Peer review and beyond: Randomisation at the margin in the selection of research grant proposals Prof. Dr. Hans-Dieter Daniel

PEERE 2018 – International Conference on Peer Review National Research Council, Rome, 7–9 March, 2018



Peer review under fire: Should peer review be abolished?

In 2015, the former *British Medical Journal* editor Richard Smith called, in *Times Higher Education*, for peer review to be abolished.

"Peer review", he wrote, "is supposed to be the quality assurance system for science (...). In reality, however, it is ineffective, largely a lottery, anti-innovatory, slow, expensive, wasteful of scientific time, inefficient, easily abused, prone to bias, unable to detect fraud and irrelevant."

(Source: "Ineffective at any dose? Why peer review simply doesn't work", Times Higher Education, 2015, 28 May, Opinion)

Plan for peer review based research funding opposed in Sweden

A proposal for a more peer review based funding of Swedish research prepared by the Swedish Research Council and three other research councils has been rejected by a majority of Swedish universities and other stakeholders (the allocation of the block grants for research at universities is currently based on bibliometric indicators). Out of 34 responses from universities and university colleges, only 3 are positive.

(Source: Myklebust, University World News Global, Issue 404, 11 March 2016)



Short outline of my own research on ...

- peer reviewing and selection of manuscripts submitted to journals for publication
- the selection of research fellowship applicants
- peer reviewing and selection of grant proposals submitted to research councils for funding











News & Network

Fellows & Alumni FUTURA News & Pictures Study on BIF

Fellowship holders of the Boehringer Ingelheim Fonds

outstanding

First comprehensive study on peer review for the allocation of fellowships to young scientists

Social science researchers at the Swiss Federal Institute of Technology Zurich (ETH Zurich), Switzerland, showed in a study published in 2004 that the Boehringer Ingelheim Fonds supports outstanding up-and-coming young scientists. Furthermore, the study demonstrates that the peer review procedure is valid for the selection of doctoral and post-doctoral fellowship recipients.

Dr Lutz Bornmann and Professor Dr Hans-Dieter Daniel analysed a total of 2,697 applications submitted to the Boehringer Ingelheim Fonds between 1985 and 2000. They examined the three most important quality criteria for selection procedures: (i) does the foundation really select the best young scientists (predictive validity)? (ii) is the selection procedure reliable? and (iii) are all groups of applicants treated equally (fairness)? On the whole, they confirm that the selection procedure of the foundation is highly valid.

The most important results of the study



Home » Research funding » Decision-making procedure » Studies of (Self-)Evaluation

FWF Programmes

Application

Overview of Calls

Applications from abroad

Information for Principal Investigators

Personnel costs

Decision-making procedure

Studies of (Self-)Evaluation of FWF's decision making procedure

The legitimacy of a funding agency in basic research depends at least on the following factors:

- 1. organisation's ability to minimise distortions in approval probability by its decision procedure
- 2. the scientific quality of results produced by funded research proposals
- 3. the acceptance of the procedures by the scientific community

To ensure these requirements, the FWF started a sequence of empirical studies in 2010 which contain analyses of the decision making procedure (peer review) as well as statistical and bibliometric analyses of FWF funded proposals.

Outline of Talk

What follows is an attempt to take stock of research on peer review with a focus on the three psychometric quality criteria for professional evaluations:

- Inter-rater agreement
- Fairness, and
- Predictive validity

The **second part** of the talk will introduce the Volkswagen Foundation's pilot study on a partially randomized selection of research projects ("randomization at the margin", Adam B. Jaffe, 2002)



Inter-Rater Reliabilities of Grant Peer Reviews at the Austrian Science Fund (FWF)



Objective: Inter-rater reliability of peer review ratings (Ex-ante evaluation of research grant proposals)

Data: N = 8,329 proposals with N = 23,414 overall ratings by reviewers

Statistical analysis: Calculation of intra-class correlation (ICC) (single-rater and mean rater reliability)

Results:

- The single-rater reliability is .259 (low)
- The **mean rater reliability** of the aggregated peer ratings (based on the application of the Spearman-Brown equation) is **.495 (moderate) with 2.8 reviews per proposal on average**
- Reliability is lowest in the biosciences

Source: Mutz, R., Bornmann, L. & Daniel, H.-D. (2016). Heterogeneity of inter-rater reliabilities of grant peer reviews and Its determinants: A general estimating equations approach. PlosOne, *7(10)*, e48509



Fairness of Peer Review

Research grant proposals are supposed to be judged solely on the basis of their scholarly quality, not on particularistic characteristics of the applicant (cf. Merton's Ethos of Science).

Personal attributes like sex, academic status, and nationality of the researcher should play no role whatsoever in assessments of quality.

A Problem that Affects Bias Research

 Does a research grant proposal receive better reviews and thus a higher funding rate due to preferential biases affecting the review and decision-making process?

or

 Are favorable reviews and higher funding rates a simple consequence of the high scientific quality of the corresponding grant proposal?



Mean predicted probabilities for the acceptance of a manuscript for publication in *Angewandte Chemie–International Edition* in dependency (a) upon share of authors having institutional affiliations in Germany and (b) number of institutions mentioned in the acknowledgements of a manuscript <u>when the</u> <u>quality of the manuscript is taken into account in the form of citation counts</u> (measured ex-post)



Source: Bornmann & Daniel, Research Evaluation, 2009, 18(4), 262–272.



Predictive Validity of Peer Review: An Editor's Doubt

"When I divide the week's contributions into two piles – one that we are going to publish and the other that we are going to return – I wonder whether it would make any real difference to the journal or its readers if I exchanged one pile for another".

> Fox, T.F. (1965). Crisis in Communication. London.



Predictive Validity of Funding Decisions: EMBO long-term fellowships to postdoctoral researchers (approval rate: 20%)

LTF programme





Predictive Validity of Funding Decisions: EMBO study



Approved and rejected applicants: Median numbers of citations for papers published prior to application (retrospective peer review)

С



Approved and rejected applicants: Median numbers of **citations for papers** published **subsequent to application** (prospective peer review)

Source: Bornmann, L., Wallon, G., & Ledin, A. (2008). Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European Molecular Biology Organization programmes. *PLoS One, 3*(10), e3480.



In conclusion

- The research of my group shows that **retrospective peer review is valid**, whereas **prospective peer review lacks predictive validity**.
- There is a lot of evidence in the literature that the peer review process lacks fairness and single-rater reliability (i.e., reviewers are not interchangeable. Acceptance of a research proposal / manuscript / applicant depends too much 'on the luck of the reviewer draw').
- Due to limited funding resources / publication space and the resulting low approval rate there are a lot of type two errors (false negatives) in the selection decisions (that is, good grant proposals / manuscripts / applicants are rejected).
- Peer review should certainly not be abolished, but research councils / publishers should be experimenting with different approaches for selecting research grant proposals and fellowship applicants for funding or manuscripts for publication.



Proposals to reform the peer review process

To improve the reliability of the peer review process

 one should increase the number of reviewers (increase to 6 or 8)

OR

 one should test the reader system proposed by Marsh and colleagues (2006, 2008)



The Reader System

In the reader system, small numbers of experts (3–4) are used for each research field. The same experts review all the proposals (between 16 and 25) in their field independently. Because all experts read all of the grant proposals in their field, each has a similar frame of reference from which to evaluate any given proposal (comparative judgements). By using a ranking procedure, differences in leniency/harshness as a source of disagreement between the ratings of different reviewers can be eliminated. In comparison to the traditional peer review approach, the reader system is substantially more reliable, timely, and cost efficient. Source: Jayasinghe, U. W., Marsh, H. W., & Bond, N. (2006). A new reader trial approach to

peer review in funding research grants: An Australian experiment. *Scientometrics*, 69, 591–606.



Proposal for a radical departure from peer review

To neutralize biases in funding decisions and to **increase grantee diversity** with regard to age, gender and research field Fang & Casadevall suggested in 2016 a radical departure from the present peer review system, that is, a modified funding lottery (cf. *mBio* 7(2):e00422-16; *Science*, 352(6282),158):



Proposed scheme for a modified funding lottery (Fang & Casadevall 2016, p. 5)



If all meritorious applications are entered in a lottery, this is also referred to as ,randomization at the margin' (Adam B. Jaffe, 1998, 2002).



Volkswagen Foundation is just starting a test with selection by lot

- The Volkswagen Foundation in Germany
- is the third largest European
- private Charitable
- **Research Foundation**
- by total giving p.a.





Experiment! – In Search of Bold Research Ideas



- science and engineering, life sciences
- research projects
- up to 120,000 Euro (flexible use)
- up to 18 months

Career stage: **Total** (2013-2016): Postdoctoral researchers & professors (in ■) € 6.7 Mio. €, 67 grants out of 2303 proposals (approval rate: 3%)



How to pick the right projects? Newly designed application and review process to reduce the workload of all people involved since 2013



- short standardized applications
- shortlist pre-selected in-house (triage)
- external jury (8–9 researcher)
- 2–stage review procedure:
- •2–3 jury member will read all (blinded) proposals in their discipline (30–40) and each member will nominate her / his favorites
- •joint meeting of the jury and selection of 15 to 20 (blinded) proposals for funding
- optional funding joker, 1 p. p.
- no additional review reports



From 2017 onwards: Partially randomised selection in 'Experiment!'



In total (jury & lot): approx. 30–40 grants p.a. (120–160 in 4 years, 2017–2020). Evaluation of the outcome of the trial in 2023 at the earliest.



Challenges for the evaluation of Volkswagen Foundation's pilot study

- How to measure 'bold research ideas' and riskiness?
- How do we know that the preselection by the Foundation's staff works? (i.e., low rate of false negatives)
- How to deal with the **heterogeneity** of disciplines, career stages, institution types (review panel with 8 to 9 members only)?
- Diverse research topics: hard or even impossible to find **matching pairs** of grants for the **comparison of outcomes**!
- A limited number of grants: how to avoid underpowered statistics?
- Are simple binary **outcome indicators** sufficient (publication: yes/no, follow-up funding: yes/no, ...)?
- **Hawthorne effect**: Should the Foundation inform by means of the Call for Proposals that some of the applications are being funded because they were selected by lot? (running of the experiment might change the applicant pool)
- Research ethics (**disclosure arrangement**): Should the funded applicants after the completion of the project know in which group (selected by jury *versus* selected by lot) they were in?
- Will the scientific community accept a partially randomised selection of research projects?
- How to assess the **political feasibility** of a partially randomised distribution of research funds?



For further information ...

http://www.psh.ethz.ch/en/

http://www.evaluation.uzh.ch/en.html

Thank you very much!