might explain why RCTs have not been used more widely in green innovation development to date appear to be highly-context specific. RCTs do not have fundamental shortcomings that would prevent their application to build evidence in this area but equally, there are no 'rules of thumb' that would make an RCT the default method of evaluation. Ultimately, the method of evaluation should be chosen for its appropriateness to the design of the policies being evaluated, the evaluation questions being asked, and the resources available (BEIS, 2020).

Barriers relating to generalisability and transferability of lessons learnt, and time and timing, are not unique to RCTs – they can affect any evaluation method for a green innovation policy. There are also concrete ways in which RCTs can be designed to deal with these issues, for instance using complementary investigations and embedding dynamic feedback loops aligned with key decision points, respectively. On the other hand, barriers to RCTs identified under the themes of awareness, skills and institutional culture; cost; fairness, equity and ethical considerations of random allocation; and political feasibility appear not to be unique to the green innovation policy landscape. Views on these barriers appear to be rooted in the wider context around the use of RCTs in policy evaluation and evidence-based policymaking in the UK. If there were better alignment across the different actors in the UK policy community on what constitutes adequate evidence and why it is important to base policies on such evidence, as well as a more unified stance on the role RCTs can play in building this evidence, we would expect to see more RCTs in the green innovation policy landscape too. These issues go beyond our question specifically on the use of RCTs to inform green innovation policies and require further investigation involving the policy profession as a whole, which can build on the analysis we have presented in this report.

Experimental design constraints and risks, and the complexity of the green innovation system, appear to present some feasibility challenges unique to the use of RCTs to inform the development of green innovation policies. These are very likely to have implications for the applicability of RCTs as a method of building evidence on the efficacy of the UK's current and future portfolio of green innovation policies. Unlocking the full learning potential from RCTs to inform the development of green innovation policies would therefore require a concerted effort from policy professionals working specifically in the green innovation landscape. This would include being proactive about identifying unique knowledge gaps in the green innovation policy landscape that RCTs can help fill, an awareness of the methodological advantages that experimentation can offer over and in addition to other investigation methods, and an open-minded and creative approach to formulating research designs that can overcome feasibility challenges.

Interview insights

Overall, interviewees suggested that a large proportion of the green innovation work in the UK lies outside the bounds of what is practical to test in an RCT format. Even so, several areas related to green innovation were pointed out repeatedly as unlikely to suffer from practical constraints and thus being potentially fertile ground for the implementation of RCTs. These were: consumer-facing programmes (e.g. involving smart meters or car-to-grid technology), trials involving heating or insulating solutions where it is possible to have a control group of buildings, public-level behavioural interventions, and company-level interventions applied to buildings and other infrastructure. These areas tend to have large enough sample sizes made up of fairly homogenous subjects and easy-to-gather data and are thus conducive to conducting RCTs.

Opportunities for using RCTs in the Net Zero Innovation Portfolio

As a thought experiment, we now return to the NZIP and ask which of the UK's priority areas for innovation may represent the greatest opportunity for RCT-based evidence-building going forward. In this thought experiment we focus purely on the two themes we identified as containing challenges unique to the use of RCTs in the green innovation policy landscape, as opposed to those that relate to RCT-based evaluations or to the evaluation of green innovation policies generally, which require conversations beyond the scope of this report.

We ask two questions – one on experimental design constraints and risks, and one on the complexity of the green innovation system – to comment on the feasibility of a potential RCT in a given NZIP priority area. It is important to note that this is an oversimplified illustration, not a recommendation in any way, since we focus narrowly on these two themes rather than on the whole range of context-specific and highly intertwined factors that will determine the feasibility of a proposed RCT in reality. Furthermore, in the absence of information on how the policies targeting these priority areas might evolve and how the overall policy mix might look in the future, we draw our conclusions purely based on the nature of the technologies under each priority area. The nature of the technologies has insights for the likelihood of a proposed RCT in each area being able to meet sampling requirements and/or relate to a policy intervention that would lend itself well to an RCT design.

Experimental design constraints and risks

Policy interventions in the RD&D stage of the path to commercialisation commonly target firms, giving rise to experimental design constraints and risks especially in relation to sampling. As we have previously discussed, in many cases the firms targeted by the UK's innovation policies will be unlikely to yield a large and homogeneous sample that lends itself well to an RCT design. In the RD&D stage of the path to commercialisation, we assume the evaluator wishes to test a discrete, direct intervention such as an R&D grant and we ask: *would the innovation area yield a large, homogeneous sample of innovators?*

Complexity of the green innovation system

On the other hand, many of the innovative technologies included in the NZIP priority areas are large-scale and will be integrated into the energy system 'upstream', decarbonising energy production or industrial processes, rather than being used directly by consumers. This gives rise to particular questions about the feasibility of RCTs for testing policy interventions designed to create a 'pull' for these technologies on the demand side, as these interventions would likely need to be placed on other products and services enabled by the technology, rather than on the technology itself. Such long, indirect causal chains between the policy intervention and technology uptake are less feasible to test under an RCT design. Therefore, on the scale-up and growth stage of the path to commercialisation, we assume a large and homogeneous sample of subjects already exists and ask: *would the demand-side intervention have a direct relationship with the level of uptake of the innovative technology?*

We find that energy storage and flexibility, homes and buildings, and hydrogen may contain fertile grounds for conducting RCT-based investigations on the efficacy of potential policies that might operate in these areas in the future, or on variations of the aspects of innovation policies that currently operate in these areas to inform further iterations or potential extensions of these policies (see Table 4.1). This is primarily because these areas involve some domestic-scale, consumer-facing technologies that are likely to be developed by a large number of innovators, in line with the types of policy areas our interviewees also pointed out as being conducive to RCT-based evidence-building.

The example of homes and buildings is reviewed following Table 4.1 below.

Table 4.1. Illustration of the potential conduciveness of the different NZIP priority areas for RCT-based evidence-building

				RD&D stage	Scale-up and growth stage	Overall
NZIP priority area *NB. Excludes 'Disruptive technologies' due to uncertainty over these types of technologies*	**Objective**	**Scale of the innovative technology**	***Relationship of the innovative technology with the consumer (individuals and/or firms)***	***Assuming randomisation is on a discrete, direct intervention:* would the innovation area yield a large, homogeneous sample of innovators?**	***Assuming a large and homogeneous sample of subjects exists:* would the demand-side intervention have a direct relationship with the level of uptake of the innovative technology?**	***How likely is the area to be conducive to RCTs?***
Advanced CCUS	To bring down the cost of capturing and sequestering CO_2 and helping UK industry to understand the opportunity for developing and deploying next generation carbon capture technologies from 2025	Large-scale technology and infrastructure	Indirect – technology integrated upstream on energy production or industrial processes	Less likely < 500 UK firms specialising in 'industrial decarbonisation and carbon capture' (Curran et al., 2022)* [*Cited firm counts based on analysis of Curran et al. of The Data City mapping of the Net Zero Economy (https://thedatacity.com/)]	Less likely Demand-side interventions would likely target the uptake of products derived using advanced CCUS technologies (e.g. low-carbon manufactured products), indirectly driving their uptake upstream	Less likely
Bioenergy	To bring down costs and reduce barriers within the full biomass to energy value chain. This includes improving the productivity of the UK's biomass supply, the availability of conversion technologies, and the generation processes for energy vectors such as biomethane, green hydrogen, biofuels and electricity	Large-scale technology, infrastructure and processes [focusing on bioenergy production only and excluding technologies for the end use of listed energy vectors, as these are difficult to disentangle from other NZIP areas]	Indirect – technology integrated upstream in energy production processes	Less likely Large-scale infrastructure likely to be led by a small number of major players	Less likely Demand-side interventions would likely target the uptake of products derived from biomass (e.g. biofuels), indirectly driving the uptake of bioenergy production technologies upstream	Less likely

				RD&D stage	Scale-up and growth stage	Overall
Direct air capture (DAC) and greenhouse gas removal (GGR)	To support R&D into DAC technologies in the UK	Large-scale technology and infrastructure	Indirect – technology integrated upstream to offset emissions from energy production or industrial processes	Less likely < 500 UK firms specialising in 'industrial decarbonisation and carbon capture' (Curran et al., 2022)	Less likely Demand-side interventions would likely target the uptake of products and services embedding offsets from DAC technologies (e.g. carbon-neutral airline tickets), indirectly driving their uptake upstream	Less likely
Energy storage and flexibility	To support flexibility services and technologies, as well as non-conventional storage at varying technology readiness levels	Mixed – utility-scale and domestic-scale technologies	Mixed – indirect with regard to utility-scale technologies (integrated upstream), but direct with regard to domestic-scale technologies (consumer-facing)	More likely > 4,500 UK firms specialising in 'energy storage' and > 8,000 UK firms specialising in 'demand side management and digital' (Curran et al., 2022)	Mixed More likely with regard to domestic-scale technologies (e.g. domestic battery storage). Demand-side interventions (e.g. feed-in-tariffs) can be placed on the technology itself to drive uptake Less likely with regard to utility-scale technologies. Demand-side interventions would likely target the uptake of products and services embedding utility-scale storage and flexibility services (e.g. green tariffs), indirectly driving their uptake upstream	Partly
Future offshore wind	To support the development and demonstration of state-of-the-art technologies and products in the future offshore wind industry (including floating offshore wind)	Large-scale technology and infrastructure	Indirect – technology integrated upstream in energy production processes	Less likely Large-scale infrastructure likely to be led by a small number of major players	Less likely Demand-side interventions would likely target the uptake of products and services enabled by offshore wind (e.g. green tariffs), indirectly driving its uptake upstream	Less likely
Homes and buildings	Innovation to support decarbonising our homes and buildings	Domestic-scale technologies	Direct – technologies are consumer-facing	More likely ~2,000 UK firms specialising in 'low-carbon heat and buildings' (Curran et al., 2022)	More likely Demond-side interventions (e.g. heat pump grants) can be placed on the technology itself to drive uptake	More likely

				RD&D stage	Scale-up and growth stage	Overall
Hydrogen	To catalyse innovation and address blockers to the uptake of hydrogen technologies across the whole hydrogen value and supply chain, from production, supply, storage to end use	Mixed – large-scale technology and infrastructure (for hydrogen production) but domestic-scale technologies (at the point of use)	Mixed – indirect with regard to hydrogen production technologies (integrated upstream), but direct with regard to domestic-scale technologies (consumer-facing)	Mixed More likely with regard to manufacturers of hydrogen-ready products Less likely with regard to hydrogen producers (large-scale infrastructure likely to be led by a small number of major players)	Mixed More likely with regard to domestic-scale technologies at the point of use (e.g. hydrogen-ready appliances, hydrogen vehicles). Demand-side interventions (e.g. boiler scrappage scheme) can be placed on the technology itself to drive uptake Less likely with regard to hydrogen production technologies (e.g. electrolysers). Demand-side interventions would likely target the uptake of products and services enabled by hydrogen technologies (e.g. hydrogen-based steel), indirectly driving their uptake upstream	Partly
Industry	To support the development and demonstration of technologies that enable industry to switch from high- to low-carbon fuels and improve energy or resource efficiencies	Large-scale technology and infrastructure	Indirect – technology integrated upstream in industrial processes	Less likely < 500 UK firms specialising in 'industrial decarbonisation and carbon capture' (Curran et al., 2022)	Less likely Demand-side interventions would likely target the uptake of products and services enabled by industrial efficiency and fuel switching technologies (e.g. manufactured low-carbon goods), indirectly driving their uptake upstream	Less likely
Nuclear advanced modular reactors (AMRs)	To demonstrate that high temperature gas reactors (HTGRs) can produce high temperature heat which could be used for low-carbon hydrogen production, process heat for industrial and domestic use and cost-competitive electricity generation	Large-scale technology and infrastructure	Indirect* – technology integrated upstream in energy production or industrial processes [*Industrial producer firms can actually be direct consumers of AMRs but this is likely to be a small pool of firms, at least initially.]	Less likely Large-scale infrastructure likely to be led by a small number of major players	Less likely Demand-side interventions would likely target the uptake of products and services enabled by nuclear AMRs (e.g. AMR-based electricity, manufactured goods produced using AMR-based heat/ electricity/ hydrogen), indirectly driving their uptake upstream	Less likely

Homes and buildings: how can an RCT inform innovation policies?

One of the most pressing challenges facing the UK's transition to net zero is reducing carbon emissions from homes and buildings, which requires a rapid rollout of energy efficiency measures alongside a transition to low-carbon heating. Home heating in the UK is currently dominated by fossil fuels, with 85% of homes supplied directly by the mains gas grid (CCC, 2016). Electrification, through a mass uptake of heat pumps, is one of the main pathways to decarbonise home heating in the UK. The Government has a target to increase the installation of heat pumps from around 35,000 per year today to 600,000 per year by 2028. However, despite their potential to offer significant lifetime savings given their higher efficiencies, heat pumps currently remain more expensive to install than gas boilers. This means that viable financing options will be a critical prerequisite for the mass uptake of heat pumps (Egli et al., 2018; Kokoni and Leach, 2021; Nesta, 2022; Marshall, 2021).

The Government is aiming to lower the cost of heat pumps by directly supporting their development and demonstration through its Heat Pump Ready Programme. This is alongside strong signals on the demand side to promote a market for heat pumps, including an expected ban in the forthcoming Future Homes Standard on new homes from connecting to the gas network from 2025 and a proposed phase-out of new gas boilers from 2035.

The Government is also offering direct financial incentives for domestic and small non-domestic properties wishing to have a heat pump installed through its £450 million Boiler Upgrade Scheme (BUS). The BUS offers grants to homeowners worth £5,000 for an air source heat pump, £5,000 for a biomass boiler, and £6,000 for a ground source heat pump.

Research to date on willingness to pay

The importance of financial support for scaling up low-carbon technologies, including in the form of grants, subsidies, tax incentives or feed-in tariffs, is widely acknowledged in the literature (Curtin et. al, 2017; Castaneda et. al, 2020; de Fuller et al., 2009). However, only a few RCTs have tested the effect of such support on the uptake of certain low-carbon technologies (see Reichert et. al, 2021 and Alcott and Sweeney, 2015). The uptake of heat pumps depends on a variety of complex factors and specific analysis is needed to understand how financial support (on its own or in combination with other mechanisms) would impact uptake. For example, one analysis aiming to understand the factors underlying uptake reveals that the adoption of heat pumps in London to date is positively correlated with average household income and is more likely in detached houses than other building forms (Munjal, 2022).

The Behavioural Insights Team and Nesta (2022) conducted a willingness-to-pay study on heat pumps with a representative sample of UK homeowners, showing that there is an 'early adopter' group of homeowners (25%) who are prepared to spend the present full price for heat pumps (£10–12,000). A bigger potential customer base of roughly one-third of households expressed readiness to spend something extra on a heat pump (compared with the cost of a gas boiler), but not enough to meet the present price of a heat pump. Overall, this analysis implied that a subsidy (or market cost reduction) of £5-7,000 on the current cost of heat pumps – as now offered by the BUS – could unlock installation levels in line with government targets. However, these are likely to be upper bound estimates of potential installation levels as they draw on stated intentions of homeowners from an online experiment.

A role for RCTs

An RCT could evaluate grant-based incentives for heat pumps in practice and investigate what happens when consumers spend real money and do not just fill out online surveys.

While the value of the current BUS grants is already set at £5,000 (or £6,000 for a ground-source heat pump), is this value enough to incentivise uptake? And for whom? What is the exact threshold at which higher monetary values of the grant no longer lead to proportionally higher installations? Answering these questions could inform the Government about how to make best use of taxpayers' money as it works to scale up heat pumps.

Since there would likely be significant fairness and equity issues around randomly allocating a grant that has implications for a household's choice of a heating system and consequently its energy bills, keeping the trial to a 'relatively modest scale' would be sensible (Roumbanis, 2019). The opportunity to conduct a randomised study of the BUS in a safe, convenient and small-scale environment is offered by the Energy Systems Catapult's (ESC) living lab. ESC has already provided numerous case studies on the viability, performance and uptake of heat pumps with real people residing in the over 1,000 homes around the UK (with a variety of tenures, property types and demographics) connected to the living lab. These households have already consented to participate in subsequent studies and trials.

Despite the Government's goal of installing 600,000 heat pumps every year by 2028, the BUS has only received £450 million in funding so far, making only 90,000 homes eligible for the grant over the next three years (BEIS, 2022). Lessons learned from an RCT on the optimal level of a grant to incentivise the uptake of heat pumps in the UK would inform a potential scale-up or extension of the BUS scheme, which will be necessary for the Government to meet its own installation targets.

5. Conclusions

With this Policy Insight we have aimed to initiate a conversation on whether RCTs could be used more widely to build evidence to inform the development of green innovation policies in the UK, and to encourage further debate and collaborative research on this issue. Talking to policymakers around the world to extend the discussions we had with policy professionals in the UK could be an especially interesting starting point for further work.

RCTs are a proven tool that can help develop policies that achieve more effective outcomes per pound spent. When complemented with other methods to reveal the contextual factors underpinning results, they can contribute to an evidence base that can be used as a public good to accelerate net zero-aligned innovation and growth. RCTs' contribution to achieving green outcomes could be maximised if they were used as a way to make innovation policy itself more innovative, shifting the focus from 'evaluating' a single policy design and asking if it works to approaching the policy as an experiment in its entirety – that is, having an open mind, testing promising ideas on a small scale and basing consequent decisions on lessons learned on the ground. Where lessons are shared, RCTs can also create opportunities to learn from the experiences and practice of others across local, regional and national boundaries.

Despite their unique strengths, RCTs are unlikely to be practicable in every context within the green innovation policy landscape. Where they are not appropriate, there are other sources of rigorous evidence – for example, well-designed quasi-experimental evaluations – that policymakers can turn to as they work to develop and improve their policies. Ultimately, RCTs are one important enabler of evidence-based policymaking rather than an end in themselves; the case for using RCTs to inform the development of green innovation policies in the UK (or elsewhere) should be evaluated through this lens.

Policy implications

- Green innovation is needed at scale and quickly to meet climate targets in the UK and globally. While we are learning what works, more evidence on the causal effects of policies designed to achieve specific technological objectives is needed – which RCTs can provide – to ensure public resources are used most efficiently.
- While RCT-based evidence on what works for driving green innovation is as yet limited, there is a lot to be gained from an open-minded approach to recognising and learning from transferable knowledge – whether that be from RCTs in other areas of innovation, sectors or geographical locations, from other rigorous methods of policy evaluation, or from policy failures as well as successes.
- Following best practice in research can ensure that knowledge from future green innovation RCTs can feed into a shared, credible evidence base that can be used as a public good to accelerate net zero-aligned innovation; this includes pre-registration of study methodology, sharing results even if they are 'null', and remembering that what has gone wrong during implementation can be equally as insightful to other researchers as what has gone well.
- An RCT-enabled experimental approach to policy development – identifying different policy alternatives with an open mind, testing promising ones at small scale and identifying those that work based on well-defined intermediate outcomes before scaling them up – might be especially relevant for policies aiming to support the RD&D of green technologies, as the potential of RCTs to generate evidence in this area appears largely unexploited.
- The heterogeneity and relatively small size of the pool of innovators targeted by many of the UK's innovation policies, as well as the long, indirect and complex

relationships between many green technologies and their ultimate users, present challenges to implementing RCT-based evaluations in the context of green innovation in the UK.

- Energy storage and flexibility, homes and buildings, and hydrogen may present fertile grounds for RCT-based evidence-building for the UK's innovation policies. Certain policies designed to drive the supply of or the demand for innovative technologies in these areas could lend themselves well to an RCT design, given many of these technologies are consumer-facing and likely to be developed by a large and sufficiently homogeneous pool of innovators.
- The method of policy evaluation is a highly context-specific choice; RCTs are one of the most rigorous but not the only rigorous method for policy evaluation. Whichever method is chosen, it would be most useful to policymakers if complemented with investigations that can extract rich information on 'why' and 'how' a policy works (or not), in order to generate knowledge that can be used in a wider range of contexts.

References

Acemoglu D, Aghion P, Bursztyn L and Hémous D (2012) The environment and directed technical change. *The American Economic Review*, 102 (1) (2012), pp. 131-166. http://dx.doi=10.1257/aer.102.1.131

Adam D (2019) Science funders gamble on grant lotteries. *Nature* 575, 574–575. https://doi.org/10.1038/d41586-019-03572-7

Aghion P, Hepburn C, Teytelboym A, and Zenghelis D (2014) *Path dependence, Innovation and the Economics of Climate change*. Working paper. Centre for Climate Change Economics and Policy/Grantham Research Institute on Climate Change and the Environment Policy Paper & Contributing Paper to New Climate Economy. https://www.lse.ac.uk/granthaminstitute/wp-content/uploads/2014/11/Aghion_et_al_policy_paper_Nov20141.pdf

Aghion P, Hepburn C, Teytelboym A and Zenghelis D (2019) *Path dependence, innovation and the economics of climate change. Handbook on green growth*. https://doi.org/10.4337/9781788110686.00011

Allcott H and Sweeney R (2015) Can retailers inform consumers about energy costs? Evidence from a field experiment. In *E2e. Working paper 008*.

Athey S and Imbens G W (2017) *The Econometrics of Randomized Experiments: A Toolkit*. 73–140. https://doi.org/10.1016/bs.hefe.2016.10.003

Atkin D, Khandelwal A K and Osman A (2017) Exporting and Firm Performance: Evidence from a Randomized Experiment. *The Quarterly Journal of Economics*, *132*(2), 551–615. https://doi.org/10.1093/qje/qjx002

Bakhshi H, Edwards J S, Roper S, Scully J, Shaw D, Morley L and Rathbone N (2015) Assessing an experimental approach to industrial policy evaluation: Applying RCT+ to the case of Creative Credits. *Research Policy*, 44(8), pp.1462-1472. https://doi.org/10.1016/j.respol.2015.04.004

Behavioural Insight Team and Nesta (2022) *Estimating the Willingness to Pay for a Heat Pump*. https://media.nesta.org.uk/documents/Estimating_the_willingness_to_pay_for_a_heat_pump_v1.pdf

Bloom N, Eifert B, Mahajan A, McKenzie D and Roberts J (2013) Does Management Matter? Evidence from India *. *The Quarterly Journal of Economics*, *128*(1), 1–51. https://doi.org/10.1093/qje/qjs044

Bloom N, Van Reenen J and Williams H (2019) A toolkit of policies to promote innovation. *Journal of Economic Perspectives*, 33(3), pp.163-84. https://cep.lse.ac.uk/pubs/download/dp1634.pdf

Bravo-Biosca A (2020) Experimental innovation policy. *Innovation Policy and the Economy*, 20(1), pp.191-232. https://doi.org/10.1086/705644

Breckon J and Sutherland A (2020) *Beyond Randomised Controlled Trials – expanding the horizon for experimental research techniques in the social sciences.* LSE Impact Blog. https://blogs.lse.ac.uk/impactofsocialsciences/2020/01/21/beyond-randomised-controlled-trials-expanding-the-horizon-for-experimental-research-techniques-in-the-social-sciences/

Cameron D B, Mishra A and Brown A N (2016) The growth of impact evaluation for international development: how much have we learned? *Journal of Development Effectiveness*, 8:1, 1-21. https://doi.org/10.1080/19439342.2015.1034156

Campos F, Coville A, Fernandes A M, Goldstein M, and McKenzie D (2014) Learning from the experiments that never happened: Lessons from trying to conduct randomized evaluations of matching grant programs in Africa. *Journal of the Japanese and International Economies,* 33, pp.4-24. https://doi.org/10.1016/j.jjie.2013.12.007

Carattini S, Levin S and Tavoni A (2020) Cooperation in the climate commons. *Review of Environmental Economics and Policy*.

Carbon Pricing Leadership Coalition (2017) *Report of the High-Level Commission on Carbon Prices*. International Bank for Reconstruction and Development and, International Development Association / The World Bank. https://openknowledge.worldbank.org/handle/10986/32419

Castaneda M, Zapata S, Cherni J, Aristizabal AJ and Dyner (2020) The long-term effects of cautious feed-in tariff reductions on photovoltaic generation in the UK residential sector. *Renewable Energy*, 155, pp.1432-1443. https://doi.org/10.1016/j.renene.2020.04.051

Cavanagh J, Duru M, Gupta S, Kopper S and Welch K (2021) *Evidence from the AEA RCT Registry on new research during Covid-19,* World Bank Blogs, Development Impact guest blog. https://blogs.worldbank.org/impactevaluations/evidence-aea-rct-registry-new-research-during-covid-19-guest-post-jack-cavanagh

Chater N and Loewenstein G (2022) *The i-Frame and the s-Frame: How Focusing on Individual-Level Solutions Has Led Behavioral Public Policy Astray*. https://doi.org/10.2139/ssrn.4046264

Climate Action Tracker (2022) *CAT Net zero target evaluations*. https://climateactiontracker.org/global/cat-net-zero-target-evaluations/

Climate Change Committee [CCC] (2020) *The Sixth Carbon Budget: The UK's path to Net Zero*. https://www.theccc.org.uk/wp-content/uploads/2020/12/TheSixth-Carbon-Budget-The-UKs-path-to-Net-Zero.pdf

Climate Change Committee [CCC] (2022) *Progress in Reducing Emissions: 2022 Report to UK Parliament*. https://www.theccc.org.uk/publication/2022-progress-report-to-parliament/

Committee on Climate Change [CCC] (2016) *Heat in UK Buildings Today, Annex 2*. https://www.theccc.org.uk/wp-content/uploads/2017/01/Annex-2-Heat-in-UK-Buildings-Today-Committee-on-Climate-Change-October-2016.pdf

Cooper A C (2018) Evaluating energy efficiency policy: understanding the 'energy policy epistemology' may explain the lack of demand for randomised controlled trials. *Energy Efficiency,* 11(4): 997-1008. https://doi.org/10.1007/s12053-018-9618-8

Curran B, Martin R, Muller S, Nguyen-Tien V, Oliveira-Cunha J, Serin E, Shah A, Valero A and Verhoeven D (2022) *Growing clean: Identifying and investing in sustainable growth opportunities across the UK*, The Resolution Foundation, Centre for Economic Performance and the Grantham Research Institute on Climate Change and the Environment. https://www.lse.ac.uk/granthaminstitute/wp-content/uploads/2022/05/Growing-clean.pdf

Curtin J, McInerney C and Gallachóir BÓ (2017) Financial incentives to mobilise local citizens as investors in low-carbon technologies: A systematic literature review. *Renewable and Sustainable Energy Reviews*, 75: 534-547. https://doi.org/10.1016/j.rser.2016.11.020

Dalziel M (2018) *Why are there (almost) no randomised controlled trial-based evaluations of business support programmes?* Palgrave Communications 4. https://doi.org/10.1057/s41599-018-0069-9

Day T E (2015) The big consequences of small biases: A simulation of peer review. *Research Policy 44*, 6: 1266-1270. https://doi.org/10.1016/j.respol.2015.01.006

Deaton A and Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210: 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Dechezleprêtre A, Martin R, and Mohnen M (2014) *Knowledge Spillovers from Clean and Dirty Technologies. CEP Discussion Papers (CEPDP1300)*, Centre for Economic Performance. https://cep.lse.ac.uk/_new/publications/abstract.asp?index=4503

DellaVigna S and Linos E (2022) RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica*, *90*(1): 81-116. https://doi.org/10.3982/ECTA18709

Department for Business, Energy & Industrial Strategy [BEIS] (2018) *Business Basics Programme: overview and objectives*. https://www.gov.uk/government/publications/business-basics-programme-overview-and-objectives

Department for Business, Energy & Industrial Strategy [BEIS] (2020) *BEIS monitoring and evaluation framework*. https://www.gov.uk/government/publications/beis-monitoring-and-evaluation-framework

Department for Business, Energy & Industrial Strategy [BEIS] (2021) *UK Innovation Strategy: leading the future by creating it*. https://www.gov.uk/government/publications/uk-innovation-strategy-leading-the-future-by-creating-it

Department for Business, Energy & Industrial Strategy [BEIS] (2022) *Future Support for Low Carbon Heat: Boiler Upgrade Scheme (BUS) – Impact Assessment*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1055336/boiler-upgrade-scheme-final-impact-assessment.pdf

Department of Energy & Climate Change [DECC] (2013) *Removing the hassle factor associated with loft insulation: Results of a behavioural trial*. HM Government, London. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/236858/DECC_loft_clearance_trial_report_final.pdf

Department of Energy & Climate Change [DECC] (2014a) *Advice on how to use heating controls: Evaluation of a trial in Newcastle*. HM Government, London. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/349855/decc_heating_controls_report.pdf

Department of Energy & Climate Change [DECC] (2014b) *Evaluation of the DECC/John Lewis energy labelling trial*. HM Government, London. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/350282/John_Lewis_trial_report_010914FINAL.pdf

De Lyon J, Martin R, Oliveira-Cunha J, Shah A, Shah K, Thwaites G and Valero A (2022) *Enduring strengths: Analysing the UK's current and potential economic strengths, and what they mean for its economic strategy, at the start of the decisive decade*. The Economy 2030 Inquiry. https://economy2030.resolutionfoundation.org/reports/enduring-strengths/

Duflo E, Glennerster R, and Kremer M (2007) *Chapter 61 Using Randomization in Development Economics Research: A Toolkit* (pp. 3895–3962). https://doi.org/10.1016/S1573-4471(07)04061-2

Economics of Energy Innovation and System Transition (EEIST) Consortium (2022) *Ten Principles for Policy Making in the Energy Transition: Lessons from Experience*. https://eeist.co.uk/download/932/

Egli F, Steffen B and Schmidt TS (2018) A dynamic analysis of financing conditions for renewable energy technologies. *Nature Energy*, 3(12): 1084-1092. https://doi.org/10.1038/s41560-018-0277-y

Ekins P and Zenghelis D (2021) The costs and benefits of environmental sustainability. *Sustainability Science*, 16(3): 949-965. https://doi.org/10.1007/s11625-021-00910-5

Ettelt S, Mays N, Allen, P (2015) Policy experiments: investigating effectiveness or confirming direction? *Evaluation*, 21(3): 292-307. https://doi.org/10.1177/1356389015590737

Farrell N, Hepburn C, Pless, J (2020) Bringing rigour to energy innovation policy evaluation. *Nature Energy*. https://doi.org/10.1038/s41560-020-0557-1

Firpo T (2017) *Barriers to experimentation: survey results*. Nesta. https://www.nesta.org.uk/blog/barriers-to-experimentation-survey-results/

Firpo T and Phipps J (2019) New Running experiments in innovation and growth policy: What can we learn from recent experience? *Journal for Research and Technology Policy Evaluation*, 47: 46-50. 10.22163/fteval.2019.328. http://repository.fteval.at/id/eprint/414

Flory J, John L and Arndt R (2021) *Argentina Solar Home System and Tariff Impact Evaluation*. AEA RCT Registry. August 18. https://doi.org/10.1257/rct.971 https://www.socialscienceregistry.org/trials/971

Fowlie, M, Greenstone M and Wolfram C (2018) Do energy efficiency investments deliver? Evidence from the weatherization assistance program. *The Quarterly Journal of Economics*, 133(3): 1597-1644. https://doi.org/10.1093/qje/qjy005

Frederiks E R, Stenner K, Hobman E V and Fischle M (2016) Evaluating energy behavior change programs using randomized controlled trials: Best practice guidelines for policymakers. *Energy Research & Social Science*, 22: 147-164. https://doi.org/10.1016/j.erss.2016.08.020

Fuller M C, Portis S C and Kammen D M (2009) Toward a low-carbon economy: municipal financing for energy efficiency and solar power. *Environment: Science and Policy for Sustainable Development*, *51*(1): 22-33. https://doi.org/10.3200/ENVT.51.1.22-33

Gillingham K and Palmer K (2014) Bridging the Energy Efficiency Gap: Policy Insights from Economic Theory and Empirical Evidence. *Review of Environmental Economics and Policy,* 8(1): 18–38. https://doi.org/10.1093/reep/ret021

Gosnell G K, List J A and Metcalfe R D (2020) The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy,* 128(4): 1195-1233. https://doi.org/10.1086/705375

Green Alliance and Nesta (2022) *Climate for growth-productivity, net zero and the cost of living*. https://green-alliance.org.uk/wp-content/uploads/2022/05/Climate-for-growth.pdf.

Grubler A, Aguayo F, Gallagher K, Hekkert M P, Jiang K, Mytelka L, Neij L, Nemet G, Wilson C, Johansson T B and Nakicenovic N (2012) *Policies for the energy technology innovation system (ETIS)*. Cambridge University Press.

Hahn R, Metcalfe R, and Rundhammer F (2017) *Power to the people: A new trend in regulation*. Working Paper, Brookings Institution. https://techpolicyinstitute.org/wp-content/uploads/2017/06/Power-to-the-People_A-New-Trend-in-Regulation.pdf.

Hall B H and Lerner J (2010) The financing of R&D and innovation. In *Handbook of the Economics of Innovation Vol. 1*. North-Holland. https://doi.org/10.1016/S0169-7218(10)01014-2

Haynes L C, Goldacre B and Torgerson D J (2012) Test, learn, adapt: developing public policy with randomised controlled trials| cabinet office. Behavioural Insights Team. https://ssrn.com/abstract=2131581 or http://dx.doi.org/10.2139/ssrn.2131581

Haynes L C, Green D P, Gallagher R, John P and Torgerson D J (2013) Collection of delinquent fines: An adaptive randomized trial to assess the effectiveness of alternative text messages. *Journal of Policy Analysis and Management,* 32(4): 718-730. https://doi.org/10.1002/pam.21717

Heiskanen E, Happonen J, Matschoss K and Mikkonen I (2022) Learning from failures - Encouraging lesson-sharing in the Finnish energy transition. *Energy Research & Social Science*, 90, 102676. https://doi.org/10.1016/j.erss.2022.102676

Hepburn C, Pless J and Popp D (2020) Policy brief—encouraging innovation that protects environmental systems: five policy proposals. *Review of Environmental Economics and Policy*.

HM Government (2020) *Energy White Paper: Powering our net zero future*. https://www.gov.uk/government/publications/energy-white-paper-powering-our-net-zero-future

HM Government (2021a) *Net Zero Strategy: Build Back Greener*. https://www.gov.uk/government/publications/net-zero-strategy

HM Government (2021b) *UK Net Zero Research and Innovation Framework*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1030656/uk-net-zero-research-innovation-framework.pdf

HM Treasury (2020a) *The Magenta Book: Central Government guidance on evaluation*. https://www.gov.uk/government/publications/the-magenta-book

HM Treasury (2020b) *The Magenta Book: Supplementary Guide: Handling Complexity in Policy Evaluation*. https://www.gov.uk/government/publications/the-magenta-book

HM Treasury (2022) *The Green Book: appraisal and evaluation in central government*. https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent/the-green-book-2020

House of Commons [HoC] Health Committee (2009) *Health Inequalities. Third Report of Session 2008–09*, vol. 1, House of Commons. https://publications.parliament.uk/pa/cm200809/cmselect/cmhealth/286/286.pdf

International Energy Agency [IEA] (2021) *Net Zero by 2050*. https://iea.blob.core.windows.net/assets/deebef5d-0c34-4539-9d0c-10b13d840027/NetZeroby2050-ARoadmapfortheGlobalEnergySector_CORR.pdf

International Energy Agency [IEA] (2022) Policies Database (filtered by Technology R&D and innovation policies). https://www.iea.org/policies?topic%5B0%5D=Technology%20R%26D%20and%20innovation

Intergovernmental Panel on Climate Change [IPCC] (2022) *Climate Change 2022: Mitigation of Climate Change*. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Shukla P R, Skea J, Slade R, Al Khourdajie A, van Diemen R, McCollum D, Pathak M, Some S, Vyas P, Fradera R, Belkacemi M, Hasija A, Lisboa G, Luz S, Malley J, (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA. doi: 10.1017/9781009157926

Kleine M, Heite J, and Huber L R (2022) Subsidized R&D collaboration: The causal effect of innovation vouchers on innovation outcomes. *Research Policy*, 51(6):104515. https://doi.org/10.1016/j.respol.2022.104515

Kokoni S, and Leach M (2021) Policy mechanisms to support heat pump deployment: A UK case study based on techno-economic modelling. *Renewable and Sustainable Energy Transition*, *1*, 100009. https://doi.org/10.1016/j.rset.2021.100009

Lee C J (2015) Commensuration Bias in Peer Review. *Philosophy of Science,* 82: 1272-1283. https://doi.org/10.1086/683652

Madaleno M and Waights S (2015) *Guide to scoring methods using the Maryland Scientific Methods Scale*. What Works Centre for Local Economic Growth. https://whatworksgrowth.org/public/files/Scoring-Guide.pdf

Meager R (2019) Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments. *American Economic Journal: Applied Economics,* 11(1): 57–91. https://doi.org/10.1257/app.20170299

Marcellesi A and Cartwright N (2013) *Modelling climate mitigation and adaptation policies to predict their effectiveness: The limits of randomised controlled trials*. Grantham Research Institute on Climate Change and the Environment. https://doi.org/10.1007/978-3-319-65058-6_15

Marshall J (2021) T*he Economy 2030 Enquiry*. *Home is where the heat (pump) is*. https://economy2030.resolutionfoundation.org/wp-content/uploads/2021/10/Home-is-where-the-heat-pump-is.pdf.

Martin R, Unsworth S, Valero A and Verhoeven D (2020) *Innovation for a strong and sustainable recovery*. Centre for Economic Performance, London School of Economics and Political Science. https://cep.lse.ac.uk/pubs/download/cepcovid-19-014.pdf

Martin S and Scott J T (2000) The nature of innovation market failure and the design of public support for private innovation. *Research Policy*, *29*(4-5): 437-447. https://doi.org/10.1016/S0048-7333(99)00084-0.

Masset E, Rathinam F, Nath M, Wood B (2018) *Successful Impact Evaluations: Lessons from DFID and 3ie. CEDIL Inception Paper*. https://cedilprogramme.org/wp-content/uploads/2019/08/Successful-evaluations-FINAL-004.pdf

Mayo N, Brophy J, Goldberg M, Klein M, Miller S, Platt R and Ritchie J (2006) Peering at peer review revealed high degree of chance associated with funding of grant applications. *Journal of Clinical Epidemiology* 59: 842–848. https://doi.org/10.1016/j.jclinepi.2005.12.007

McKenzie D (2010) Impact Assessments in Finance and Private Sector Development: What have we learned and what should we learn? *The World Bank Research Observer*, 25(2): 209-233. https://doi.org/10.1093/wbro/lkp011

Morey L D (2022) *Randomised controlled trials in green innovation policy: What are the barriers to the introduction of rigorous methods to this field?* Submission to the Department of Geography and Environment to fulfil the requirements for the MSc in Environmental Policy and Regulation at the London School of Economics and Political Science. (unpublished)

Munjal A (2022) *What are the factors that impact the adoption of heat pumps in the UK?* Submission to the Department of Geography and Environment to fulfil the requirements for the MSc in Environmental Economics and Climate Change at the London School of Economics and Political Science. (unpublished)

National Audit Office [NAO] (2013) *Evaluation in Government*. National Audit Office. https://www.nao.org.uk/wp-content/uploads/2013/12/10331-001-Evaluation-in-government_NEW.pdf

Nesta (2022) *How to reduce the cost of heat pumps*. https://media.nesta.org.uk/documents/How_to_reduce_the_cost_of_heat_pumps_v4_1.pdf.

Nisa C F, Bélanger J J, Schumpe B M and Faller D G (2019) Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. *Nature Communications* 10(1): 1-13. https://doi.org/10.1038/s41467-019-12457-2

Norton B J and Strube M J (2001) Understanding Statistical Power. *Journal of Orthopaedic & Sports Physical Therapy*, 31(6), 307–315. https://doi.org/10.2519/jospt.2001.31.6.307

Observatory of Public Sector Innovation (2022) *Behavioural Insights Units*. https://oecd-opsi.org/bi-units/.

OECD and the World Bank (2014) *Making Innovation Policy Work: Learning from Experimentation,* OECD Publishing. https://dx.doi.org/10.1787/9789264185739-en

Oliveira-Cunha J, Kozler J, Shah P, Thwaites G and Valero A (2021) *Business time: How ready are UK firms for the decisive decade?* The Economy 2030 Inquiry. https://economy2030.resolutionfoundation.org/reports/business-time/

Pawson R and Tilley N (1997) An introduction to scientific realist evaluation. In E. Chelimsky & W. R. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 405–418). Sage Publications, Inc. https://doi.org/10.4135/9781483348896.n29

Phipps J (2017) *Innovation Growth Lab. Taking the first steps in business policy experimentation*. https://innovationgrowthlab.org/blog/taking-first-steps-business-policy-experimentation.

Pier E L, Brauer, M, Filut A, Kaatz A, Raclaw J, Nathan M J, Ford C E, Carnes M (2018) Low agreement among reviewers evaluating the same NIH grant applications. *Proc Natl Acad Sci U S A* 115: 2952–2957. https://doi.org/10.1073/pnas.1714379115

Pless J, Hepburn C and Farrell N (2020) Bringing rigour to energy innovation policy evaluation. *Nature Energy*, 5(4): 284-290. https://doi.org/10.1038/s41560-020-0557-1

Popp D (2019) *Environmental Policy and Innovation: A Decade of Research*. https://doi.org/10.3386/w25631

Priceonomics (2015) *What Happened When Nigeria Created the World's Largest Business Plan Competition?* Priceonomics. https://priceonomics.com/what-happened-when-nigeria-created-the-worlds/

Pritchett L, Samji S, Hammer J (2013) It's All About MeE: Using Structured Experiential Learning ('e') to Crawl the Design Space. *SSRN Electronic Journal* 322. https://doi.org/10.2139/ssrn.2248785

Protocol (2022) *Weatherization is the low-tech climate solution we need, and the White House wants to help*. https://www.protocol.com/bulletins/weatherization-assistance-program-influx

Riom C, Valero A and Oliveira-Cunha J (2022) *Lessons on designing tech RCTs with SMEs*. *Innovation Growth Lab*. https://innovationgrowthlab.org/blog/lessons-designing-tech-rcts-smes

Rogers P (2014) *Overview of Impact Evaluation: Methodological Briefs-Impact Evaluation No. 1* (No. innpub746). https://ideas.repec.org/p/ucf/metbri/innpub746.html

Roumbanis L (2019) Peer Review or Lottery? A Critical Analysis of Two Different Forms of Decision-making Mechanisms for Allocation of Research Grants. *Science, Technology, & Human Values* 44, 994–1019. https://doi.org/10.1177/0162243918822744

Rydge J, Martin R and Valero A (2018) *Sustainable Growth in the UK: Seizing opportunities from technological change and the transition to a low-carbon economy*. Grantham Research Institute on Climate Change and the Environment and the Centre for Economic Performance, LSE.
https://www.lse.ac.uk/granthaminstitute/publication/sustainablegrowth/

Sabet S, Heard A C, Neilitz S and Brown A N (2017) Assessing the evidence base on science, technology, innovation and partnerships for accelerating development outcomes in low- and middleincome countries. *New Delhi, India: International Initiative for Impact Evaluation*. https://www.3ieimpact.org/sites/default/files/2019-01/sp6-stip_0.pdf

Seto K C, Davis S J, Mitchell R B, Stokes E C, Unruh G, and Ürge-Vorsatz D (2016) Carbon lock-in: types, causes, and policy implications. *Annual Review of Environment and Resources*, 41(1): 425-452. 10.1146/annurev-environ-110615-085934

Shafik M (2017) In experts we trust?. Bank of England. https://www.bis.org/review/r170223h.pdf

Sovacool B K, Daniels C and AbdulRafiu A (2022) Science for whom? Examining the data quality, themes, and trends in 30 years of public funding for global climate change and energy research. *Energy Research & Social Science*, 89: 102645. https://doi.org/10.1016/j.erss.2022.102645

Stern N and Valero A (2021) Innovation, growth and the transition to net-zero emissions. *Research Policy,* 50(9): 104293. https://doi.org/10.1016/j.respol.2021.104293

Stern N, Unsworth S, Valero A, Zenghelis D, Rydge J and Robins N (2020) *Strategy, Investment and Policy for a Strong and Sustainable Recovery: An Action Plan*. Centre for Economic Performance: CEP COVID-19 Analysis. https://www.lse.ac.uk/granthaminstitute/publication/strategy-investment-and-policy-for-a-strong-and-sustainable-recovery-an-action-plan/

US Energy Department (2021) *Weatherization Assistance Program*. https://www.energy.gov/sites/default/files/2021/01/f82/WAP-fact-sheet_2021_0.pdf

Valters C and Whitty B (2017) *The politics of the results agenda in DFID: 1997–2017*. London: Overseas Development Institute. https://cdn.odi.org/media/documents/11730.pdf

Vine E, Sullivan M, Lutzenhiser L, Blumstein C, and Miller B (2014) Experimentation and the evaluation of energy efficiency programs. *Energy Efficiency* 7, 627–640. https://doi.org/10.1007/s12053-013-9244-4

Vine E, Sullivan M, Lutzenhiser L, Blumstein C and Miller B (2018) Response to: "Evaluating energy efficiency policy: understanding the 'energy policy epistemology' may explain the lack of demand for randomized controlled trials," by Adam Cooper, Energy Efficiency, published online 26 January 2018. *Energy Efficiency,* 11(8):2179-2180. https://doi.org/10.1007/s12053-018-9650-8

White H (2019) *The twenty-first century experimenting society: the four waves of the evidence revolution*. Palgrave Communications, 5(1), 47. https://doi.org/10.1057/s41599-019-0253-6

White H and Sabarwal S (2014) What Works. *Quasi-experimental design and methods. Methodological briefs: impact evaluation*, 8(2014): 1-16.

Wikkelsø A J, Edwards H M, Afshari A, Stensballe J, Langhoff-Roos J, Albrechtsen C, Ekelund K, Hanke G, Secher E L, Sharif H F and Pedersen L M (2015) Pre-emptive treatment with fibrinogen concentrate for postpartum haemorrhage: randomized controlled trial. *British Journal of Anaesthesia*, 114(4): 623-633. https://doi.org/10.1093/bja/aeu444

Appendix

Table A1. RCTs relating to the scale-up of green technologies

#	Title	Researchers	Primary affiliations of researchers	Trial location	Research objective[10]	Type of intervention tested* and type of outcome data measured *[*See Table 2.1]*	Registration date/ Stated trial period/ Stated trial status	Post-trial information / analysis published?
1	A survey experimental study for stimulating individual action to combat climate change	Jinwook Shin, Syngjoo Choi, Booyuel Kim, Sungmin Lim, Heerae Lee	University	South Korea	To examine the effect of different information treatments on the acceptance of 'green' electricity tariffs	Information, networks and collaboration Stated preferences	12 Oct 2021 (pre-trial)/ 12-22 Oct 2021/ Completed	No[11]
2	Are Consumers Poorly-Informed about Fuel Economy?	Hunt Allcott, Christopher Knittel	University	US	To test the effect of the provision of fuel economy information to new vehicle shoppers on the average fuel economy of vehicles purchased	Information, networks and collaboration Revealed outcomes	4 Sept 2016 (post-trial)/ 1 Dec 2012-31 July 2016/ Completed	Yes (available here)

10 Many studies presented here have broadly defined and/or multiple objectives. The 'research objective' presented here may be a selected segment of the overall objective of a given study which is found to be most relevant for the purposes of this review. It may not be representative of the full scope of the given study.

11 All reasonable effort has been made to scan sources where this information could possibly be published. Nevertheless, omissions may have been made; any such errors are the authors' alone. This applies to the whole table.

3	Are experienced people affected by a pre-set default option – Results from a field experiment	Åsa Löfgren, Peter Martinsson, Magnus Hennlock, Thomas Sterner	University, Research Institute	Sweden	To investigate whether pre-set default options affect the uptake of carbon offsetting among people with existing knowledge on the subject (environmental economists)	**Information, networks and collaboration**[12] **Revealed outcomes**	Unspecified/ 25–28 June 2008/ Completed	Yes (available **here**)
4	Argentina Solar Home System and Tariff Impact Evaluation	Arndt Reichert, Jeffrey Flory, John List	University (relates to a World Bank project)	Argenti na	To test the effects of varying tariff subsidy levels on households' willingness to pay for solar home systems, how long it takes them to reach their "true valuation" of electricity access, and what factors affect the rate at which this happens	**Financial incentives** **Stated preferences**	3 Dec 2015 (during trial)/ 15 Mar 2015-31 Dec 2019/ Completed	No
5	Barriers to Consumer Choice of Energy Efficient Products	C. Dennis Anderson, John D. Claxton	University	Canada	To assess the impact of energy labels and energy information emphasis by sales staff on refrigerator purchasing decisions	**Information, networks and collaboration** **Revealed outcomes**	Unspecified/ Unspecified (pre-1982)/ Completed	Yes (available **here**)
6	Bridging the Energy Efficiency Gap: A Field Experiment on Lifetime Energy Costs and	Steffen Kallbekken, Håkon Sælen, Erlend Hermansen	Research Centre	Norway	To examine how information on lifetime energy cost (provided through labelling and sales staff at retail store) affects consumers' appliance choices	**Information, networks and collaboration** **Revealed outcomes**	Unspecified/ 2011/ Completed	Yes (available **here**)

[12] Defaults do not directly fall into any policy intervention type we had defined earlier. This type appears to be the closest fit given defaults need to be tied with the provision of appropriate information.

	Household Appliances				(fridge-freezers and tumble driers)			
7	Can product demonstrations create markets for sustainable energy technology? A randomized controlled trial in rural India	Johannes Urpelainen, Semee Yoon	University	India	To examine the effect of village demonstrations of off-grid solar technology on sales, awareness and perceptions of solar technology	**Information, networks and collaboration** **Revealed outcomes**	Unspecified/ 2014/ Completed	Yes (available **here**)
8	Clean Development Mechanism	Raunak Kalra, Nicholas Ryan	Research Centre, University	India	To test whether industrial energy audits affect technology choices among small and medium sized energy-intensive industrial plants	**Information, networks and collaboration** **Measured outcome variable unspecified**	5 Mar 2015 (post-trial)/ 15 June 2010-23 May 2014/ Completed	No
9	Consumer-Driven Virtual Power Plants: A Field Experiment on the Adoption and Use of a Prosocial Technology	Quentin Coutellier, Greer Gosnell, Zeynep Gürgüç, Ralf Martin and Mirabelle Muûls	University	UK	To test whether distributing monetary vouchers via a lottery system based on usage level could overcome reservations around switch-off events related to smart plugs and encourage their adoption	**Financial incentives** **Revealed outcomes**	26 Sept 2017 (pre-trial)/ 2019-2020/ Completed	Yes (available **here**)
10	Default Effects and Follow-On Behavior: Evidence	Meredith Fowlie, Catherine Wolfram, C.	University, Government -run laboratory	2012-2014, US	To study how default effects impact enrolment into a time-based electricity pricing plan	**Information, networks and collaboration**[13] **Revealed outcomes**	Unspecified/ 2012-2014/ Completed	Yes (available **here**)

[13] As above, defaults do not directly fall into any policy intervention type we had defined earlier. This type appears to be the closest fit given defaults need to be tied with the provision of appropriate information.

	from an Electricity Pricing Program	Anna Spurlock, Annika Todd, Patrick Baylis, Peter Cappers						
11	Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program	Meredith Fowlie, Michael Greenstone, Catherine Wolfram	University, Research Centre	2011-2012, US	To investigate the effect of encouragement (e.g. home visits, robo-calls, personal calls) and application assistance in the take-up of energy efficiency support through the Weatherization Assistance Program[14]	**Financial incentives** **Revealed outcomes**	27 Jan 2015 (post-trial)/ 1 Jan 2010-30 Nov 2014/ Completed	Yes (available **here**)
12	Equilibrium Effects of Competition on Solar Photovoltaic Demand and Pricing	Kenneth Gillingham, Stefan Lamp, Bryan Bollinger	University	US	To test how increased competition in local market (by varying the number of competitors allowed to operate) affects the adoption of residential solar PV	**Capacity building** **Revealed outcomes**	Unspecified/ 2012-2015/ Completed	Yes (available **here**)
13	Findings from a behavioural trial conducted with John Lewis	Department of Energy and Climate Change, Cabinet Office Behavioural	Government (partnered with private sector)	2013-2014, UK	To examine whether providing information on the electricity lifetime running costs at the point of sale changed purchasing behaviour towards appliances with	**Information, networks and collaboration** **Revealed outcomes**	August 2013 (pre-trial i.e. ONS review of trial design)/ Sept 2013-June 2014/	Yes (available **here**)

[14] Note that this is an encouragement design, meaning the randomisation is on the efforts to encourage households to participate in the programme. The control group is equally able to access the programme but just does not receive active encouragement.

		Insights Team (individual researchers unspecified)			lower energy consumption		Completed	
14	How Do Managers' Beliefs about New Technologies Evolve? Informational Interventions and the Adoption of Energy-Efficient Stitching Motors in Bangladesh	Eric Verhoogen, Ritam Chaurey, Yunfan Gu, Gaurav Nayyar, Siddharth Sharma	University	Bangladesh	To understand the role of information as well as free installation in the adoption of energy-efficient motors for stitching machines	**Information, networks and collaboration** **Stated preferences**	19 Apr 2021 (pre-trial)/ 20 May 2021-20 May 2023/ In development	No
15	How experience of use influences mass-market drivers' willingness to consider a battery electric vehicle: A randomised controlled trial	Stephen M. Skippon, Neale Kinnear, Louise Lloyd, Jenny Stannard	Private consultancy (funded by an international oil company)	UK	To investigate how direct experience of use affects mass-market consumer drivers' willingness to adopt battery electric vehicles (BEVs)	**Information, networks and collaboration** **Stated preferences**	Unspecified/ Unspecified (pre-2016)/ Completed	Yes (available **here**)
16	Information Disclosure, Incentives, and Energy Costs in the United States	Hunt Alcott, Richard Sweeney	University	US	To test the impact of rebates for customers (as well as of information disclosure and sales incentives for agents) on the sale of energy efficient models of natural gas water heaters	**Financial incentives; and information, networks and collaboration** **Revealed outcomes**	Unspecified/ 2012-2013/ Completed	Yes (available **here**)

17	Market failures and willingness to accept the smart energy transition: Experimental evidence from the UK	Greer Gosnell, Daire McCoy	University	UK	To elicit households' willingness to accept smart meter installations in response to different information treatments (regarding expected personal and social benefits) and potential subsidy values	**Information, networks and collaboration; and financial incentives** **Stated preferences**	2 May 2019 (pre-trial)/ 2019/ Completed	Yes (available **here**)
18	Measuring the Welfare Effects of Residential Energy Efficiency Programs	Hunt Allcott, Michael Greenstone	University	US	To examine the informational, behavioural, and monetary drivers (tested via audit subsidies and cash reward) behind the take-up of residential energy efficiency audits and subsequent investments	**Information, networks and collaboration; and financial incentives** **Revealed outcomes**	22 Aug 2017 (post-trial)/ 1 June 2012-30 Sept 2013/ Completed	Yes (available **here**)
19	Messaging and Low-income Solar Adoption	Kenneth Gillingham, Steven Sexton, Bryan Bollinger	University	2017-2018, US	To examine the effect of information provision (pro-social messaging or financial-based messaging) on the adoption of residential solar panels by low and moderate income households	**Information, networks and collaboration** **Revealed outcomes**	15 May 2017 (pre-trial)/ 15 May 2017-15 July 2018/ In development	No

20	Nudge me if you can - how defaults and attitude strength interact to change behavior	Max Vetter, Florian Kutzner	University	Germany	To test how defaults affect consumer choices of a green or gray electricity provider	**Information, networks and collaboration**[15] **Stated preferences**	Specified as pre-registered (no date given)/ Pre-2014/ Completed	Yes (available **here**)
21	Peer effects, Rewards, and Image Concerns in Energy Decision (PRICED)	Greer Gosnell, Stefano Carattini, Alessandro Tavoni	University	UK	To examine whether early adopters of 100% renewable energy tariffs are motivated to display their climate-friendly behaviour	**Information, networks and collaboration** **Revealed outcomes**	27 Feb 2019 (pre-trial)/ 27 Feb 2019-30 Jun 2020/ In development (stated status outdated	Yes (available **here**)
22	Powering Small Retailers: The Adoption of Solar Energy under Different Pricing Schemes in Kenya	Tavneet Suri, William Jack	University	Kenya	To test how price and payment method affect the adoption of off-grid solar power among small retailers	**Financial incentives** **Revealed outcomes**	March 2013 (during trial, available **here**)/ 2013-?/ Unclear	No
23	Social Learning and Solar Photovoltaic Adoption	Kenneth Gillingham, Bryan Bollinger	University	US	To investigate the effect of social learning and peer interactions (through a group pricing offer and informational campaign) on the adoption of residential solar PV systems	**Information, networks and collaboration** **Revealed outcomes**	Unspecified/ Unspecified (pre-2015)/ Completed	Yes (available **here**)

[15] As above, defaults do not directly fall into any policy intervention type we had defined earlier. This type appears to be the closest fit given defaults need to be tied with the provision of appropriate information.

24	Solar Adoption, Local Initiatives, Exchanges among Neighbors, and Conspicuous Environmentalism (SALIENCE)	Stefano Carattini, Erez Yoeli, Kenneth Gillingham	University	US	To examine the effect of information campaigns (emphasising either community benefits, individual benefits or both) on the adoption of solar panels through peer-to-peer system	**Information, networks and collaboration** **Revealed outcomes**	28 Sept 2018 (pre-trial)/ 28 Sept 2018-30 June 2020/ In development	No
25	The Effectiveness and Persistency of Information Disclosure for Adopting Climate-Friendly Goods	Yu Gao	-	China	To measure the short- and long-term effects of an informational campaign (relating to monetary and climate benefits) on the purchase of climate-friendly goods (low-carbon lightbulbs)	**Information, networks and collaboration** **Stated preferences**	7 Jan 2021 (pre-trial)/ 8 Jan 2021-8 Feb 2021/ In development	No
26	The Effect of Messaging on Electric Vehicle Purchases	Kenneth Gillingham, Stephanie Weber	University	US	To analyse the effect of email messages with different emphases (economic or environmental) on the decision to purchase an electric vehicle, including through the random allocation of an alert to a $10,000 rebate. (This is a phase-in design, so the randomisation is on the timing of the alert to the rebate rather than the rebate itself. The control group is also informed of the rebate after the experiment.)	**Information, networks and collaboration; and financial incentives** **Revealed outcomes**	31 Aug 2017 (during trial)/ 31 May 2017-30 Sept 2017/ In development	No

27	The Effects of Energy Cost Labelling on Appliance Purchasing Decisions: Trial Results from Ireland	James Carroll, Eleanor Denny, Sean Lyons	University	Ireland	To examine the effects of lifetime energy cost labelling on appliance purchasing decisions (i.e. tumble dryers with different efficiencies)	**Information, networks and collaboration** **Revealed outcomes**	Unspecified/ Aug-Oct 2013/ Completed	Yes (available **here**)