Author’s response to reviews

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

Authors:
Mengyao Liu (TLiu@hrc.govt.nz)
Vernon Choy (vchoy@hrc.govt.nz)
Philip Clarke (philip.clarke@ndph.ox.ac.uk)
Adrian Barnett (a.barnett@qut.edu.au)
Tony Blakely (antony.blakely@unimelb.edu.au)
Lucy Pomeroy (lpomeroy@hrc.govt.nz)

Version: 1 Date: 16 Dec 2019

Author’s response to reviews:

Editor’s comments.

Abstract

1. Remove Surprisingly, - as that is an interpretation – Start the sentence – Most app….
RESPONSE: Changed as suggested.

2. Conclusion should be more in light of the goal – if the goal of the survey was to see if the lottery system should be continued, then please state so in the background and aims, otherwise rewrite it.
RESPONSE: Yes, this was an important reason for doing the survey and we have added that to the abstract.

3. Please include the number of respondents and response rate in the abstract (overall, including 2013-2018 and 2019).
RESPONSE: Added as suggested.
Background

1. Please also mention that lottery also reduces the cost of peer review for funders, which can be quite high (and if you have data on how much this cost was pre 2013, compared to after please include it).

RESPONSE: We’ve added funders as suggested, but the New Zealand HRC do not collect data on the time spent on peer review, so we cannot compare costs.

2. Not compatible with funding through other HRC schemes – please list in this sentence a few examples of things that cannot be funded – (e.g. …), or alternatively in the sentence after this list a few examples of type of studies that were most commonly funded through this scheme

RESPONSE: We have added some further explanation (bottom of page 4).

3. Please specify are Explorer Grants aimed at single researchers or larger teams/consortia, and if there are any specific eligibility criteria regarding the PI.

RESPONSE: We have added a sentence on page 5.

4. In Results, Table 1 seems to imply that you fund 6 projects per year. Can you please add this info it true, or of the actual (average) number to intro after the money that is allocated per project (line 43)

RESPONSE: We have added this information to a new table on page 5.

The Explorer Grants assessment process

1. Please specify the 4 panels areas, and how are the three panel members chosen? How many panel members are there for each 4 areas?

RESPONSE: We have added this information to page 6.
2. Where and how does the 3 panel member scoring look like, is it yes no mark of eligibility for each of the two questions through an online system or in writing? Do both need to be yes for a project to be randomized?

RESPONSE: We have added this information to page 7.

3. Could you add information on agreement scores for eligibility (this may be out of the scope of this paper, but a slight indication of this would be welcome) to explain why 3 panel members are required.

RESPONSE: This is an interesting exercise, but it better tackled by an investigation of the scoring process whereas the focus of this paper was the reaction of applicants. Three reviewers were used to avoid being reliant on the opinion of one reviewer.

4. Can you disclose the exact random number generator that you use? And which “number/method” is yes – for funding (see comment 6 below).

RESPONSE: We have added this information to page 7.

5. Please move the sentence: The HRC considered random funding to be a fair and transparent approach… to the background, and leave this section to contain only specifics of the method.

RESPONSE: Moved as suggested.

6. I would also welcome if you would specify how do you order eligible applications in the system – by date when they were submitted or when they were found eligible, i.e. a little explanation on how the final order of those that will be assigned a random number is compiled.

RESPONSE: Do you mean the order in which the applications are peer reviewed? Perhaps our new text on page 7 answers this question.

7. Please move the sentence: All applicants are informed of the outcome as being either “Declined”, “Fundable but not funded” or “Funded”, to follow the sentence: Those meeting the criteria enter the pool of fundable applications, whilst those that do not are declined.
RESPONSE: Moved as suggested.

Methods

1. You specified in the paragraph above the methods section, that applications are called annually, but state 9 calls from 2013 to 2019 – please clarify.

RESPONSE: This should have said 7 calls and we have fixed that now (page 8).

2. Please list if there were reminders sent to applicants, and how many emails bounced/were invalid.

RESPONSE: We have added the information on reminders to page 8. Information on bounced or invalid emails was not collected.

3. You said: For applicants who had more than one application, they were represented once using their best outcome in the order of – can you try to determine if the applicants were more likely to respond to your survey if they had more than one application? Additionally, can you provide the median number of applications per unique user with min and max values (i.e. how many max times so far has an individual applied for the scheme)

RESPONSE: We do not have linked identifiers for this analysis.

4. Can you please also confirm if at the time of the survey for 2013-2018 – all respondents already knew their outcomes? Or did the 2018 did not know them (in that case should they be grouped with 2019 applicants?

RESPONSE: Only the 2019 applicants were unaware of their outcome.

5. You listed that 7 applicants completed the survey in 2018 and 2019 – am I right to presumed that in 2018 they did not receive the grants, and in 2019 – they were still not aware of that decision – can you please provide in the results a sub-analysis of these 7 (with possible separate table for them in the appendix for their reposes – as it is very interesting to see if their opinions in any way changed from 2018 to 2019? And if their responses are very different from responses of those who did not get awarded funding.
RESPONSE: It is possible that those applying in 2018 and 2019 did receive funding in 2018. We think this sample size is too small to warrant any new analyses.

6. Please provide how you coded the open-ended responses, and was this done by one or more authors?
RESPONSE: The open-ended responses were not coded, instead we include illustrative comments. We have added this as a limitation (page 15).

7. Finally, could you possibly also add information on the number of male and female applicants per year? And gender responders in the survey?
RESPONSE: Gender was not collected as part of the survey.

Results:
1. Please reformat the table 1 to:

<table>
<thead>
<tr>
<th>Time period</th>
<th>Outcome</th>
<th>Response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013 to 2018</td>
<td>x% (x of y)</td>
<td></td>
</tr>
<tr>
<td>Ineligible</td>
<td>24% (48 of 199)</td>
<td></td>
</tr>
<tr>
<td>Not funded</td>
<td>(RNG not allocated)</td>
<td></td>
</tr>
<tr>
<td>Funded</td>
<td>(RNG allocated)</td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>68% (50 of 74)</td>
<td></td>
</tr>
</tbody>
</table>

*All applicants of 2013 to 2018 period were emailed in September 2018 (and were aware of the funding allocation), while 2019 applicants were emailed in January 2019 (and were not aware of the allocation).

RESPONSE: Changed as suggested.

2. Table 1 implies that there were more applications in 2019, compared to the other years, can you please provide a summary or table of applications per year or call (see also my Q1 for results)
RESPONSE: We have added a new table 1 with the annual number of applications.

3. Please delete table 3, and include responses to all 8 questions of the survey in one big table. I also suggest you format it in the following way to highlight differences between 2013-2018 and 2019 survey (you don’t need to list the questions in order they were asked in, you can group them as you did for tables 2 and 3):

<table>
<thead>
<tr>
<th>Question</th>
<th>2013-2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>Unsure No</td>
</tr>
<tr>
<td>Q2 – Text</td>
<td>115 (91)</td>
<td></td>
</tr>
<tr>
<td>Q3 – Text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6- Text</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RESPONSE: We have merged the table 3 and 4 as suggested. We have added a table to the appendix that splits the responses by time, we think the combined tables provide a clearer picture of results.

5. Please list first the number of open-ended responses you received per each question, you can add this to the table 2 I suggested above, or in a table in the appendix – if latter then provide a sentence on average or median no. per question in the results (with accompanying SD or IQRs).

RESPONSE: We have a new table as an additional file and summary statistics on the percent of comments in the text.

6. I understand that your sample size is small for those you can link to outcomes, but I do not think that pairing of the declined and not-funded is appropriate – as I would expect that those who were declined, would have harsher words for the system than those not funded. Can you qualitatively say this is not the case? If yes say so in the methods, and then combine them, otherwise I would refrain from doing so. Do you have information on how many of those that were declined reapplied next year, compared to those that were not funded?
Please split the table 4 in the 3 categories, to see how many are per each of the 3 categories, and then say in text that you merged them into two for analysis – alongside chi-square, you could do either an ordinal or binary regression (if you merge them – again which I think you don’t need to).

RESPONSE: We have now used three categories for the funding outcome as suggested and updated the table and used a non-parametric statistical test instead of the chi-squared test.

Discussion:

1. It has been using random funding since 2013 and has grown in size with the most recent round allocating five times as many grants as the original round. – please list in brackets (xy in the first call in 2013 vs xy in the last in 2019).

RESPONSE: We have added a new table 1 with this information.

2. Given this rare innovation in funding, there should be wide interest in hearing from the researchers involved. – delete this sentence (as you have in intro), and rather say: Having surveyed the researchers that applied for this funding we found xy in support of the scheme, and 25% opposing the lottery system.

RESPONSE: We have deleted this sentence from the discussion.

3. Only 1 in 4 survey respondents thought – I would not say only, as this is 25%, that is not a per se a small percentage - see my above suggestion on rephrasing this.

RESPONSE: We have reworded this start of this sentence.

4. The outcomes from potentially transformative research are almost impossible to predict, and hence a lottery is a logical and evidence-based approach to funding where expert prediction models have failed (20). – Please rephrase, as this sentence seems to imply there is evidence for this – rather the cit. 20 you used is an opinion paper of a prestigious individual. Not all of your survey responders agree – and they are also scientist. So rather say, prof Ioannidis in 2018 paper saw lottery as ….

RESPONSE: We have changed the key part of the sentence to “Ioannidis argued that a lottery is a logical approach to funding that could save time and would better spread funding across more researchers”.


My hope is that you will take the reviewers and my comments as benign and enthusiastic suggestions and requirements of those working in this field who wish to see it grow and improved, so that your resubmitted version is soon published in our journal.

RESPONSE: We are very grateful for the thoughtful reviews and many useful suggestions.

Reviewer #1: This is a very original and informative study that will be of great interest to the field. It is very interesting that this is the first example, world-wide, of the use of a lottery scheme to award research funding. The fact that it commenced in 2013 and it is only now that this study has been conducted makes me wonder why this wasn't evaluated and reported on earlier? (though, granted, there would have been fewer applicants available and thus fewer data to analyse). Nonetheless, this article should have a good impact in the field of research funding and peer review in the health sciences, and beyond. My suggestions below are to enhance the clarity of the text.

Page 3 line 36. It isn't immediately obvious why using a lottery would mean shorter application forms. I understand what you mean here but presumably some funders would still require a sufficient level of detail to be given in order to assess eligibility for the lottery. Later in the discussion section it is commented that a surprising finding is that the lottery didn't reduce the time spent writing applications. Also it is suggested that for larger grant schemes random allocation would only be used for applications in the grey zone between not fundable and outstanding. Thus, there would remain a requirement for sufficiently detailed proposals. Please consider how this issue might be better introduced, for example by citing the literature (refs 22 and 23) earlier.

RESPONSE: We have added more detail on page 3 and moved the two references to this page.

Page 3 line 47. You could also consider mentioning other biases that occur in peer review, such as ageism, and professional/disciplinary prejudices in the sciences (e.g. certain specialities or sub-disciplines seen as 'the poor relation', which sometimes happens in the health sciences, and beyond).

RESPONSE: We had already mentioned ageism. The issues with disciplinary prejudices is interesting, but the only evidence we are aware of is anecdotal.
Page 3 line 52. The end of the sentence "which occurs because the review process is likely to have an element of randomness". I am not sure what is meant here. I imagine it is a reference to the often unpredictable views of peer reviewers and lack of consensus between peer reviewers on applications which could be seen as happening at random. However, the sentence as written is ambiguous, so please clarify.

RESPONSE: We have clarified this sentence.

Page 4 line 2 - I think the sentence would be better if it said "peer review score was a poor predictor of the quality/productivity of the subsequent research output". You could also consider using "impact" instead of "output".

RESPONSE: We have now said “number of research outputs”. Impact can be difficult to measure because it includes returns that can happen years later and via indirect benefits. Outputs is the better term here, because the immediate outputs of a project (e.g., papers, conferences) are much easier to measure.

Page 5 line 48, if applications are selected randomly and thus treated equal how can they be ranked as the sentence states?

RESPONSE: We have reworded this section (page 7).

Page 6 line 57. It would be informative if the sentence could mention what type of questions these were i.e. fixed response categories of yes/no/unsure (even though the questions themselves are reported later in the manuscript).

RESPONSE: We have added that there were seven closed-ended and one open-ended question.

Page 7 line 54 state in full the CHERRIES acronym.

RESPONSE: We have added the full text.
Page 8 line 41 there was less support for random allocation of funds for other grant types - do you know how many applicants had successfully applied for funding from these other grant types? Could there be a bias here because applicants are less informed about the other grant types? Perhaps there could be a similar survey done of applicants to the other grant types to assess their views?

RESPONSE: This is interesting, but we do not have information from the respondents about what other schemes they applied to. We think any bias would be small, because the major point of difference here is random allocation versus standard allocation using peer review which would be the dominant model for almost all the other schemes they would have applied to.

Page 14 - Limitations. You could consider whether there is any bias from the survey being conducted by the funder, as opposed to a more neutral agent. The respondents (particularly those who have received Explorer Grant funding) may have been inclined to give more desirable, positive, responses about the random method of funding (notwithstanding the anonymous nature of the survey responses). I think the risk of this bias is probably low, but you might like to consider whether you think it is plausible.

RESPONSE: We have added this as a potential source of bias (page 15).

Page 14 - line 59 add the word "funding" before "success rate".

RESPONSE: Added as suggested.

There isn't much in the way of recommendations for funding agencies and applicants, other than there should be further uptake of lottery systems for awarding research funding. You might consider qualifying this statement with the fact that the evidence presented only currently supports the use of a lottery for transformative research funding schemes. Further research is needed in the use of lottery schemes in other research funding schemes, particularly larger scale funding programmes. Also, in relation to innovation in health research funding processes you could, if space allows, refer to other funding models such as 'sandpits' which have been used in recent years as an alternative to the standard peer review model.

RESPONSE: Sandpits are interesting, but it is outside the scope of this paper to discuss other innovative funding schemes as that would be better achieved by an in-depth review paper. We agree that our recommendations are limited in terms of number, but we think the implications are large because if the paper motivates other funding agencies to experiment with lotteries, then that is a big change to current practice.
Reviewer #2: The authors present survey evidence of applicant perceptions of the acceptability of a lottery to allocate research funds. They find that, of the 126 applicants, 63% felt that a lottery was an acceptable way to allocate funds for Explorer Grants.

Overall, I felt this was an interesting paper that deserves to be made more publicly available. Scientists have voiced increasing levels of concern that current models of funding are inefficient and arbitrary; information on whether applicants perceive lotteries as acceptable/fair is valuable. I do have some thoughts and concerns that I would like to see addressed in some form, which I list in bullet points below.

* It would be useful to know the percentage of people who make it past the first step in the review process for Explorer Grants. If a sufficient number of applicants are filtered out at the first step, the Explorer Grant scheme could de facto be operating so similarly to a more traditional peer review scheme that it is perceived that way by applicants.

RESPONSE: We have added a new Table 1 with this information.

* Do the authors have access to any data on the perceived acceptability of more traditional grant schemes? Or data on the perceived acceptability of the Explorer Grant scheme prior to the implementation of the lottery? I found myself wondering if the 63% represents a low value, a high value, or par for the course (and whether the lottery may have improved perceived acceptability for this particular granting mechanism).

RESPONSE: We are not aware of any previously published surveys. We have added this is a point for future studies in the limitations section (page 16).

* I would like some data on the representativeness of the people who returned their surveys given the 39% response rate. Do the authors have a way to compare the demographics of the people who returned their surveys to the general pool of applicants? Or, alternatively, can they compare the funding rate of the surveyed applicants to the normal funding rate of Explorer Grants (see https://bmjopen.bmj.com/content/3/5/e002800 for an example of this approach)?

RESPONSE: This would be a useful analysis, but we cannot use data on applicants who did not return the survey because they did not consent.
* Similarly, I would like to see the authors temper their interpretations with some reflection on the possibility that their results may have been affected by selection bias. For example, is it possible that most of the people who returned their surveys were irritated at not receiving funding and were expressing "sour grapes" at the funding scheme? (This is just one possibility - I'd just like to see some discussion of selection bias more generally)

RESPONSE: We had a higher response rate amongst successful applicants, hence we think the overall bias is more likely to be towards a favourable opinion.

* People are not terribly accurate at retrospectively estimating how much time they spend on a task (https://www.researchgate.net/profile/Ann_Bostrom2/publication/247937960_The_Overestimated_Work_Week_What_Time_Diary_Measures_Suggest/links/0c96053bc136625223000000.pdf). I think the authors should temper their interpretations of the time estimate data accordingly

RESPONSE: Yes, this is an issue and we say, “Retrospective questions about time spent are hard to complete accurately, and four applicants mentioned this difficulty in their comments.”

* Please put the study N and response rate in the abstract

RESPONSE: We have added this as the first sentence in the abstract results section.

* The authors might consider mentioning theoretical evidence suggesting that typical "contest" models of allocating funding are inherently inefficient https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000065

RESPONSE: We did already reference this paper in discussing the idea that lotteries would reduce application times.

* To facilitate reproducibility, the authors should deposit their data, materials, and analysis code on a public repository (I recommend the Open Science Framework (https://osf.io/), but there are other options

RESPONSE: Unfortunately the datasets are not publicly available because consent was not obtained for the data’s wider use.
Reviewer #3

This is a well-written manuscript on an important topic: introducing a lottery element to the funding evaluation system. There is currently an international debate regarding how to improve the funding system, and one important suggestion has been to introduce a lottery element to grant evaluation. Right now, external and/or panel peer review is the international standard for nearly all funding agencies and funding schemes. One New Zealand funder has introduced an element of randomness into one of their funding schemes in 2013. This manuscript describes the results of a survey done amongst the applicants to this particular funding scheme (in the years 2013-2019). It is obviously of high importance to evaluate how applicants are experiencing such funder experiments, and whether they would recommend its continuation and possible extension to other schemes. I therefore congratulate the authors to perform such a survey and to make the outcomes of the survey public. As such, the manuscript is thus of high importance and will likely receive broad interest. The manuscript also fits well into the journal's scope. Overall, I am therefore looking forward to see the manuscript being published in due time. That said, I also have some extensive criticism with the manuscript, and I would recommend that the authors first revise their manuscript accordingly.

Major criticisms:

1. The paper lacks some crucial information as to how this specific grant scheme had been evaluated previously, before 2013. This is important information because from what I can guess, the randomisation element seems to not have been the only change introduced in this particular funding scheme in 2013: Alongside, (1) the length of the application may have been changed (towards shorter and/or more simplified applications), (2) there might perhaps have been a switch at the evaluation step one, from external peer-review to panel peer review (but this is not clear), and (3) the assessment might have changed from being applicant-explicit grant proposals to being anonymous grant proposals. If this all had indeed been the case, please state so explicitly.

RESPONSE: The Explorer Grant scheme was started in 2013 with a short form, panel review and anonymous applications.
Our aim was to examine this unique approach to funding with no specific reference system. We assume that respondents reflected on the difference based on the past Explorer Grant scheme and the other schemes they apply to.

2. Make some important figures more explicit. From table 1, I extracted that - after removing the cohort of 2019 (for which funded/non-funded data were not available yet) - a mere 15% of the applicants had received funding between 2013 and 2018. Really interesting was to calculate that a total of 79% (!) of the applications had been declined at step 1, meaning via panel review. In contrast, at step 2 (the lottery step) 73% had received funding - which is a near-reversed situation. Indeed, one really wonders then about the overall actual impact of the lottery element, as only 14 out of the in total 251 proposals (5.6%) had been removed by chance. Please be more explicit in the manuscript that this was in fact mostly a panel review with subsequently only a very minor lottery element, and provide these numbers clearly in the text. One wonders also about the criteria for declining at step 1 - were really 79% of applications completely unfundable? That seems to be a quite high rate to me, and my guess is that it may have been connected to how the two criteria were interpreted by the panel (that consisted only of three persons). Alternatively, perhaps panel members were even instructed to sieve out as much as ¾ of applications at this stage? One wonders. Indeed, once the importance of step 1 becomes obvious, some of the responses are not so surprising any more - e.g. that researchers prepare their applications equally carefully and spend a near-equal amount of time on them.

RESPONSE: The scheme does not use the word “unfundable” rather they are “ineligible” because they don’t meet the aims of the scheme. We agree that the responses from applicants make sense given the high failure rate at the first hurdle.

Panel members were not given numerical targets for filtering. The high number of ineligible may be due to the high threshold for “transformative research”.

3. Abstract and page 8: I find it confusing to consider a 63% agreement to be a "general agreement" - this seems to be more like a slight majority to me. Perhaps say so, and then also mention in the abstract that only ¼ were really against the idea of a lottery element? Even then, as the authors clearly showed and discussed, there is a bias by the "lucky winners" as well, making one doubt whether the overall conclusion for support of the lottery element is actually warranted. Similarly, I doubt that 40% can be called "less support", in particular because 37% were against the idea - which seems more like a quite well-divided opinion to me. Please rephrase, also in the conclusions.
RESPONSE: We have added the 25% against the lottery in the abstract and have removed the word “general” in order to let the 63% figure speak for itself. The reviewer thinks this figure could be higher, but we see it as a good majority and enough to provide ongoing support for a lottery. As reviewer #2 says, it would be useful to have the results for alternative funding schemes, as no scheme is likely to get 100% support when success rates are generally below 20%.

We used “less support” to describe the drop in support for using a lottery from 63% to 40% for the Explorer grant scheme to other schemes. We agree that the 40 to 37 percent split of positive to negative responses for other schemes is an almost perfectly divided sample, and we have changed our wording in the abstract and results (page 9), calling it, “a close to perfect split in opinion”.

4. It would be important to put even more focus on some of the other elements that emerged from the survey (besides the lottery). For example, there was a really broad consensus (89% & 91%, with only 4% & 2% against!) that the length and the format (simplification?) of the application were appropriate and that anonymization of applicants was a good idea. I would think that both types of insights are of at least equally big interest to the readers and should be taken up much more prominently (e.g. in the abstract and conclusions as well). Thus, the authors should consider making these two insights two of the main results of the survey, besides the far more moderate enthusiasm for the lottery element (also because the lottery element was very limited in its execution). A broader discussion of the anonymization element could e.g. usefully make a connection to the "Matthew effect" that has been widely detected also at the funder level (see e.g. Bol et al., The Matthew effect in science funding, PNAS 2018).

RESPONSE: Yes, we agree these are interesting. We used the word “almost all” and “strong support” to describe the 91% and 89% results.

The Matthew effect paper is interesting, and it could be that the knowledge that previous funds were allocated randomly did reduce any Matthew effect. We aim to ask the reviewers about this in a further planned survey. Also, we can directly observe any Matthew effect as we have a study planned to compare the outcomes of those that were and were not randomly funded, with future funding as an outcome.
Minor criticism:

- Page 3: Funders do not exclusively fund "rigorous experiments". Please rephrase.

RESPONSE: This was stated as the aim of research not funders.

- Page 3: Please add "lotteries may also minimize" (this is only a suggestion so far)

RESPONSE: We do not think that we do need to add “may” here given the knowledge that funding systems have shown such biases and lottery would be unbiased.

- Page 5: "big picture". I would rephrase to "focus of assessment to the project idea, while…"

RESPONSE: Changed as suggested.

- Page 9: Is it possible to specify the (approximate) numbers? ("a number", "some", "others" is a bit vague, perhaps)

RESPONSE: We are happy with this wording as the reader has already seen the numbers and our aim here is to provide an easy to read summary.

- Page 9/10: "Stand-out applicants" - this idea should be discussed more, also in the discussion.

RESPONSE: We reviewed this section and are happy with our discussion.

- P. 12: It might be better to mention the growth of the scheme in the introduction, not in the discussion.

RESPONSE: See new Table 1 in the background section.

- Page 13: This is not surprising, really (see Major criticism 2). Please rephrase.

RESPONSE: We were surprised by these results at the time, so we prefer to keep our phrasing.