



REVIEW

REVISED **Where next for partial randomisation of research funding? The feasibility of RCTs and alternatives [version 2; peer review: 3 approved]**

Tom Stafford ¹, Ines Rombach¹, Dan Hind¹, Bilal Mateen ²,
Helen Buckley Woods ³, Munya Dimario¹, James Wilsdon ⁴

¹The University of Sheffield, Sheffield, England, UK

²Wellcome Trust, London, England, UK

³Research on Research Institute, London, England, UK

⁴University College London, London, England, UK

v2 **First published:** 17 Jul 2023, **8**:309
<https://doi.org/10.12688/wellcomeopenres.19565.1>
Latest published: 30 May 2024, **8**:309
<https://doi.org/10.12688/wellcomeopenres.19565.2>

Abstract

We outline essential considerations for any study of partial randomisation of research funding, and consider scenarios in which randomised controlled trials (RCTs) would be feasible and appropriate. We highlight the interdependence of target outcomes, sample availability and statistical power for determining the cost and feasibility of a trial. For many choices of target outcome, RCTs may be less practical and more expensive than they at first appear (in large part due to issues pertaining to sample size and statistical power). As such, we briefly discuss alternatives to RCTs. It is worth noting that many of the considerations relevant to experiments on partial randomisation may also apply to other potential experiments on funding processes (as described in The Experimental Research Funder's Handbook. RoRI, June 2022).

Keywords

metascience, metaresearch, review, experiments, lottery

Open Peer Review

Approval Status

	1	2	3
version 2 (revision) 30 May 2024			 view
version 1 17 Jul 2023	 view	 view	 view

1. **Becky Ioppolo** , University of Cambridge, Cambridge, UK

2. **Rachel Heyard** , University of Zurich, Zürich, Switzerland

3. **Merle Jacob**, Lund University, Lund, Sweden

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Tom Stafford (t.stafford@sheffield.ac.uk)

Author roles: **Stafford T:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Rombach I:** Writing – Original Draft Preparation, Writing – Review & Editing; **Hind D:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Mateen B:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Woods HB:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Dimario M:** Writing – Original Draft Preparation, Writing – Review & Editing; **Wilsdon J:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: TS and HBW are members of the Research on Research Institute which has received funding from the Wellcome Trust. IR's post has been funded by income from different NIHR, MRC and charity research funding streams. DH serves as a member of the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Clinical Evaluation and Trials (CET) award committee. His post has been funded by income from different NIHR and MRC research funding streams since 2008. BAM is an employee of the Wellcome Trust, which has funded the Research on Research Institute. MD serves as a member of the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Clinical Evaluation and Trials (CET) award committee. His post has been funded by income from different NIHR and MRC research funding streams. JRW is Director of the Research on Research Institute which has received funding from the Wellcome Trust.

Grant information: This work was supported by Wellcome [221297]

Copyright: © 2024 Stafford T *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Stafford T, Rombach I, Hind D *et al.* **Where next for partial randomisation of research funding? The feasibility of RCTs and alternatives [version 2; peer review: 3 approved]** Wellcome Open Research 2024, 8:309 <https://doi.org/10.12688/wellcomeopenres.19565.2>

First published: 17 Jul 2023, 8:309 <https://doi.org/10.12688/wellcomeopenres.19565.1>

REVISED Amendments from Version 1

We thank the reviewers for their thoughtful and informed comments. We address these comments point by point in our response to reviewers. Our summary of the changes is

- we have clarified the wording where indicated
- added additional references on the suggested issues
- clarified the desirability of being clear on motivations for adopting a partial randomisation and adhering to trial best practice (with, e.g. pre-registration, transparency)
- noted the limitations of RDD designs
- noted the bias introduced by the selection of participants in any funder trial: application sample is out of our control, but could be a target measure for a trial

Any further responses from the reviewers can be found at the end of the article

Introduction

In recent years, applications of partial randomisation to research funding processes have received growing attention from research funders, meta-researchers and the wider research community (Nature editorial, 2022; Woods & Wilsdon, 2021a). Partial randomisation, also known as focal randomisation or random selection, is a method for allocating research funding. It is used in addition to peer review, where peer review has reached the limits of efficacy and fairness, and the comparative qualities of applications are largely indistinguishable or ‘equally good’ (Bedessem, 2020). In its partial form, only some applications are subject to random selection, once those which are evaluated as clearly fundable or clearly non-fundable have been removed.

A recent study (Woods & Wilsdon, 2021b) found that the strongest motivator for funding institutions to use partial randomisation is fairness: a fairer decision-making process when peer review had reached its limits; fairer to applicants, as it is blind to institution, geographical location, race, gender, discipline and methodology; and also a transparent process and therefore easier to communicate and understand funding decisions. Other organisational motivators are the desire to break deadlocks in, or reduce time spent on panel decision making, and to ameliorate risk aversion or other concentrations of awards so as to facilitate the funding of a greater plurality of research topics and methodological types.

Pilots of steadily increasing volume and sophistication have been conducted (Bendisoli *et al.*, 2022). There are some clear emerging lessons, but also much that remains unknown. This includes the extent of any ultimate benefits in terms of reduction of biases or gains in efficiency, as well as assurance that harms, such as trust in and acceptance of funding allocation that involves partial randomisation, are acceptable. Strategies for enhancing the evidence based around partial randomisation could be divided into three general categories:

a) “Steady as she goes”

Conduct more smaller scale pilots. Concerns regarding this approach are that another five years of small-scale pilots will

not aggregate to a compelling evidence base. By the same logic of “the plural of anecdote is not data”, smaller, potentially flawed, studies never add up to the evidential power of a more comprehensive, systematic trial (Beets *et al.*, 2021).

b) “From model to implementation”

Some, e.g. Gross & Bergstrom (2019), have asserted that the costs to research time are sufficient to warrant major overhauls of research funding allocation processes, including partial randomisation. By this account, we have little to lose and much to gain by a larger scale implementation of partial randomisation. However, the concern is that abstract models under-estimate system complexity, particularly with respect to stakeholder aspects (see Barlösius & Philipps, 2022; Liu *et al.*, 2020), namely the perception of, and reaction to, of funding allocation by lottery among those who apply for funding and those who are awarded funding. Additionally, larger scale implementation assumes that the motivations for partial randomisation are agreed, something which is not clear (Woods & Wilsdon, 2021a).

c) Funder experiments

A third option, which is the focus of this paper, is to conduct larger scale experiments, across multiple funding agencies if necessary, to produce a compelling test of the benefits of partial randomisation. An obvious candidate method is a randomised controlled trial (RCT) of the effects of partial randomisations, honouring the “gold standard” of evidence for medical innovations.

Outline of work

In this article we outline essential considerations for any study of partial randomisation of research funding, and consider scenarios in which RCTs would be relevant. We highlight the interdependence of target outcomes, sample availability and statistical power for determining the cost and feasibility of a trial. For many choices of target outcome RCTs may be less practical and more expensive than they first appear (in large part due to issues pertaining to sample size and statistical power). As such, we also introduce and briefly discuss alternatives to RCTs. It is worth noting that many of the considerations relevant to experiments on partial randomisation may also apply to other potential experiments on funding processes (see Bendiscioli *et al.*, 2022). Similarly, investigations of partial randomisation can be usefully informed by other experiments on research funding (e.g. Lane *et al.*, 2021; Pier *et al.*, 2018; Strolger & Natarajan, 2019).

Considerations for studies of the potential benefits of partial randomisation

Designing a robust study requires a clear understanding of the research question being asked (Riva *et al.*, 2012). Only with clarity on the precise question that is being asked can the answer that is sought be defined. In the language of trials this answer is the ‘treatment effect’ that one seeks to estimate. This ‘estimand’ is best described by its five constituent attributes: (i) the population of interest, (ii) the ‘treatment’ conditions compared, (iii) the outcome measure of interest, (iv) the population-level summary measure used

to describe how outcomes differ under the aforementioned conditions (e.g., a risk ratio, odds ratio, etc.), and (v) acknowledgment and management of intercurrent events that can impact interpretation of the results (e.g., career change or securing funding from another source). For more detail see the [ICH E9 \(R1\) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials \(2020\)](#)

Defining the estimand for studies of partial randomisation of research funding

Below is a brief discussion of the nuances that need to be considered when defining the estimand in this context, and particularly those pertaining to outcome selection.

Focusing initially on describing the population of interest, most examples of partial randomisation that have been undertaken to date have focused on the applicants or reviewers associated with a specific scheme run by an individual funding organisation ([Woods & Wilsdon, 2021a](#); [Woods & Wilsdon, 2021b](#)). To assume that these results are generalisable to other funders and national research systems, even when there is overlap in archetypal applicants (e.g., schemes limited to early career researchers in a specific area of biomedical science), is probably not appropriate given the non-overlapping pools (whether geographic or otherwise) that funders receive applications from. Instead, to make a definitive claim would require a trial that by design did not limit its population of interest to an individual funding organisation.

Similarly, there are nuances to the operationalisation of the ‘partial randomisation’ concept (see [Woods & Wilsdon, 2021a](#) for details on pre-existing funder experiments), which in turn influence the nature of the treatment conditions being compared.

Finally, there are a number of different outcomes that can be measured which assess the potential benefits of partial randomisation. Critical to effective study design is the declaration, **in advance**, of the way the target outcome measure(s) will be operationalised. It is this prespecification (also known as ‘preregistration’; [Nosek et al., 2018](#)) which defines the study as a ‘Trial’, and is as important to successful inference interpretation as the ‘Randomised Control’ aspect of an RCT ([Simmons et al., 2011](#)). Candidate outcomes include those that measure impact on the funded portfolio, on diversity of applicants, or efficiency of review and/or decision process. Notably, these outcomes either do not have fully satisfactory outcome measures, or they afford a choice between different outcome measures which trade-off ease and accuracy. For example, the target outcome of enhanced diversity of applicants could be operationalised using the demographic parity (mathematical) definition of fairness (i.e., trying to ensure that a diversity of applicants in the funded portfolio reflects the application rate for each demographic group in the applicant pool).

In practice, this would be implemented by requesting applicant ethnicity data and comparing the impact of partial randomisation, and comparing that to the diversity in the portfolio funded under standard operating procedures. Importantly, any specific operationalisation will be limited by the principle on which it is based (i.e., there are many definitions of fairness, and satisfying several simultaneously can be impossible; [Kleinberg et al., 2016](#)), what it does not ask (e.g. other dimensions of diversity such as sex or socioeconomic status) and how it asks (e.g. the categories for ethnicity which define how applicants are asked to report).

We note that diversity implications of changes to selection among applicants are downstream of the self-selection of applicants into applying for a funding scheme. This necessarily limits the potential effects of partial randomisation. Effects on applicants rather than selection of applicants may be as, or more, important as target measures for trials of partial randomisation.

In addition, both potential benefits and harms will need to be assessed by any outcome measure. Considerations of potential target outcomes and outcome measures are shown in [Table 1](#). Further evidence on researchers’ views of partial randomisation can be found in [Liu et al. \(2020\)](#), [Philipps \(2021\)](#) and [Philipps \(2022\)](#).

In essence, partial randomisation of research funding represents a broad church of meta-research questions and associated approaches, and thus, there are many estimands that one might consider running a study to address. In the next section we explore how study design might be influenced by the choice of estimand. Regardless though, collaboration and coordination amongst funders to undertake any study which conforms to a set of standard estimand definition is a non-trivial endeavour, and this is where organisations such as RoRI can play a role in facilitating collaborations and exchange across different funders.

How study design impacts partial randomisation meta-research

In partial randomisation research, the choice of outcome measures, particularly the primary outcome measure, plays an important role in determining the study design. Some outcomes may not require full implementation of allocation of awards via partial allocation (see “shadow experiments” below). On the other hand, if the outcomes of primary interest can only be measured post-funding selection, or even project completion, such as appeals against funding decisions, the impact of funding on career development, or the impact on scientific advancement, then an RCT may be the most suitable design option.

Once stakeholders have chosen a ‘primary outcome’ - the target outcome that key stakeholders (e.g., funders, researchers, funding panels, patients) would agree is the most important - and a plausible operationalisation of it in an outcome measure,

Table 1. Possible outcomes: a non-exhaustive list.

Outcome	Issues
Benefit outcomes	
Fairness	Lack of a relatively objective criterion (gold standard) measure (Brezis & Birukou, 2020). Operationalised as distributive fairness (fairness of outcomes), may require large numbers/long timescales, to spot differences in clustering of grants. Procedural / informational fairness would require researchers to observe and code committee work or an adjudication committee to assess content of rejection letters.
Efficiency: time to deliberation	Objective, continuous (therefore efficient) measure and has been used successfully (Bendischioli <i>et al.</i> , 2022) but may not be seen as important to the public.
Efficiency: appeals	Objective, dichotomous measure. May require large sample sizes, depending on base rate. Not universally applicable and not every funder permits appeals.
Diversity	Requires operationalisation by applicant demographic (gender, ethnicity, etc.) or topic (academic disciplines and research methodologies). The latter might require coding manuals and coders / adjudication committees to resolve.
High-risk, high reward projects	Risk is subjective and would require researchers to observe and code applications or an adjudication committee. Reward would require long timescales (and open complex definitional/measurement issues)
Exceptional scientific advances	Requires a long timescale and large numbers (very rare event). Would require researchers to observe and code applications or an adjudication committee running over years.
Harm outcomes	
Lower Application quality	Requires coding manuals and coders / adjudication committees to resolve.
Questionable research practices	Subjective. Would require advertisement of the RCT between lottery and usual practice, as well as two researchers to code grant applications against a framework for QRP.
Reputational damage to funder	Not subject to experimental design. Can perhaps be operationalised via perceptions of individual scheme and/or individual scheme applicants.
Stigmatisation of awardee as unworthy/merely lucky	Timescales likely to be undesirable. Measurement likely to be problematic.

it is necessary to consider the target effect size. This is the difference between the two experimental conditions (e.g., partial randomisation and funder standard practice) which would be worth detecting if it existed.

The statistical power of a trial is the probability of detecting the target effect, should one exist. Given the time, cost and effort of RCTs this probability should be high (e.g., at least 90%, if not prevented by feasibility constraints). The target effect size, along with elements of trial design such as the sample size and number of conditions, define statistical power. All other things being equal, larger effects can be detected with smaller sample sizes.

The primary outcome also defines the unit of analysis. In medical trials, the unit of analysis is often patients. For our purposes it may be grant applicants, grant applications, peer reviewers, funder panellists, awarded grants or successful awardees. The questions being asked influence the outcome, the unit of analysis, and the nature of outcome assessment. Table 2 illustrates how different potential target outcomes

affect trial practicality via determination of available sample sizes.

These considerations show that considerable range exists under the headline call to conduct RCTs of partial randomisation. Different choices of target outcome(s), and so of unit of analysis, have large implications for the ease, rate and cost of recruitment for an adequately powered trial. In addition, different target outcomes afford outcome measures which are more or less satisfactory in the terms of capturing the true value of the outcome and delay required to collect them (at one extreme being the target outcome of selecting for high-risk, high-value discovery-mode research. While obviously laudable, the delay between adjustments to any funding process and the outcome of increased rates of fundamental breakthroughs alone makes this a less practical target outcome).

Indicative protocol

As a thought-experiment and to illustrate the aforementioned issues in more detail, we present an indicative protocol for a RCT of partial randomisation in the context of a major UK

Table 2. Target outcome, unit of analysis and sample availability for one funding call.

Target outcome	applicant diversity, beliefs about partial randomisation	proposal novelty, ambition/risk	reviewer burden, review consistency	project productivity, diversity characteristics of awardees, awardee reaction to award by partial randomisation
Unit of analysis	APPLICANTS	APPLICATIONS	REVIEWS	AWARDS
Sample available	number of investigators	number of applications	number of applications x reviews per application	number of applications x proportion funded
Illustrative numbers assuming 100 applications, 3 investigators, 4 reviews per applications, and a 10% success rate	300	100	400	10

health research scheme, the National Institute for Health and Care Research (NIHR) Health Technology Assessment (HTA) programme. This scheme typically receives several hundred applications per year¹.

Research question: Does partial randomisation enhance the impact of the funded portfolio?

Population: Research applications to the NIHR HTA programme.

Intervention: Receipt of funding will be allocated via lottery for proposals rated as fundable by external peer review, without going to panel discussion.

Control: “standard practice”, i.e., adjudication by committee discussion along funder specific guidelines.

Outcome: Patients benefit arising from both portfolios, calculated by the number treated with the HTA-approved products (assuming approved) × Quality Adjusted Life Years (QALY). Impact per £ can be calculated by dividing by total portfolio value.

Logistics:

Practicalities around recruiting and randomising study participants (i.e., grant applications) require careful consideration.

- How will grant applicants be informed that funding decisions may be based on partial randomisation? Would all the usual candidates be willing to participate in the trial, or apply for alternative funding elsewhere?
- Timing of randomisation: all funding applications would be required to pass some minimum quality standard,

and would then be randomised to either be allocated funding via a lottery, or by the committee.

When would be the most appropriate time to perform the randomisation, and who would do this?

Online randomisation systems, with appropriate stratification or minimisation by funder, funding round and other important factors can easily be implemented by a clinical trials unit (CTU). Specific guidance would have to be drawn up to ensure delegates of the funders are able to perform the randomisation.

- Details on how grant applicants are informed of the outcomes need to be considered, including how details of the stage which each application reached will be communicated (e.g. if an application was rejected at the minimum quality threshold stage or at the randomisation stage).

Sample size:

For the purposes of this example, assume that the allocation of research funding by lottery would be deemed successful if the total patient benefit was raised - on some assumed, continuous, outcome measure - by 0.4 of a standard deviation (actually a moderate, rather than small effect). Based on these expected event rates, using a two-sided 5% significance level and 90% power, a total sample of 264 funded studies (132 per arm) would be required. The use of a two-sided test is common practice in trials and reflects that although we might hope for a positive (beneficial) effect of the intervention, we account for the potential for a negative (harmful) effect.

Enough studies would have to be randomised to either approach of funding allocation to result in the required number of funded studies, i.e., allowing for the proportion of studies not receiving funding. If the success rate is 20%, this means 660 applications would be required *in each arm* (1320

¹<https://www.nihr.ac.uk/documents/hta-programme-success-rates/23178>

total applications). Larger effect sizes would decrease the required sample size, though might be less realistic.

Feasibility:

Finally, the total duration and cost of the trial can be estimated based on the recruitment rate. Let us imagine a funding panel that adjudicates on around 30 applications at each of three time points per year. Let us further imagine that 20/30 grants are neither clearly fundable nor clearly unfundable and that applications in this middle segment are randomised to lottery or usual practice. As the committee sits three times per year, a total of 60 grants might be entered into an RCT comparing randomisation lottery or usual practice. This number may be optimistic depending on: (1) how programme-level funding limits apply; or, (2) whether panels only make funding recommendations to a government department rather than directly allocate grant income themselves. Let us imagine this funder has three other panels, so during the course of one year all their funding panels will contribute 180 grants. Given these considerations, it would take (applications * number of treatment arms) / applications per year years to allocate sufficient grants for an adequately powered RCT of partial randomisation for this target outcome and trial design from this large multi-panel funder alone. In our example this calculation is $(660 * 2 / 180)$, which is 7.3 years for the allocation stage alone.

Following allocation we might assume 3–6 years from funding to publication. This *may* be followed by approval for wider use (e.g., by the National Institute for Health and Care Excellence [NICE] in the UK) and then commissioning by healthcare delivery agencies. A speedy estimate of these aspects is 10 years (Morris *et al.*, 2011), which means that final outcome data for this trial would be available at approximately year 17 of its existence, at the earliest. Many, but not all, non-healthcare funding domains will have equally lengthy delays between research completion and feasible outcome measurement.

This makes clear the need for funder collaboration to support timely recruitment to trials, as well as - perhaps - to focus on outcome measures which afford early assessment. It also suggests that there may be benefits to explore alternatives to RCTs (see below).

Typical costs:

The average cost per participant in medical industry trials is around US\$41,413 (or £29,744), based on FDA data from 2015 to 2017 (Moore *et al.*, 2020). By comparison, the average cost per participant in NIHR HTA-funded trials is probably around £3000 per participant. An RCT comparing allocation by lottery versus usual funding panel practice would have grant applications, not humans as its ‘participants’. Depending on the outcomes of interest (some of which would require expensive data collection infrastructure) the trial could be relatively cheap, perhaps £1m.

Costs to run, whether supplied directly to a third-party to administer the trial or provided “in kind” by funding agency

staff who administer the trial is separate from the funding allocation to grant awards. Funding allocated to awards by partial randomisation, while necessary for a trial, would be allocated anyway, even if the RCT was not conducted. The money will just be allocated differently because some of it will be allocated at random rather than by funding panel adjudication.

Alternative to RCTs

There exist a number of alternative methods which are currently used where RCTs are less feasible, or inappropriate. We briefly review their strengths and weaknesses here.

Causal inference methods

The baseline assumption of many researchers, and the impetus for RCTs, is that observational studies may not yield reliable inferences of treatment effects (Young & Karr, 2011). Statistically controlling for confounders in observational data is challenging (Westfall & Yarkoni, 2016). Moreover, natural assumptions, such as balanced ‘unmeasured’ covariates that arise from formal randomisation are more difficult to justify in observational settings, thus leaving the door open to residual confounding which can produce misleading/inaccurate results (unbeknownst to the analyst).

However, a new generation of causal inference methods have recently gained attention which purport to allow more reliable inference from observational data (Pearl *et al.*, 2016). For example, Hernán & Robins (2016, see also Hernán *et al.*, 2016) propose guidelines for causal inference from large observational databases. Like an RCT, it is necessary to specify the population, intervention, comparator and outcome, as well as any important mediators or moderators, with these being used to attempt mitigation of selection biases. For an exemplar demonstration of how to compare these observational methods to an RCT see Lodi *et al.*, 2019. These approaches have proven robust in estimating average treatment effects, as demonstrated by Hernán and colleagues in their observational data-based confirmation of the effectiveness of the mRNA COVID-19 vaccine (Dagan *et al.*, 2021).

Caveats include that very large data sets are required, so the application of such methods would require data collection and harmonisation across multiple decades and funding agencies. Additionally, successful inference can only be done if the population of grant-schemes observed includes some occurrence of allocation by partial randomisation.

Finally, it should be humbling that studies by researchers at social media platforms with access to both billions of data points and the ability to run true experiments, akin to the RCTs discussed here, have shown that even the latest generation of causal inference methods may not be successful at accurately revealing the true causes of things (Gordon *et al.*, 2019; Gordon *et al.*, 2022), or may have restricted applicability to novel phenomena (Eckles & Bakshy, 2021).

Qualitative comparative analysis

Qualitative Comparative Analysis (QCA; Marx *et al.*, 2014) is a formal method of studying causality in a simple data

table of binary or ordinal variables from small-medium ‘N’ samples (8–200). The method uses Boolean algebra to understand the necessary or sufficient conditions for outcomes to occur. Methods such as QCA might be attractive if the numbers required by probabilistic causal inference approaches are deemed infeasible, due to the sample sizes required, or if funders want exploratory studies of the effects of different factors on outcomes.

Natural experiments

A special case of causal inference with observational data is the existence of natural experiments (Dunning, 2012). An example is regression discontinuity designs, which take advantage of arbitrary thresholds which divide cases near that threshold into two groups, despite being essentially similar. This analysis has been applied to study the effect of funding success on longer term research career outcomes (Bol *et al.*, 2018; Wang *et al.*, 2019 but see de la Cuesta & Imai, 2016). It may be that there are similar natural experiments possible within the funding system.

Shadow experiments via simulated outcomes

Outcome measures that can be assessed without implementing selection, such as diversity of awardees, or time taken to make funding decisions may not require the full process of randomisation and subsequent awards. Instead, the “as if” impact of partial randomisation can be simulated and the outcome compared directly with outcomes from the standard procedure. In other words, funders could choose to base their funding decisions on their usual decision making process while the experiment is ongoing, using standard procedure as data collection for a “shadow experiment” on partial randomisation. Examples of shadow experiments using funding decisions include Graves, Barnett & Clarke (2011) and Bieri *et al.* (2021).

This approach is statistically powerful, since the entire universe of outcomes which would be produced by partial randomisation can be simulated and used as a basis for comparison. An advantage of partial randomisation is that it is a process which can be easily modelled. Panel review is non-reducible, it exists because the selection of projects to fund made

by a panel is not knowable in advance. In contrast, partial randomisation is a minimal process which could be applied to the population of grants under consideration by a panel, without implementing it as the selection process. So, for example, if a panel funded 10 grants from 100 fundable grants, it is possible to identify the precise statistical distribution of awards which would have been made by partial randomisation.

Conclusions

It is challenging to trial novel methods of funding allocation and evaluation. Peer networks of funders offer a route to sharing lessons from pilots and trials, and to building a more robust evidence base. There is more scope for funders to work together—including through the RoRI consortium—to deepen our shared understanding of the value and limitations of partial randomisation and other experimental methods; to share data and methods and coordinate with the growing community of metaresearchers.

There is a need for more robust experimental studies, with defined baselines and controls: ideally involving multiple funders, which will allow for comparison across funding systems. The potential of early pilots will not be realised without more rigorous, long-term experiments which can generate transferable evidence of the pros and cons, opportunities and limitations of specific interventions. Moving beyond analysis of changes to applicant and allocation processes to study the full impacts of different funding methods on research outcomes will require sustained analysis, as these will take several years to become apparent.

At the same time, a move towards formal RCTs in this arena is not straightforward. RCTs are complex design objects. A defining consideration is deciding in advance what target outcome you wish to focus on. This choice may render the necessary sample size for adequate statistical power, and related cost and ease of recruitment, prohibitive. Larger samples will need coordination between multiple funding agencies.

Data availability

No data are associated with this article.

References

- Barlösius E, Philipps A: **Random grant allocation from the researchers’ perspective: Introducing the distinction into legitimate and illegitimate problems in Bourdieu’s field theory.** *Soc Sci Inf.* 2022; **61**(1): 154–178. [Publisher Full Text](#)
- Bedessem B: **Should we fund research randomly? An epistemological criticism of the lottery model as an alternative to peer review for the funding of science.** *Res Eval.* 2020; **29**(2): 150–157. [Publisher Full Text](#)
- Beets MW, von Klingraeff L, Weaver RG, *et al.*: **Small studies, big decisions: the role of pilot/feasibility studies in incremental science and premature**

- scale-up of behavioral interventions.** *Pilot Feasibility Stud.* 2021; **7**(1): 173. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bendiscoli S, Firpo T, Bravo-Biosca A, *et al.*: **The experimental research funder’s handbook (Revised edition, June 2022, ISBN 978-1-7397102-0-0).** *Research on Research Institute.* Report, 2022. [Publisher Full Text](#)
- Bieri M, Roser K, Heyard R, *et al.*: **Face-to-face panel meetings versus remote evaluation of fellowship applications: simulation study at the Swiss National Science Foundation.** *BMJ open.* 2021; **11**(5): e047386. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

- Bol T, de Vaan M, van de Rijdt A: **The Matthew effect in science funding.** *Proc Natl Acad Sci U S A.* 2018; **115**(19): 4887–4890.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brezis ES, Birukou A: **Arbitrariness in the peer review process.** *Scientometrics.* 2020(1); **123**: 393–411.
[Publisher Full Text](#)
- Dagan N, Barda N, Kepten E, et al.: **BNT162b2 mRNA Covid-19 vaccine in a nationwide mass Vaccination setting.** *N Engl J Med.* 2021; **384**(15): 1412–1423.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- de la Cuesta B, Imai K: **Misunderstandings about the Regression Discontinuity Design in the study of close elections.** *Annu Rev Polit Sci.* 2016; **19**(1): 375–396.
[Publisher Full Text](#)
- Dunning T: **Natural experiments in the social sciences: a design-based approach.** Cambridge University Press, 2012; 358.
[Reference Source](#)
- Eckles D, Bakshy E: **Bias and high-dimensional adjustment in observational studies of peer effects.** *J Am Stat Assoc.* 2021; **116**(534): 507–517.
[Publisher Full Text](#)
- Gordon BR, Moakler R, Zettelmeyer F: **Close enough? A large-scale exploration of non-experimental approaches to advertising measurement.** *arXiv preprint arXiv: 2201.07055.* 2022.
[Publisher Full Text](#)
- Gordon BR, Zettelmeyer F, Bhargava N, et al.: **A comparison of approaches to advertising measurement: evidence from big field experiments at Facebook.** *Market Sci.* 2019; **38**(2): 193–225.
[Publisher Full Text](#)
- Graves N, Barnett AG, Clarke P, et al.: **Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel.** *BMJ.* 2011; **343**: d4797.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gross K, Bergstrom CT: **Contest models highlight inherent inefficiencies of scientific funding competitions.** *PLoS Biol.* 2019; **17**(1): e3000065.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hernán MA, Robins JM: **Using big data to emulate a target trial when a randomized trial is not available.** *Am J Epidemiol.* 2016; **183**(8): 758–64.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hernán MA, Sauer BC, Hernández-Díaz S, et al.: **Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses.** *J Clin Epidemiol.* 2016; **79**: 70–75.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 17 February, 2020.
[Reference Source](#)
- Kleinberg J, Mullainathan S, Raghavan M: **Inherent trade-offs in the fair determination of risk scores.** *arXiv preprint arXiv: 1609.05807.* 2016.
[Publisher Full Text](#)
- Lane JN, Teplitskiy M, Gray G, et al.: **Conservatism gets funded? A field experiment on the role of negative information in novel project evaluation.** *Manage Sci.* 2021; **68**(6): 4478–4495.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Liu M, Choy V, Clarke P, et al.: **The acceptability of using a lottery to allocate research funding: a survey of applicants.** *Res Integr Peer Rev.* 2020; **5**(1): 3.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lodi S, Phillips A, Lundgren J, et al.: **Effect estimates in randomized trials and observational studies: comparing apples with apples.** *Am J Epidemiol.* 2019; **188**(8): 1569–1577.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Marx A, Rihoux B, Ragin C: **The origins, development, and application of Qualitative Comparative Analysis: the first 25 years.** *Eur Polit Sci Rev.* 2014; **6**(1): 115–142.
- Moore TJ, Heyward J, Anderson G, et al.: **Variation in the estimated costs of pivotal clinical benefit trials supporting the US approval of new therapeutic agents, 2015-2017: a cross-sectional study.** *BMJ Open.* 2020; **10**(6): e038863.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Morris ZS, Wooding S, Grant J: **The answer is 17 years, what is the question: understanding time lags in translational research.** *J R Soc Med.* 2011; **104**(12): 510–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nature editorial: **The case for lotteries as a tiebreaker of quality in research funding.** *Nature.* 2022; **609**(7928): 653.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Nosek BA, Ebersole CR, DeHaven AC, et al.: **The preregistration revolution.** *Proc Natl Acad Sci U S A.* 2018; **115**(11): 2600–2606.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pearl J, Glymour M, Jewell NP: **Causal inference in statistics: a primer.** John Wiley & Sons, 2016; 160.
[Reference Source](#)
- Phillips A: **Research funding randomly allocated? A survey of scientists' views on peer review and lottery.** *Sci Publ Policy.* 2022; **49**(3): 365–377.
[Publisher Full Text](#)
- Phillips A: **Science rules! A qualitative study of scientists' approaches to grant lottery.** *Res Eval.* 2021; **30**(1): 102–111.
[Publisher Full Text](#)
- Pier EL, Brauer M, Filut A, et al.: **Low agreement among reviewers evaluating the same NIH grant applications.** *Proc Natl Acad Sci U S A.* 2018; **115**(12): 2952–2957.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Riva JJ, Malik KM, Burnie SJ, et al.: **What is your research question? An introduction to the PICOT format for clinicians.** *J Can Chiropr Assoc.* 2012; **56**(6): 167–71.
[PubMed Abstract](#) | [Free Full Text](#)
- Simmons JP, Nelson LD, Simonsohn U: **False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant.** *Psychol Sci.* 2011; **22**(11): 1359–66.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Strolger L, Natarajan P: **Doling out Hubble time with dual-anonymous evaluation.** *Phys Today.* 2019.
[Publisher Full Text](#)
- Wang Y, Jones BF, Wang D: **Early-career setback and future career impact.** *Nat Commun.* 2019; **10**(1): 4331.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Westfall J, Yarkoni T: **Statistically controlling for confounding constructs is harder than you think.** *PLoS One.* 2016; **11**(3): e0152719.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Woods HB, Wilsdon J: **Experiments with randomisation in research funding: scoping and workshop report (RoRI Working Paper No.4).** *Research on Research Institute.* Report, 2021a.
[Publisher Full Text](#)
- Woods HB, Wilsdon J: **Why draw lots? Funder motivations for using partial randomisation to allocate research grants (RoRI Working Paper No.7).** *Research on Research Institute.* Report, 2021b.
[Publisher Full Text](#)
- Young SS, Karr A: **Deming, data and observational studies: a process out of control and needing fixing.** *Significance.* 2011; **8**(3): 116–120.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:   

Version 2

Reviewer Report 10 June 2024

<https://doi.org/10.21956/wellcomeopenres.24789.r85279>

© 2024 Jacob M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Merle Jacob

Professor in Research Policy, Lund University, Lund, Sweden

I am happy with the revised version and recommend indexing.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Research funding, research policy

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 04 September 2023

<https://doi.org/10.21956/wellcomeopenres.21675.r65750>

© 2023 Jacob M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Merle Jacob

Professor in Research Policy, Lund University, Lund, Sweden

The paper addresses the problem of how to design a study to implement RCTs in research funding. This is a narrow area of research outside of the discourse about clinical trials so the literature is rather sparse. In my view the language of the paper may not be entirely accessible to the rest of the research community who work with research funding.

The paper might also benefit from considering the fact that RCTs need to focus on funder objectives. Increasing the perception of procedural fairness? Reducing selection bias, etc. are all potential objectives. Some outcomes may be more or less amenable to being grouped together in the same trial. The current paper is not clear on which objectives it takes on board in its worked example or thought experiment.

I agree that a proper experiment would require multiple funders but there are very few funding situations to which the results of this experiment would be useful. Real world cases with multiple funders are rare and often further complications such as different authority rights and jurisdictions arise.

Perhaps considering applying the method to more than one funding instrument may be a potential alternative.

All research on funding faces a problem of selection bias. We can only select from those who apply and the character of the applicant sample can only be partially controlled for by qualification criteria. Who applies is therefore dependent not on instrument design and likewise who gets funded is only partially determined by the character of the selection process. Here the analogy with clinical trials is only partially helpful because the sample of patients who are potential candidates is non random in the sense that they must be selected from those afflicted. With funding, not all those eligible are likely to apply and those who apply may not represent a random sample of all the characteristics of the set of potential applicants.

Generally, the paper is though provoking and I could comment reams on this but the key issue here is that RCTs are a solution in search of a problem. If the problem can be identified then we can better discuss the efficacy of RCT as a solution.

Is the topic of the review discussed comprehensively in the context of the current literature?

Yes

Are all factual statements correct and adequately supported by citations?

Yes

Is the review written in accessible language?

Partly

Are the conclusions drawn appropriate in the context of the current research literature?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Research funding, research policy

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have

significant reservations, as outlined above.

Reviewer Report 01 September 2023

<https://doi.org/10.21956/wellcomeopenres.21675.r65759>

© 2023 Heyard R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Rachel Heyard 

University of Zurich, Zürich, Switzerland

Stafford and colleagues wrote a comprehensive review on the potential of RCTs to evaluate the effect or the effectiveness of the partial randomisation of funding as an intervention.

In the past, many of the decisions done by funders have indeed only partly been backed-up by evidence in the sense that funding allocation processes have been changed without a properly planned experiment to test whether the intervention actually worked. Often it is also not clear what the intervention was supposed to change in the first place. This is mainly because of this experimentation culture the authors refer too is quite new in research funding. Therefore I believe reviews as this one are extremely valuable to introduce the topic to funders. As such, I think it would be crucial to highlight a couple more examples of experiments that have already been done. The authors do refer to The Experimental Research Funder's Handbook in the abstract, but could cite more of the current research on the funding evaluation process. In the end, partial randomisation is an intervention to improve specific aspects of the process so any type of intervention study in research funding could be used as example.

More specifically, I have the following comments:

- [p3 right] In my opinion, the very first step before even deciding on the estimand or outcome of the study is to define a clear and narrow research question, maybe even following the PICOT guidelines. Having a clear research question, will help find the best suited outcome and analysis. (see Riva *et al.*, 2012¹).
- [p4...] Since the authors mention that target outcomes etc should be prespecified, I thought they should also mention (pre)registration of study protocols as it is f.ex. mandatory in clinical trials (to reduce QRPs and other biases). As the authors are suggesting to do more proper experiments and studies in research funding, they should also explicitly highlight best practices in planning such experiments to ensure robustness, reproducibility and transparency.
- [p4] The most suitable outcome should align with the goal of the studied funding scheme or the strategy of the funder. Could something like this be added?
- [regarding outcome reward and similar] There is no universally agreed on measure of research or societal impact yet - could this be added in the discussion.

- [sample size of the indicative protocol] The first sentence in the paragraph indicates a clear direction of effect that will be tested - an increase in the total patient benefit. Hence, you would not use a two-sided but a one-sided test (usually with a 2.5%-level) which leads to a sample size of 156 studies per arm. Could you also say something about how one would estimate the target effect (here 0.4)?
- [regarding natural experiments] The effect estimates from regression discontinuity designs (RDD) are unfortunately only valid for applications around the threshold. Plus this threshold is not constant and depends on outside factors such as the number of applications to a call, the available budget, ...(see de la Cuesta & Imai, 2016²).
- [regarding shadow experiments] Could you add some citations of shadow-like experiments that have been conducted in research allocation? (see maybe Graves *et al.*, 2011³ or Bieri *et al.*, 2021⁴).
- [Conclusion] To not overload funding agencies who might not have the resources nor the expertise to run such experiments by themselves with more tasks it is important to (1) harmonise data among funders and then (2) share it with or collaborate with meta-researchers to get to the best possible outcome. In general it is important to improve transparency of funders when pursuing experiments and involve researchers and panel members/reviewers (i.e. the subjects of study) whenever possible to improve the trust and relationship between funders and the research community.

References

1. Riva JJ, Malik KM, Burnie SJ, Endicott AR, et al.: What is your research question? An introduction to the PICOT format for clinicians. *J Can Chiropr Assoc.* 2012; **56** (3): 167-71 [PubMed Abstract](#)
2. de la Cuesta B, Imai K: Misunderstandings About the Regression Discontinuity Design in the Study of Close Elections. *Annual Review of Political Science.* 2016; **19** (1): 375-396 [Publisher Full Text](#)
3. Graves N, Barnett AG, Clarke P: Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. *BMJ.* 2011; **343**: d4797 [PubMed Abstract](#) | [Publisher Full Text](#)
4. 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.* 2021; **11** (5): e047386 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the topic of the review discussed comprehensively in the context of the current literature?

Yes

Are all factual statements correct and adequately supported by citations?

Partly

Is the review written in accessible language?

Yes

Are the conclusions drawn appropriate in the context of the current research literature?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistics, meta-research, funding evaluation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 17 August 2023

<https://doi.org/10.21956/wellcomeopenres.21675.r64331>

© 2023 Ioppolo B. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Becky Ioppolo 

University of Cambridge, Cambridge, England, UK

Thank you for the opportunity to review this interesting manuscript, which was a pleasure to review. The authors explain that while there is currently a lot of interest in determining research funding by partial randomisation, gathering robust evidence as to whether partial lottery is “better” than other selection strategies would be costly, time-consuming, or otherwise infeasible. Therefore, alternatives to the most robust form of testing (RCTs) are proposed and discussed.

The manuscript is clear with a logical flow between arguments. The authors clearly outline the necessary planning considerations for funders and metascientists interested in testing or evaluating the possible success of a new decision mechanism for funding. The inclusion of the example of potential partial randomisation in the HTA scheme was particularly illuminating for the challenges that would lie ahead.

A few opportunities where I would like to suggest amendment, clarification or possible addition:

1. “Steady as she goes” section: Are there no citations for the concerns about small scale studies not aggregating to a compelling evidence base?
2. Table 1: Are there not more citations for the concerns outlined here? Maybe there aren't, but the final two harm outcomes (reputational damage to funder & stigmatisation of awardee) might be mentioned in Liu *et al.* 2020, for example, or similar papers.
3. Table 1: Could you define or illustrate what stigmatisation of awardee would look like? Do you mean specifically the risk that awardees in partial randomisation could be viewed as “not worthy” of funding by peers, or other forms of non-acceptance by scientific community?
4. Table 1: Would it improve clarity in the table if you specified that the harm outcome is “**Low** quality application”? I assume the intent is that the items in the harm category are all legible

as “Potential increase in”

5. Perhaps Table 1 would be more informative to readers if it also included the stated intention(s) of the funders who have undertaken partial randomisation schemes of funders to date (where available).
6. In the final bullet point in the Logistics section of the HTA example, perhaps it is worth adding some clarification such as: *e.g., if an application is rejected, was it rejected at minimum quality threshold stage, at lottery, or via standard procedure committee adjudication?*
7. In the sample size section of the HTA example, it would be great if you could mention how many applicants typically apply for this scheme, which would give readers a sense of how feasible this is.
8. In the feasibility section of the HTA example, should it not be “(**applications** * number of treatment arms)”? In the example you have laid out, applicants have not been mentioned so I’m wondering whether there is a typo in the equation or an omission from a previous section.

There are a few possible typos that may be worth correcting:

1. Introduction – “subset of applications are” should be “subset of applications is”: “In its partial form, only a subset of applications **is** subject to random selection, once those which are evaluated as clearly fundable or clearly non-fundable have been removed.”
2. Considerations... – add comma after e.g.: (**e.g.**, schemes limited to early career researchers in a specific area of biomedical science)
3. Table 1 – remove] in “Efficiency: Time to deliberation” row.
4. Considerations... - remove comma after hyphen: “...would agree is the most **important - and** a plausible operationalisation...”
5. Considerations... - add comma after e.g.: “(**e.g.**, partial randomisation and funder standard practice) which would be worth detecting if it existed.”
6. Considerations... - Rephrase: “Outcome: **Patient** benefit arising from both portfolios, calculated by the number **of patients** treated with the HTA-approved products”
7. Alternative to RCTs – add accent for author’s name: “For example, **Hernán** & Robins (2016, see also [Hernán et al., 2016](#)) propose guidelines for causal inference from large observational databases.”

Is the topic of the review discussed comprehensively in the context of the current literature?

Yes

Are all factual statements correct and adequately supported by citations?

Partly

Is the review written in accessible language?

Yes

Are the conclusions drawn appropriate in the context of the current research literature?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: metaresearch, research policy, research funding, research evaluation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
