



INNOVATION &  
RESEARCH  
CAUCUS

# REBUILDING THE ASSESSMENT SYSTEM: AN ACCOUNT FROM INSIDE INNOVATE UK

IRC Report No: 086

## REPORT PREPARED BY

**Ammon Salter**

Warwick Business School, University of Warwick

**Rossella Salandra**

School of Management, University of Bath



Delivered with  
ESRC and  
Innovate UK

## Authors

The core members of the research team for this project were as follows:

- » Professor Ammon Salter - Warwick Business School, University of Warwick
- » Dr Rossella Salandra - School of Management, University of Bath

This document relates to IRC Project FFEE004: Testing Innovate UK's new assessment questions

## Acknowledgements

This work was supported by Economic and Social Research Council (ESRC) grant ES/X010759/1 to the Innovation and Research Caucus (IRC).

We are very grateful to the project sponsors at UK Research & Innovation (UKRI) for their input into this research. The interpretations and opinions within this report are those of the authors and may not reflect the policy positions of UKRI or its constituent councils. We are indebted to James Currer at Innovate UK, our key point of contact throughout, who commissioned and championed this work across both phases.

We would also like to acknowledge and appreciate the efforts of the IRC Project Administration Team for their support in preparing this report. Finally, we thank Nadia Maamoun for her contribution to Phase 2 of the research, in particular the assessor re-assessment interviews.

## AI declaration

Parts of this report were drafted with the assistance of AI (Claude, Anthropic). AI was used to transcribe dictated text and to edit drafts for clarity. All analysis, judgements and conclusions are the authors' own; all AI-generated material was reviewed and verified by the authors.

## About the Innovation and Research Caucus

The Innovation and Research Caucus supports the use of robust evidence and insights in UKRI's strategies and investments, as well as undertaking a co-produced programme of research. Our members are leading academics from across the social sciences, other disciplines and sectors, who are engaged in different aspects of innovation and research systems. We connect academic experts, UKRI, IUK and the (ESRC), by providing research insights to inform policy and practice. Professor Tim Vorley and Professor Stephen Roper are Co-Directors. The IRC is funded by UKRI via the ESRC and IUK, grant number ES/X010759/1. The support of the funders is acknowledged. The views expressed in this piece are those of the authors and do not necessarily represent those of the funders.

## Contact

You are also welcome to email us if you have any questions about this report or the work of the IRC generally: [info@ircaucus.ac.uk](mailto:info@ircaucus.ac.uk)

Cite as: Salter, A. & Salandra, R. 2026. *Rebuilding the Assessment System: An Account from Inside Innovate UK*. Oxford, UK: Innovation and Research Caucus

## CONTENTS

|                                                                             |    |
|-----------------------------------------------------------------------------|----|
| EXECUTIVE SUMMARY .....                                                     | 4  |
| 1. Introduction: Innovation Policy at the Street Level.....                 | 5  |
| 2. Choosing What to Fund: The Design of R&D Assessment.....                 | 7  |
| 3. IUK in Brief .....                                                       | 8  |
| 4. Timeline: Touchpoints in a Multi-Year Relationship .....                 | 9  |
| 5. What the Diagnosis and Testing Revealed .....                            | 12 |
| 5.1 Findings about the organisation .....                                   | 12 |
| 5.2 Diagnosing the original system: heterogeneity in acceptance rates ..... | 13 |
| 5.3 Diagnosing the original system: number of criteria .....                | 14 |
| 5.4 Designing the new framework .....                                       | 15 |
| 5.5 Testing the new question set.....                                       | 16 |
| 6. Working as Advisors .....                                                | 17 |
| 6.1 Building the relationship over time.....                                | 17 |
| 6.2 The engagements in practice .....                                       | 17 |
| 6.3 Frictions and trade-offs .....                                          | 18 |
| 6.4 Engagement and impact.....                                              | 18 |
| 7. Conclusions and Lessons .....                                            | 19 |
| REFERENCES .....                                                            | 23 |

## EXECUTIVE SUMMARY

Each year, thousands of businesses apply to Innovate UK (IUK), the UK's national innovation agency, for innovation funding. Every application is judged through a broadly similar assessment process: questions, criteria, scores and guidance. This assessment process has been in place since at least the early 2000s, with only minor modifications. This report is an insider account of how IUK attempted to redesign its assessment process from 2017 to 2026, and the role our research team played in this process. The report draws on a research project commissioned through the Innovation and Research Caucus (IRC), but the redesign itself was led by IUK.

Between 2017 and 2024, across a series of engagements, we reviewed IUK's assessment records and practices at other international innovation agencies. Our research found little indication of favouritism – by location, gender, and other factors – in how applications were scored. It also showed that IUK's original ten criteria often produced a single judgement, and that comparable agencies managed with far fewer criteria. We then facilitated workshops in which IUK staff drafted a new five-area assessment framework.

In 2025 and 2026, we tested how the new assessment framework and question set performed in practice through interviews with assessors, who re-assessed applications under the new questions, and through surveys of assessors and applicants in the Women in Innovation Awards. Both groups broadly welcomed the new approach. Most applicants now use Generative AI tools in preparing their applications, and the growing ease of producing plausible applications is the major challenge the redesign leaves open.

This report adopts a 'street level' perspective on innovation policy, seeking to describe the assessment process through the experiences and practices of administrators directly involved in developing and overseeing it. Rather than evaluating a policy, we try to surface the challenges of implementing change within a complex, delivery-focused organisation. We draw implications for researchers working alongside these processes and for the development of new metascience tools for innovation policy.

## 1. Introduction: Innovation Policy at the Street Level

Most research on innovation policy focuses on strategy at the top or outcomes at the bottom. Strategies and budgets are debated, programmes are evaluated, projects are assessed in their thousands. The layer in between, where policies are turned into actions, guidelines and programmes, is far less understood. A long tradition in policy research explains why that matters. Michael Lipsky's work on street-level bureaucracy (Lipsky, 2010) showed that policy, as it is experienced, is made in large part by the people who deliver it: teachers, social workers, and caseworkers, working under pressure, with limited information, and with considerable discretion. Related literature on implementation has traced how policy intentions are transformed as they move into practice (Pressman & Wildavsky, 1973). Policies are not self-implementing: they are enacted through guidance documents, application forms, scoring rubrics, panels and information systems, and through the everyday judgements of the people who operate them.

Innovation policy is no exception, though it is rarely studied this way. A substantial literature examines the choice of innovation policy instruments – grants, loans, tax incentives, procurement –, the mixes in which governments deploy them (Flanagan et al., 2011; Borrás & Edquist, 2013) and their downstream effects (e.g., Lerner, 2000; Bianchi et al., 2019; Hottenrott & Richstein, 2020; Kok et al., 2022). With a few exceptions (e.g., Biegelbauer et al., 2020), the selection processes at innovation agencies have received little attention, and far less is written about what happens after an instrument is chosen: how a grant scheme, once decided upon, is turned into competitions, questions, criteria and scores, and how those choices shape who and what is funded. Taking a street-level perspective on an innovation agency therefore means studying its administrators and the processes, practices and routines through which they work, including the artefacts through which judgement is exercised.

Agencies allocate public funding to firms through competitive grant competitions, and every competition runs on such artefacts: a set of questions applicants must answer, criteria by which assessors judge those answers, and guidance that tries to keep those judgements consistent. This is what Lamont (2009) called the 'technologies of evaluation'. Though policy is framed at the strategic level, assessment is the agency's key allocation device at the level of the individual project: it is here that public money is matched to ideas. If the assessment process is flawed, the allocation of treatment is flawed as well. Poor assessment means poor treatment, however well designed the programme around it.

This report examines these mechanics at IUK, describing how IUK's assessment process was redesigned and the advisory relationship between the agency and our research team from 2017 to 2026. Within this system, the street-level bureaucrats are IUK's own project managers and administrators: the people who design and run competitions, recruit and brief assessors, and decide how the shared assessment framework is applied in each case. The assessors who score applications have a voice too, and exercise real judgement, but the assessment process sets the rules and principles on which they must decide. In addition, applicants also

shape these assessment systems, choosing whether to participate in a competition, writing applications, receiving feedback, and winning or not funding and support. What the design of the process shapes is how that assessment unfolds: the criteria applied, the information collected from applicants, and how it is assessed. The report views the agency at this level, though what follows is a study of its redesign rather than of the administrators themselves.

Redesigning an assessment system is an exercise in organisational change, and the literature on strategic change explains why it is demanding (Balogun & Hope Hailey, 2015). People need to be brought on board and persuaded that the change is worth believing in, and stakeholders need to be consulted and prepared for the changes ahead. This takes sustained effort from those who want the change and a degree of engagement from those it touches, whose reactions, in turn, reshape the change itself. What follows in this report, the workshops, the pauses, the renegotiations, is best read as change management in a complex public organisation.

This kind of attention to the workings of funding systems has recently gained an institutional home in the UK. The Metascience Unit, established jointly by the Department for Science, Innovation and Technology (DSIT) and UKRI, applies scientific methods to the design and delivery of research and innovation funding, funding experiments with review processes, indicators, and funding mechanisms, so far mostly in the science system (DSIT, 2025). Beginning before the term “metascience” had currency in UK policy, we worked with a funding institution, using scientific methods, to understand how its assessment system behaved and how it might be improved, building on the work of science and innovation policy researchers before us.

The immediate occasion for this report is a project commissioned through the Innovation and Research Caucus (IRC), the network of researchers that supports the use of robust evidence and insights in UKRI’s strategies and investments. That project, the second phase of the work described here, tested IUK’s new question set in practice in 2025-2026. The project cannot be understood on its own, though. This report therefore covers the ground before it, from the first contact in 2017 through the audit, the benchmarking and the workshops that produced the new framework and uses the whole arc to draw lessons for innovation and research policy and for research practice: for the agencies that run assessment systems, the bodies that commission evidence about them, and the researchers who provide it.

This report describes the redesign from the inside and treats IUK as a case with lessons beyond it. The account is not linear. The agency’s priorities and personnel shifted more than once over the period; such changes are negotiated across many teams and programmes rather than decided centrally.

## 2. Choosing What to Fund: The Design of R&D Assessment

Selection sits at the heart of innovation funding. Agencies receive many more proposals than they can support and must choose among them under uncertainty: applications span technologies and markets that no single expert can command, and the ideas with the greatest potential are often the hardest to evaluate. Research on the selection of R&D projects and scientific proposals has documented how evaluators struggle with this task: highly novel proposals tend to be penalised in expert review (Boudreau et al., 2016; Criscuolo et al., 2017), panels weigh negative information more heavily than positive (Lane et al., 2022) and the order in which projects are considered shapes their chances (Criscuolo et al., 2021). A selection system is an agency's answer to these difficulties: a process designed for turning uncertain claims about the future into a defensible set of funding decisions.

Selection is also what distinguishes grant funding from the other main channel of public support for business R&D. For example, while an R&D tax credit is claimed by any firm whose spending meets auditable accounting definitions, a grant is awarded, and someone must decide which proposals should receive public money. A grant programme must therefore run a selection process, and what passes through it is public money provided to businesses for commercial ends, unlike public research funding, which is concentrated in universities. Even when randomisation is used, the assessment process still determines who is considered fundable and who is not. Evaluations of R&D grant programmes internationally, including studies of IUK's own grants, find evidence of benefits to funded firms in performance, innovation and access to further finance (Howell, 2017; Vanino et al., 2019; Santoleri et al., 2024), while analysis of the European Union's SME Instrument indicates that outcomes also depend on which firms apply and how funds are allocated (Mina et al., 2021). Whether the right projects are selected remains the harder and much less studied question.

Building an assessment system means making a series of choices, summarised in Figure 1. Together they determine how effective and efficient a selection system is, e.g., how much time and effort it demands of assessors and applicants, and the quality of the judgements it produces. Should the criteria be spelt out formally, or left to assessors' judgement? Should they be the same for every project or flexible across them? How many assessors should read each application, and should they work alone or together? Should assessment happen in one stage or several, in scores or in words, behind closed doors or in the open, and how much is the agency willing to spend on it? Each of these choices carries costs. For example, discussion between assessors can surface arguments but risks conformity; formal, codified criteria aid consistency and accountability but can penalise unusual projects.

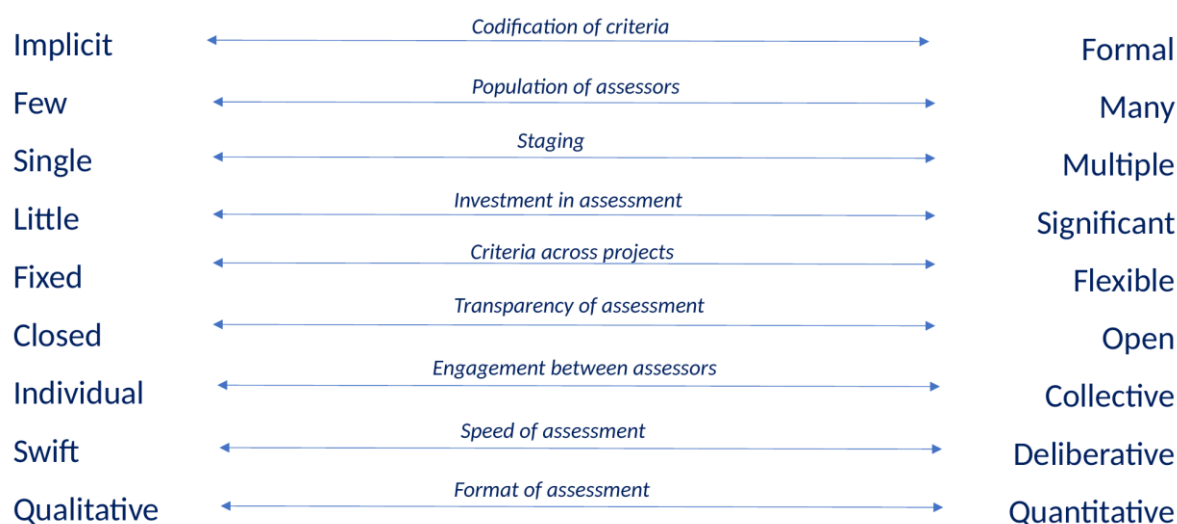


Figure 1. Key choices in the design of R&D selection systems.

Seen against these choices, IUK's system sat at a distinctive point: criteria formally codified and fixed within a competition; a large assessor population; each application scored by up to five experts, working individually and without discussion between them; and a format that combined quantitative scoring with qualitative comments. Within this framework, however, individual competitions retained some freedom to design their own selection processes, so these choices varied from one competition to the next (for instance, the Women in Innovation Awards examined later in this report included a video pitch alongside the written application).

This report describes changes to several of these choices at once: the question set and its guidance were rebuilt, the format of feedback to applicants changed, and the agency separately reduced the number of assessors used on many grants.

### 3. IUK in Brief

IUK is part of UKRI, the UK's national funding agency for science and research, alongside seven Research Councils and Research England. In 2025-26, IUK operated with an annual budget of £948 million, distributed through grant funding competitions open to UK-based businesses seeking to test the feasibility of an idea, develop a new or improved product, process or service, or collaborate with other businesses or research organisations (UKRI, 2025).

IUK runs many different competitions at once: over 800 between 2017 and 2024, each managed by its own project manager. Each competition has its own criteria for inclusion and area of focus. Some competitions are open, whereas others are targeted to specific groups or technologies. Some are one-off, and others run a continuous temporal cycle. After submission, applications are checked for eligibility by competition managers and for fit with the competition's

scope by an innovation lead. Applications that pass are assessed by up to five independent experts, drawn from a pool of over 1,000 assessors spanning business and academia, each of whom is remunerated a standard fee per application. Under the original framework, each assessor scored ten questions on a 1-10 scale, provided written comments on each, and gave an overall funding recommendation; assessors could not see one another's scores. Applications had to meet a quality threshold, and final decisions rested with a funders panel of the agency's domain experts, which could also apply a portfolio approach balancing scope areas, project sizes and locations. Successful and unsuccessful applicants alike received the assessors' written feedback after decisions were notified. The receipt of funding is dependent on a further due diligence process. The time from application to final funding can be eight months, but IUK has tried to create more 'fast funding' mechanisms to address concerns about the speed of the assessment process. In 2025, IUK committed to improving its funding pathways, including a 90-day target from the competition close to the Grant Offer Letter.<sup>1</sup>

The assessment process sits within a wide circle of stakeholders, each with different interests. There are the firms considering whether to apply, the individual applicants themselves, and the people who support applications, from consultants to the universities that collaborate on grants. There are assessors who judge the proposals, providing the applicants with their assessments and feedback. And there is the government that funds the exercise, looking to ensure it is run efficiently and generates benefit and value for society by stimulating innovation in the UK. These interests do not always point in the same direction. Applicants would prefer lower rejection rates; assessors want to work within a process that is efficient from their perspective; the government wants resources used well and to get as much public value as possible from the funding. An assessment system must balance these interests, which is part of what makes changing it so difficult.

## 4. Timeline: Touchpoints in a Multi-Year Relationship

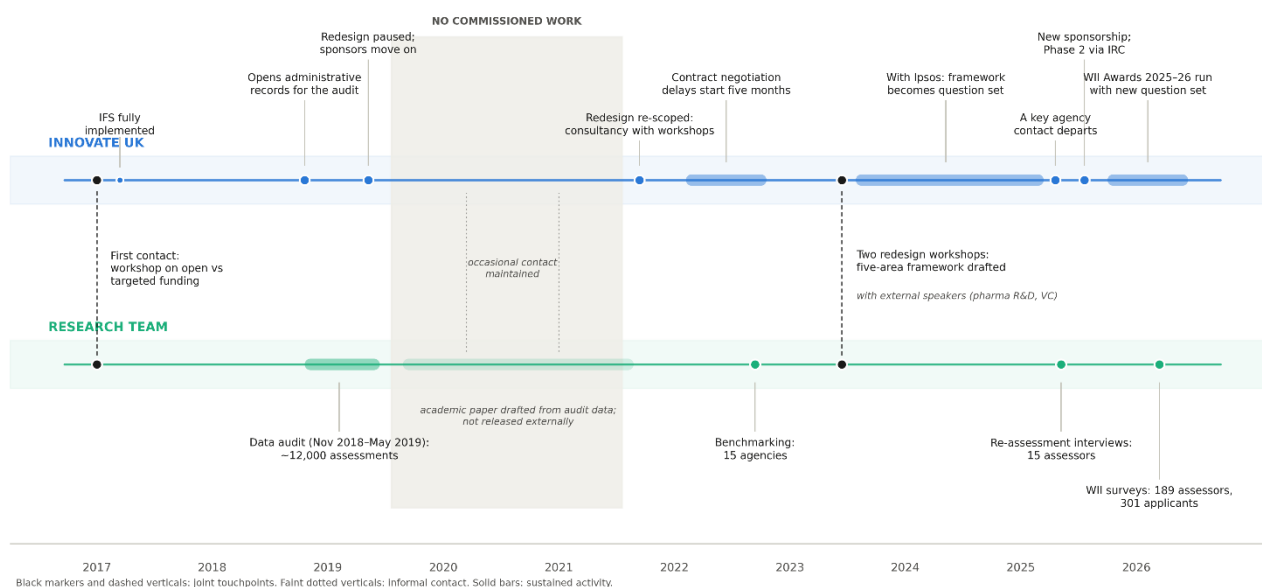
Table 1 sets out the touchpoints in our relationship with IUK between 2017 and 2026, and Figure 2 shows the same period as parallel timelines for the agency and the research team. There were quieter periods when internal sponsors had to re-establish the case for this work, when institutional priorities were being renegotiated, and when, from our side, we had very little visibility into what, if anything, was happening internally.

---

<sup>1</sup> see <https://iuk-business-connect.org.uk/news/innovate-uk-improves-their-path-to-funding/>, accessed July 2026.

**Table 1. Touchpoints in the IUK advisory relationship, 2017-2026.**

| Period              | Touchpoint                                           | Nature of engagement                                                                                                                                                                                                                                                                             |
|---------------------|------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2017                | First contact: workshop on open vs. targeted funding | An initial workshop exploring whether open or targeted funding competitions produced better applications — the question that, in effect, opened the relationship.                                                                                                                                |
| Nov 2018 – May 2019 | Data audit                                           | Quantitative audit of around 12,000 application assessments; a 70-page internal report; presentation to IUK staff and IRC members; a public blog post.                                                                                                                                           |
| 2019 – 2022         | Quieter period                                       | No formally commissioned project work. Groundwork already laid was shelved for a period, and contact with our original sponsors lapsed, while other institutional priorities took precedence at IUK. The United Kingdom also experienced three major nationwide lockdowns between 2020 and 2021. |
| Sept 2022           | Benchmarking review                                  | Review of assessment practices at 15 international innovation agencies; findings delivered as a presentation and supporting workbook.                                                                                                                                                            |
| June 2023           | Two redesign workshops                               | ‘Divergence Day’ and ‘Convergence Day’ with IUK’s internal working group and competition managers; a new five-area framework drafted and stress-tested.                                                                                                                                          |
| 2023 – 2025         | Internal development (limited visibility to us)      | IUK worked with Ipsos to translate the agreed framework into a live questionnaire. We were not closely involved in this stage and cannot fully account for the choices made within it.                                                                                                           |
| 2025                | Phase 2 commissioned via the IRC                     | New senior sponsorship renewed attention on the redesign; an independent testing phase was agreed and ethics approval secured from the University of Warwick Research Ethics Committee.                                                                                                          |
| May 2025            | Re-assessment exercise and interviews                | 15 assessors re-evaluated a previously assessed application under the new questions, followed by semi-structured interviews.                                                                                                                                                                     |
| Feb 2026            | WII assessor and applicant surveys                   | Live testing in the Women in Innovation Awards 2025–26 competition, surveying 189 assessors and 301 applicants.                                                                                                                                                                                  |

**Figure 2. Timelines of the advisory relationship, 2017-2026.**

The project's correspondence, which we reconstructed for this report from just over 600 emails of one of the authors spanning September 2018 to May 2026, records two additional features of the relationship. The first is that the starting point sits further back than participants later recalled, and that the questions changed shape along the way. What became the redesign of the question set began in 2018 as a study of whether open or targeted competitions produced better applications and of how assessors' characteristics shaped scoring. It was reframed as an analysis of the assessment process, then as a redesign of the criteria, then as the development and testing of a new question set and finally as a pilot assessment in a competition for women innovators. Each reframing tracked a change in sponsor, funding vehicle, or institutional home. The second is the weight of administrative repair required to keep the relationship going. Contracting, scheduling, expenses and data-sharing agreements recur across the years rather than being settled once: a negotiation over standard contract terms delayed the start of substantive work by five months in 2022, a data-sharing agreement went astray between institutions for months in 2024-25, and when one of us moved universities in 2023, the change later required a subcontract, a disbursement agreement and fresh ethics arrangements.

Some of this friction produced institutional learning of its own: we raised the publication restrictions attached to the first project with senior IUK officials at a public meeting, and, informed by this and by others' experience of similar difficulties, the IRC's directors and UKRI subsequently developed a new framework agreement to support productive collaboration between the parties. Even the smallest thread in the archive follows this pattern: a laptop issued to one of us for data-access work in 2019 stopped working in 2022, by which time the agency had forgotten it was still being used. None of this is unusual. A long advisory relationship is sustained as much by this largely invisible administrative work on both sides and by goodwill, although conventional project accounts rarely record it.

## 5. What the Diagnosis and Testing Revealed

This section summarises the evidence generated across the phases. This is brief by design: the story of this report is the process that produced this evidence, and what we can and cannot infer from it, rather than the evidence in isolation.

### 5.1 Findings about the organisation

IUK is best understood as a federation. Its decentralised structure gives it responsiveness and local ownership across very different sectors, technologies, and missions, and that same structure produces real internal complexity: local variation in shared systems, plural constituencies with competing expectations, information systems that resist integration, and parallel efforts that no one convenes. We regard these as ordinary consequences of a federated design, common to many public agencies, rather than as failings particular to IUK. But any attempt to change a shared system, such as assessment, has to work through them.

IUK runs a diverse set of competitions, often simultaneously, managed by different people with relatively limited coordination between them. The assessment framework was broadly the same across them, but project managers retained real autonomy over how it was implemented in practice: how assessors were recruited to a given competition, and how the information assessors provided was treated in the review process. The assessment system, although intended as a coordinating device across the organisation, therefore carried substantial local variation in how it was applied, used, and understood from one competition to the next. There is also a structural tension here that funding agencies generally live with: directed programmes appeal as a way of mobilising attention towards specific technology challenges, but they then face very different dynamics from open ones.

This plurality also helps explain a pattern we observed directly: the tendency of assessment criteria to expand over time. Different constituencies within IUK held different views about what assessment should prioritise; for instance, some programme teams cared particularly about sustainability, while others ran cross-sector programmes with different priorities. In a federation, a shared assessment system becomes the place where the priorities of many constituencies meet, which makes it a coordination device. The practical consequence is a tendency for the list of topics the assessment is expected to cover to grow longer, as each constituency seeks to have its priority reflected, until the system becomes noticeably more complex than any single part of the organisation intended. We suggest that this dynamic, more than any single decision, explains much of why the original ten-question framework had grown as complex as the audit found it to be in the last stages of the assessment redesign.

A related complexity concerns information. IUK holds many datasets and tools, built for different purposes: systems for managing communications with applicant businesses, systems for administering grants, and systems for the assessment process itself, each maintained by different parts of the organisation with varying priorities around access. IUK's online Innovation

Funding Service (IFS) was fully implemented in 2017, and the older system it replaced was phased out from 2017 onwards. Our audit drew on data from the IFS only: extracting data from the older system was a major challenge, which places the agency's earlier activity beyond the reach of our analysis. Information systems in most large organisations grow up around the practical needs of distinct teams rather than as a single deliberate design.

## 5.2 Diagnosing the original system: heterogeneity in acceptance rates

In any assessment process with a high rejection rate, such as IUK's, most applicants will be disappointed, and some of that disappointment may lead to the belief that the process was unfair or that certain applicants were favoured. Our research on IUK's assessment records found little direct evidence to support suspicion of favouritism.

Using the IFS data, we investigated whether there were unexplained differences in acceptance rates across certain groups of applicants. Starting with firm size, we found that the acceptance rate was lower for small firms. This is likely due to differences in administrative capability and internal selection processes, so little can be drawn regarding assessors' preferences. Looking at geography, the acceptance rates across home nations (England, Wales, Northern Ireland and Scotland) were broadly similar; this may be due to the nature of the assessment process or be a product of geographical 'portfolio management'. Open competitions had lower acceptance rates than targeted ones; however, large, targeted programmes had acceptance rates similar to the overall portfolio, suggesting these programmes are highly competitive. Finally, looking at the gender of the applicants (as determined by the lead applicants' names), we found no difference between men and women in acceptances and rejections. A caveat worth flagging is that gender detection tools cannot handle non-binary identities, and their accuracy drops sharply for names outside the training data's geographic scope.

To better understand how funding decisions are shaped by potential applicants' characteristics, including gender, we also examined projects at the margins of the funding threshold. The intuition behind this approach is that, in the case of very poor or very strong applications, assessors are likely to pay little attention to applicant characteristics, since the quality of the applications is readily observable. However, when a proposal is more marginal, assessors might rely on cues from applicants' characteristics to guide their views on the project's merits. We focused on differences across the score bands ('bins') immediately before and after the threshold level of 7.0. Focusing on the 7.0-7.25 bin, we noticed that the rejection rate for female applicants was higher than for male applicants. However, the opposite appears to be true in the next bin (7.50-7.75), where women were more likely to be funded. Taken together, these results suggest no clear statistically significant pattern of favouritism among assessors towards either men or women applicants on the margin. We used a similar approach for other potential sources of bias and again found little evidence for favouritism to specific groups of applicants. It must be said that our analysis was limited by incomplete information on assessors and applicants.

### 5.3 Diagnosing the original system: number of criteria

Correlation analysis of 34,869 assessor evaluations found that IUK's original ten criteria loaded onto a single statistical factor: assessors appeared to be forming one overall judgement of quality rather than ten independent assessments. A modelling exercise also showed that the 'all criteria equal' weighting approach was a design choice with real consequences for funding decisions, not a neutral default. Variations in weighting to different criteria, such as focusing on the team or novelty of the application, might have yielded different outcomes.

The audit also revealed the limits of what could be observed. The final stage of selection, the funders panel and portfolio-balancing process in which funded projects are checked against goals such as regional representation, was only sparsely recorded in the data available to us. We had rich, time-stamped information on who assessed what and what scores they gave, but comparatively little on assessors who were considered but not used. We also lacked information on how each funder's panel dealt with variance in assessors' scores. However, our simulation of removing extreme positive and negative assessor scores from the evaluation did not notably change the pool of funded applications.

During the analysis of assessors' scoring, the audit also surfaced a difference that bears on a design choice upstream of the assessment system: whether a competition is open to any innovative idea or targeted at a particular sector, technology, or group. That is a competition-design decision, shaped by project managers and by the program objectives, rather than a feature of the assessment system itself. It nonetheless changes the environment in which the same system operates: funded applicants in open competitions required an exceptionally high assessment score, above 80%, while in targeted competitions the effective threshold was closer to 70%. The assessment system was identical in both cases; what differed was the level of competition around it. Oversubscription raises the effective threshold for funding. IUK's recent decision to suspend the SMART programme, which was a very competitive open competition, to some degree reflects this concern about contrasting thresholds across different competitions.

The benchmarking review, covering fifteen agencies across fourteen countries (Table 2), found that most agencies do not publish their assessment criteria in detail, but that among those that do, the number of criteria used ranged from three to seven, with most using four to five (Table 3), against IUK's ten at the time.

**Table 2. Innovation agencies included in the benchmarking review.**

| European agencies                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Non-European agencies                                                                                                                                                                                                                                                                                                             |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. European Innovation Council – EIC (EU)<br>2. Vinnova (Sweden)<br>3. Business Finland (Finland)<br>4. Invitalia (Italy)<br>5. Innosuisse (Switzerland)<br>6. Estonian Business and Innovation Agency (Estonia)<br>7. Centro para el Desarrollo Tecnológico Industrial – CDTI (Spain)<br>8. Austrian Research Promotion Agency – FFG (Austria)<br>9. Banque Publique d'Investissement – Bpifrance (France)<br>10. Technology Agency of the Czech Republic – TACR (Czech Republic) | 1. A*STAR – Agency for Science, Technology and Research (Singapore)<br>2. Callaghan Innovation (New Zealand)<br>3. Japan Science and Technology Agency – JST (Japan)<br>4. Commonwealth Scientific and Industrial Research Organisation – CSIRO (Australia)<br>5. Canadian Industrial Research Assistance Program – IRAP (Canada) |

**Table 3. Number of assessment criteria at comparator agencies.**

| Agency                                            | Programme                         | Type       | No. of criteria |
|---------------------------------------------------|-----------------------------------|------------|-----------------|
| Vinnova (Sweden)                                  | Innovative Impact Startups        | Grant      | 7               |
| European Innovation Council (EU)                  | Pathfinder Open/Challenge         | Grant      | 3               |
| A*STAR (Singapore)                                | Industry Alignment Fund           | Grant      | 6               |
| Business Finland (Finland)                        | Co-innovation                     | Loan       | 4               |
| Callaghan Innovation (New Zealand)                | RDTI                              | Tax credit | 5               |
| Invitalia (Italy)                                 | Smart&Start Italy                 | Grant/Loan | 4               |
| Innosuisse (Switzerland)                          | Innovation projects with partners | Grant      | 4–5             |
| Estonian Business and Innovation Agency (Estonia) | Programme for applied research    | Grant      | 4               |

## 5.4 Designing the new framework

To develop a new assessment framework, we helped to organise two workshops with IUK in June 2023, which were designed on a divergence-convergence model. The first, Divergence Day, aimed to broaden the frame of reference before any design decisions were made: our internal sponsor set out the strategic rationale for the redesign, we presented the benchmarking findings, and two external speakers, a pharmaceutical R&D manager and the founder of a

venture capital fund, described how selection works in their settings. Working groups then generated candidate assessment areas. The second, Convergence Day, held two working days later, moved from ideas to drafts: groups drafted criteria for each candidate area, exchanged structured peer feedback, and stress-tested the draft framework by scoring the same project with the new approach and discussing the results together before the workshops closed.

The outcome was the five-area framework: why, what, how, commercial intent, and finance and value for money. Agreement on the first three areas came quickly; commercial intent and finance provoked more debate, including how much financial information should be sought at the application stage rather than at due diligence, and whether one framework could serve both small and large applications. Two practices from the workshop room carried forward into the design: specifying, for each criterion, the supporting evidence required alongside a plain-language summary; and assessing team capability on skills rather than credentials, to improve inclusivity. The stress test revealed clarity problems that prompted a final round of revision.

The five-area framework then had to become a working questionnaire. From September 2024, IUK contracted Ipsos to develop and cognitively test a new question set, and the two organisations worked through an in-person application redesign workshop in December 2024 and a programme of questionnaire development running into 2025. We attended meetings and workshops with the Ipsos team, commented on drafts, and sought to align our own research with their timeline, including asking that the assessor interviews be postponed until the question bank had stabilised. However, responsibility for the new framework passed to Ipsos and then back to IUK.

## 5.5 Testing the new question set

Once the new framework had reached the draft stage, in May 2025, 15 assessors from a range of disciplines and experience levels each re-assessed an application they had previously assessed under the old question set, now using the new questions, and were then interviewed about the experience. Views of the operationalised question set were mixed: assessors valued the sharper problem-definition framing, but the expansion from around 10 questions to a larger number of sub-questions was the most common concern.

The Women in Innovation Awards competition was then used as a pilot to test the framework at scale. We supported IUK in conducting a paired survey of applicants and assessors. We prepared a draft questionnaire based on IUK's concerns, and they implemented this survey. The survey included a range of quantitative and qualitative questions about the new assessment system, including the measure of perceived justice. We were then provided access to the anonymised data from these surveys for analysis. Of 432 assessors, 189 responded (43.7%); most were satisfied overall, though guidance and scoring rubrics drew some disagreement, chiefly because the guidance was in a separate document rather than in the portal, and the scoring descriptors used undefined terms. Of roughly 1,500 applicants, 301

responded (about 20%); most were satisfied, and returning applicants were significantly more positive about the change. An unanticipated finding was that 80% of applicants used AI tools somewhere in their application, most often to gauge whether it was worth applying and to prepare their answers; 94% used no external grant-writing consultant. The redesign did not anticipate this, and we return to what it means later. The interview schedule, the survey instruments and the full tables of themes from the interviews and surveys are available from the authors on request.

## 6. Working as Advisors

This section focuses on the relationship between the research team and IUK and seeks to draw lessons on engagement in innovation policy development.

### 6.1 Building the relationship over time

Our engagement was a sustained, collaborative process in which we contributed evidence and an external perspective while being shaped by the organisation's own rhythms, priorities and constraints. In broad terms, we did the research, while IUK ran the assessment system and decided what to do with its own processes. In practice, the division of labour was not always so clear and had to be worked out as the collaboration developed. We were commissioned at particular points for specific pieces of work and invited back repeatedly to bring an independent eye to an internal problem.

The project rested on a pairing of outsiders and an insider. We, the authors, are academic researchers with no employment relationship with IUK; inside the agency, the work was championed and coordinated across both phases by an internal sponsor, our key point of contact, who leads its Strategic Reporting function. The pairing gave us access and institutional candour that a purely external evaluation would be unlikely to receive, and it gave the sponsor an independent evidence base to draw on internally. It also means this account is not a fully independent audit: the early consultancy work was contracted by IUK, and the IRC-commissioned project was conducted in close partnership with the agency. The team operated under a non-disclosure agreement in Phase 1, and Phase 2 ran under IRC governance with ethics approval from the University of Warwick. Throughout, we reported findings as we found them, including unwelcome ones, and the relationship was built through the audit, the workshops, and a willingness on both sides to keep going when no formal project was active.

### 6.2 The engagements in practice

In practice, Phase 1 unfolded as three distinct pieces of work, each dependent on the last. The data audit, conducted between November 2018 and May 2019, required IUK to open its administrative records to us and to spend considerable staff time helping us understand data that had never been assembled for this kind of analysis before. The benchmarking review in September 2022 was delayed by the scarcity of published information on other agencies'

assessment processes. The two workshops in June 2023 were the most concentrated moment of collaborative work in the whole engagement. Around 11 representatives of IUK's internal working group, together with project and competition managers from across the agency's sector programmes, spent two days together. It was the first time some of these colleagues had met in person. The workshops therefore did two things at once: they produced the five-area framework, and they created a forum for people managing different, siloed competitions to compare notes on how they each approached assessment.

Phase 2, in 2025-26, examined how the framework operated once IUK had translated it into a live question set. The re-assessment exercise and interviews required IUK to approach working assessors and ask them to give up their time, a second time, to help us understand how they were interpreting the new questions. The Women in Innovation Awards pilot depended on IUK agreeing to embed two research surveys inside a live, high-profile competition, with all the operational coordination that involved: co-designing survey instruments with IUK staff, timing distribution around the competition's own communications, and working within constraints, such as surveying applicants before outcomes were known. Throughout both phases, our role was closer to that of a facilitator and "critical friend" than a conventional external evaluator. In the 2023 workshops, we combined facilitation with observation and analysis; in the Phase 2 interviews, we tried to create a space in which assessors felt able to be candid about a process their own employer had designed. Both depend on relationships built over time.

### 6.3 Frictions and trade-offs

Not everything translated cleanly from the workshop room to live competition. The five-area framework was, by design, a broad structure, and turning it into the detailed question set used in 2025-26 was IUK's and Ipsos's work; we attended meetings and commented on drafts, but even so, we could not easily reconstruct how the question was developed as it did.

This way of working carried its own trade-offs. We were researchers rather than consultants, unpaid apart from a small consultancy arrangement around the 2023 workshops, and our designs had to stay adaptive: IUK paused the project at various points, we changed approach several times in response, and a planned randomised controlled trial, a shadow assessment of old against new questions, and a pilot of AI tools were all dropped. Usefulness and timeliness came at the cost of academic output, an exchange that other researchers considering such work should weigh.

### 6.4 Engagement and impact

Whether any of this changed what IUK does is difficult to establish. There is no counterfactual, and the organisation absorbed evidence from many directions while leadership, government priorities and sponsors changed around the work; at one point, the redesign was shelved, and the issue was taken up again only when a new management team enabled it. Even where change followed our findings, as when the agency reduced the number of assessors per

application after we had shown how strongly assessors' scores correlated, we cannot show that our work caused it.

Our contribution is more readily described than our impact. We supplied analysis of the agency's own records, comparative material on the practice of other agencies, facilitation of the workshops at which the new framework was drafted, and an independent examination of how the new question set performed in use. How far these inputs shaped the decisions that followed cannot be established from the available evidence, and the decisions themselves were, throughout, IUK's to take.

## 7. Conclusions and Lessons

IUK redesigned the way it assesses applications for R&D funding, moving from ten scored questions to a framework of five areas, and this report has described that change from the inside: the commissioned project that tested the new question set, and the audit, benchmarking and workshops that came before it. The main findings can be stated briefly. The original ten criteria were producing a single judgement rather than ten; equal weighting was shaping funding outcomes; comparable agencies use far fewer criteria. The new framework was broadly welcomed by assessors and applicants in its first live competition. The system is still being piloted, and the testing surfaced the kinds of refinement piloting exists to find: a question set assessors find demanding in its current volume, scoring guidance that needs sharper anchors, the persistent difficulty of recognising novelty, and the fact that most applicants are already drafting with AI. Change of this kind, inside a large, decentralised public organisation, is rarely linear, and the lessons below draw on both the evidence and the experience of contributing to it.

### Lessons from the research

These lessons concern less the mechanics of assessment than how an innovation agency learns: from its own data, from its own people, from its stakeholders, and from the experience of others.

- » *Data as a mirror.* The audit told IUK things about its system that were not fully visible to the organisation itself: that ten criteria were functioning as one judgement, that some important nuances provided by the scores were overlooked, and that oversubscription had raised the effective threshold for funding.
- » *Learning from itself.* Much of what an agency knows is held by its people and surfaces when they meet: the workshops brought together staff from different competitions, and the redesign advanced as much through their comparing of practice as through any external evidence. A shared system such as assessment is where an organisation's plural priorities become visible to itself.

- » *Learning with stakeholders.* Assessors and applicants experience the system from positions the agency cannot occupy, and the interviews and surveys gave them a structured way of telling the agency how its evaluation was operating in practice.
- » *Learning from others.* The benchmarking of fifteen agencies and the perspectives of pharmaceutical and venture capital practitioners supplied no ready-made solution, yet they legitimised change and gave the agency reference points against which to see its own choices as choices.
- » *Learning requires occasions:* audits, workshops and pilots. Learning of this kind can be convened with outside help, but the learning itself takes place within the agency as it tries to make sense of what it is currently doing and what it wants to try to achieve.
- » *Keeping track of perceived justice:* for innovation agencies, regularly monitoring applicants' and assessors' perceptions of the system's fairness is as important as monitoring its actual accuracy, since perceptions of injustice can hinder trust and participation even when the process is sound.

### **Lessons from our experience of the process**

- » *Plan for non-linearity.* Pauses, leadership changes and renamings proved to be the normal condition of the work rather than interruptions to it, and the useful measure of a long advisory relationship is its capacity to resume rather than its momentum.
- » *Back up continuity.* The work survived its gaps because specific individuals championed it; agencies and research teams should document and hand over relationships of this kind so that progress does not depend on any one person staying in post.
- » *Stay adaptive.* We dropped a planned RCT and a shadow assessment, designed surveys just in time against a live competition's timetable, and paid for usefulness with a potential academic output; researchers should weigh that exchange before entering this kind of work, and those who fund it should recognise the cost.
- » *Claim contribution, not outcomes.* There is no counterfactual; evidence enters an organisation as one ingredient among many, and the claim available to engaged researchers is to have better informed the decisions, rather than to have made them.

### **Metascience in practice**

The project also bears on the metascience agenda described in the introduction. The Metascience Unit now pursues, at programme scale, evidence-based experimentation with the workings of funding systems. This collaboration involved much the same activities: administrative data used to examine a live system, external benchmarks to widen the options

considered, a redesign developed with the people who run the system, and a live competition in which the result was tested.

It also speaks to a question the metascience agenda will increasingly face: what happens to findings in practice when an agency has to understand them, take them up and implement them? Our experience suggests the answer lies in the organisation as much as in the evidence. Assessment systems have many stakeholders inside and outside an agency, and they need consulting through whatever mechanisms fit them: workshops, focus groups, interviews and surveys. Decision-making in agencies of this kind is decentralised, so no single actor can determine the behaviour of a whole programme, and shifts of strategy at the top can redirect, limit or derail a project already underway. Organisations will also, quite reasonably, act on their own perspectives before the researchers' findings are absorbed into an agency's priorities, not simply adopted from a report. If there is a lesson here for metascience, it is that projects of this kind succeed slowly, through sustained effort, coordination and embedding in the context, so that the solutions developed align with the needs, capabilities and priorities of UKRI and agencies like it, and not only with the research. They also require extensive consultation and engagement with stakeholders, such as applicants and assessors, who are not merely subjects to be manipulated in an experiment, but equal partners in development and operations of an assessment system.

### **The challenge ahead: assessment in the AI era**

When the redesign project began, large language models were neither widely available nor well understood. They arrived in the middle of it, with the release of ChatGPT in November 2022, and have improved in quality and performance ever since. By the time the new question set faced its first live test in the Women in Innovation Awards 2025-26, four in five applicants were using AI tools somewhere in the process, most often to work out whether applying was worth their time and to prepare their answers, and almost none were paying for grant-writing help. The redesign did not necessarily anticipate this. The question set, the guidance and the scoring rubrics were all built on the assumption that an application's text is written by the applicant and that its quality carries information about them.

The economics of the system make the challenge concrete. The cost of preparing an application has fallen sharply, while the cost of assessing one has not. In the legacy system, each application incurred around £600 in assessor fees (based on five readings at £120), before IUK's own administrative costs. A system in which plausible text costs applicants little, while every submission costs the agency the same to judge, invites more applications than it can readily assess, and unsuccessful applicants can return at little expense. AI tools for preparing proposals and for generating pre-emptive assessments will improve the hygiene factors of application quality, but they will not lower the volume.

Generative AI also unsettles what the scores measure. If the gap in presentation between strong and weak proposals is narrowed, assessors lose one of the signals they have

traditionally relied on, and the difficulty the Women in Innovation Awards assessors already reported in distinguishing genuinely novel projects from incremental ones becomes harder still.

We raise the challenge without resolving it: it was not addressed in this project, and IUK is far from alone in facing it. Potential remedies exist, video and verbal pitches among them, and they may help, but they do not solve the issue. The questions it opens go beyond this report. How should the costs and incentives of applying be set so that volume does not outrun assessment capacity? What should assessors be told about AI-assisted text, and how should their guidance and rubrics change? How should AI use be measured as it continues to rise? What is the value of human assessment when AI tools might generate similar levels of quality? Will those assessed (at least partly) by AI tools perceive their assessment as fair?

## Summary

This report began from the observation that innovation policy is enacted through its delivery: through the competitions, questions, criteria and scores by which an agency selects among proposals. The redesign examined here illustrates that process. Analysis of IUK's administrative records indicated that the ten original criteria functioned as a single evaluative judgement, and that the equal weighting of criteria had consequences for which projects were funded, while comparison with other national agencies showed that most operate with substantially fewer criteria. On this evidence, and through workshops with its own staff, the agency developed a five-area framework, which was subsequently turned into a new question set and tested in a live competition. The testing found broad support for the new approach among assessors and applicants, together with areas for refinement of the kind piloting is intended to reveal. The question that remains open, for IUK as for other funders of R&D, is how assessment should be organised as applications become progressively cheaper to produce than to judge.

## REFERENCES

- Balogun, J., & Hope Hailey, V. (2015). *Exploring strategic change* (4th ed.). Harlow: Pearson.
- Bianchi, M., Murtinu, S., & Scalera, V. G. (2019). R&D subsidies as dual signals in technological collaborations. *Research Policy*, 48(9), 103821.
- Biegelbauer, P., Palfinger, T., & Mayer, S. (2020). How to select the best: Selection procedures of innovation agencies. *Research Evaluation*, 29(3), 289-299.
- Borrás, S., & Edquist, C. (2013). The choice of innovation policy instruments. *Technological Forecasting and Social Change*, 80(8), 1513-1522.
- Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. (2016). Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science*, 62(10), 2765-2783.
- Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2017). Evaluating novelty: The role of panels in the selection of R&D projects. *Academy of Management Journal*, 60(2), 433-460.
- Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2021). The sequence effect in panel decisions: Evidence from the evaluation of research and development projects. *Organization Science*, 32(4), 987-1008.
- DSIT. (2025). *A year in metascience 2025*. London: Department for Science, Innovation and Technology. Available at: <https://www.gov.uk/government/publications/a-year-in-metascience-2025>.
- Flanagan, K., Uyerra, E., & Laranja, M. (2011). Reconceptualising the 'policy mix' for innovation. *Research Policy*, 40(5), 702-713.
- Hottenrott, H., & Richstein, R. (2020). Start-up subsidies: Does the policy instrument matter? *Research Policy*, 49(1), 103888.
- Howell, S. T. (2017). Financing innovation: Evidence from R&D grants. *American Economic Review*, 107(4), 1136-1164.
- Kok, H., Faems, D., & de Faria, P. (2022). Pork barrel or barrel of gold? Examining the performance implications of earmarking in public R&D grants. *Research Policy*, 51(7), 104514.
- Lane, J. N., Teplitskiy, M., Gray, G., Ranu, H., Menietti, M., Guinan, E. C., & Lakhani, K. R. (2022). Conservatism gets funded? A field experiment on the role of negative information in novel project evaluation. *Management Science*, 68(6), 4478-4495.

Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Harvard University Press.

Lerner, J. (2000). The government as venture capitalist: the long-run impact of the SBIR program. *The Journal of Private Equity*, 55-78.

Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public services* (30th anniversary expanded edition). New York: Russell Sage Foundation.

Mina, A., Di Minin, A., Martelli, I., Testa, G., & Santoleri, P. (2021). Public funding of innovation: Exploring applications and allocations of the European SME Instrument. *Research Policy*, 50(1), 104131.

Pressman, J. L., & Wildavsky, A. (1973). *Implementation: How great expectations in Washington are dashed in Oakland*. Berkeley: University of California Press.

Santoleri, P., Mina, A., Di Minin, A., & Martelli, I. (2024). The causal effects of R&D grants: Evidence from a regression discontinuity. *Review of Economics and Statistics*, 106(6), 1495-1510.

UKRI (2025). 2025-26 budget allocations for UK Research and Innovation. Available at: <https://www.ukri.org/wp-content/uploads/2025/04/UKRI-030425-BudgetAllocations25-26-Explainer.pdf>

Vanino, E., Roper, S., & Becker, B. (2019). Knowledge to money: Assessing the business performance effects of publicly-funded R&D grants. *Research Policy*, 48(7), 1714-1737.



[www.ircaucus.ac.uk](http://www.ircaucus.ac.uk)

Email [info@ircaucus.co.uk](mailto:info@ircaucus.co.uk) Twitter [@IRCaucus](https://twitter.com/IRCaucus)



Delivered with  
ESRC and  
Innovate UK