RoRI Working Paper No. 6

The experimental research funder's handbook

Sandra Bendiscioli, Teo Firpo, Albert Bravo-Biosca, Eszter Czibor, Michele Garfinkel, Tom Stafford, James Wilsdon and Helen Buckley Woods

December 2021

Produced in partnership with the European Molecular Biology Organisation (EMBO), the Innovation Growth Lab at Nesta and the Swiss National Science Foundation (SNSF).

About the authors

Sandra Bendiscioli is a Senior Officer in the Policy Programme at the EMBO (European Molecular Biology Organization). bendisci@embo.org; @sbendi1; http://www.embo.org/policy/

Teo Firpo is Senior Researcher at the Innovation Growth Lab at Nesta. Teo.Firpo@nesta.org.uk; @TeoFirpo; http://www.innovationgrowthlab.org

Albert Bravo-Biosca is Director of the Innovation Growth Lab at Nesta, Guest Professor at the Barcelona School of Economics and Visiting Fellow at the Laboratory for Innovation Science at Harvard. abravobiosca@nesta.org.uk; @abravobiosca; http://www.innovationgrowthlab.org

Eszter Czibor is Affiliate at ConCIV, Department of Economics, University of Reykjavik. czibore@gmail.com @eczibor https://orcid.org/0000-0002-3182-6564

Michele Garfinkel is Head of the Policy Programme at the EMBO (European Molecular Biology Organization) and chaired the Steering Group for this project. michele.garfinkel@embo.org; @michgarfinkel; https://orcid.org/0000-0001-8946-7907; http://www.embo.org/policy/

Tom Stafford is University Research Practice Lead and Senior Lecturer in Psychology and Cognitive Science at the University of Sheffield. t.stafford@sheffield.ac.uk @tomstafford https://orcid.org/0000-0002-8089-9479 http://tomstafford.staff.shef.ac.uk/

James Wilsdon is Director of RoRI and Digital Science Professor of Research Policy at the University of Sheffield. j.wilsdon@sheffield.ac.uk; @jameswilsdon; https://orcid.org/0000-0002-5395-5949; https://www.sheffield.ac.uk/is/people/academic/james-wilsdon

Helen Buckley Woods is a Postdoctoral Research Associate at RoRI and in the Information School at the University of Sheffield. h.b.woods@sheffield.ac.uk; https://orcid.org/0000-0002-0653-9803; @HelenBWoods; https://www.sheffield.ac.uk/is/people/research/helen-buckley-woods

For citation: Bendiscioli, S, Firpo, T, Bravo-Biosca, A, Czibor, E, Garfinkel, M, Stafford, T, Wilsdon, J, Woods, H B (2021) *The experimental research funder's handbook*.. RoRI Working Paper No.6, December 2021. **DOI: 10.6084/m9.figshare.17102426**

Acknowledgements. The authors would like to record their gratitude to members of the project steering group for their advice and guidance at every stage: Gert Balling (Novo Nordisk Foundation); Marco Bieri (Swiss National Science Foundation); Amanda Blatch-Jones (National Institute for Health Research; Jon Holm (Research Council of Norway; and Vincent Traag (CWTS-Leiden). Thanks also for invaluable contributions to the case studies in Part 2 from Marco Bieri (SNSF); Ulrike Bischler and Pavel Dutow (Volkswagen Foundation); Lucy Pomeroy (Health Research Council of New Zealand); Uwe von Ahsen, Elisabeth Nindl, Falk Reckling, Tina Olteanu and Ralph Reimann (Austrian Science Fund—FWF). Any errors or omissions remain our own.

Contents

RESEARCH ON RESEARCH INSTITUTE

https://researchonresearch.org/

Foreword

Across the international research funding community, there is a growing appetite for more sophisticated approaches to evidence gathering and experimentation, to inform evaluation, allocation and decision-making. Different research funders are at different stages of this journey, and don't always benefit as much as they might from the insights and experiences of others, or from the latest academic studies in this field. This Handbook aims to provide a practical resource for funders looking to move further or faster down the experimental path.

Here we have collated and synthesised insights from funder partners in the RoRI consortium that have conducted trials with aspects of peer review and evaluation. From these accounts, we have assembled practical descriptions of designing, implementing, and evaluating changes to peer review processes. By supplementing these with more foundational information about experimental processes and theories of peer review, we aim to present a comprehensive description of end-to-end processes that might be used by funders considering similar experiments in future.

Detailed descriptions of the implementation of such processes and of their effects have been largely hidden from those outside of the organisations carrying out the experiments. By providing these here, thanks to direct input from those funders, we highlight a variety of approaches that funders may take, challenges they have experienced and lessons learned.

In line with RoRI's goal to achieve more efficient, dynamic, diverse and inclusive research systems through rigorous evidence and experimentation, our hope is that this Handbook will support funders to design effective interventions that can be properly evaluated, and yield robust findings of relevance to others.

This is an initial working paper version of the Handbook, for discussion at a workshop in December 2021, after which we will update it for final publication in spring 2022. We warmly invite comments at hello@researchonresearch.org, and look forward to seeing a growing diversity of experiments with research funding in the months and years to come.

Matthias Egger
President, SNSF

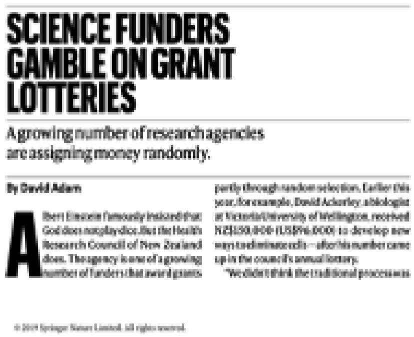
Michele Garfinkel
Head of Policy, EMBO

James Wilsdon
Director, RoRI

Albert Bravo-Biosca
Director, IGL at Nesta

Part 1:
The case for experimental research funding

1. Summary


SCIENCE FUNDERS GAMBLE ON GRANT LOTTERIES

A growing number of research agencies are assigning money randomly.

By David Adam

Albert Einstein famously insisted that God does not play dice. But the Health Research Council of New Zealand does. The agency is one of a growing number of funders that award grants partly through random selection. Earlier this year, for example, David Ackerley, a biologist at Victoria University of Wellington, received NZ$150,000 (US$96,000) to develop new ways to eliminate cells — after his number came up in the council's annual lottery.

"We didn't think the traditional process was

This recent Nature article describes a growing number of funder experiments with randomisation in research funding.

Over the past five years, in response to longstanding debates over the strengths and weaknesses of conventional peer review-based approaches to the evaluation of research proposals and allocation of resources (Guthrie et al., 2013), a growing number of research funding organisations have been piloting and experimenting with new methods (Adam, 2019; Bendiscioli and Garfinkel, 2021). However, in many cases, detailed descriptions of the design and implementation of such interventions, and of their effects, have not yet been widely shared outside of the organisations involved.

This handbook collects and synthesises information from funder partners in the RoRI consortium. Its purpose is to provide practical and evidence-based guidance for funding professionals willing or interested in trialling changes to their peer review processes. It includes practical instructions on how to design, implement and evaluate trials of partial randomisation and other interventions in funding allocation.

The aim is to support funders to design effective interventions that are fit for purpose and rigorously evaluated. Decision-makers and broader research communities that oversee or participate in funding processes will also find useful information here.

The handbook complements other outputs from this RoRI workstream, including a study of funder motivations for partial randomisation, bespoke partner support opportunities, and a workshop on experimental research funding in December 2021.

This initiative is part of a wider movement to create a culture of evidence-based evaluation and funding practice (Azoulay and Li, 2021) by (i) designing effective interventions that are fit for purpose and rigorously evaluated; and (ii) making this information available in an accessible and usable form. These goals are consonant with broader moves towards evidence-informed policy and practice as championed by numerous initiatives in recent years (Boaz et al. (eds.), 2019).

2. Why experiment?[1]

"If I look back on many years of involvement in political decision-making and policy-making around science, innovation and R&D, I am struck by how much of it tends to turn on gut feel of the individuals involved, than on hard evidence and analysis. This is of course ironic, since good science is all about testing hypotheses against data, empirical results and facts." ***Sir John Kingman***, *Reflections on his time as Chair of UK Research and Innovation, 14 July 2021.*

Maximising the societal and economic returns from limited research and innovation funding is important, as is distributing that funding in ways that are fair and support dynamic, diverse and inclusive research cultures. Despite this, it is still relatively rare to see research funders applying scientific methods to test and evaluate the best ways to achieve their goals. As Floyd Bloom, former editor--in-chief of *Science*, wrote in a 1998 editorial: "*Does it make sense to be scientific about everything in our universe except the future course of science?*".

Experimentation is a cornerstone of the scientific method, and as such, research funders are of course intimately familiar with it. They devote billions every year to fund countless experiments across many different disciplines. But look inside funding organisations and a very different picture often emerges. Despite the many challenges they face, research funders fail to invest even a tiny fraction of that spending into R&D activities to increase the impact of their own funding, and very rarely systematically experiment with different ways to design and run their funding competitions. As a result, research funders are missing out on opportunities to achieve their goals in a more cost-effective way, and hence to further accelerate the progress of science by maximising the impact of their funding.

While there are many types of experiments, all of them have one characteristic in common: learning. They are intentionally set up to learn and thus have a clearly structured learning strategy, defined ex-ante, that generates new information, evidence and data. Therefore, a pilot "trying something new" is not a real experiment, unless the systems and processes required to learn from it are also put in place. Experimental organisations systematically set out hypotheses, design experiments to test them, and gather data to validate or reject them.

In the same way as scientific research experiments, policy experiments can seek to achieve different objectives. They can be focused on exploration and discovery, in order to understand how the world works (e.g., to diagnose whether there is bias in peer review). Alternatively, they can also be used with an impact evaluation objective, in order to find out what works, when, and for whom. Sometimes these may evaluate the impact of a single programme, test the impact of small tweaks in a programme, or compare the impact of two or

[1] This section partly builds on Bravo-Biosca, A (2019) "Experimental Innovation Policy," in NBER Innovation Policy and the Economy, Volume 20, pages 191-232, edited by Josh Lerner and Scott Stern, University of Chicago Press.

more different programmes (e.g., which programme design is more effective at increasing interdisciplinary collaborations?).

Lastly, it is becoming more common to use experiments to optimise the processes used in the delivery of a programme. These experiments do not seek to measure whether a programme's ultimate objectives are achieved, but rather to improve one of the steps involved in the delivery of the programme (e.g., what messages are more effective at increasing the number of funding applicants from minoritised groups?).

The experimental process typically involves three phases, and in this handbook we provide tools, methods and case studies for each of them:

1. Diagnosing the issue: before jumping into solutions, an experimental funder will spend time and resources to understand what really is the issue – which will help in the next step;
2. Designing a solution: ensuring the proposed solution actually address the challenge;
3. Testing the solution: evaluating the solution using robust experimental methods, paving the way to scaling the ones that work

The experimentation process does not end when the results of the test become available. Instead, organisations that have successfully embraced a culture of experimentation not only set up experiments, but they also make sure the resulting learning and evidence is used in decision-making, scaling-up successful ideas while continuing to iterate and experiment.

Embracing experimental approaches into an organisation provides multiple benefits. First, the evidence created can help save money because, despite investing a little more upfront in learning and evaluation, experiments allow funders to "weed out" ineffective activities early on. Experiments can also help increase the impact of existing programmes by constantly testing tweaks in the way they are delivered. Experimenting with new programmes can strengthen their design from the outset, by testing different versions or components of a programme, and understanding how they fit together. When it comes to deciding which programmes to scale, randomised trials are especially well-suited to inform decisions, because their results typically come in the form of a robust quantitative estimate that can be easily used to do a cost-benefit analysis.

Lastly, the discipline required to undertake experiments encourages a much more fine-grained look at an organisation's data and processes, which provides substantial benefits on its own (such as helping to explore why organisations do things in a particular way and whether the assumptions underlying their decisions are actually justified). Organisations that can provide better evidence on their decisions and their impact can put forward a better case to their funders, and equally important, this also supports enhanced trust and buy-in from stakeholders (including their own funding applicants).

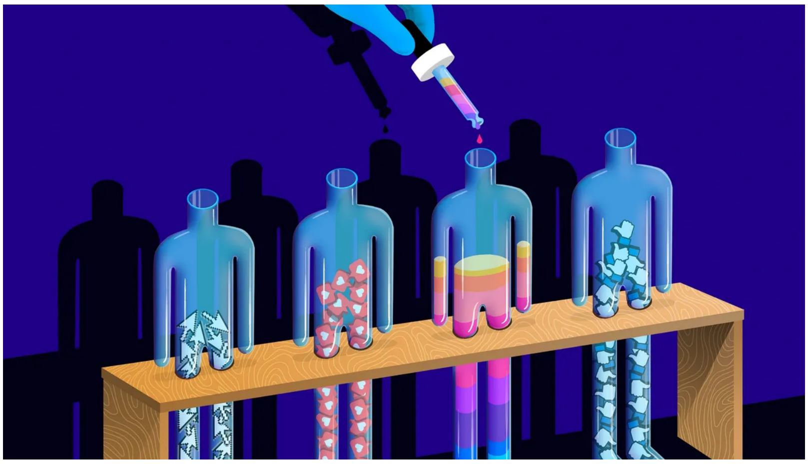
An often overlooked benefit of experimentation is how it encourages organisations to become more agile and innovative, continuously searching for new ideas to test rather than defaulting to the status quo (the "what if" question). In addition, experimentation helps de-risk the process of exploring new ideas. By starting small and testing effectiveness early, experiments can in fact make it easier for risk-averse organisations to sample new approaches and venture into more innovative fields, without having to commit large amounts of resources (or reputation) in the process. As with any other innovation, some experiments will undoubtedly fail, but these are "good failures" that create useful knowledge and prevent "bad failures" from happening. In other words, they are small-scale, controlled, and essential if we want to learn about what works in an uncertain and complex world.

While embracing policy experimentation is a substantial change from business as usual, starting the experimentation journey requires only a few small steps, which even organisations with limited resources can take. Many fear that experiments are too complex and disruptive of the status quo - assuming any trial must set out to randomise large sums of funding or radically alter the way a programme is run. Yet as this handbook shows, there are many ways to become experimental, and many potential reasons to do so.

The value of embracing experimentation has been demonstrated in many fields, such as health, development or education. A well-known example is the MIT Poverty Action Lab (JPAL), founded by Nobel prize winner's Abhijit Banerjee, Esther Duflo, and Michael Kremer, which has set up over 1000 experiments to find out what are the most effective interventions to fight poverty, helping to improve the lives of millions and radically transforming research in this field. Similarly, the UK-based Education Endowment Foundation has conducted over 150 experiments involving more than 14,000 schools and 1,500,000 pupils in order to test different ways to improve educational outcomes. Both of these experiences have shown how our priors, beliefs, hypotheses or models can be wrong, and therefore that there is no substitute for testing new programmes rigorously in the real world.

Despite frequent calls from the research community to increase experimentation in the field of science funding (e.g., Azoulay, 2012; Azoulay and Li, 2020; Boudreau and Lakhani, 2016), and initiatives like RoRI, the Innovation Growth Lab or the Laboratory for Innovation Science at Harvard, we have not (yet) seen a similar take up of experimentation in this field. But support for increased experimentation exists in the research community at both the level of researchers and government. Experimentation offers a path to address dissatisfactions with processes of funding allocation and improve its effectiveness and efficiency. The examples and case studies presented in this section demonstrate both the feasibility and value of this approach.

3. Tools for experimenting with research funding



This section will introduce the main tools and methods you can use to experiment in your funding processes. As described in the previous section, we have broken down the process of experimenting into three phases. We start with **diagnose**, which revolves around understanding what the problem *really* is about. In the diagnose section, you will learn approaches to learn more about the challenge you are facing, interrogate your assumptions, and get to the root causes behind the issue. Next, the **design** phase focuses on developing an intervention to address the problem you have identified. The final section, **evaluate**, presents a number of approaches to track whether your solution worked as intended.

Experimentation needs not be a linear process – you might go back and forth between each phase, as you discover more about the challenge you are tackling. The tools presented are also not mutually exclusive; in fact, you will often benefit from mixing and matching them. For each method or approach introduced, we have provided examples to illustrate how it might be used in the science funding context. While we hope these are realistic, they remain largely hypothetical. We have also added external links where you can find out more.

3.1 Tools to diagnose

Section 2 discussed a number of issues that funding organisations might encounter. While it is natural to strive to find fixes for them, people can often be too quick to jump to solutions, being confident that they understand the underlying problem. However, it is usually better to spend time really understanding what the issue is. Doing so can help you:

- Target your solution so that it addresses the right problem;
- Understand the limitations of your solution, as you will likely be unable to fix everything at once;
- Not break what's not broken!

In this section we present a range of tools and methods to diagnose the issue you are facing. Which ones you apply will depend on context, but will likely involve three steps: first, *gathering information about the problem*; then, *analysing and synthesizing the problem*; and finally, *defining the problem*. We cover each step in turn.

Gathering information about the problem

3.1.1 Interviews: a useful first step is to conduct a few interviews of users or key actors to check your initial assumptions, and uncover additional ones. This can seem quite basic, but it is often quite crucial to ensure you are headed in the right direction. For instance, imagine your organisation is convinced the reason why early-career researchers do not apply to a specific funding stream is because of the complexity of the application process. Before jumping to any solutions, a simple first step would be to reach out to a number of researchers in that group, and ask them whether this is the case. Read more about interviews **here**.

3.1.2 Shadowing: some insights won't come out in interviews, e.g., because the actors aren't consciously aware of the phenomenon. As a result, it could be more useful to observe actors (researchers, reviewers) 'in the field'. One way to do this is to shadow a process, taking notes as it happens. For example, if you are interested in understanding dynamics among peer reviewers ('do most funding decisions get made by senior members?'), you could 'shadow' a number of panel meetings, using a pre-established plan to track certain actions. Read more about shadowing **here**.

3.1.3 Collecting new data: it might seem obvious, but often the data to check our assumptions isn't there. There are no substitutes for evidence directly addressing a question; collecting the necessary data, while burdensome, can both surprise and clarify. For example, if your organisation is trying to reduce the burden of assessing grant applications on reviewers, it is useful to have high-quality data on how long the reviews take. You could ask reviewers to report times immediately after a review, or automate data collection by tracking the time taken on the review website.

Sometimes this will need additional steps, such as setting up small scale experiments to collect data on alternative approaches. For instance, if you wanted to know the extent to which peer review scores depended on the match between reviewer and applicant characteristics, simply collecting data might not be sufficient, especially if there is not enough variation in the characteristics of your current reviewers. To address this, you could set up a 'shadow panel', introducing more reviewers with different characteristics, and collect data on their scores. Read more about shadow experiments in Section 3.3.2.

3.1.4 Matching assessment and outcome data: many useful questions about the assessment process can only be answered by matching data from peer review to funding or research outcomes (publications, subsequent funding, etc.). You might, for instance, be interested in finding out which is better (in terms of research outcomes): a proposal where reviewers strongly disagree, or one where they agree (assuming the two proposals share the same average reviewer score)? This question can only be answered by matching data on reviewer scores to ultimate project outcomes. Alternatively, your organisation might be concerned with biases in the review process. You might ask whether there are applicant or proposal characteristics that predict project success but not funding outcomes, as studied by Banal-Estañol et al. (2019).

3.1.5 Simulations: for some questions you might already have the data, but you would like to understand better what would happen if things were organised differently. Simulations can be a useful tool here. For example, you might be interested in finding out whether adding additional peer reviewers would improve the accuracy of funding decisions. You could run a simulation, like the one conducted by Graves et al. (2011), to investigate how average scores would change with additional reviewers, based on data from past assessments.

Synthesising the problem

3.1.6 Creating 'personas': a useful way to synthesise the information gathered with the previous tools is to create 'personas'. The idea is to map the different incentives, motivations, and behaviours of relevant actors (applicants, grant managers, peer reviewers, etc.) based on evidence you have collected, categorising them into distinct groups. Doing so can help target solutions to the right actor. For instance, imagine you wanted to foster collaborations between researchers in related disciplines. After conducting a number of interviews and collecting some data through survey, you could taxonomise researchers into 4-5 personas (the 'lone wolf', the 'timid collaborator', the 'enthusiastic partner', etc.). Read more about personas **here**
.

3.1.7 User journey: to identify the pain points faced by those you are trying to help, a useful approach is to map out each step required of them, and come up with reasons at each stage that might explain the problem. For instance, if you're trying to understand why most applications you fund are from a specific type of researcher (e.g. senior, male), you could map the journey of applicants from finding out about the programme to getting the funds, collecting data to identify the steps where other types of applicants drop off. Read more about user journeys **here**.

Defining a problem statement

3.1.8 Point of View (problem statement): having synthesised the key aspects of your problem it is useful to come up with a clear statement of the problem. An approach commonly used in the design thinking process, a Point of View (POV) is a problem statement that is actionable. The POV should never contain any indication to the solution and be wide enough to allow for solutions beyond the status quo. Read more about POV and problem statements **here**.

3.1.9 Problem definition: if you are working in a team to define the problem based on all the facts gathered or multiple Points of Views, the problem definition tool is a helpful way to get everyone on the same page by focusing on what is most important. Read more about problem definition **here**.

Further resources

- Design Thinking Canvas (link)
- Nesta/IDEO Guide (pdf)
- Human Centred Design resources (link)

3.2 Tools to design

Once you have spent time really understanding the problem, it is time to devise an intervention to address it. The beauty of experiments is that you don't need to stick to one specific solution: you can compare multiple approaches – and test which one is best (see next section for more on evaluating interventions). The key is to ensure that the intervention you've come up with will actually address the challenge. In this section we present a number of approaches to help you do that; once again, they can be combined or used individually.

3.2.1 The 'Double Diamond' framework: a useful starting point is to consider the Design Council's *Double Diamond* framework. As discussed in Nesta's *Experimenter's Inventory*, this framework "proposes that creative processes involve a number of possible ideas being created ('divergent thinking') before refining and narrowing down to the best idea ('convergent thinking'), and this can be represented by a diamond shape. But the double diamond indicates that this happens twice – once to confirm the problem definition and once to create the solution. This means that ideas are developed and refined a number of times, with weak ideas dropped in the process". In the previous section, you learned about approaches to open up your thinking to better understand the problem. Having defined the problem, you can again make use of 'divergent thinking' to develop a number of potential solutions. Read more about the double diamond **here**.

3.2.2 Using behavioural insights: one source of inspiration for solutions can be to use insights from behavioural science. If you are trying to influence behaviour (e.g., of reviewers, of potential applicants), it can be useful to rely on behavioural mechanisms that have worked before in other contexts. There are a number of resources available such as BIT's EAST framework, the Behavioral Evidence Hub (B-Hub), and more. An example of using behavioural insight could be as follows: imagine that you are trying to encourage applicants to make their data publicly available by committing to it in advance. You can test making this the 'default' option (whilst allowing them to opt out if they really want), or providing them a 'social norm' nudge highlighting how many of their colleagues have already done so. Read more about using behavioural insights **here**.

3.2.3 Involving participants in the design: as in the 'diagnose' phase, it is always useful to include a variety of relevant actors as you design an intervention. This can help you quickly decide which ideas are promising and which aren't, by taking into account their perspective. A hypothetical example could be as follows: in an effort to increase applications from minoritised groups, you have identified a lack of first-time applicant support as a key barrier to be addressed. You consider designing a programme to provide group-based support to new applicants to harness the power of peer networks. But after interviewing stakeholders – including some from the target group of researchers – you discover that the peer support element would do more harm than good. Read more about participatory design **here**.

3.2.4 Prototyping: once your idea(s) for a solution is starting to take form, it is useful to start prototyping it. The idea of a prototype is to create a simple or mock-up version of whatever the solution will look like. For instance, if you're designing a new application website interface, create wireframes for it; if you're delivering a training programme, outline a basic version of the syllabus. The advantage of prototyping is that it allows you to build a small version of your solution – using fewer resources than the real thing – so you can learn and improve it before investing more in it. There are a number of methods you can use as part of prototyping. See, for instance, this toolkit, as well as this prototype testing plan. This will not be always possible – if your innovation is an entirely new grant call, it will be difficult to come up with a mockup– but it can be a useful approach. Read more about prototyping **here**.

3.2.5 Developing a theory of change: once you have a clear intervention in mind, it is useful to create a theory of change for it. A theory of change is a document that connects inputs and activities to outputs and outcomes. The idea is to sketch out how each activity leads to the outputs and outcomes you are after, which can help you uncover tacit assumptions (e.g. 'if we shorten the grant application form, applicants will need to spend less time on it'). A theory of change can help you plan for the testing phase by collecting the right data and asking the most important questions. Read more about developing a theory of change **here**.

3.2.6 Piloting your solution: when you have narrowed down your options to one specific intervention, it is useful to start with a small pilot before scaling it to a full-fledged experiment. This is especially useful for larger interventions, such as launching a new programme. Piloting helps you iron out implementation glitches and discover issues you might not have been able to anticipate. Depending on the success of the pilot, you may decide to set up an experiment to test it, or to go back to the drawing board to design a different solution, re-starting the process with the new information that you've collated. However, a pilot won't tell you about the overall effect of your intervention – for that, you will need to the tools in the next section.

Further resources

- EAST framework
- B-Hub
- Behavioural interventions by OPSI
- BASIC Toolkit
- Nesta's 20 Tools for innovation in government

- The Experimenter's Inventory – section 3.18 on Prototyping

3.3 Tools to evaluate

By this stage we assume you've spent some time understanding the problem and designing an intervention to address it. Now – before launching into implementing the solution – you will need to think about how you will find out whether it worked.

There are many ways to do this – and often the best approach is a mix of different methods. Here, following the taxonomy in Nesta's *Experimenter's Inventory*, we have divided them into three categories: randomised evaluations; quasi-experimental designs; and other (non-randomised) methods.

3.3.1 Randomised evaluations

Randomised controlled trials (RCTs) are familiar to most researchers, and are widely used in the health sciences (e.g., to test new medical interventions). In essence, in a RCT participants are randomly allocated to receive the intervention or not (or they may receive different interventions). Because (on average) the only difference between participants is the randomisation, researchers can estimate the causal average impact of the intervention on participants.

In the scientific funding context, randomising an intervention might not always be possible. However, there are lots of ways to make the most of the advantages of randomisation without denying any applicant an intervention or treating any of them unfairly. Below we describe different ways to use randomisation to evaluate changes to funding processes:

- **Randomising reviewers:** this is sometimes done as a matter of course, but randomising assessors can be a useful way to estimate their impact on the funding decision. For instance, Boudreau et al. (2016) randomised reviewers to applications to study the effect of 'intellectual distance' (i.e., the proximity between a reviewer's sub-field and that of an applicant) on assessment scores for medical research.

- **Randomising elements of the review process:** this can be done within calls, while maintaining fairness. For example, to understand how an applicant's reputation affects the assessment of their scientific ideas, a funding agency might randomise reviewers such that some would see the proposals anonymised. Each proposal would be read exactly the same amount of times anonymised vs not – so they wouldn't be treated differently. The funder could then compare anonymised and non-anonymised scores, and determine whether there are systematic differences (including for specific groups, e.g., research from less prestigious institutions). Other elements that could be randomised within calls are the evaluation of certain sections (e.g., policy impact), the order in which reviewers view proposals, or the stages at which certain information is shared with reviewers (e.g., the amount of funding being asked).

- **Randomising elements across funding calls:** similar to the above, elements of the grant process could be randomised by call. This would likely only be possible for large funders or for collaborations across funders. It might be especially useful when rolling out a new programme or intervention that needs to be staggered by funding calls anyway. For instance, if you are rolling out an applicant support programme for certain types of researchers, you could start with only half the funding calls in your portfolio. You could then track whether the support increases applications from your target applicant group in those calls compared to the others.

- **Randomising funding (such as partial randomisation):** this is the focus of Part 2 of this Handbook, where we cover how it can be used to address a number of shortcomings in the peer review process. However, it can also be used as an evaluation technique. For instance, if you want to compare two 'evaluation regimes' (e.g., one where decisions are made based on an interview, another where they are made by reviewing a paper application), you could run both processes in parallel (i.e., two separate panels each with their own evaluation regime), randomising funding for applications on which the two regimes disagree. This would allow you to compare the two regimes, while funding high quality proposals (because they would be funded under both regimes) and avoiding proposals that both regimes would reject.

- **Randomising matches:** it is often possible to randomise how different actors are matched to each other, and compare their outcomes to those of matches that never happened. For example, Boudreau et al. (2017) randomly allocated researchers interested in applying to a grant call to sit in the same information session with other researchers (see Box 4.3 in Section 4 for more on this example). They tracked whether researchers who attended the same session were then more likely to co-author an application.

- **Randomising messages:** sometimes we want to find out what is the best way of framing a funding call, or an element of it. A simple way to do this is to randomly vary the messages you send out to potential applicants. For example, a funder might vary the language used to encourage researchers to apply for a programme, using existing evidence to encourage applications submitted by researchers from minoritised groups. (See Box 4.1 in Section 4 for a case study on this type of messaging experiments). Similarly, you could experiment with how you communicate with reviewers (e.g., how criteria are communicated to them), or grantees (e.g., how flexibility around the grant is presented), or even unsuccessful applicants.

3.3.2 Shadow experiments

Sometimes changing the funding process might not be feasible – for instance, because the organisation is reluctant to change anything before evidence is available. In these cases, it might be easier to set up a 'shadow' review panel on which to safely test a new intervention. For instance, imagine you are interested in significantly reducing the length of a grant application to reduce the burden on reviewers. One step could be to ask a shadow panel of qualified reviewers to assess only certain sections of applications – the ones considered crucial to make funding decisions – without using their scores to make funding decisions.

You could then compare their rankings to the actual rankings of the 'status quo' assessment process. If they don't differ by much, it would be a good indication that the shorter version of the application is valid. Notice that this could be done by applying randomisation (in the example above, by first creating a pool of reviewers, and then randomising which form the 'status quo' and 'shadow' panels) or not. This experiment does come with several caveats: reviewers need to be from the same pool; shortening the application as a result might change how applicants write their grants, etc.

3.3.3 Non-randomised and quasi-experimental designs

Sometimes it will not be possible to randomise at all, because of ethical, operational, reputational or legal constraints. There are alternatives that attempt to mimic random allocation. Because they do not actually randomise, they rely on additional assumptions. These approaches can often be used ex-post; but it is often best to plan for their use in advance, to ensure that everything is in place – such as collecting the right data from the outset. One key limitation is that they often require even larger samples compared to the randomised approaches outlined above, and very high quality data. Examples include:

- **Regression discontinuity designs (RDD):** this method leverages the fact that participation in a programme is often ruled by arbitrary cut-off points, such as application or scoring thresholds for funding. On the basis that participants (e.g., applicants) on either side of the threshold are otherwise quite similar, RDDs allow you to estimate the impact of taking part in the programme or receiving the intervention. For instance, if your organisation provides application support to early-career researchers, defined as having completed their PhDs no more than eight years before applying, you could compare applicants just above and below that threshold to estimate the impact of providing support.

- **Matching:** this method involves finding a group of individuals that are similar to the participants based on a number of observable characteristics, to which the participants can be compared. The quality of the comparison will depend on how well the two sets of individuals can be matched. An underlying assumption is that the two groups do not systematically differ in terms of 'unobservable' characteristics (such as motivation). For instance, to test an applicant support programme, you could find a group of researchers that match the applicants receiving assistance (e.g., in terms of field of expertise, seniority, previous publications, etc.).

3.3.4 Other (non-counterfactual) methods

In all the methods described above, the goal was to capture the effectiveness of an intervention by comparing those who got it to a *counterfactual* – i.e., a similar individual or organisation that was similar in many ways but did not get the intervention. Sometimes this type of comparison might not be feasible at all, or it might not be suitable to the circumstances, such as when the timelines are very short and it is more useful to get quick, suggestive results now rather than robust results later. In these cases, other approaches might be more appropriate.

3.3.5 Pre-post tests

This approach simply compares the effect of an intervention on the individuals who took part. For instance, a funder running a funding programme for early career researchers might want to add a mentoring intervention. It might not be possible to randomise this intervention (due to reputation concerns), and the funder might conclude that quasi-experimental evaluation designs might also not be suitable (e.g., because the sample is too small). In this case, the funder could simply compare the researchers before and after receiving the mentoring, for instance by assessing whether it has helped them employ better research strategies.

Moreover, we are not always exclusively interested in the impact of an intervention. Often we (also) care about how a new policy affects stakeholders (e.g., are reviewers happy with the new change in assessment procedures?) or whether it can be implemented as intended (e.g., does the new process lengthen review timelines and increase organisational costs?). These types of questions can often be addressed using some of the 'Design' tools we discussed above, such as prototyping.

3.3.6 Ensuring that you learn from your experiment

The crucial difference between an experiment and 'just trying something' is the learning process. Sometimes, experiments will have null results – you find that your proposed solution did not have the effect you hoped for. This need not be a failure; after all, you will have learned something useful. But you *do* want to avoid null results that are due to other factors, such as a failure to implement your solution properly. For instance, imagine you are testing the effect of anonymising proposals by removing the title page in applications. If, at the end of the experiment, you find that most reviewers saw the applicant names because they were often mentioned in other sections, you will not have learned much about the advantages of blinding. To avoid this type of failure, it is useful to think about the following:

- What could go wrong? In particular, if what I hope will happen does *not* happen, what might explain that?

- How can we change the design of the experiment to avoid that?
- Is there any data or information I should be collecting to track alternative explanations?

To do this, a useful exercise to conduct before an experiment is a 'pre-mortem' – where you assume the intervention was a failure, brainstorm why that might happen, and make contingency plans. Read more about 'pre-mortems' **here**.

Further resources

- The Experimenter's Inventory
- IGL's Guide to RCTs in innovation

4. Becoming more experimental

Icons by Scribble Liners. Illustration by Kathleen Krolowec.

In this section, we provide a number of examples and case studies that showcase how the different methods and tools described in Section 3 could be used in the context of research funding. While some of the examples focus on how you could set up and evaluate an experiment, others relate instead to other stages of the process, such as improving your diagnosis of the problem. The case studies will sometimes come from other contexts, but we have tried to ensure that they are related enough to provide a realistic example of something a research funder might want to try.

We have broken down the grant making process into four steps, from attracting the right kind of applicant and proposal, to making funding decisions. This is not an exhaustive list – there are almost unlimited experiments one could conceive of in this field. Experimentation could also be helpful in setting the scope and priorities of funding calls, or even in devising interventions to foster good science beyond funding.

4.1 Attracting applicants

For a funding programme to be successful, the quality and diversity of applicants and their proposals is key. But how do we make sure that we are attracting a strong mix of applicants, and that they are putting in the best version of their proposal? Experiments can help funders test ways to answer these questions, from simple messaging trials to varying structural incentives. Here, we provide case studies based on three relevant questions.

How the framing of a funding call influences who applies (and how)

Funders will often seek to elicit applications from specific groups of researchers (early-career, researchers from minoritized groups, etc.). They might also want to encourage certain *types* of proposals – e.g. 'high-risk, high-reward'. Often small choices regarding the way the funding programme is framed (e.g., the language used to describe it), could have potentially large consequences on these outcomes. There is growing evidence of this from other fields, e.g., in recruitment, where gendered language (Bohnet, 2016) or emphasising the competitive nature of a job (Flory et al., 2015) may also inadvertently discourage qualified women from applying.

If this is a concern for your organisation, you could:

- Conduct interviews with (a) applicants who started but did not finish their applications or (b) researchers in your target group who have not applied, to explore what elements of your call (would have) discouraged them.
- Carry out a messaging experiment, where you vary the language used to describe the call, to test which alternatives are more successful at receiving applications from your target group. You could then also track how those applications fare in the peer review process, and their ultimate outcomes if they are funded (see Box 4.1 for a real-life case study in the social innovation context).

Box 4.1 – Case study: A messaging experiment on applicant motivations
In the context of a grant programme for social entrepreneurs, a group of researchers studied the effect of various messages on applicants and their proposals. After entrepreneurs submitted an Expression of Interest (EoI), the funder sent them an email with information on how to apply. The email was the same for all applicants, except a short paragraph emphasizing either: A. The **social aspect** of being a social entrepreneur – highlighting the 'opportunity to make a difference by helping transform communities and tackle the many social challenges we face". B. The **cash element** of the programme – reminding them of the amount of money they could receive if successful. C. The **support** they would receive – in the form of 1-1 meetings with an 'Award Manager' – if they were accepted. The researchers were interested in how these messages would affect who would end up submitting a full application, and how much effort they would put in. Because the emails were randomised, and the three groups receiving them were large enough, average differences could be attributed to the messages themselves. They found that the groups exposed to the 'extrinsic' rewards (cash and support) had fewer candidates, and fewer applications targeting disadvantaged groups; however, their proposals were more likely to be successful in the review process. But when they tracked the funded projects on a number of relevant outcomes, they found that the social enterprises in the extrinsic reward groups were less likely to be successful compared to the 'social' group. **Read the full paper here.**

How incentive structures and funding criteria affect applicants

Researchers decide to apply to a funding call partly based on the expected risk/reward ratio, which varies according to their characteristics and the call. Varying the requirements (e.g., how long the application should be) or the expected reward (e.g., the expected likelihood of getting funded) can change this calculation – possibly in different ways for different kinds of applicants. This can ultimately determine who applies, and the content of their proposals.

There are different ways to use experimentation to find out how changing your incentives and evaluation would affect your applicant pool:

- **Create personas**, based on qualitative evidence you collect, on the type of potential applicants and what motivates them. Use the personas to **prototype** alternative funding schemes.
- Use **behavioural insights** to understand what obstacles might be restricting the diversity of your applicant group. For instance, one study (Gee, 2018) which attempted to increase the number of women applicants for a role found that clearly communicating the number of people that had already applied was effective – potentially by signalling the value of the job role (a *social norm*).
- Experiment with **evaluation criteria.** Of course, all applicants should be subject to the same criteria. However, you could conduct a messaging experiment where varied the *salience* of particular criteria.
- Experiment with the **scope of the funding call**, by e.g. opening different streams. For instance, Howell et al. (2021) found that when the US Air Force's SBIR programme launched an 'Open topics' funding stream (where applicants could put forward their own topics) along its existing 'Conventional topics' (where topics were imposed by the agency) the type of applicants and their outcomes when funded changed radically. This type of changes can be evaluated with a pre-post test.
- Change the **incentives** by altering the funding amounts in different calls.

How funders can encourage new collaborations among applicants

Often funders may struggle to attract applications from collaborations beyond the 'usual suspects'. This could be due to researchers facing frictions when forming collaborations (e.g., they stick to a circle of co-authors they know, they work primarily within their department, etc.). This effect could potentially have stronger negative consequences for specific groups of researchers (e.g. women, or under-represented minorities, see Ductor et al. 2020). Moreover, funders might want to promote multidisciplinary collaborations – which are often hard to create in the first place. To promote new or different collaborations, you could:

- Collect and analyse **data on existing collaborations** – especially more unusual ones (such as between very different disciplines) – to identify potential opportunities and gaps. If the data can be matched to outcomes, you could estimate which possible collaborations might have the greatest potential.

- Organise **pre-application sessions** that facilitate collaborations. In these sessions, randomise the match between researchers (see Box 4.2 for a real-life example of an experiment conducted at Harvard Medical school).

- You could also **randomise matches** between researchers with different characteristics. This could be specifically aimed at creating multidisciplinary collaborations, by inviting those from different departments within an institution (or across them). You could require (or just strongly encourage) researchers interested in applying to your programme to attend a pre-grant 'coffee and research chat' with a matched researcher from a different discipline, tracking whether this leads them to co-author an application (and if so, whether they are more/less successful).

- You could **test matching algorithms** based on different types of proximity (e.g, based on the fields and technologies they work on, the challenges they are trying to tackle, or the researcher characteristics, such as language, culture, gender or seniority). By randomising whether you share information on the suggested matches with researchers (e.g., showing them only half of the algorithm's recommendations), you could estimate how useful they are.

- You could also **use behavioural insights** to promote more collaborations at existing events. For instance, Bapna and Funk (2020) experimentally tested two interventions to reduce networking barriers to women professionals in an IT event, designed to help participants (both men and women) search for contacts similar to them, and to connect with others. The interventions were found to increase the number of contacts women participants made, and even their odds of changing jobs subsequently.

Box 4.2 – Case study: Randomising researchers' matches to foster collaborations
In the context of a grant opportunity for medical researchers at Harvard University, a group of researchers conducted a field experiment to test the impact of reducing frictions to collaboration. As part of the grant programme, would-be applicants were required to take part in one of three information sessions. During these events, participants were asked to share their research ideas in small groups of other potential applicants. Each researcher was randomly assigned to a group. This allowed the authors of the study to track whether any pair of participants in the same group ended up co-submitting a grant application, and compare them to pairs of researchers that were not matched. In other words, was *Researcher X* more likely to collaborate with *Researcher Y* – who was in the same information-sharing group – compared to *Researcher Z* who was randomly allocated to another group? The experiment found that this simple intervention increased the probability of collaboration by 75%. What's most striking is that all participating researchers already worked in the same institutional context (Harvard University or the Harvard Medical School system of hospitals and centres), and were based in the same geographical area. **Read the full paper** here

4.2 Selecting reviewers

Peer reviewers are central to the grant assessment process, and so selecting who is involved will determine to a large extent who you end up funding. Experiments could help your organisation navigate a number of questions related to how the reviewers are selected.

Choosing the right number of reviewers

Funders have to navigate a fine balance between having too many reviewers (which can be costly and burdensome to organise) and too few (which might be detrimental to the quality of proposals selected). To explore this trade-off, you could:

- **Leverage existing data** on your past funding calls to run a simulation estimating the optimal number of reviewers in your context. This is what Graves et al. (2011) did in the in the context of a medical research programme, finding that additional evaluators would not significantly decrease the percentage of applications in the 'grey zone' (i.e., for which the funding decision depended on who sat on their review committee).
- If your starting number of reviewers is low compared to similar programmes, you could **experiment with adding additional reviewers** through a shadow experiment – i.e. maintaining the same amount of reviewers for the official selection, but asking a number of additional (qualified) evaluators to assess the proposals. At the end you could check whether using their scores would have led to 'better' funding decisions (in terms of programme outcomes, such as publications).

How reviewers' expertise affects funding decisions

Evaluators are typically chosen for peer-review based on their subject matter expertise. But what is the optimal level of proficiency in the proposal's sub-field? Proximity to the field might introduce bias – positive, if the evaluator elevates their topic area above others; or negative, if they are better able to spot minor shortcomings. Sometimes more than subject-matter expertise might be sought, e.g., if one of the goals is for the research to be commercialised, or have policy impact, knowledge in those areas could be sought.

To answer these questions, you could:

- **Conduct interviews with your assessors**, e.g. as cited in Li (2017), where reviewers acknowledge that if a proposal is not in their own subfield, it's "not doing science".
- As in Li (2017), **match review scores to project outcomes** (such as publications) to test whether assessors are biased based on their expertise. Using an observational approach makes it more difficult to fully capture the causal link between the two, which is why randomisation – see next point – would be useful.

- **Randomise assessors to proposals**, allowing for sufficient variation in the expertise of each proposal's reviewers. Leverage this random variation to causally estimate the impact of expertise on funding decisions. This is what Boudreau et al. (2016) do in the context of a medical research grant programme (see Box 4.3 for their findings).
- Conduct a **shadow experiment** where you ask evaluators with commercial or policy expertise for their views on proposals (but don't use their scores to make decisions). You could then follow up to see whether funded proposals they scored highly did better along relevant metrics (e.g. citations in policy reports).

Box 4.3 – Case study: Randomising reviewers to understand the importance of 'intellectual distance'
Working with a 'research-intensive US medical school', a group of researchers set out to understand the relationship between the funding decisions and 'intellectual distance' – i.e. how far apart the fields of expertise of applicants and reviewers are. To do so, they adapted the assessment processes of a grant for endocrine-related disease research. The number of reviewers was increased to 142, with varying degrees of expertise in the grant's field. Single-author proposals were solicited and randomised to the reviewers, generating a large number of evaluator-proposal pairs (2,130). By randomly manipulating the distance in expertise between applicant and reviewer, this design allowed the study authors to estimate the effects of intellectual distance on funding scores and decisions. The results showed that evaluators "gave systematically lower scores to research proposals that were closer to their own areas of expertise", with large average effects. The researchers also found that "more novel proposals [were] associated with lower evaluations, with magnitude of effects comparable to those associated with intellectual distance". **Read the full paper** here

How the match between reviewers and applicants affects reviews

Another crucial aspect of peer-review is the specific match between reviewers' and applicant's characteristics. For instance, in a forthcoming paper, Banal-Estañol et al show that panels of a particular funder tend to favour applicants with whom they share certain characteristics (e.g., in terms of prior research performance). This could be an issue in particular for certain groups and research topics. For instance, there is growing evidence that women and Black researchers choose to study different topics than their white male counterparts, and having a homogeneous reviewer pool might lead to lower scores for topics proposed by members of underrepresented groups (Hoppe et al., 2019; Koning et al., 2020). One way to counteract this effect could be to construct panels that include assessors with a wide range of lived experience, not just subject matter expertise.

You could explore these dynamics using experimentation to:

- **Shadow panel dynamics** to track how proposals from specific groups or types of applicants are discussed by different panel members (e.g., applications from junior researchers mentioned by more senior reviewers).
- In a shadow experiment, you could **oversample the number of reviewers** you need for the assessment, ensuring enough variation in the characteristics of interest (gender, institution, seniority, etc.). You could then randomise which assessments are actually used (following usual practice, e.g., three scores per proposal). By tracking funding decisions by applicant characteristics, you could study how different evaluator-applicant matches impact funding decisions.

4.3 Assessing proposals

Once a pool of evaluators is selected, what input should they base their decisions on? And what process should be followed to score the proposals? There is wide variation in the types of questions, personal information and even format used by science funders worldwive (see, for instance, Janger et al. (2019) for a review). Given these differences, what are the optimal choices? The answer likely depends on the context and specific goals of your organisation, which is why you should explore your options through experimentation. Here we propose a few ideas in response to four key questions.

What information is requested from applicants

As in the case of choosing the number of reviewers, there is a trade off between asking too much information (which will increase the amount of time applicants spend – and is thus mostly 'wasted' for unsuccessful ones), and asking too little (which might make it impossible to discern the quality of proposals). It is also important that the way the grant questions are framed matches the information reviewers need to make decisions. Moreover, linked to the question of biases discussed above, the type of personal information requested (and the way it is presented) can have implications for which researchers get selected.

To explore these questions, your organisation could:

- **Use a range of tools** from the 'Diagnose' and 'Design' sections to better align the information that is asked of applicants, and especially the way it is presented. For instance, this could help you pilot and improve on ways to present research quality information (such as the 'Narrative CV').
- If you are concerned that the length of **proposal requirements** is putting off good candidates from applying, you could map the journey of an applicant and collect

additional data on the amount of time and effort each section requires, to better assess which areas of the proposal could be cut out.

- These improvements could then be tested either through a **regression discontinuity design** (e.g., by leveraging applicants just below and above the threshold to provide certain information), or shadow experiments, such as by asking certain reviewers to assess a paired-down version of a proposal (and checking whether their scores align with the original ones).

What information is disclosed to assessors

Not all information collected from applicants needs to be reviewed; indeed, often peer-review is double-blind. Some funders are already studying the impact of blinding or 'hiding' personal information (e.g., at the NIH, or in crowdfunding). Blinding won't always work, as biases will often manifest themselves in multiple ways (e.g., gender differences in writing styles, as examined by Kolev et al. (2019). However, it can be a powerful way to correct for reputation effects so that proposals from less famous applicants and institutions are scored fairly. There might also be interactions between the impact of blinding and proposal novelty or radicalness (e.g., would a radical idea proposed by a young post-doc receive the same technical score as if it had been submitted by a famous professor, and if not, are we missing out on novel research ideas?). There are a number of ways to experiment with these questions:

- **Randomise reviewers to view different information** (e.g. half single-blind and half double-blind) – ensuring for fairness that each proposal is assessed by an equal number of reviewers in each condition (the proposal could be blinded either fully or partially - e.g., blinded evaluation for some criteria, such as proposal novelty, but unblinded for other criteria, such as ability to deliver, for which the researcher track record is relevant). You could then track whether their reviews vary systematically. This is what Tomkins et al. (2017) do in the context of a Computer Science conference paper review (see Box 4.4 for more information).

- Use a shadow experiment to **retroactively re-assess proposals**, altering key characteristics of the proposals. This is what Forscher et al. (2016) did with NIH applications to investigate the existence of race and gender bias.

Box 4.4 – Case study: Comparing single- and double-blinded review
To study the effect of reviewers knowing the identity of applicants, a group of researchers conducted an experiment in the context of a computer science (CS) conference. (In CS, conferences are similar to publications elsewhere – i.e. research appears first there, so full manuscript submissions with peer review are the norm). The researchers experimented with the peer review committee's approach to *blinding*, i.e. whether

the applicant's identity was revealed to reviewers. In the context of this conference, reviewers see all papers beforehand and can 'bid' for the ones they want to review. The study authors randomly divided evaluators into two pools: half were *blinded*, the other half was not. The blinding persisted throughout the process – bidding for, reviewing and scoring proposals. Each proposal was assigned four reviewers – two blinded and two unblinded.

The study found that unblinded reviewers bid for fewer proposals (22% less on average). These reviewers were also more likely to bid for papers from "top universities and top companies". Finally, unblinded reviewers were relatively more likely to give better scores to "papers with a famous author and for papers from a top university or top company". **Read the full paper** [here]

What evaluation format is used

Beyond the information used, there are a number of questions relating to the format in which the application is presented and evaluated. For instance, certain funding calls require multiple stages, or an in-person interview. These choices have implications for the type of proposals that get funded. There are different ways to experiment with the format of applications:

- For multi-stage processes, **match assessment and project outcome data** to check how decisions change at each step. For instance, if your application process has two stages, how predictive are scores from stage-1 of funding decisions *after* stage 2? If it turns out that the top 10% of stage-1 proposals always end up getting funded, is there a way to fast-track them earlier?

- **Compare different evaluation 'regimes'** (e.g. paper only vs paper plus interviews), running each in parallel. You can then track which one performs better (on whichever metric you are most interested in, e.g., burden on reviewers, diversity of the pool of funded researchers, outcomes of selected proposals) in a number of different ways.

- With a **shadow experiment**: use the current 'regime' to make funding decisions, but compare them with the alternative, shadow 'regime' (e.g., would the pool of funded researchers be different if you had ignored the assessment that came out of the interviews?).

- With **partial randomisation**: run two evaluation 'regimes' in parallel, randomising among those deemed eligible by either or both. This is the approach taken by this ongoing RCT (in a different context – funding for high-growth firms). Alternatively, you could only randomise proposals where the two regimes disagree (i.e., reject applications with low scores by both, accept those with high scores, and randomise those for which the two regimes have very different scores).

How the order of reviews affects funding decisions

When assessing applications, organisations make a number of decisions regarding the order in which things are done. Some of these decisions might seem trivial – such as which application should be discussed first – but could have potentially large consequences. These arrangements can be studied through experiments, for instance:

- **Explore potential ordering effects**. For instance, research on evaluations of R&D projects has shown (Criscuolo et al., 2021) that a proposal's chances to get funded decrease if they are discussed immediately after another successful project. You could investigate similar 'ordering' effects simply by using data analysis.
- If you suspect there are ordering effects, you could also **randomise the order** in which proposals are reviewed and/or discussed. For instance, each reviewer could receive their proposals in shuffled order, to ensure no proposal is unfairly disadvantaged.
- Other important ordering effects may occur when **disclosing information to reviewers.** For instance, if panel members eventually find out the other reviewers' scores, at what stage do you disclose this? (See Box 4.5 for a real-life example of how reviewers change their assessments based on other reviewers' scores).

Box 4.5 – Case study: How reviewers update their scores based on others' assessments
How do peer reviewers react to finding out whether their evaluation gave a lower or higher score to a proposal, compared to other reviewers? To address this question, a group of researchers worked with a grant administrator in medical research. In two related experiments, they asked reviewers to assess proposals. Once reviewers submitted their scores, they were randomised to see (a) that other reviewers had given a *lower* score to the proposal, or (b) other reviewers had given a *higher* score, or (c) again their own scores. They were then given the chance to update their score. (Only the original scores were used by the funder). By comparing scores of reviewers in each of the three groups, they were able to better understand the effect of disclosing information on other reviewers' assessments on an evaluator. They found that assessors who saw other reviewers' scores tended to change them *in the same direction*. But the updating wasn't symmetric: reviewers shown scores more critical than their own lowered their scores by a bigger margin than the increase by reviewers shown more favorable scores. **Read the full paper here.**

4.4 Making funding decisions

After reviewers have provided their scores and comments, all funding committees have met, how are decisions actually made? The answer will depend on the specific goals of the funding call, the risk preference of the organisation and other factors. In this section, we discuss a number of ways in which funders can experiment with the ways they make decisions.

Aggregating reviewer inputs

The first step before making any decision is to aggregate the inputs prepared by reviewers. These might include numerical scores, assessor comments, funding recommendations by review panels, and more. There is ample room for experimenting with innovative approaches in this area. For instance, funders will often rely on consensus-based approaches to aggregate the reviewers' disparate choices. If seeking to fund highly innovative projects, this might risk putting more novel proposals at a disadvantage.

For a private sector comparison, venture capital funds often select promising ventures through an advocacy/champion model (i.e. only one member of the committee needs to favour a proposal) for early investments, but migrate to a consensus model with later stages of investment Malenko et al. (2021)). Below we provide a number of examples.

- You could **simulate how alternative score aggregation mechanisms would fare,** and in particular which types of proposals would be more (dis)advantaged. This could require matching assessment and project outcome data (publications, further funding), to track not just changes in the portfolio of funded projects, but also their ultimate impact (at least for funded projects).
- To experiment with **'champion-based mechanisms'**, you could ask reviewers to provide a 'gold star' (as the Gates Foundation does) to their top proposals, alongside traditional scores. You could then evaluate whether this would have funded more novel proposals through a shadow experiment, i.e. still using scores for decisions but tracking how the funding allocation would have differed if you had used the gold stars.
- Experiment with **quadratic voting** (as suggested by Azoulay and Li, 2020), a system that incorporates intensity of reviewer preferences by assigning them a fixed number of 'votes' they can distribute across proposals as they see fit. You could test this with a shadow panel run in parallel or retrospectively, or with partial randomisation between two panels.

Testing different decision-making approaches

The final step in the peer-review process is to make funding decisions. This is often left to administrators or a budget committee. Here again, experimentation can help explore crucial questions affecting the quality of funded projects.

- There is often a **tension between following strict rules** (e.g. fund all proposals above a certain score threshold) **and discretion** (allowing administrators to shift proposals between the successful and rejected proposals). While there are likely various competing rationales (so no system will be perfect), you could test the effect different approaches have by matching assessment data to outcomes (this may require asking administrators what they would do differently if given discretion, if they don't already have it), checking how each performs on metrics of interest.
- Similarly, you could test the value of **using algorithms** in the decision-making process. This could be tested via a shadow experiment or partial randomisation.

5. References for Part 1

Azoulay, P (2012). Turn the scientific method on ourselves. Nature, volume 484, pp 31-32. https://doi.org/10.1038/484031a

Azoulay, P, and Li, D (2020). Scientific Grant Funding. NBER Working Paper 26889. https://doi.org/10.3386/w26889

Banal-Estañol, A, Liu, Q, Macho-Stadler, I and Pérez-Castrillo, D (2021). Similar-to-me effects in the grant application process: Applicants, panelists, and the likelihood of obtaining funds. Economics Working Papers 1801, Department of Economics and Business, Universitat Pompeu Fabra.

Banal-Estañol, A, Macho-Stadler, I and Pérez-Castrillo, D (2019). Evaluation in research funding agencies: Are structurally diverse teams biased against? Research Policy, Volume 48, Issue 7, pp 1823-1840. https://doi.org/10.1016/j.respol.2019.04.008

Bapna, S, and Funk, R J (2021). Interventions for Improving Professional Networking for Women: Experimental Evidence from the IT Sector. MIS Quarterly, volume 45 (2), pp 593-636.

Boaz, A, Davies, H, Fraser, A and Nutley, S (Eds) (2019) What Works Now? Evidence-Informed Policy and Practice. Policy Press

Bohnet, I (2016). What Works: Gender Equality by Design. Harvard University Press.

Boudreau, K J, Brady, T, Ganguli, I, Gaule, P, Guinan, E C, Hollenberg, A, Lakhani, K R (2017). A Field Experiment on Search Costs and the Formation of Scientific Collaborations. The Review of Economics and Statistics, volume 99(4), pp 565–576. https://doi.org/10.1162/REST_a_00676

Boudreau, K J, Guinan, E C, Lakhani, K R, and Riedl, C (2016). Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science. Management Science, volume 62(10), pp 2765-2783. https://doi.org/10.1287/mnsc.2015.2285

Boudreau, K J, and Lakhani, K R (2016). "Innovation Experiments: Researching Technical Advance, Knowledge Production, and the Design of Supporting Institutions", in NBER Innovation Policy and the Economy, volume 16, pp 135-167, edited by J Lerner and S Stern, University Chicago Press. https://doi.org/10.1086/684988

Bravo-Biosca, A (2019). "Experimental Innovation Policy", in NBER Innovation Policy and the Economy, volume 20, pp 191-232, edited by J Lerner and S Stern, University Chicago Press. https://doi.org/10.1086/705644

Criscuolo, P, Dahlander, L, Grohsjean, T, and Salter, T (2021). The Sequence Effect in Panel Decisions: Evidence from the Evaluation of Research and Development Projects. Organization Science, volume 32(4), pp 987-1008. https://doi.org/10.1287/orsc.2020.1413

Ductor, L, Goyal, S, and Prummer, A (2020). Gender and Collaboration. Cambridge-INET Working Paper Series No: 2018/07.

Flory, J A, Leibbrandt, A, List, J A (2015). Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions. The Review of Economic Studies, volume 82(1), pp 122–155. https://doi.org/10.1093/restud/rdu030

Forscher, P S, Cox, W T L, Brauer, M (2019). Little race or gender bias in an experiment of initial review of NIH R01 grant proposals. Nature Human Behavior, volume 3, pp 257–264. https://doi.org/10.1038/s41562-018-0517-y

Ganguli, I, Huysentruyt, M, Le Coq, C (2021). How Do Nascent Social Entrepreneurs Respond to Rewards? A Field Experiment on Motivations in a Grant Competition. Management Science, volume 67(10), pp 6294-6316. https://doi.org/10.1287/mnsc.2020.3793

Gee, L K (2018). The More You Know: Information Effects on Job Application Rates in a Large Field Experiment. Management Science, vol 65(5), pp 2077-2094. https://doi.org/10.1287/mnsc.2017.2994

Graves, N, Barnett, AG, and Clarke P (2011). Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. BMJ, volume 343:d4797. https://doi.org/10.1136/bmj.d4797

Hoppe, T A, Litovitz, A, Willis, K A, Meseroll, R A, Perkins, M J, Hutchins, B I, Davis, A F, Lauer, M S, Valantine, H A, Anderson, J M, and Santangelo, G M 2019). Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. Science Advances, volume 5(10). https://doi.org/10.1126/sciadv.aaw7238

Howell, S T, Rathje, J, Van Reenen, J, and Wong, J (2021). Opening up Military Innovation: Causal Effects of ‘Bottom-Up’ Reforms to U.S. Defense Research. NBER Working Paper 28700. https://doi.org/10.3386/w28700

Janger, J, Schmidt, N, and Strauss, A (2019). International Differences in Basic Research Grant Funding. A Systematic Comparison. WIFO Studies, WIFO, number 61664, June.

Kolev, J, Fuentes-Medel, Y, and Murray, F (2020). Gender Differences in Scientific Communication and Their Impact on Grant Funding Decisions. AEA Papers and Proceedings 2020, volume 110, pp 245–249. https://doi.org/10.1257/pandp.20201043

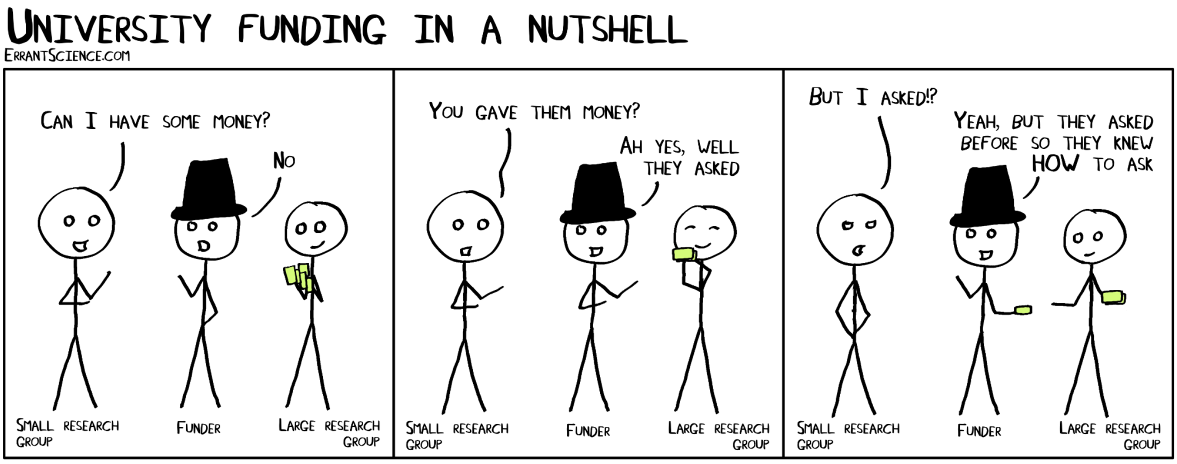
Koning, R, Samila, S, and Ferguson, J (2020). Inventor Gender and the Direction of Invention. AEA Papers and Proceedings, volume 110, 250-54. https://doi.org/10.1257/pandp.20201045

Li, D (2017). Expertise versus Bias in Evaluation: Evidence from the NIH. American Economic Journal: Applied Economics, volume 9(2), pp 60-92. https://doi.org/10.1257/app.20150421

Malenko, A, Nanda, R, Rhodes-Kropf, M, Sundaresan, S (2021). Investment Committee Voting and the Financing of Innovation. Harvard Business School Working Papers 21-131, Harvard Business School.

Tomkins, A, Zhang, M, & Heavlin, W D (2017). Reviewer bias in single- versus double-blind peer review. Proceedings of the National Academy of Sciences of the United States of America, volume 114(48), pp 12708–12713. https://doi.org/10.1073/pnas.1707323114

Part 2:
Funder experiments with partial randomisation



1. Summary

The second part of the handbook collects, summarises and synthesizes information from funder partners in the RoRI consortium that have experimented with partial randomisation of their funding processes. In a subsequent version of the handbook, further experiments planned by members of the consortium will be included.

Our aim is to encourage experimentation among funders and to support them to design effective interventions that are fit for purpose and rigorously evaluated. This is in line with RoRI's broader goal to achieve more efficient, dynamic, diverse and equitable funding systems via rigorous approaches to experimentation and evaluation.

Part 2 first covers problems in peer review that served as early indicators that whether or not changes to processes are required. Next, it describes four experiments with partial randomisation carried out by a small group of funding organisations. These case studies describe in detail each stage of the process and the results, where applicable. A checklist of essential steps completes the Handbook.

Concerns about peer review

Peer review is a long-established mechanism to allocate research funding world-wide, and the main mechanism of quality control and self-regulation in research. But despite the trust of the scientific community that peer review is "the only effective way of properly assessing the quality of research proposals" (Royal Society, 2007), it is not a perfect system. Its limitations have been widely acknowledged, as have concerns about its efficacy and effectiveness (Roy, 1985; Guthrie et al., 2018). Trends in the research system, such as the increase of the number of researchers, high levels of competition, low success rates, and increased attention to the results of public-funded research by policymakers and the wider public, have exacerbated some of these concerns (Bendiscioli, 2019).

Time consuming and slow

Peer review is time consuming, for applicants, reviewers and funding agencies. High levels of competition and reduced success rates in funding schemes (ERC 2021), have the effect that reviewers are asked to evaluate an ever increasing number of applications, and so have less time to dedicate to each evaluation. Reviewers also increasingly deny agencies' requests, so funding managers need to approach many potential reviewers for each evaluation round, which contributes to making the process increasingly arduous and slow.

Conscious and unconscious biases

There is increasing evidence that bias related to applicants' affiliation, ethnicity, age, previous funding success, research area and gender, affect objectivity in selection processes (Ginter et al., 2011; Hoppe et al., 2019; Li, 2017; Murray at al., 2016; Severin et al., 2020; McAllister et al., 2015). These risks increase when reviewers have limited time to dedicate to evaluations.

Lack of efficacy

Studies on whether peer review is successful in identifying the best researchers and the best projects have provided contradictory evidence. Some of these studies have looked at the career trajectories of awarded and rejected applicants, others at their productivity in terms of publications and citations (Li and Agha, 2015; Bornmann et al., 2008; Klaus and Del Álamo, 2018; Fang et al. 2016).

Lack of reliability

Studies have shown that the level of agreement between reviewers evaluating the same proposals is often rather low: ratings and scoring can vary significantly (Sattler et al. 2015; Mutz et al., 2012; Pier at al., 2018). This raises concerns about the validity of funding decisions.

Conservatism and hindering innovation

Because of the limited funding available, peer reviewers have been suspected of playing it safe and selecting the applications that are most likely to succeed, rather than risky and innovative ones. A reason for conservatism is also that it is easier to obtain consensus in a panel with more traditional applications (Benda and Engels, 2011; Luukkonen, 2012).

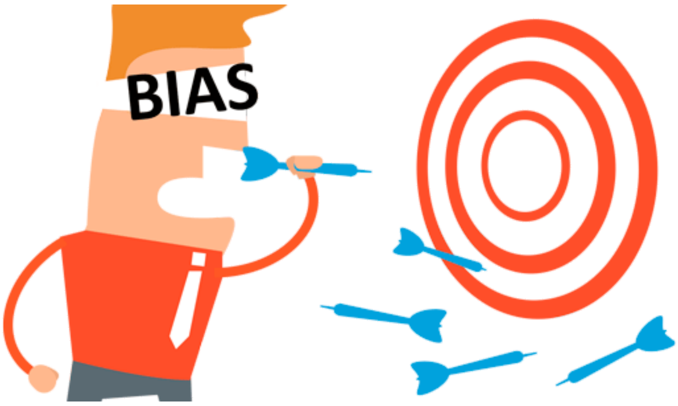
Lack of transparency

The criteria used by funders and reviewers for the evaluation and selection of projects and researchers are not always clearly stated, nor how funding decisions are made. Many funders do not give feedback to rejected applicants about the reasons for rejections. Moreover, review reports are typically not made public, reviewers' names are not always disclosed, and final reports of funded projects are rarely publicly available (Gurwitz et al., 2014).

Unable to make fine distinctions between applicants

While reviewers are generally good at selecting very good from very bad applications, they have difficulties in distinguishing among qualitatively very similar applications. In competitive funding schemes, these applications are typically found in a middle 'grey zone', where reviewers face a daunting task to select among them (Crossley, 2015). In multidisciplinary schemes, the grey zone might consist of applications that are so different that cannot be compared and selected against each other. The forced selection increases the risk of biased decisions. It also increases the time spent by individual reviewers on evaluating and the time needed by panels to discuss and reach a consensus on grey zone applications.

2. An introduction to partial randomisation



Partial randomisation (also called 'focal' or 'targeted' randomisation, or 'modified lottery') is a mechanism that complements peer review for allocating research funding. Unlike randomisation in more general ways, this mechanism is applied only to a subset of already selected applications; hence, the 'partial' or 'targeted' aspect. It relies on peer reviewers' expertise to first identify applications that are worthy of funding and those that are not. Randomisation is then applied to select among the worthy applications. There are slight variations of this procedure, with randomisation being applied after one or more rounds of peer reviewed selection depending on the funder.

2.1 What is partial randomisation?

Selection processes with partial randomisation include different stages, starting with a **pre-selection** by the funders' board or an external review panel to identify applications eligible for funding, and those of poor quality that should be discarded. The applications that are worth funding either are all entered in a draw directly, or are further evaluated by the review panel to define a cut off-line.

When a cut-off line is defined, the top applications within it are selected for funding by the review panel, and the remaining ones enter a **randomisation process**. There is also variability in the tools used for the randomisation procedure, from a manual lottery drum to different kinds of software. Some funders reveal to the successful applicants whether they have been selected by peer review or by partial randomisation, while others have decided not to disclose it.

The idea of introducing an aspect of randomisation in the allocation of research funding is not new (Greenberg, 1998; Brezis, 2007; Graves et al., 2011), but so far only a small number of funding agencies have tried it out. The first public funding agency to introduce it was the **Health Research Council of New Zealand (HRC)** for its Explorer Grants in 2013, followed by another New Zealand funding initiative, the Science for Technological Innovation National Science Challenge (SfTI) in 2014. In Europe, the German private **Volkswagen Foundation (VWF)** used it in a pilot funding scheme, the Experiment! initiative, from 2017 to 2020. An analysis of the experiments in these three funding schemes is given by Avin (Avin, 2015).

The **Swiss National Science Foundation (SNSF)** piloted partial randomisation in its Postdoc.Mobility fellowship scheme from 2018 and 2020, and has since introduced it as a possible support to peer review in all its funding programmes. In 2019, the **Austrian Science Fund (FWF)** started an experiment with partial randomisation in its 1000 Ideas grant programme. Other European funders have shown interest in trying out this modification of the traditional peer review process.

The first funding schemes in which partial randomisation has been trialled are i**nterdisciplinary small-scale grant or fellowship schemes**, with relatively small budgets and a short time duration. Also, with the exception of the SNSF Postdoc.Mobility fellowship scheme, they provide seed funding for **high-risk, bold and potentially transformative research ideas** at an early stage that have less chance of being funded through traditional schemes.

Smaller schemes usually attract a higher number of applications and are more expensive for funders in terms of costs and efforts than larger schemes, while the benefits they bring are smaller. Applying randomisation to small schemes gives the opportunity to funders to offer an important resource for researchers without incurring excessive organisational costs (Bishop, 2018). Also, rather than trialling partial randomisation in already established schemes, all funders have introduced them first as pilots, either at the start of the scheme, or after a few rounds of submissions. At HRC the pilot initiative has then become a standard scheme; SNSF trialled partial randomisation first in a pilot scheme and then decided to include the possibility of using it in all its schemes when necessary.

Other **common elements** of the schemes that have used partial randomisation so far relate to the application process, which can be **simplified**, requiring short applications and a limited amount of information; and **anonymised**, such that applicants have to submit whole or parts of their applications without disclosing personal and professional information. These elements contribute, together with the use of a randomised element, to the funders' aim to reduce bias and limit applicants' and reviewers' work.

2.2 Why partial randomisation?

The funders who have trialled partial randomisation came to this decision by observing the decision-making processes in their funding schemes and acknowledging some of the limitations of traditional peer review procedures. Partial randomisation offered a tool to improve their selection procedures while maintaining the core mechanism for the necessary quality control: expert reviewers' judgment.

High level of competition. The limitations observed derive mainly from the increasing high level of competition in the search for funding: all funders have experienced a general increase in the number of applications for their funding schemes, which is not matched to an increase in their budgets.

Reviewer fatigue. Reviewers are asked to look at an increasing number of applications in most funding schemes, so the time and attention they can dedicate to each application is necessarily shorter. This contributes to what has been called 'reviewer fatigue', adding to the evaluation work reviewers do for scientific journals, award committees and institutional hiring and selection committees.

Reviewers' limited ability to make fine distinctions between similar applications. Moreover, the high number of applications makes peer reviewers' evaluation and selection of the applications increasingly difficult, and highlights peer reviewers' limited ability to make fine distinctions among qualitatively similar applications. These applications are found in the so-called grey area between the top ones that will be funded and the ones that do not reach the funding threshold. In these cases, scientific arguments pro or contra applications are no longer convincing, and the risk that conscious or unconscious biases influence the selection is particularly high.

A panel of reviewers cannot cover all research areas. In the case of multidisciplinary or cross-disciplinary funding schemes, the problem is often that applications might be so different that it becomes impossible to compare them, and the reviewers' expertise does not cover all the research areas of the applications. Thus, niche or underrepresented research areas have fewer chances to be funded.

Peer review is averse to risk-taking. Another challenge addressed by the funders with partial randomisation is the conservative effect of the peer review process. When budgets are tight, bold, risky or unconventional proposals have fewer chances to be funded, as reviewers seem to favour projects with higher chances of succeeding. This is also the case when consensus in a review panel is needed to reach funding decisions. Partial randomisation seemed to be the appropriate selection tool for funding schemes specifically aimed at encouraging high-risk, unconventional and potentially revolutionary research ideas.

The effects of these limitations of peer review can be a combination of bias, unfairness, lack of diversity in projects and researchers, lack of innovative research, and funding agencies' difficulty to find qualified reviewers.

2.3 Arguments for and against partial randomisation

This section summarises arguments about partial randomisation suggested and discussed in the literature. They do not correspond entirely with the actual motivations of funders to trial partial randomisation (which are the subject of a linked study—see Box 2.3 below). More has been written on lotteries in general, but we refer here mainly to what has been written on partial randomisation.

Box 2.3 Why draw lots? Funder motivations for using partial randomisation to allocate research grants

Alongside this Handbook, the RoRI team has recently undertaken a small-scale qualitative study exploring motivations and restraining factors around experimental approaches to grant allocation. Based on interviews with practitioners and leaders from six funding organisations either planning or implementing partial randomisation, the study shows the diversity of motivations at play, and the ways in which funders rank and prioritise these differently. **Read the full study here:** https://figshare.com/s/393eb91c415121b33c9a

Arguments in favour of partial randomisation

- **It eliminates bias and increases diversity.** One of the most popular arguments in favour of partial randomisation is that it would make the allocation of funding objective and reduce the risk of bias related to, e.g., age, gender, ethnicity, affiliation, or research area, as all meritorious applications have equal chances to receive research funding. This would increase fairness and diversity in workforce (Fang and Casadevall, 2016; Frith, 2017; Avin, 2018).
- **It fosters innovation and creativity.** Compared to only peer reviewed selection, advocates of partial randomisation argue that it can bring forward more risky, unconventional or innovative research proposals, which are thought more likely to lead to progress than conventional, lower-risk projects. This would also increase thematic and methodological diversity in the research funded (Brezis, 2007; Gillis, 2014; Fang and Casadevall, 2016; Humphries, 2017).
- **It reduces reviewers' and applicants' burden.** 'Reviewer fatigue' is a concern voiced increasingly due the exponential increase in the number of applications for funding. With the use of partial randomisation, reviewers still have to evaluate submissions, they are at least relieved from the task of trying to stratify qualitatively similar proposals, which are often in a wide variety of subject areas. Partial randomisation might also reduce the time researchers spend writing applications in the schemes with a simplified submission process (Gross and Bergstrom, 2019; Roumbany, 2020).
- **It increases transparency.** Partial randomisation makes the selection process more transparent because all applications considered equally deserving are subject to the same known treatment. This presupposes that the process is clearly explained to all applicants (Fang and Casadevall, 2016; Bendiscioli and Garfinkel, 2021).
- **It is more efficient and saves costs.** Selecting applications through randomisation takes less time so it would increase efficiency in selection processes (Greenberg, 1998; Gross and Bergstrom, 2019). It would also be less expensive to manage for

funders, and the saved costs could be applied to other priorities of the funder (Fang and Casadevall, 2016; Humphries, 2017).

- **Rejected applicants are less disappointed.** Applicants rejected via a partial randomisation system would be less disappointed as they know they have been rejected by bad luck and not by lack of merit (Adam, 2019; Frith, 2017; Bishop, 2018).

- **It makes explicit the role of chance in peer review**. Elements of randomness are intrinsic to peer-reviewed selection processes, such as the chance composition of evaluation panels, the time of the day or the order on the list in which applications are evaluated. Making randomisation a formal part of the process would normalise the fact that chance plays a role (Greenberg, 1998; Graves et al., 2011 a; Humphries, 2017; Bishop, 2018; Roumbany, 2020).

- **It addresses the lack of reliability in peer review.** Given the observed high variability in reviewers' scoring of the same applications (Graves et al., 2011 b), partial randomisation would solve this problem in the middle group of proposals that are most difficult to discriminate (Graves at al., 2011 a; Humphries, 2017; Bishop, 2018).

Arguments against partial randomisation

- **It undermines merit-based decision making**. The main argument against the use of partial randomisation in funding allocation is that it undercuts the fundamental principle in science that recognition is based on merit (Reinhart and Schendzielorz, 2020).

- **It lowers the quality of applications.** In the schemes that have used partial randomisation so far applications are shorter by requirement. This is seen by some as diminishing the value and the quality of applications (Adam, 2019). Another concern is that applicants might be disincentivised to write good proposals knowing that chance plays a big role in the selection process, although applicants know that a first stringent peer reviewed selection will be applied, so they would likely try to write good applications (Bendiscioli and Garfinkel, 2021).

- **It may create stigma and reputational damage.** Researchers funded via partial randomisation might be stigmatised; their merit or quality could be undermined, and this would have negative consequences for their career. Similarly, funding schemes using partial randomisation may be considered less worthy, and so classes of funding schemes with different prestige would be created. The reputation of the funding agency could also be damaged (Reinhart and Schendzielorz, 2020; Vindin, 2020).

- **It undermines trust and credibility.** Admitting that reviewers cannot distinguish among similarly good applications would undermine the trust of policymakers and the public in researchers' and funders' ability to objectively recognise quality and in the validity of the peer review system (Reinhart and Schendzielorz, 2020; Anderson, 2015). It would also give the impression of a lack of will to carry out a difficult task (Barnett, 2016; Reinhart and Schendzielorz, 2020). Using randomisation might not capture deserving researchers, and so reduce credibility of the funding scheme or of the successful applicants (Ioannidis, 2011).

- **It undermines organised scepticism.** Reading and evaluating applications and debating their value with peers contributes to maintaining standards of quality, which would be endangered by selecting applications randomly (Reinhart and Schendzielorz, 2020).
- **It creates new bureaucratic burdens.** For some researchers, the use of randomisation would turn the intellectual process of reviewing and evaluating research proposals into a bureaucratic process, and it would take agency and power out of the hands of researchers and into the hands of administrators (Beattie, 2020).

3. Case studies: funder experiments with partial randomisation

To date, only a small number of funding agencies have tried partial randomisation, although the idea of introducing an aspect of randomisation in the allocation of research funding is not new (Greenberg, 1998; Brezis, 2007; Graves et al., 2011). The case studies below describe in detail the experiments with partial randomisation by **three partners in the RoRI consortium** (Volkswagen Foundation, SNSF and FWF) **and one project partner** (Health Research Council of New Zealand). They provide practical and evidence-based guidance for funding professionals willing or interested in trialling changes to their funding systems.

3.1 Volkswagen Foundation: Experiment! funding initiative[2]

In 2017, the Volkswagen Foundation pioneered a randomized selection within its funding initiative 'Experiment! – In Search of Bold Research Ideas' to support high-risk exploratory research. The Foundation took the challenge to tailor an appropriate selection process for radically new ideas lacking reliable preliminary work. The key features of this scheme were short proposals, double-blind review by a jury, a decisive wildcard for promoting great ideas that cannot find a majority, and as a test run from 2017 to 2020/21, a partially randomized selection element with prior quality assurance. The introduction of partial randomization is considered a success. The scientific community was open for a change, other funders have adopted the idea, and an accompanying research (still ongoing until December 2022) has not found major drawbacks.

The funder and the funding scheme

The Volkswagen Foundation (VWF; VolkswagenStiftung) is the largest German private funder of academic research and education in the humanities, social sciences and science and technology. The Foundation's funds come partially from its capital and assets of about 3.5 billion euros, and partially from the dividends earned from the automobile company

[2] With thanks to Ulrike Bischler and Pavel Dutow of Volkswagen Foundation for their contributions and advice on the text of this case study.

Volkswagen AG shares held by the Federal State of Lower Saxony with the Foundation as beneficiary. The Foundation is not affiliated to the automobile company, but independent and autonomous in its decisions. The Foundation's Board of Trustees and, on its behalf, the Secretary General decide how the funding is allocated.

The focus of its funding was realigned in 2021 with the **three profile areas** "Exploration", "Understanding Research – Evaluation and Science Practice", and "Societal Transformations", as well as the cross-sectional area "Science in Society". Its annual funding volume is about 200 million euro. The Foundation receives about 1000 proposal outlines and applications per year and engages about 500 external German and international leading researchers for written reviews and as panel members.

The Foundation was the first funding agency in Germany, and one of the few funders worldwide, to introduce partial randomisation as a trial in one of its funding schemes, the Experiment! funding initiative, from 2017 to 2020/21. This experiment was in line with the Foundation's goal to be innovative and drive development in the way scientific research is evaluated and supported.

Start and end of the initiative. The scheme Experiment! – In Search of Bold Research Ideas was developed by the VWF Funding division – with feedback from the scientific community – and approved by the Foundation's Board of Trustees in 2012. Funding started in 2013. Partial randomisation was introduced in 2017 after four calls and after the Board of Trustees' approval. The funding initiative ended in December 2020 after eight calls for proposals. It is the Foundation's practice to stop successful initiatives once an impulse is set. The ending of the scheme coincided with the development of the Foundation's new funding strategy.

Aims. The scheme aimed at supporting unconventional, bold, or risky research ideas in a wide range of areas within the life sciences, natural sciences, and engineering. With this scheme, the Foundation wanted to fill a gap in the national funding system, where creative and unconventional research projects do not easily pass conventional peer review.

Eligibility and expected outcomes. Eligible for funding were researchers, from young postdocs to full professors, working at German universities and research institutes, wanting to follow ideas that challenge established knowledge, explore new fields of research, or try out new hypotheses or methods. The scheme gave them the possibility to demonstrate the potentiality of their preliminary research ideas in a short exploratory period of maximum 18 months. The fact that a funded project might fail or obtain negative results was accepted and recognized as a possible outcome and as a valuable learning experience.

Amount and duration of grants. Funding was provided for up to 120,000 Euro (or 100,000 Euro until 2016) for a maximum of 18 months. The funds could be used flexibly for personnel and non-personnel costs.

Success rates. The scheme was very competitive, with 5.051 applications in total and 183 grants during eight calls. Application numbers per call varied between 425 and 824 (last call), and the number of successful ones varied between 13 and 37. The success rate was rather low (about 3.6 %).

Novel aspects. The selection process was rather unconventional from the start. It involved standardized short applications, shortlisting by experienced Foundation staff (with doctoral degrees in the respective or close disciplines), and a double-blind review by an interdisciplinary jury, including an optional decisive vote (a 'wildcard' or 'funding joke') to override the majority opinion.

VWF's motivation to trial partial randomisation in the Experiment! scheme

From its initial launch in 2012, the Foundation observed some limitations of the scheme:

- **Low success rate**. Since its start, the Experiment! funding scheme was very successful in attracting a high number of proposals. However, because of a strict selection and the allocated budget, the success rate was extremely low (around 3.6%).
- **Jury's difficult task.** The jury had a difficult task to select among many high risk-approaches, often without preliminary work and of equally high quality. Most funding recommendations were based on consensus. The wildcard was used only in a few cases. In particular, decisions on proposals from niche disciplines were challenging, since not all subject areas could be covered equally well by the jury of 8-10 reviewers.
- **Lack of diversity in successful projects.** Consequently, the diversity in the funded topics was lower than expected. Further, it was apparent that young investigators and women – though on the shortlist – were underrepresented among the successful applications, supposedly because they lacked experience in proposal writing.

These problems are common in the peer review selection in competitive funding schemes, and other funders are trying out possible solutions to overcome them.

Benefits of partial randomisation in this scheme. Randomisation seemed more suitable than quota for young investigators or for neglected disciplines, because an a priori knowledge of possible underrepresented groups is not necessary: by definition the lot takes a representative sample. With this experimental procedure, the Foundation intended to contribute to overcoming some of the recognised limits of the traditional peer review system, such as conservatism and bias. By shaping a funding scheme that ensures a fair and unbiased selection, all eligible applications received the same chances to succeed, and at the same time, the jury was relieved from taking difficult decisions.

Box 3.1 VWF's motivations to introduce partial randomisation

- It ensures diversity in topics funded (only a limited range of subject/expertise areas can be represented by the jury); increases the success chances of neglected topics and unconventional ideas.
- It increases diversity in the scientific workforce (minority groups have the same chance to succeed as other applicants).
- It avoids implicit bias (the lottery is blind).
- It is an alternative to the jury's difficult decisions in the grey zone, easing the jury's work.
- Motivation to come up with unconventional approaches.
- Unsuccessful applicants are not stigmatised as they did not know whether they had been discarded by the reviewers or by the lot.

VWF's decision process

In 2017, the Foundation decided to tackle these difficulties by modifying the selection process. Using randomisation had already been suggested in the scientific literature. As a private funder, there was enough leeway for VWF to take the risk of introducing a completely new and disputed selection element. The funding budget was also increased to double the number of funded applications per year.

Program directors and the executive management of the Foundation. The idea to introduce partial randomization was first discussed internally among the program directors and the executive management of the Foundation. A randomisation element in peer review after a (coarse) quality check had been met with an open mind, but not everybody was convinced. In particular, the program directors of the social sciences and the humanities still had some reservations.

Jury members. The randomisation approach was also discussed with the Experiment! jury at a regular review meeting. The resonance was mixed: some jury members were positive and even thrilled in taking part in such an experiment, while others opted for staying with the previous scheme and double the number of selected proposals. After the meeting, some hesitant jury members approached the Foundation and mailed their approval after having given some more thoughts. In the end, not a single jury member quit the panel because of the new procedure.

The Board of Trustees. The Board (14 members from academia, politics, economy) had to take the final decision on the deviation from pure peer review as the accepted gold standard. A confidential paper written by a program director was used as the basis of the Board's discussion. The paper included an analysis of the course of the Experiment! funding initiative, its shortcomings (low success rates, underrepresentation of minority disciplines, women, and young investigators among the grantees), potential positive effects of randomisation (no bias, in total over many rounds equal representation, encouraging truly bold ideas) and ways to overcome negative implications (expert-guided pre-selection, as the lot is blind to quality).

After a lively discussion, the Board approved the randomization trial for four rounds and accepted the funding division's suggestion to arrange for independent accompanying research.

The application and selection process

There was one application deadline per year. The selection procedure lasted 3 to 5 months, depending on the total number of applications. Reapplications of proposals rejected in previous rounds were not accepted.

Standardised short applications. The application process was simplified to save applicants' and reviewers' time. The proposal included a 3-page pre-structured outline of the project (max. 1000 words) and a 1-page self-assessment (maximum 300 words) in English, in which the applicants justified along three given questions how their project proposal was in line with the aims of the scheme.

Anonymised applications. In order to focus the jury's attention on the ideas and not on the identity or the affiliation of the applicants, the information submitted was anonymised: applicants were asked to omit in the text and in figures any names of applicants, project partners, participating institutes, and self-citations as references. A short CV with a list of publications was submitted only for internal verification by the Foundation's staff but was not considered by the jury. The process was double blind, as the jury did not know the identity of the applicants and the applicants did not know the identity of the members of the jury.

Three actors were involved in the selection process: qualified funding division staff, an international and interdisciplinary jury, and, since 2017, a lottery. No written reviewers' reports were collected. The jury's funding recommendations were made after a one-day meeting with intensive discussion at the Foundation's premises.

The selection process was in three-stages:

1. **Shortlisting by experienced Foundation staff:** WVF program directors with an academic background in natural sciences, life sciences, and engineering checked all applications and screened out those that were not formally correct or did not meet the funding criteria. A short list of 100-130 anonymised applications was created.
2. **Evaluation and selection by scientific jury.** In a meeting at the Foundation's premises, the jury evaluated and discussed the short-listed applications. The jury consisted of between 8 to 10 international reviewers from different subject areas and with interdisciplinary experiences. Based on the anonymised documentation, they identified a pool of scientifically sound applications that were in principle eligible for funding and rejected weak proposals. This selection was based on five criteria:
    - scientific originality, vision, and unorthodoxy of the research idea;

- the anticipated impact if successful;
- the potential learning effect in case of a failure;
- the high-risk character and
- the appropriateness of a limited exploratory phase to advance the idea.

3. **Selection by the jury and partial randomisation**. The third stage implied a decision by the jury (until 2016), and an additional decision by lot from 2017 onwards. From the pool of between 45 and 85 scientifically sound applications, the jury selected the most convincing ideas, usually between 15 and 20, which were directly earmarked for funding. Each reviewer had one **wild card** (or 'funding joker') to select a project he/she was convinced of and liked to see funded, but on which no consensus could be reached in the discussion. However, the wild card was used only moderately, and the funding recommendations were mostly made by consensus. Afterwards, the pool of fundable proposals, including those already selected by the jury, was entered in a raffle draw. The same number of applications as those selected by the jury (15-20) were drawn and proposed to the Foundation for funding. It might happen that applications were selected twice, by the jury and by the lottery. In that case, these applications were subtracted from the ones selected by the lottery and no further application was drawn. This was done to ensure that bias did not influence the selection at all, and to allow the comparison between the two selection procedures (by jury and by lottery). Finally, The Secretary General (on behalf of the Board of Trustees) approved the proposals selected by the jury and by the lottery for funding.

The preselection by the Foundation staff, and the evaluation and selection by the jury of a pool of worth-funding applications were maintained as part of the process, as the aim was not to substitute peer review, but to complement it with a new mechanism.

In total, up to 37 projects were funded annually, depending on how many projects were selected in the review process. Since the beginning of the trial with partial randomisation in 2017 until 2020/21, 117 projects were funded by the scheme.

Description of the randomisation procedure

The procedure was carried out under supervision of the Foundation legal officer who signed a protocol stating an orderly process or noting any deviations. A **physical lottery drum** was used, a standard toy ('Bingo set') purchased online for about 50 USD.

Once the jury had selected the pool of applications eligible for funding, **balls with the number corresponding to the selected application numbers** were entered in the lottery drum by a staff member. Staff members and reviewers took turns to activate the drum in order to first mix the balls for a certain time, and then softly reverse the rotation direction for scooping up and expelling a ball from the drum.

The number of balls to draw corresponded to the number of applications already selected by the jury. The reasoning behind this choice was twofold: (i) to build two cohorts of the same size, (ii) the assumption was that a high/low number of top proposals by the jury is a good measure for the average proposal quality in the drum and, thus, it was used to leverage quality differences among the calls.

The process lasted about one hour and was public within the Foundation. It was photographed for record keeping.

Communication of the selection results

The applicants received a grant or a rejection letter as usual, but successful applicants were not told whether they had been selected by the review panel or by the lottery. In this way, the concerns that recipients selected by the lottery could be stigmatised because considered less meritorious, were resolved.

Because of the high number of applications and the lack of written reviews, no feedback could be given to the applicants. The details of the selection procedure, including partial randomisation, were published online in the 'Information for Applicants'. Thus, at least in principle, all applicants knew about the procedure. The summaries of all accepted projects were published on the Foundation's website and then the identities of the applicants were disclosed.

Evaluation of the experiment: Accompanying research

When the Foundation decided to trial partial randomisation, it had been tried out only by a few funders, such as the Health Research Council of New Zealand, and not much information on its effects was available. The Foundation decided to commission a **long-term evaluation** of the effects of the modification during a trial period until 2022, when the outcomes of the majority of projects would be clear. The Board of Trustees granted the necessary funds for the evaluation of the experiment, called accompanying research, amounting to about 1.5% of the total grant in Experiment!.

Two concepts were solicited from experienced science of science researchers, from an evaluation agency, and from two university groups who made a joint proposal. Based on a written draft, a presentation and an interview, the Foundation favoured the proposal that involved experienced staff (instead of PhD students) and promised to deliver timely results. Other than that, both proposals had their own merits and appeared to be effective.

In 2018, the Foundation commissioned the external consultancy **EvaConsult GbR** to carry out a long-term evaluation of funding initiative, which will run until December 2022. Unfortunately, due to the Coronavirus pandemic, many projects have been delayed and will run until 2023. The results of the evaluation will be published later, and the data made available so that other funders will be able to make evidence-based decisions on whether to experiment with partial randomisation.

Aim. The purpose of this accompanying research is to find out whether the Foundation was successful in identifying high-risk research and how the partially randomized scheme has contributed to reach this goal. The evaluation focuses on the partial randomised procedure and the whole peer-review process for the funding scheme. Additionally, by comparing the two procedures, jury's decision and lottery, it seeks to understand their effects, and to gain insights into the possible future use of partial randomisation in other funding schemes.

Methodology. As failure of the projects is acceptable as a learning experience as well, the evaluation cannot be done using the criteria usually adopted to measure success such as publications. Therefore, the evaluation is mainly based on information elicited from the grant recipients and the reviewers: online surveys of recipients before randomisation were introduced; online surveys of recipients of projects funded using partial randomisation; and semi-structured interviews with a limited number of recipients at the beginning and the end of their grants. The researchers are asked about their understanding of high-risk research, their assessment of the efficacy of the Experiment! scheme in this respect, and about the outcomes of their projects. The accompanying research also used participatory observation of a review meeting. Furthermore, the researchers took part in the regular 'Forum Experiment!' meetings where grantees had been invited towards the end of their projects to report on their progress as well as disappointments.

Survey participants. The grant recipients were invited to participate in the online surveys in three rounds. In 2018, all 67 recipients of grants between 2013 and 2016 were invited and 50 of them responded (75 % response rate). In 2019, the 28 recipients of the 2017 grants were invited and 25 responded (89 %). In the recent survey, only 25 out of 37 grantees of 2018 filled out the online survey (68 %). This low response rate might reflect time constraints during the Covid-19 pandemic.

Challenge. Particularly challenging for the evaluation of this scheme is the heterogeneity of the funded projects in respect to discipline, research topics, career stage of the applicants, and the inclusion of both individual and collaborative projects. This will make the comparison of the effects of the partial randomised process difficult. In order to reflect different research cultures, the accompanying research is evaluating the natural sciences, life sciences, and engineering separately. The data on the respective career stage (postdoc, not tenured/tenured position, professorship) is collected in the online surveys for distinct evaluation.

Preliminary findings of the evaluation

Fo Forschung über Forschung

Martina Röbbecke & Dagmar Simon

Die Macht des Zufalls

Neue Wege für die Förderung riskanter Forschungsideen?

Martina Röbbecke Dagmar Simon

Peer review as the central assessment instrument in science is often criticised in the scientific debate as risk-averse and structurally conservative. However, it is precisely the funding of "risky" research that the Volkswagen Foundation pursues with its "Experiment!" funding line and has therefore equipped this instrument with a partially randomised procedure. The article presents selected data and results of the accompanying research: How do the funding recipients evaluate risky research and the randomised procedure? Subsequently, the peer review procedure is discussed, particularly with regard to its selectivity in research and post-doc funding. Finally, the article deals with the perspectives for research funding and the question under which conditions the use of a lottery procedure can be recommended.

In der Wissenschaft stellt *Peer Review*, also die Begutachtung durch fachnahe Kolleginnen und Kollegen, das zentrale Element der wissenschaftlichen Qualitätskontrolle dar. Es hat eine lange Tradition und ist eine wichtige Grundlage, um beispielsweise Manuskripte für eine Veröffentlichung in Zeitschriften oder Sammelwerken auszuwählen, für Berufungsverfahren, für Förderentscheidungen über Projekte und für die Verleihung von Stipendien oder wissenschaftlichen Preisen. Zugleich gibt es eine anhaltende und weit verbreitete Diskussion

„Experiment!" geschaffen, mit der ausdrücklich gewagte und „riskante" Forschungsideen in den Natur-, Ingenieur- und Lebenswissenschaften gefördert werden sollen. Antragsberechtigt sind promovierte Wissenschaftlerinnen und Wissenschaftler aller Karrierestufen, die im Erfolgsfall über einen Zeitraum von 18 Monaten und mit 120.000 Euro gefördert werden. Für die Antragstellung sind eine dreiseitige Projektskizze und eine einseitige Selbsteinschätzung erforderlich, die anonymisiert eingereicht werden müssen. Auf dieser Grundlage werden die

The preliminary results of the evaluation of the Experiment! initiative were published in a paper in 2020 (Röbbecke and Simon, 2020—see picture on the left) and are summarised on the Foundation's website.In general, the preliminary evaluation indicated an initial successful outcome:

- The introduction of the partial randomised selection **increased the submission of risky proposals** and the diversity of the projects funded.
- Moreover, it **relieved the burden on the reviewers**. Thus, it seemed to have reached the main goals that the Foundation had set from the beginning of the funding initiative.
- Another main outcome was a **general agreement on the value of the initial peer reviewers' selection** as a quality assurance. Hence, partial randomisation should be used as a complement to peer reviewed selection, not to substitute it.
- Interestingly, the evaluation also showed that **the scope of the funding scheme filled a gap in the funding system,** which is the lack of funding opportunities to develop unconventional or risky research ideas still at a preliminary stage. Although this is not relevant to the concerns about partial randomisation for this case study, this was part of what the evaluation was looking for, and it is an important contribution to the understanding of peer review processes.

Funding recipients' reaction

The preliminary evaluation showed a **high degree of acceptance** of the introduction of partial randomisation by the recipients of the funding. Most of the recipients acknowledged the benefits of the partially randomised selection process in terms of encouraging risky proposals (84%), promoting equal opportunity (92%), reducing bias and conflicts of interest in the selection process (88%), and increasing the chances that risky research gets funded (80%). The majority of respondents was also positive about the simplified application process that did not require too much preparation time.

However, concerns were raised that partial randomisation could lead to the **selection of projects of lower quality** (56%) and could have negative consequences on the recipients' reputation (48%). For the respondents it was irrelevant whether they had been selected by the reviewers or by the lottery, but they appreciated that the Foundation does not disclose this.

Peer reviewers' reaction

The preliminary evaluation found a **high degree of acceptance** of the partial randomisation process also among the members of the jury, as it relieves them from the difficult task of having to differentiate among many equally high-quality applications in the grey zone. They also mentioned that the procedure ensures more fairness in the selection.

Limitations of the evaluation

The positive assessment of the partial randomisation procedure could have been influenced by the fact that the respondents did not know whether they had been selected by the reviewers or by the lottery. More importantly, the evaluation did not include unsuccessful applicants. The sheer number of 5.051 applicants plus the required compliance with the new European GDPR, unfortunately, put a hindrance to this meaningful exercise.

Other stakeholders' reaction to partial randomisation

- **Scientific community.** The reaction of the scientific community was mixed, but predominantly positive. The Wissenschaftsrat (the science advisory body to the German Federal and State governments) endorsed in its white paper on peer review the option of randomization, in particular if used for exploratory research schemes.
- **Applicants.** The reaction of the applicants was positive. Even researchers whose proposals were rejected urged the Foundation to keep the lottery.
- **Press**. The Volkswagen Foundation experiment with partial randomization has been the subject of over 30 publications by scholars and science journalists.

Observation

The high level of acceptance could also depend on the type of scheme it is applied to. Experiment! was a small grant scheme, with a short duration and limited budget and aimed at testing preliminary and risky research ideas, for which not many funding opportunities exist. It is not clear whether researchers would regard positively a partial randomisation with a simplified application procedure used in the selection of larger grants. In the evaluation of the Health Research Council in New Zealand, respondents were sceptical about its use for larger funding schemes not focused on innovative or risky research (Liu et al., 2020).

Box 3.2 Summary of the Results of the preliminary evaluation:

Effects on the selection process:

- It relieves the burden on the review panel (grey zone problem) but not the workload in the Experiment! setting
- Decisions are free from bias and influences due to group dynamics.
- An initial evaluation by reviewers is still necessary, as the lottery would not be able to discern between high and lower quality applications.

Effects on the research funded:

- It increases the diversity in the range of projects funded.

Effects on the kind of applications:

- It encourages the submission of risk-taking proposals.

Reactions:

- It was generally welcomed by the research community, including reviewers and funders.

Lessons learned

Transparency in communication is vital. It is of utmost importance to clearly communicate the motivation, goal and precise procedure of a randomization element in peer review to the applicants and the scientific community as large.

There was no apparent negative impact in introducing randomization – neither in the quality of proposals or in the outcomes of the grants. The following theses are based on practical experience and initial results from the accompanying research:

- In the case of highly competitive procedures, the lottery r**elieves the burden on reviewers** faced with the problem of differentiating quality among a large number of equally high-ranking proposals.
- **Decisions by lot are free of any bias** and of any influences caused by group dynamics.
- **The lot is blind to quality. T**o work properly, therefore, the process requires an initial quality assurance.
- In the event that a reviewer panel does not cover all topics equally well, such a procedure **ensures fairness** among eligible applications.
- Regarding the outcome, **diversity is enhanced by drawing lots.** Often, procedures based on consensus tend to favour established topics and conventional methods ("mainstreaming").
- Accompanying research shows that the partially randomized selection, i.e. including a lottery, **encourages the submission of risk-taking research proposals**.
- The introduction of a randomized element is **broadly welcomed by the research community**, including reviewers, and by an increasing number of funding organizations.
- Selection by lot has to be regarded as a **useful supplement – but not an alternative – to peer review** and cannot replace other forms of scientific discourse.

Box 3.3 References and further reading:

Liu et al. (2020) The acceptability of using a lottery to allocate research funding: a survey of applicants. Research Integrity and Peer Review, 5(3)

Röbbecke M & Simon D (2020) Die Macht des Zufalls. Neue Wege für die Förderung riskanter Forschungsideen?. Forschung über Forschung

Apart from the articles above, most of the information in this case study draws on information collected from the Volkswagen Foundation website:

https://www.volkswagenstiftung.de/en/foundation/about-us

https://www.volkswagenstiftung.de/en/funding/our-funding-portfolio-at-a-glance/experiment?tx_itaofundinginitiative_itaofundinginitiativelist%5Bcontroller%5D=FundingInitiative&cHash=29d4f3d9556a5d7f02d3a438b7a91ac7

https://www.volkswagenstiftung.de/en/funding/our-funding-portfolio-at-a-glance/experiment/partially-randomized-procedure

https://www.volkswagenstiftung.de/en/news-press/news/experiment-mehr-geld-fürs-risiko-und-test-eines-losentscheids

https://www.volkswagenstiftung.de/en/news-press/funding-stories/give-chance-a-chance-—-a-lottery-decides-which-daring-research-ideas-receive-funding

https://www.volkswagenstiftung.de/en/funding/our-funding-portfolio-at-a-glance/experiment

https://www.volkswagenstiftung.de/en/funding/our-funding-portfolio-at-a-glance/experiment/infografic-partially-randomized-selection

https://www.volkswagenstiftung.de/sites/default/files/downloads/MB_100_e.pdf

https://www.volkswagenstiftung.de/en/foundation/governance/accompanying-scientific-research/partially-randomised-method

https://www.volkswagenstiftung.de/en/news-press/news/high-level-of-acceptance-for-project-selection-by-lot

https://www.volkswagenstiftung.de/en/news-press/news/8-theses-for-a-lot-element-in-research-funding

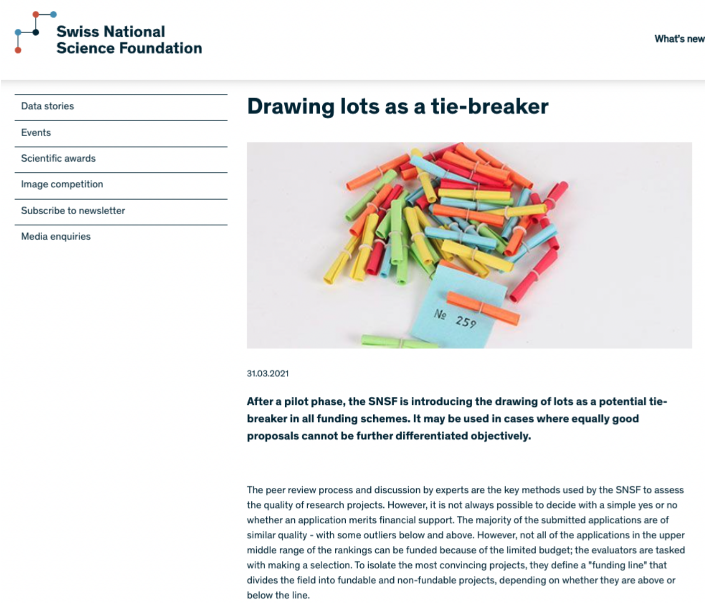
https://www.volkswagenstiftung.de/veranstaltungen/veranstaltungskalender/einzelveranstaltung/forum-experiment-2021

https://www.slideshare.net/RoRInstitute/rorilaunch-5-decisions-krull

Examples of funded projects, by topic

https://www.volkswagenstiftung.de/veranstaltungen/veranstaltungskalender/einzelveranstaltung/forum-experiment-2017

3.2 Swiss National Science Foundation: Postdoc.Mobility fellowships[3]



The Swiss National Science Foundation (SNSF) is the main funding agency for research in all academic disciplines in Switzerland. It is an independent private foundation with a government mandate. It awards annually public resources in the range of about CHF 900 million (about 850 Euro) to more than 6000 research projects through competitive funding schemes.

The SNSF Foundation Council is the highest government body and makes strategic decisions. Peer review for the SNSF funding schemes is carried out by the National Research Council, which is composed of researchers who are mostly based in Switzerland. Divided in four disciplinary divisions, they assess research proposals and make funding decisions. The administrative offices support and coordinate the activities of the Foundation Council, the National Research Council and of local Research Commissions, and are in charge of organising and evaluating the SNSF peer review processes.

Start of the initiative. SNSF trialled and evaluated the use of partial randomisation between 2018 and 2019 in an already existing fellowship programme for postdoctoral researchers, the Postdoc.Mobility scheme. The pilot was evaluated and ended at the beginning of 2020. The fellowship scheme is still ongoing, with partial randomization as an optional tool. A final report was submitted to the Presiding Board of the National Research Council, which subsequently decided in 2021 to allow the use of partial randomisation in all the funding schemes when deemed necessary by the reviewers.

Partial randomisation as an optional tool. At SNFS partial randomisation is used as an optional tool to support reviewers' funding decisions related to grey zone applications, used only if reviewers cannot reach a decision. This differentiates the SNSF partial randomisation procedure from the one trialled by the other funders described in this manual, for which it was or is a regular part of the funding selection.

---

[3] With thanks to Marco Bieri of the SNSF for his extensive inputs to the drafting of this case study. A range of text sources from the SNSF (website, regulations, guidelines), a blog on using lotteries, and a publication on the Postdoc.Mobility pilot were also used.

The Postdoc.mobility fellowship scheme

Aim. The fellowship scheme Postdoc.Mobility is aimed at junior postdocs based in Switzerland who wish to deepen their scientific knowledge and increase their scientific independence during a research stay abroad.

Amount and duration of the funding. The average fellowship amount is about 100,000 Swiss Francs (about 94,000 Euro) for 12 to 24 months and includes a grant for subsistence costs, a flat rate for travel expenses, and a possible contribution to research and conference costs. In addition, fellowship holders can apply for a return grant to finance their initial research period after returning to Switzerland. The return grant is awarded for 3 to 12 months and includes a salary and social security contributions.

Success rate. About 800 proposals are submitted annually, covering all research disciplines. The success rate is currently just below 50%. During the pilot phase involving two calls, 296 proposals were evaluated, 12 (4%) entered the draw, of which eight were funded (four in the February and four in the August selection round).

Novel aspects in the pilot phase. The planning phase of the pilot started in spring 2018 and lasted almost one year. The execution phase started in 2019 and involved two Early Postdoc.Mobility calls. After the pilot, the assessment and selection procedure was changed slightly in a new Unified Evaluation Procedure (UEP) framework aimed at harmonising all SNSF funding schemes. Partial randomisation is still foreseen in the current changed procedure as an optional tool. During the pilot, the following features were trialled:

- **Triage.** Outstanding proposals and those clearly out of the running were recommended by the reviewers for funding or rejection, without additional panel discussion. The remaining proposals in a middle group, that were neither clearly excellent nor poor, were discussed and ranked in meetings of the different disciplinary divisions.

- **Random selection.** If proposals around the funding cut-off could not be objectively differentiated any further, the review panels could decide which ones would be funded by drawing lots.

- **Remote evaluation.** A remote evaluation based on expert review was simulated and its outcomes were compared to the official outcome of the traditional evaluation procedure relying on panel meetings. Partial randomisation could be applied also in the simulated remote evaluation if the proposals around the funding cut-off could not be differentiated any further.

Motivation to trial partial randomisation in this scheme

Partial randomisation was not an element of the Postdoc.Mobility evaluation process before it was introduced as pilot in the 2019 calls. The motivations to trial it were as follows:

- **Increase in the number of applications.** The decision to start the pilot was triggered by changes in SNSF career funding portfolio which would significantly increase the number of proposals for this scheme. SNSF decided to consider ways to adapt the evaluation process to this change.

- **Acknowledged limits of peer review.** Changes to the evaluation and selection process would offer the possibility to mitigate other limits of peer review that had been highlighted in the scientific literature (Guthrie et al., 2018; Fang and Casadevall, 2016). Studies had also evidenced a certain degree of randomness in the review process, for example showing that two independent panels evaluating the same set of proposals reached different decisions (Cole et al., 1981; Hodgson, 1997; Fogelholm et al.,g 2012). Thus, the outcome depended not only on the proposal's scientific content but also on which panel, or reviewer, assessed it. While evaluating its own peer review processes, SNSF had also noted, similarly to other funders, the reviewers' difficulties in discriminating among qualitatively very similar proposals around the funding cut-off (the so-called "grey area"). Because reviewers are forced to decide on which of these proposals to fund, there is a danger of biased decisions being made (Severin and Egger, 2020), of non-relevant criteria being applied, or of existing criteria being weighted inconsistently.

- **Benefits of partial randomisation in this scheme.** The decision was then taken to trial partial randomisation as a "tiebreaker" in such cases where proposals around the funding cut-off cannot be objectively differentiated any further (SNSF, 2021). Partial randomisation was thus also expected to reduce some pressure on the panellists and avoid lengthy discussions. In addition, partial randomisation would make the decisions more transparent for applicants in the grey area, who would know that their proposal – although rejected – was of high quality and would have been funded if the budget had allowed it.

SNSF's decision process

To obtain approval for the trial by the SNSF decision-makers in-depth documentation around partial randomisation was prepared, including a legal assessment and an explanation of the key arguments for using a random selection.

There were no reservations from a legal point of view, as in 2011 the Federal Supreme Court had allowed the use of randomisation and set out the requirements for it: the procedure must be transparent and credible, and can be carried out either physically or electronically, provided that there is a level playing field for all contestants. The Specialized Committee Careers (SC CAR) of SNSF, which oversees the SNSF Career funding schemes, formally approved the trial. The Presiding Board of the National Research Council, the Management of the Administrative Offices, and other internal boards were informed about it.

Among the Postdoc.Mobility **panellists and reviewers**, the introduction of partial randomisation led to some discussion. Specifically, there were concerns that partial randomisation could harm SNSF's reputation and that the evaluation could be done less rigorously. A clear framing and communication strategy was key. SNSF repeatedly pointed to the documented limitations of peer review and stressed that expert peer review would remain the core of the evaluation. Eventually, the use of partial randomisation as a last resort or "tiebreaker" reached a good level of acceptance, which grew with practical experience. This might be due also to the panellists' own experience with having to discriminate between proposals that are neither clearly excellent nor poor.

The application and selection process

The Postdoc.Mobility scheme has two calls for applications per year. Proposals for a Postdoc.Mobility fellowship must be submitted electronically via the mySNF platform. Proposals consist of an administrative part including personal information and information about the research project (e.g., designated host institution, research discipline, abstract, keywords, and requested funding). Further, key documents for the evaluation must be provided: the research plan (max. 8 pages), a standardized CV of 3 pages with a 1-page description of the major scientific achievements, the research output list, a document declaring the net academic age, a statement of mobility, the career plan, and a confirmation of the host institute.

Single-blinded process. The Postdoc.Mobility evaluation process lasts about five months and is single-blinded: the identity of the applicants is known to the reviewers, but the identity of reviewers who evaluate a proposal is not disclosed, in line with the Federal Act on the Promotion of Research and Innovation. However, the members of the evaluation panels are announced on the Postdoc.Mobility website.

The **Administrative Offices** first check the proposals. The ones that do not fulfil the eligibility criteria (personal and formal requirements) or that are obviously inadequate in relation to the scientific content are rejected at this stage. SNSF also verifies whether the requirements relative to research integrity and good scientific practice are met.

Postdoc.Mobility evaluation process during the pilot in 2019

The evaluation process used during the pilot phase of the Postdoc.Mobility fellowships in 2019 was in four steps. A more detailed description was published in 2021 (Bieri et al., 2021). Proposals were evaluated by one of five review panels: Humanities, Social Sciences, STEM (Science, Technology, Engineering and Mathematics), Biology or Medicine.

Latest content Archive For authors

Article Text
Article info
Citation Tools
Share
Rapid Responses
Article metrics
Alerts

Research methods
Original research

Face-to-face panel meetings versus remote evaluation of fellowship applications: simulation study at the Swiss National Science Foundation

PDF
XML

Marco Bieri [1], Katharina Roser [1, 2], Rachel Heyard [3], Matthias Egger [4, 5, 6]
Correspondence to Dr Marco Bieri; marco.bieri@snf.ch

Abstract

Objectives To trial a simplified, time and cost-saving method for remote evaluation of fellowship applications and compare this with existing panel review processes by analysing concordance between funding decisions, and the use of a lottery-based decision method for proposals of similar quality.

Design The study involved 134 junior fellowship proposals for postdoctoral research ('Postdoc.Mobility'). The official method used two panel reviewers who independently scored the application, followed by triage and discussion of selected applications in a panel. Very competitive/uncompetitive proposals were directly funded/rejected without discussion. The simplified procedure used the scores of the two panel members, with or without the score of an additional, third expert. Both methods could further use a lottery to decide on applications of similar quality close to the funding threshold. The same funding rate was applied, and the agreement between the two methods analysed.

Setting Swiss National Science Foundation (SNSF).

Participants Postdoc.Mobility panel reviewers and additional expert reviewers.

Primary outcome measure Per cent agreement between the simplified and official evaluation method with 95% CIs.

Results The simplified procedure based on three reviews agreed in 80.6% (95% CI: 73.9% to 87.3%) of applicants with the official funding outcome. The agreement was 86.6% (95% CI: 80.6% to 91.8%) when using the two reviews of the panel members. The agreement between the two methods was lower for the group of applications discussed in the panel (64.2% and 73.1%, respectively), and higher for directly funded/rejected applications (range: 96.7%–100%). The lottery was used in 8 (6.0%) of 134

1. **Evaluation and scoring by individual panel members.** In the first step, each proposal was independently evaluated and scored by two panel members. The objects of the evaluation were the applicant, the proposed research project, and the host institution, which were assessed using the criteria in the guidelines. Panel members scored proposals on a 6-point scale.
2. **Triage.** In the second step, based on the ranking of the mean scores given to each proposal, three groups of proposals were formed: fund without panel discussion (top group), discuss in a panel meeting (middle group), reject without discussion (bottom group). Each panel member could request discussion of any proposal allocated to the fund or reject group.
3. **Group discussion and ranking of middle-group proposals.** In the third step, members of the review panels met to discuss and rank the middle group proposals. The same scoring system as for the written assessments were used. A final average score was calculated for each proposal, which determined its rank in the final list.
4. **Optional partial randomisation.** Partial randomisation could be used in a fourth step to fund or reject proposals of similar quality close to the funding cut-off, if the panel could not reach a decision. This process was changed slightly at the end of the pilot as part of SNSF's new Unified Evaluation Procedure (UEP) framework (described below).

Description of the randomisation procedure

The tool used for partial randomisation during the pilot in 2019 was a **manual drawing of lots.** Review panels analysed the scores around the funding cut-off and identified potential groups of proposals for partial randomisation. In case two or more proposals could not be differentiated any further based on objective criteria, the Offices wrote their numbers on pieces of paper. Each piece of paper was inserted into an opaque capsule. A member of the evaluation panel then drew the capsules from a transparent bowl one by one. This all was documented in the meeting minutes and video recorded. Partial randomisation was applied only to a few proposals around the funding cut-off (four in the February and eight in the August selection round).

Modified evaluation procedure. A Unified Evaluation Procedure (UEP) will be introduced at SNSF in 2022 to harmonise evaluation procedures across funding schemes. Under the UEP, the scientific evaluation and the funding decisions will be clearly separated. This means that evaluation panels will no longer set the funding threshold or assign proposals for partial randomisation. Their task will be to rank proposals based on their scientific quality and identify those worthy of funding. Deciding the funding line and possible allocation of proposals to partial randomisation will be the task of the SNSF committee that organises the evaluation panel.

Under the UEP, the allocation of proposals to partial randomisation will be based on credible intervals using a Bayesian ranking method that SNSF adapted for this purpose. In contrast to using simple average scores, Bayesian ranking considers missing votes of panel members and their voting behaviour and will allow for more consistent identification of proposals to enter partial randomisation. Due to the separation of scientific evaluation and funding decisions, drawing lots will be carried out by the Administrative Offices. The process will be documented in the minutes and video recorded. A digital tool for partial randomisation, involving a random number generator, might be used in the future after having ensured that no bias or manipulation can occur.

Communication of the selection results

SNSF informs applicants whether their proposal was selected or refused by drawing lots. This approach increases transparency and complies with the San Francisco Declaration on Research Assessment (DORA), which states that funders must be explicit about assessment criteria. The rejection letter outlines why the proposal was not among the best ones, allowing applicants to resubmit a revision. In addition, the rejection letter acknowledges the scientific quality of applications excluded by drawing lots, stating that if sufficient financial resources had been available, the proposal would have been funded.

Stakeholders' reaction to partial randomisation

Applicants. The announcement that funding decisions could be partially reached by drawing lots prompted only a few technical questions by the applicants. No applicant objected to the use of partial randomisation as such, and some explicitly welcomed its use. The experience during and after the trial indicates that the acceptance of applicants is generally high.

Media and politicians. So far, SNSF has not received negative criticism in the media or by politicians for using partial randomisation in research funding allocation. This is likely to be the result of the clear framing and communication strategy adopted, based on clarification and transparency.

Evaluation of the experiment

The Postdoc.Mobility pilot was evaluated by the Careers division of the SNSF Administrative Offices. The division submitted a comprehensive report focused on the lessons learnt to the Presiding Board of the National Research Council. The report was based on comments on partial randomisation gathered from applicants and panellists. The extent to which the review panels applied partial randomisation was also analysed. The aim was to evaluate whether partial randomisation could reach a good level of acceptance among panellists and applicants that would allow SNSF to keep this process for the Postdoc.Mobility and even extend it across other SNSF funding schemes.

The evaluation found that **applicants generally accepted the procedure.** This might be due to the fact that partial randomisation is applied only on a small share of proposals after a rigorous scientific assessment by peer reviewers.

No increase in proposals after the introduction of partial randomisation was observed. Thus, partial randomisation does not seem to attract more applicants, as it is clear that the process is based on merit and not on chance, and that proposals continue to be evaluated based on the same rigorous scientific assessment, whether or not partial randomisation is used.

In general, **the small sample of the proposals that have entered the draw so far precludes the drawing of meaningful conclusions**. Since partial randomisation continues to be used in the scheme, an evaluation of its impact will be possible in the near future.

Based on the experiences gained by the Postdoc.Mobility trial, the Presiding Board of the Research Council decided at the beginning of 2020 to **expand the option** of using partial randomisation for all SNSF funding instruments. **SNSF is thus the first national funder that allows applying partial randomisation in all its schemes.** In any case, peer review remains the basis for scientific evaluation at SNSF, which continuously strives to improve its peer review processes.

An evaluation of the impact of funding on career development is planned comparing applicants who received funding with those who were rejected by partial randomisation. This will be carried out within the Career Tracker Cohorts (SNSF-CTC) project, which started in 2018. The main goal is to gain a better understanding of the career paths of applicants to post-doctoral SNSF career funding schemes, as well as the medium and long-term impact of all SNSF career funding schemes. The results will serve as a basis for the future development of career funding policies and instruments.

Lessons learned

- **Preparation and communication**: SNSF learned from the Postdoc.Mobility experiment that preparing the pilot carefully, clarifying all relevant aspects including the legal context, paid off in obtaining understanding, acceptance and internal approval for the trial. Communication and the framing of partial randomisation were key, and it was fundamental to address any potential misunderstandings readily, e.g., to avoid concerns that random selection could replace peer review.
- **Terminology is important:** Which term should be used to describe the partial randomisation process was discussed by the SNSF legal services and the

Communication division. In general, SNSF is obliged to communicate in a clear, understandable and unmistakable manner. For this reason, it was decided to use the term "drawing lots" or "decisions reached by drawing lots" for external communication (applicants, decision letters, call texts, guidelines, etc.). The process and the term "drawing lots" are also used in the Organisational Regulations of the National Research Council (Article 23, paragraph 5). Internally or for communication with external expert groups, the term "random selection" is frequently used. SNSF avoids using the term "lottery" for its negative connotation (e.g., gambling), which is, of course, not compatible with the rigorous scientific assessment applied at the SNSF.

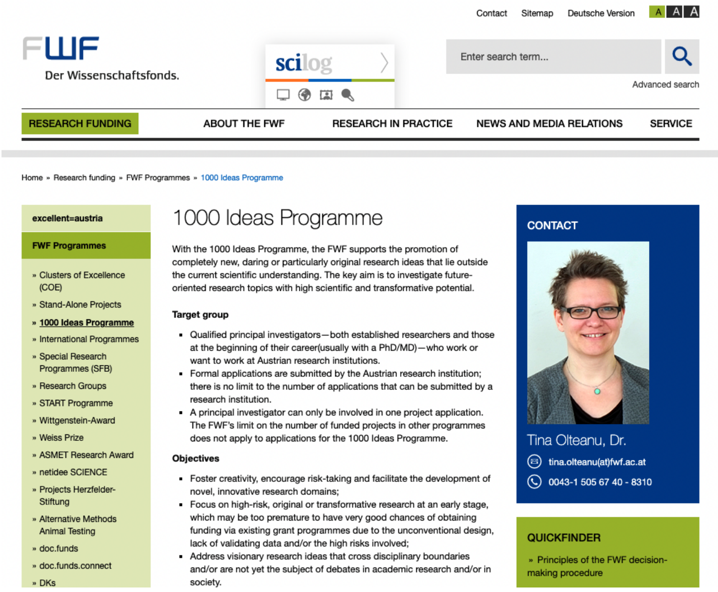
- **Defining the cut-off:** The review panels had to allocate proposals to the partial randomisation group based on visual inspection of the scores, which turned out to be difficult, especially when there was no clear "gap" or "step" around the funding cut-off. Therefore, a method based on Bayesian statistics (Heyard et al., 2021) was developed that allows to rank proposals starting from the votes of the panellists. Instead of establishing a ranking based only on simple average scores, the Bayesian ranking considers missing votes (e.g., when a panellist must abstain due to a conflict of interest) and panellists' voting behaviours. It provides credible intervals for the delineation of groups of proposals for partial randomisation. The availability of this tool at the beginning of the pilot would have been helpful.
- **Experimenting led to further changes:** The Postdoc.Mobility pilot contributed to internal discussions of the SNSF's evaluation procedures. The trial on partial randomisation triggered the development of the Bayesian ranking, a new procedure that ensures that the allocation of proposals to partial randomisation is consistent. Moreover, the triage, which was also trialled during the Postdoc.Mobility pilot, has now become an option in all SNSF funding schemes.

Box 3.4 References and further reading:

1. Organisational Regulations of the National Research Council.
https://media.snf.ch/3j0cWN6L3JPY3ha/por_org_rec_reglement_e.pdf
2. Swiss National Science Foundation Postdoc.Mobility.
https://www.snf.ch/en/XIZpfY3iVS5KRRoD/funding/careers/postdoc-mobility
3. Bieri, M., Roser, K., Heyard, R. & Egger, M. Face-to-face panel meetings versus remote evaluation of fellowship applications: simulation study at the Swiss National Science Foundation. BMJ Open 11, e047386 (2021).
4. Severin, A. et al. Gender and other potential biases in peer review: cross-sectional analysis of 38 250 external peer review reports. BMJ Open 10, e035058 (2020).
5. Swiss National Science Foundation - Drawing lots as a tiebreaker
https://www.snf.ch/en/JyifP2I9SUo8CPxI/news/news-210331-drawing-lots-as-a-tie-breaker.
6. Adam, D. Science funders gamble on grant lotteries. Nature 575, 574–575 (2019).

7. Singh Chawla, D. Swiss funder draws lots to make grant decisions.
Nature d41586-021-01232–3 (2021) doi:10.1038/d41586-021-01232-3.
8. Bieri, M. Using lotteries to allocate research funding: Perspectives from Switzerland
https://www.rug.nl/library/open-access/blog/using-lotteries-to-allocate-research-funding-perspectives-from-switzerland?lang=en
9. Postdoc.Mobility fellowships and return grants: Guidelines to submit a proposal via mySNF
https://www.snf.ch/api/media/en/Q61O7mz9Im905hJX/weisungen-pm-en.pdf
10. Federal Act on the Promotion of Research and Innovation
https://www.fedlex.admin.ch/eli/cc/2013/786/en.
11. SNSF Postdoc.Mobility Regulations
http://www.snf.ch/SiteCollectionDocuments/Reglement_PM_ab2021_en.pdf
12. Heyard, R., Ott, M., Salanti, G. & Egger, M. Rethinking the Funding Line at the Swiss National Science Foundation: Bayesian Ranking and Lottery. arXiv:2102.09958 [stat] (2021).
13. San Francisco Declaration on Research Assessment (DORA). https://sfdora.org/
14. Swiss National Science Foundation Career Tracker Cohorts (CTC)
https://careertrackercohorts.ch/
15. SNSF Funding Regulations
https://www.snf.ch/de/api/media/en/lCCrvpOHZ38Hbq5Y/allg_reglement_16_e.pdf
16. SNSF Database P3. https://p3.snf.ch/Default.aspx

3.3 Austrian Science Fund (FWF): 1000 Ideas Programme[4]



The Austrian Science Fund (FWF - Fonds zur Förderung der wissenschaftlichen Forschung) is the main public funding agency for basic research in science and the humanities in Austria. In 2020 its funding budget, allocated by the Federal Ministry of Education, Science and Research, was about 243 million euro. A new three-year agreement has increased the FWF budget and will allow 806 million euro to be invested in funding basic research projects between 2021 and 2023. The FWF Board takes the final decisions on the funding of more than 2000 research proposals per year that are selected by about 5000 international scientific reviewers. The FWF's funding approach is strictly bottom-up, i.e., the research topics are proposed by the investigators in all areas and all schemes and there are no thematic calls. FWF's mission is to support the development of Austrian basic research at a high international level.

The 1000 Ideas Programme

Start of the initiative. FWF was among the first public funders in Europe to experiment with partial randomisation in 2019 when it introduced it in a pilot grant funding scheme, the 1000 Ideas Programme, which provides seed funding for radical new and bold research ideas that have the potential to transform established scientific knowledge in all disciplines. There have been two calls for applications so far, one in November 2019 and one in November 2020. The scheme is still running.

Aims. The aims of the scheme are to encourage creativity and risk-taking; to foster the development of radically new and innovative research domains; and to support research ideas at an early stage. Project proposals of this kind have a hard time obtaining funding in the traditional funding schemes. The 1000 Ideas Programme gives investigators willing to take risks the opportunity to provide initial evidence of the feasibility of bold and potentially transformative projects in all disciplines and fields. The possibility that a project might fail is taken into account as an acceptable result of risk-taking.

---

[4] With thanks to Uwe von Ahsen, Elisabeth Nindl, Falk Reckling, Tina Olteanu and Ralph Reimann of the Austrian Science Fund—FWF for their contributions and advice on the text of this case study.

Amount and duration of the grant. The annual budget of the scheme is of 3 million euros for a period of 3 years. Projects can be funded for a minimum of 6 months and a maximum of 2 years and receive funding up to 150,000 euro in total. The costs that can be covered include personnel costs for the investigator (up to max. 50 %), employees, material, equipment, travel, other costs and 5 % general project costs.

Eligibility. Austrian research institutions, not individual researchers, can submit, and there is no limit to the numbers of applications that can be submitted by the same institution.

Novel aspects. The evaluation and selection process, which is described in detail below, engages peer reviewers in choosing the top applications to be funded directly and in identifying a group of high-quality applications from which further projects to be funded are chosen by random. Thus, half of the funded projects are directly selected by peer reviewers, and half are randomly drawn from a group of equally worth funding applications identified by the same reviewers.

Success rate. In the first call for applications for this scheme, 401 applications were received, 95 were taken out for formal reasons and 306 were evaluated by the FWF Board, who selected 122 to enter the second evaluation stage. The jury identified 43 projects worth funding and selected 12 applications to propose to the FWF Board for funding. Additional 12 applications were drawn randomly by the FWF Board. In total, 24 projects were funded, receiving a total funding of 3,4 million euro.

FWF's motivation to trial partial randomisation in this scheme

- **Limits of peer review.** Reflecting and seeking to improve the process of decision making is a usual part of business at FWF. The organisation has experienced, like many other funders, an increasing number of applications for its funding schemes; in 2018 alone, it received 2,500 submissions across all schemes. FWF programme managers and Board were aware of the limits of decision making in highly competitive selection processes: because the budget allows only a limited number of proposals to be funded, reviewers are faced with the difficult task of differentiating among qualitatively very similar applications within a group of worth-funding applications. In these cases, the risk that biases affect funding decisions for projects at the margin is high, and risky, unconventional and niche proposals have little chance to succeed. The programme managers also noted the danger that other factors might influence the selection, such as the communication skills of the applicant, or the time of day and level of tiredness of the reviewers. The organisation deploys a variety of application and evaluation mechanisms to mitigate these problems and was aware of the possible use of partial randomisation in funding allocation.

- **Benefits of partial randomisation in this scheme.** Applying randomisation in the late stage of the process, after peer reviewers have selected the projects worth funding, was seen by the FWF managers and Board as a possible useful additional procedure to reduce bias in decision making and worth piloting in their new 1000 Ideas Programme. Thus, the new funding scheme was designed to include partial randomisation from the start.

The application and selection process

Applications can be submitted once a year via the electronic application portal of the FWF. Only unsuccessful applications from the pool of initially considered worth funding can be re-submitted, while other applications cannot be submitted again.

Anonymous and short applications. The process is double-blind, i.e., applicants' and reviewers' identities are hidden. Applicants are required to write the academic abstract (of maximum 700 words) and the project description (maximum 7 pages) in an anonymous way, such that their affiliation, identity, sex, and career stage and that of their partners cannot be identified. Reference to own publications is only possible if no conclusions can be drawn about the identity of the principal investigator or the participating researchers. This is done to enable the reviewers to focus solely on the research idea and the degree of innovation of the proposal. The anonymised part of the application includes an implementation plan, an assessment of the riskiest aspects of the proposal and a description of the potential learnings if the project fails. Applications that do not meet these requirements are rejected without review. Only the anonymised parts of the application are used by the reviewers to select the projects to fund. Other information, including a list of publications, is used by the scientific project officers of the FWF Office to check that the formal criteria are met. All information must be submitted in English such that the international review panel can assess it.

Evaluation and selection

The evaluation and selection process takes about five months. All applications are initially checked for completeness by the scientific project officers of the FWF Office. Those who do not meet formal criteria are rejected without review. The assessment and selection are then performed in three stages:

1. **Pre-evaluation by FWF Board.** In the first stage, members of the FWF Board pre-evaluate the anonymous applications. Board members are 60 to 65 Austrian researchers from the Natural and Technical Sciences, the Biological and Medical Sciences, and the Humanities and Social Sciences. Each application is evaluated by two Board members from the relevant discipline; if the application is interdisciplinary, the two Board members are experts in the two main disciplines of the proposal. **The criteria used** are: (a) Transformative potential of the research idea; and (b) Suitability of the proposed research approach and description of the risk assessment and the possible learning potential. Each criterion is evaluated using a 5-point rating scale,

where 1 is the lowest and 5 the highest score. The applications are then ranked according to the total number of points received.

2. **Assessment, scoring, ranking and selection by jury.** In the second stage, an international jury of 20 experts from a wide range of academic disciplines assesses all eligible proposals. The proposals are **split into three groups**: one with positive reviews, one with negative reviews, and one with heterogeneous reviews. The jury evaluates in depth all proposals with two positive reviews; the rest of the applications are proposed for later discussion by the jury. Each application with two positive reviews is evaluated by two jury members using **three criteria**: (a) The transformative potential; (b) The suitability of the methods, and (3) The coherence and the strength of the project idea. Each criterion is scored using a 5-point scale, and the applications are then re-ranked based on the new scores. In a meeting, the jury then discusses the ranking and defines a cut-off line. All applications above this line are considered worth funding. From this pool, the **jury can directly select a maximum of 12 applications** that are proposed to the FWF Board for the final funding decision. To limit the risk of conservatism deriving from group thinking, each jury member has one **wild card** or joker (also called "golden ticket") to include in the pool of funded projects an application against the judgement of the other panel members. The maximum number of selected projects, however, cannot exceed 12. To avoid difficulties in case more members want to use their wild card, this procedure was changed slightly in the second round of selections: a proposal needs now to be supported by the wild cards of two members to be moved into the group of funded projects. This possibility, however, was neither used in the first or second selection round.

3. **Partial randomisation.** The third and last stage involves partial randomisation, which is carried out by the FWF Board in a meeting. From the pool of remaining applications worth funding, the same number of applications as those selected by the jury (maximum 12) are drawn randomly. The Board then approves both the projects selected by the jury and by randomisation. The identities of the applicants are revealed after the funding is approved.

Description of the randomisation procedure

The tool used for the random selection of the projects to be funded from the pool of worthy applications is the **R software (package 'dplyr'),** a freely available programming language used for statistical computing. The first time that the software was used, an alternative procedure with pieces of papers for a manual draft was prepared in case of any technical problems.

The procedure using R is carried out live during a meeting of the FWF Board (this took place virtually in 2020 due to the Covid pandemic). In the meeting the FWF Board is first presented with the anonymous list of projects selected for funding by the jury. The list of the projects to be subjected to randomisation is then presented, again only showing the ID and title of the projects. The randomised process is run and produces a positive and a negative list. The lists are shown on a screen so everybody can see them. They are also saved in an Excel table for record keeping. The identity of the applicants is made visible only after the Board approves the positive list of projects to fund.

Communication of the selection results

Because of the high number of applications, no feedback from the reviewers is given to the applicants, and only two standard reasons for rejection are given. Successful applicants are not told whether their project has been selected by the review panel or by randomisation. This is done to avoid potential discrimination against the researchers selected by the lot rather than by peer review panel, and to protect the panel from being accused of bias (against any of the applicants who were not selected by the panel).

FWF's decision process

The FWF Board was asked to approve the use of the software R as a tool for the randomised selection. Although not all Board members were in favour of using partial randomisation, the procedure used received positive appraisal. The most critical comment from a Board member was: "I am still against randomisation, but if we want to do it, then this procedure is really excellent."

Initially some reviewers and institutions had concerns about the idea of applying partial randomisation, whilst others were immediately open to trying it out. A long process of discussion and consultation took place to fully explain the process, the aims of the pilot, and answer questions from stakeholders on how it would work.

Feedback from applicants and the review panel

- **Applicants.** Applicants' feedback after the first round of applications was generally positive about the selection procedure, particularly from early career researchers. Some investigators enquired out of curiosity whether their project had been selected by the jury or by randomisation, but no complaints were received.

- **Review panel.** Feedback from the review panel and its chair was elicited at the meeting just after they had selected the applications to be proposed for funding, and focused on the quality of the applications received, rather than on the selection procedure. The applications appeared to be still somewhat conservative and "too safe" compared to what was expected.

- **Board.** Similar feedback was received from the Board in discussions before the second call was launched. One of the Board members, however, noted that all the projects she had reviewed within the biomedical sciences had no preliminary results within the research proposal – a clear difference from applications received in all other existing FWF funding schemes. Thus, the first exercise did to some extent, though not entirely, meet the aim of the programme to encourage high-risk and potentially transformative research ideas.

Evaluation of the experiment

A formal evaluation of the funding scheme will be carried out when a larger sample of applications is available.

Box 3.5 References and further reading:

Most of the information in this case study draws on material from the FWF website:
https://www.fwf.ac.at/en/about-the-fwf/corporate-policy
https://www.fwf.ac.at/en/news-and-media-relations/news/detail/nid/20210922
https://www.fwf.ac.at/en/research-funding/fwf-programmes/1000-ideas-programme
https://www.fwf.ac.at/en/research-funding/application/1000-ideas-programme
https://www.fwf.ac.at/fileadmin/files/Dokumente/Antragstellung/1000-Ideen-Programm/tai_application-guidelines.pdf
https://scilog.fwf.ac.at/en/article/10632/support-risk-takers
Other sources of information:
https://rori.figshare.com/articles/report/Experiments_with_randomisation_in_research_funding__scoping_and_workshop_report/16553067
https://www.meduniwien.ac.at/web/en/about-us/news/detailsite/2020/news-im-juli-2020/1000-ideen-programm-des-fwf-foerderungen-fuer-drei-forscherinnen-der-meduni-wien/

3.4 Health Research Council of New Zealand: Explorer Grants[5]


Research Integrity and Peer Review

Home About Articles Submission Guidelines

Research | Open Access | Published: 03 February 2020

The acceptability of using a lottery to allocate research funding: a survey of applicants

Mengyao Liu, Vernon Choy, Philip Clarke, Adrian Barnett, Tony Blakely & Lucy Pomeroy

Research Integrity and Peer Review 5, Article number: 3 (2020) | Cite this article

8534 Accesses | 9 Citations | 202 Altmetric | Metrics

Abstract

Background

The Health Research Council of New Zealand is the first major government funding agency to use a lottery to allocate research funding for their Explorer Grant scheme. This is a somewhat controversial approach because, despite the documented problems of peer review, many researchers believe that funding should be allocated solely using peer review, and peer review is used almost ubiquitously by funding agencies around the world. Given the rarity of alternative funding schemes, there is interest in hearing from the first cohort of researchers to ever experience a lottery. Additionally, the Health Research Council of New Zealand wanted to hear from applicants about the acceptability of the randomisation process and anonymity of applicants.

The Health Research Council of New Zealand (HRC) is the main governmental funder of health research in New Zealand. It invests about 120 million NZD (about 74 million Euro) per year in basic and applied research projects and research careers aimed at improving health equity, advancing Māori health, strengthening the national health system and promoting the health of people, whānau and communities. It receives annually about 900 applications for research funding, which are mostly evaluated by expert committees, as well as about 700 national and international reviewers. It allocates most of its funding through an annual contestable funding round. HRC was the **first public funder in the world to experiment with partial randomisation** in its Explorer Grants scheme.

The Explorer Grants

Aim. The Explorer Grants scheme trial started in 2013 and continues to date. The aim of the scheme is to fund potentially transformative ideas still at an early development stage, in any health research discipline. Applications can include novel hypotheses, methodologies, tools, technologies, and conceptual frameworks. The scheme is designed to support applications not fundable via other HRC schemes because of unpredictability and lack of supporting data.

Amount and duration of the grant. Each funded project receives 150.00 NZD (about 92.000 Euro) for a maximum of 2 years. The grant covers only research working expenses (direct costs); the host institution is expected to cover all other costs. Individual researchers or groups based in New Zealand are eligible for funding.

Success rate. The number of funded projects has steadily increased from 3 out of 116 applications (success rate 2,6%) in 2003, to 15 out of 65 applications (success rate 23%) in 2021. The budget has steadily increased too, such that the number of projects that can be funded is now much higher than in the first rounds. Unsuccessful applicants are allowed to reapply in the next call.

Novel aspects. The application and selection process for this scheme is different from other HRC schemes, as it requires shorter applications, it is anonymous to peer reviewers, and it uses partial randomisation to select the projects to fund. In the HRC partial randomisation

[5] With thanks to Lucy Pomeroy of the Health Research Council of New Zealand for her inputs and advice on the text of this case study.

selection process the reviewers evaluate applications only to judge whether they are fundable but without scoring or ranking them. All applications judged in-scope (see specifications below) by the reviewers are then considered equally deserving, i.e., potentially fundable. The selection of all those that will be funded is done through a randomisation procedure. No application is selected for funding directly by the reviewers. This is different from the partial randomisation procedure used by VWF and FWF.

HRC's motivation to trial partial randomisation in this scheme

- **Supporting potentially transformative ideas.** HRC is a relatively small funding agency, open to trying out new mechanisms and evaluating their effects to improve their mechanisms to distribute research funding. In 2012/2013, HRC wanted to start a new funding scheme to support potentially transformative research ideas in all health-related fields.

- **Peer reviewers' difficulties in identifying high-risk research.** HRC staff had been aware of the discussions internationally about the limits of peer review, including limitations to appropriately identify high-risk research to be funded. It considered partial randomisation to be particularly suited to the planned Explorer Grants, as it would be difficult for reviewers to compare and score high-risk applications with unpredictable results in many different areas of health research.

- **Benefits of partial randomisation in this scheme.** Random allocation after an initial selection of worthy funding applications would be a more transparent approach and would limit reviewers' burdens. HRC considered partial randomisation to be particularly suited to the planned Explorer Grants, as it would be difficult for reviewers to compare and score high-risk applications with unpredictable results in many different areas of health research.

HRC's decision process

HRC staff prepared the rationale and recommendation for the introduction of the new scheme and selection process, which was approved and supported by the HRC Governing council. Partial randomisation was used in the Explorer Grants since the first call for applications and it has now become part of the standard selection procedure in this funding scheme. At the original approval of the scheme, the Council did not consider there were any legal matters related to the use of randomisation in the distribution of public research funding.

The application and selection process

The evaluation and selection process takes about six months. Applications can be submitted once a year via the electronic application portal of the HRC.

Short and anonymous applications. The part of the application describing the proposal, its transformational nature, viability, and potential impact is short (maximum six pages) and is submitted in an anonymised form. Applicants are asked to "describe skills, experience and details of the research environment without direct reference to the people and host organisation involved."[6]

This is to ensure that the assessment of the applications focuses on the research idea. Personal information contained in other parts of the applications are not disclosed to members of the HRC's committees involved in the review process. The assessment process follows three steps:

- **Eligibility check**. First, the HRC Research Investment Manager and the independent Chair of the Assessing Committee check the applications to confirm their eligibility. Only the anonymous parts of the applications are considered from this step on. The criteria used are: the proposals are within the scope of the areas relevant to the call; they have host institution support; and they comply with formatting requirements. Applications that do not comply with these criteria are excluded from further assessment.

- **Evaluation and triage by the Assessment Committee.** In the second step, an Assessing Committee judges whether the anonymous eligible applications are transformative and potentially viable. The Committee is composed of about 12 New Zealand and Australian biomedical, kaupapa Māori, clinical, public health, social science and interdisciplinary researchers who are appointed by the HRC Research Investment Manager for their research expertise and ability to effectively assess the applications received in the particular funding round. Balance related to a variety of factors, including gender, age, location and affiliation, is also considered. Each proposal is assigned to a **subpanel of three reviewers**. They do not score or rank the applications but are asked to confirm whether two main criteria are met: transformative potential and viability of the proposals. To help reviewers reach more reliable and objective conclusions, the definition of these criteria is specified in the reviewers' guidelines.[7] The proposals for which there is unanimous agreement that the transformative criterion is met and majority agreement that the viability criterion is met enter the pool of equally potentially fundable proposals. Applications for which there are divergent opinions can be discussed and re-evaluated and can be added to the pool of fundable ones if there is majority agreement that the criteria are then met.

- **Partial randomisation.** In the third step, all fundable applications are assigned a random number using the RAND function of Microsoft Excel (Microsoft, ND). They are funded in the order of smallest to largest random number until the available budget is exhausted. Afterwards, the Governing Council needs to approve the selection. The identities of the applicants are revealed after the funding is approved.

[6] https://gateway.hrc.govt.nz/funding/downloads/2021_Explorer_Grant_Application_Guidelines.pdf
[7] https://gateway.hrc.govt.nz/funding/downloads/2021_Peer_Review_Manual.pdf

Description of the randomisation procedure

The randomisation procedure is overseen by the Chief Financial Officer (CFO) who is independent from the actual assessment process. The Head or Director of Investments generates the RAND function to match to the list of fundable applications (identified by reference number only), and the CFO generates the RAND function to match to the list of ranked numbers. The two lists are then matched together and sorted and form the prioritised list for funding.

Communication of the selection results

Applicants are informed whether their application was "declined", "fundable but not funded", or "funded". No further feedback is given. Since the entire selection process, including partial randomisation, is explained in detail in the application guidelines, it should be clear that the successful applications were drawn randomly from the pool of fundable ones.

Reaction of different stakeholders to partial randomisation

There was a high level of acceptance of the procedure across all stakeholders. An early evaluation indicated that the scheme was successfully targeting a range of different applicants which reinforced the nature of the scheme.

The HRC Council has remained supportive as evidenced by increased budget allocation over the years. Moreover, it approved to expand the use of partial randomisation into two other schemes, Consolidators Grants and Health Delivery Activation Grants, although the availability of sufficient funding has not made its use necessary yet.

Applicants are generally accepting as per the evaluation referenced below. A survey of applicants is carried out every year. The Committee Chair reports have consistently supported the process.

Evaluation of the experiment

A formal evaluation of the funding scheme was carried out in 2019 by the HRC Research Investment team in collaboration with Philip Clarke, University of Oxford, UK, Adrian Barnett, Queensland University of Technology, Australia, and Tony Blakely, University of Melbourne, Australia. Its results were published in 2020 (Liu et al., 2020).

- **Aim.** The evaluation focussed on the effect of the partial randomisation procedure on the applicants. In particular, the aim was to assess the applicants' level of acceptance of the procedure and whether their approach to the application had changed knowing that the final selection was made by randomisation.
- **Methodology.** Information was collected using an anonymous online questionnaire (SurveyMonkey). All successful and unsuccessful applicants between 2013 and 2018 were surveyed, as well the applicants in 2019 who did not know about the selection results yet.

- **Limits of the evaluation.** The evaluation showed a general applicants' acceptance of partial randomisation in the Explorer Grant scheme. However, as the authors point out, the evaluation had several limitations, mostly because of the low response rate and because most of the respondents were applicants who had been successful.

Evaluation results

- **Acceptability.** The evaluation found that for most respondents (63%) partial randomisation was an acceptable method for allocating Explorer Grant funds, but only conditional to a first selection of worthy applications by the reviewers. 40% of respondents supported implementing partial randomisation in other funding schemes. However, respondents whose application was successful were far more positive about its use. Similarly, more successful applicants supported its implementation in other funding schemes. As Liu et al. conclude, the acceptability of partial randomisation, or other methods of assessment, would probably increase with higher success rates.
- **No reduction of time spent on applications.** An unexpected result from the survey was that the majority of respondents (75%) stated that they did not spend less time preparing their applications knowing that funding would be randomly allocated. Liu et al. link this to the fact that partial randomisation involves an initial peer reviewed selection, and applicants make efforts to pass that. Although reducing applicants' burden was one of the hoped effects for the HRC, this shows that concerns about the possible negative effects of partial randomisation on the quality of applications (see Chapter 1.4 of this handbook) are probably not founded.
- **Reduced administrative burden.** The HRC staff had reported a reduced administrative burden in running the latest round of applications and evaluations compared to other rounds.
- An evaluation of the **effects on the Assessing committee** has been carried out in the last two application rounds, but the results have not been published yet.

Lessons learned

HRC decided to use the term randomisation to describe the process, rather than lottery, due to perceptions of the latter not being robust and distracting from the real rationale for using partial randomisation.

Box 3.6 References and further reading:

1. https://gateway.hrc.govt.nz/funding/researcher-initiated-proposals/2022-explorer-grants
2. https://gateway.hrc.govt.nz/funding/downloads/2021_Explorer_Grant_Application_Guidelines.pdf
3. https://gateway.hrc.govt.nz/funding/downloads/2021_Peer_Review_Manual.pdf
4. Avin S. (2018) Mavericks and lotteries. Stud Hist Philos Sci Part A. 2018; Available from: https://doi.org/10.1016/j.shpsa.2018.11.006.

5. Liu, M., Choy, V., Clarke, P. et al. (2020) The acceptability of using a lottery to allocate research funding: a survey of applicants. Res Integr Peer Rev 5, 3 https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-019-0089-z

6. Herbert, D.L., Barnett, A.G., Clarke, P., Graves, N. (2013) On the time spent preparing grant proposals: an observational study of Australian researchers. BMJ Open 3. https://bmjopen.bmj.com/content/3/5/e002800

7. Barnett, A.G., Herbert, D.L., Clarke, P., Graves, N. (2014) The pressures on the Australian grant system. Eprints. https://eprints.qut.edu.au/66659/15/66659.pdf

8. Microsoft. (ND) RAND function https://support.microsoft.com/en-us/office/rand-function-4cbfa695-8869-4788-8d90-021ea9f5be73

4. A checklist for funder experiments with partial randomization

This checklist is meant as a support for funders to experiment with partial randomisation. It helps to make sure that all elements have been considered in the three phases of the process of experimenting.

1. Diagnose the problem What is the need you are addressing, or the problem you are solving? See Part 1, 3.1	Define the problems you want to solve in relation to ● Applicants ● Applications ● Funded researchers ● Funded projects ● Reviewers' ability to distinguish middle field proposals	
2. Design the solution using partial randomisation Make sure partial randomisation is the right measure for the problems you have identified. See Part 1, 3.2	What are the expected outcomes? ● Keep in mind that partial randomisation mitigates: ● Conservatism, lack of diversity in awarded applicants and applications ● Reviewers' conscious or unconscious bias ● Limited scientific expertise of review panel ● Reviewers' fatigue ● Grey zone dilemma	
	You might want to use partial randomization together with other changes to your evaluation and selection procedure, e.g., anonymizing applications and short applications (see the case studies for some examples).	
	Legal issues: check that applying partial randomisation to the distribution of public funding is legal in your country (particularly for public funders).	
	Choose the tool you will use for the randomised procedure and find out its cost.	

	Seek approval of your experiment from your Board or Council: ● Prepare slides or bulleted documents including the facts, e.g., statistics on the gender, ethnicity, discipline, and geographical distribution of applicants and of funded projects. ● Summarise the main pros and cons about partial randomization for your board. You might want to share the case studies from this manual.	
	Define how many reviewers will be involved (see Volume I, 4.2)	
	Define how the reviewers will be selected (see Volume I, 4.2)	
	Explain your motivation, goals, and selection procedure to your reviewers.	
	Explain the motivation, goals and selection procedure in the application guidelines.	
	Decide who will physically do the draw or run the randomisation software.	
	You might want to film the physical draw so everyone can see it in case of complaints.	
	If you are using R or a similar software, make sure you save the list of funded projects also in a document, e.g., an Excel table, so it is available for later checking.	
	Decide whether you will disclose to the successful applicants who was selected by the reviewers and who by the lot.	
3. Evaluate How will you assess or verify the success of the experiment? See Part 1, 3.3	Decide ● When the evaluation will start and end ● Which aspects to evaluate ● Whether the results of the evaluation will affect the continuation of the experiment ● Who will carry out the evaluation.	
	Collect feedback on the procedure from the reviewers and the applicants.	
	Share your experience in publications or on your website for others to learn from it.	

5. A comparison of experiments with partial randomisation

Funding agency and funding scheme	**Health Research Council (NZ), *Explorer grants***	**Volkswagen Foundation (DE), *Experiment!* grants**	**Austrian Science Fund (AT), *1000 Ideas* grants**	**Swiss National Science Foundation (CH), *Postdoc.Mobility* fellowships**
Started when	In 2013, still ongoing. From the start with partial randomisation	In 2012 as traditional process, since 2017 with partial randomisation, ended in 2020.	Started in November 2019, still ongoing as a pilot.	Started in 2018, still ongoing. Since 2020, partial randomisation can be applied to all funding schemes, when needed.
Status	Pilot first, now standard	Pilot, now ended	Still pilot	Pilot first, now standard
Term used for the partial randomised procedure	Random selection, modified lottery.	Partially randomised procedure/selection	Partially random procedure	Drawing lots, random selection.
Kind of funding scheme	*Research grants* for innovative ideas and high-risk projects. Ideas at an early stage, with little supporting data. Not compatible with other funding schemes. Failure accepted.	*Research grants* for innovative ideas and high-risk projects, still in an exploratory phase. Failure accepted	*Research grants* for original or transformative research, high-risk projects. At an early stage, still in exploratory phase, with little supporting data. Not fundable by other schemes.	*Post-doctoral mobility fellowships* for a research stay abroad. In addition, fellowship holders can apply for a return grant.
Duration of funding	Max 2 years	Max 1,5 years	Between 6 months and 2 years	Between 1 and 2 years (fellowship) and 3 to 12 months (return phase).
Budget	About 92.000 Euro per project in total.	Up to 120.000 Euro per project in total.	Between 50.000 and 150.000 Euro per project in total.	In average, about 94.000 Euro per fellowship in total.
Research area	Any health research discipline	Natural sciences, life sciences and engineering.	All areas and disciplines, also Cross-disciplinary. Basic research only.	All disciplines.
Length of selection process	About 6 months	3 to 5 months	About 5 months	About 5 months
Kind of selection procedure	Anonymous applications	Double blind	Double blind	Single blind: reviewers are not known.
Assessment process: how many steps?	2 steps (plus initial pre-screen by funder)	2 steps (plus initial pre-screen by funder)	3 steps	4 steps
When is partial randomisation applied?	In the second step, to all eligible applications with two or more "yes" in the peer reviewed selection.	In the second step, to all eligible applications after peer reviewed selection, and after reviewers select the top applications directly.	In the third step, to worth funding applications after 2 rounds of peer reviewed selection, and after reviewers select the top applications directly.	In the fourth step, only to those applications for which the reviewers cannot make a decision. Not a fixed number, only rarely used.
Randomisation tool	Microsoft Excel =Rand() function.	Physical lottery drum, a Bingo set purchased	Randomisation code (R Software: R, package	A physical drawing of lots. Pieces of paper are inserted in a plastic

		online.	dplyr)	capsule and drawn out of a bowl.
No. of applications	65 in 2021	Between 425 and 824 per year	About 400 in the first call	269 in total the fist two calls.
How many funded	15 in 2021. Success rate: 23%	Between 13 and 37 per year. Success rate: 3,6%	Max 24 every year (max 12 only funded by lottery)	Only 8 of the 269 evaluated proposals were selected by the lottery.
Communication of selection results	Grantees are not told whether they were selected by the jury or randomisation.	Grantees are not told whether they were selected by the jury or draw.	Grantees are not told whether they were selected by reviewers or by randomisation.	Winners and losers are told whether the decision was obtained by partial randomisation or not.