

Potential Bias in Peer Review of Grant Applications at the Swiss National Science Foundation

Anna Severin^{1,2}, João Martins^{1,3}, François Delavy¹, Anne Jorstad¹, and Matthias Egger^{1,2}

¹Swiss National Science Foundation, Bern, Switzerland

² Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

³ European Research Council Executive Agency, Brussels, Belgium

Corresponding Authors:

Anna Severin, Swiss National Science Foundation, Wildhainweg 3, 3001 Bern, Switzerland and Institute of Social and Preventive Medicine, Mittelstrasse 43, 3012 Bern, anna.severin@snf.ch, ORCID: 0000-0002-6231-5695

João Martins, Swiss National Science Foundation, Wildhainweg 3, 3001 Bern, Switzerland and European Research Council Executive Agency, Place Charles Rogier 16, 1210 Brussels, joao.martins@snf.ch, ORCID: 0000-0001-7961-4280

Abstract

Background. The Swiss National Science Foundation (SNSF) supports fundamental and use-inspired research in all academic disciplines. As part of the evaluation procedure, grant applications to the SNSF are reviewed by external reviewers. The legitimacy of funding decisions depends on its ability to base funding decisions solely on the scientific merit of grant applications.

Aim. We examined whether the following factors influenced the scores given to grant applications submitted to the SNSF: (1) source of nomination of the reviewer, (2) the gender of the applicant and the reviewer, and (3) the country of affiliation of the reviewer.

Methods. We gathered data on 38,250 external reviews of 12,294 unique grant applications across all disciplines between 2006 and 2016. Proposals were rated on a scale from 1 (=poor) to 6 (=outstanding) by 26,836 reviewers. We used linear mixed effects regression models adjusted for research topic, applicant's age, nationality and affiliation.

Results. We found that in univariable analysis applicant-nominated reviewers awarded higher evaluation scores than reviewers nominated by the SNSF. Further, reviewers affiliated with research institutions outside of Switzerland gave more favourable evaluation scores than reviewers affiliated with Swiss institutions. Finally, male reviewers awarded higher evaluation scores than female reviewers and male applicants received more favourable evaluation scores than female applicants. When we controlled for confounding factors, adjusted differences changed little for source of nomination and country of affiliation. In contrast, the gender differences nearly disappeared, which indicates that most of the gender effects observed in univariable analysis is explained by differences in scores between research topics and applicant affiliations.

Conclusions. Our study showed that peer review of grant applications at SNSF may be prone to biases stemming from different applicant and reviewer characteristics. Based on this study the SNSF abandoned nomination of reviewers by applicants, and made members of panels aware of the other systematic differences in scores. We encourage other public funding bodies to conduct similar studies.

Introduction

In public research funding, peer review of proposals by experts in each field is the accepted best practice for determining which projects are allocated funding. Peer review is an important element of quality assurance in the scientific community (Harman 1998). Against this background, a wealth of literature is concerned with the question of the legitimacy of peer review decisions. Generally speaking, the legitimacy of funding decisions relies on a funder's ability to minimize distortions in grant evaluations resulting from the influence of factors that are unrelated to the actual quality of the grant applications (Lutz Bornmann and Daniel 2007). Empirical studies usually examine whether and to what degree such factors influence funding decisions (Demicheli and Di Pietrantonj 2007). These studies suggest that the evaluation of applications is prone to biases that can stem from a number of sources that are related to both applicants' and reviewers' characteristics, including, but not limited to, age or institutional affiliation (Lutz Bornmann and Daniel 2007).

Mandated by the Swiss Confederation, the Swiss National Science Foundation (SNSF) supports basic research and use-inspired basic research in all academic disciplines. With the aim of detecting and reducing potential biases in funding allocation, the SNSF started monitoring its evaluation processes in 2006. The main funding scheme of the SNSF is project funding, which provides support to independent researchers who propose research on self-chosen topics (Swiss National Science Foundation 2016). The final decision on funding is taken by the National Research Council (NRC), a body consisting of pre-eminent researchers based in Switzerland, which takes into account the reports of two or more external expert reviewers. Several factors are of concern in the context of external peer review of grant applications at the SNSF, including:

1. *Nomination of reviewers:* The SNSF previously allowed grant applicants to suggest reviewers to evaluate submissions via a "positive list". A study of the Australian Research Council found that applicant-nominated reviewers tend to give better ratings than panel-nominated reviewers (Marsh, Bond, and Jayasinghe 2007). Similarly, a study of peer review in biomedical journals found that author-nominated reviewers made more favourable recommendations than editor-nominated reviewers (Schroter 2006).
2. *Country of affiliation of reviewers:* The SNSF frequently invites reviewers from abroad to review grant applications. The Australian study found that reviewers affiliated with an US research institution were more lenient than reviewers affiliated with institutions located in the United Kingdom, Germany and Australia (Wood and Australian Research Council 1997). In contrast, a study of the Austrian Science Fund suggested that reviewers affiliated with research institutions located in countries known for high scientific productivity were generally more stringent (Fischer and Reckling 2010).
3. *Gender of principle applicants and reviewers:* Potential discrimination against women is the most frequently investigated bias in the context of grant peer review (Mutz,

Bornmann, and Daniel 2012). A meta-analysis of gender bias studies showed small gender differences in grant awards where men were more likely to receive research funding than women. This meta-analysis also showed that gender differences in evaluation scores vary substantially between funding schemes and funding bodies and may be explained by other, confounding variables (L Bornmann, Mutz, and Daniel 2007).

We analyzed the database of the SNSF to examine whether scores from applicant-nominated external reviewers differed from those from reviewers proposed by the SNSF, whether they differed between reviewers affiliated with a Swiss research institution and reviewers from foreign institutions, and whether they differed depending on the gender of the applicant or reviewer.

Materials & Methods

Evaluation of Grant Applications at the SNSF

The evaluation of grant applications at the SNSF consists of four steps. After researchers have submitted their applications, the administrative office performs a formal verification in the first step and assign grant applications to two members of the NRC (referee and co-referee) based on their field of expertise. In a second step, eligible proposals are peer-reviewed by external experts according to the following criteria (Swiss National Science Foundation 2016): with regard to the applicant: 1) *Scientific track-record and ability to carry out the research project*; with regard to the project: 2) *Scientific relevance, originality and topicality* and 3) *Suitability of methods and feasibility*. Reviewers score each criteria on a scale from 1 to 6: (1) *poor*, (2) *fair*, (3) *average*, (4) *good*, (5) *excellent*, and (6) *outstanding*. In addition, reviewers submit an overall score of the proposal.

During the study period, expert reviewers were identified in several ways: (1) grant applicants suggested experts via a “positive list”, (2) NRC referees suggested reviewers, (3) the SNSF administrative offices proposed experts, and (4) experts may have declined to review but suggested other reviewers (Swiss National Science Foundation 2016). Applicants could also submit a “negative list” of reviewers who, because of possible conflicts of interest, should not be contacted. For each application, at least two external independent reviews were required.

In the third step of the evaluation, the two members of the NRC (referee and co-referee) assessed the usefulness of the peer review reports and took them into account when ranking the application relative to other proposals. In the fourth and final step, referee and co-referee present their assessment at the meeting of the corresponding section of the NRC. Each application is voted on and approved or rejected (Swiss National Science Foundation 2016).

Data and Variables

The outcome variable of interest was the overall evaluation score of a grant application ranging from 1 (=poor) to 6 (=outstanding). Explanatory variables included meta-data on applicants and external reviewers, including *source of reviewer nomination* (applicant-nominated vs. SNSF-nominated), *gender of the applicant and gender of the reviewer* (female vs. male) and *country of affiliation of the reviewer* (Switzerland vs. other). The category of SNSF-nominated experts includes reviewers who were proposed by the referee, the SNSF office or by experts who were initially contacted but declined to review. The latter three sources of reviewers were grouped together and categorized as “SNSF-nominated” in the analysis. We also considered meta-data regarding the *research topic* of a grant application, *type of affiliation* and *age of the applicant*. Last, because the SNSF introduced new guidelines for reviewers in October 2011, which had an effect on the distribution of evaluation scores, we introduced a dummy variable *application call deadline* that groups applications submitted before and after October 2011.

Statistical Analysis

We used a linear mixed effects model to examine the effect of explanatory variables on the overall peer-review scores (Bates et al. 2015). This model was chosen because our data are clustered and hierarchical (Jayasinghe, Marsh, and Bond 2003). Grant applications received two or more independent reviews, some reviewers had reviewed more than one application and many applicants had submitted more than one grant application over the study period, causing evaluation scores to be clustered at the levels of research projects, reviewers and applicants. We therefore introduced random effects on the IDs of the reviewer, the applicants and the project in the model, thus taking into account the non-independence between clustered scores (Harrison et al. 2018).¹ We present regression coefficients, which reflect differences in peer review scores, and coefficients adjusted for research topic, applicant's age, nationality and affiliation, with their 95% confidence interval (CI).

¹ The notebook of the analysis, including a summary of the different statistical models, is available online at www.git.io/fhaJx.

Results

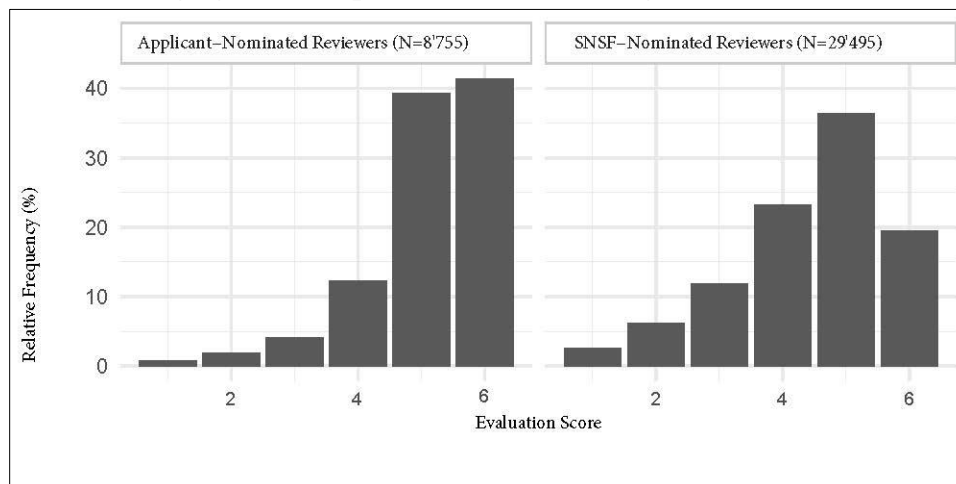
Descriptive Analysis

We analyzed the overall assessment scores of 38,250 external peer review reports on 12,294 project grant applications across all disciplines that were submitted 2006 to 2016 by 26,836 external experts from Switzerland and abroad. The average number of reviews per grant application was 3.1, applicants submitted an average of 2.1 grant applications and reviewers reviewed an average of 1.4 applications.

In a first step, we examined overall score distributions for the different reviewer and applicant characteristics that represent potential sources of bias in the external evaluation of grant applications at the SNSF. Frequency distributions of external evaluation scores are shown in Figures 1, 2, 3 and 4. Distributions were skewed for all variables, with grant applications more frequently being awarded high evaluation scores than low scores.

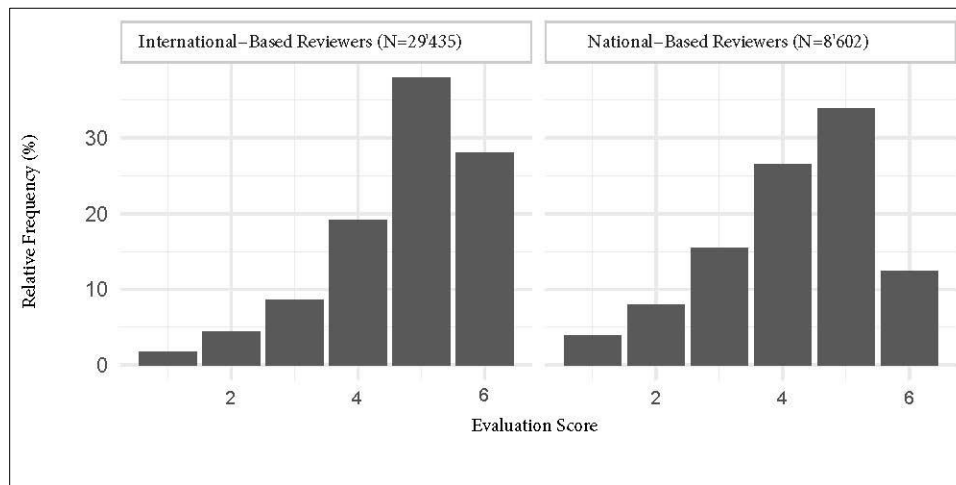
Applicant-nominated reviewers awarded higher scores than SNSF-nominated reviewers. The *source of external reviewer nomination* thus influenced evaluation scores (Fig. 1).

Figure 1: Frequency distributions of external evaluation scores by source of nomination of the reviewer, ranging from 1 (=poor) to 6 (=outstanding).



Similarly, reviewers affiliated with foreign research institutions awarded higher evaluation scores than reviewers affiliated with Swiss research institutions: *country of reviewer affiliation* could thus also influence the evaluation (Fig. 2).

Figure 2: Frequency distributions of external evaluation scores by country of affiliation of the reviewer.



The distribution of evaluation scores by *gender of the principle applicant* (Fig. 3) shows that male principle applicants received higher evaluation scores than female principle applicants. Similarly, analysis of evaluation scores by *gender of the reviewer* showed that male reviewers tended to award higher scores than female reviewers (Fig. 4).

Figure 3: Frequency distributions of external evaluation scores by gender of the applicant.

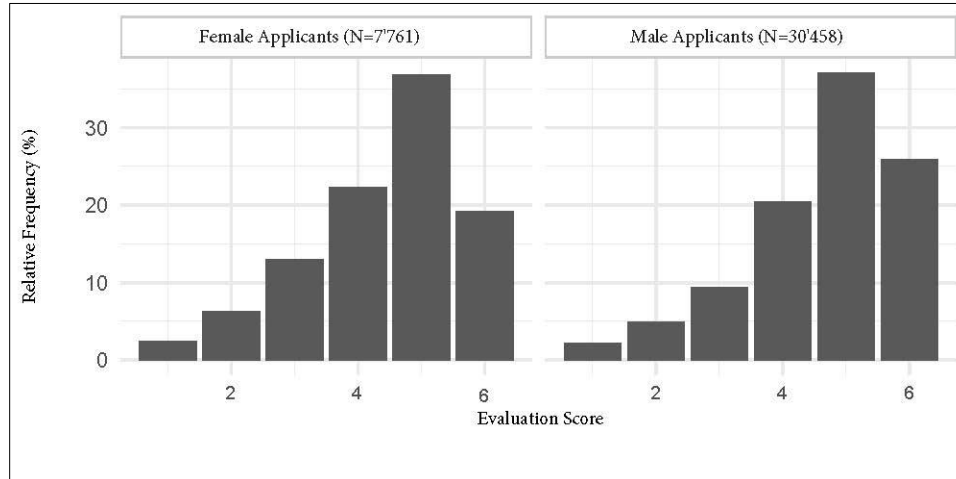
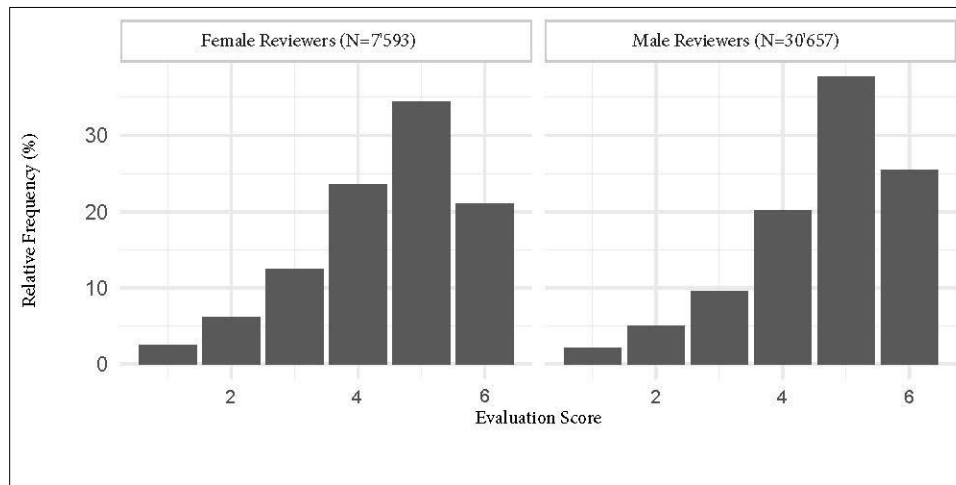


Figure 4: Frequency distributions of external evaluation scores by gender of the reviewer.



To further explore gender differences in applicant scores, we stratified analyses by *research topic* (Fig. 5, supplementary files), *applicant age* (Fig. 6, supplementary files) and *applicant affiliation* (Fig. 7, supplementary files).

There were important differences in evaluation scores across *research topics*. For example, grant applications in the natural and technical sciences or in linguistics and history received higher evaluation scores than applications covering other topics. Gender differences in evaluation scores were more pronounced for some research topics (for example mathematics and physics and engineering, biology and medicine, sociology) than others (for example geology, history, psychology). Female applicants were underrepresented (below 50 percent) in all research topics (lower panel of Fig. 5, supplementary files).

Regarding *applicant age*, applicants aged 60 years or older received the highest evaluation scores, independent of their gender. For the younger age groups, female applicants consistently received lower evaluation scores than male applicants. For groups aged 65 or older, female applicants received higher evaluation scores than male applicants. Female applicants were underrepresented across all age groups, except for the youngest age group, and representation was particularly low in older age groups (lower panel of Fig. 6, supplementary files).

Regarding *applicant affiliation*, applications submitted by applicants who are affiliated with the Swiss Federal Institutes of Technology and associated research institutions (“ETH Domain”) received higher evaluation scores than applications from Cantonal universities or from other research institutions. Gender differences in scores were evident for all three affiliations, and women were underrepresented for all affiliations (lower panel of Fig. 7, supplementary files).

Analysis of the *nationality of the applicant* showed that grant applications submitted by applicants with a Swiss nationality received slightly lower scores than those submitted by

applicants with other nationalities, with a similar gap between genders (Fig. 8, supplementary files). Finally, Figure 9 (supplementary files) shows the effect of the *new evaluation guidelines for reviewers* that were introduced in October 2011. Grant applications submitted before October 2011 received higher average scores than applications evaluated under the new guidelines.

Linear Modeling of the Effects

Table 1 presents the final, adjusted model and Table 2 (supplementary files) compared crude and adjusted differences in average scores.

Table 1: Differences in external peer review evaluation scores between characteristics of reviewers, applicants and research proposals.

Variable	Difference in score (95% CI)	P-Value
Source of nomination of reviewer		<0.001
Applicant	0.49 (0.46 – 0.51)	
Office	0	
Country of affiliation of the reviewer		<0.001
Outside Switzerland	0.47 (0.44 – 0.50)	
Switzerland	0	
Gender of the applicant		<0.001
Male	0.09 (0.04 – 0.13)	
Female	0	
Gender of the reviewer		<0.001
Male	0.08 (0.05 – 0.11)	
Female	0	
Age of the applicant		<0.001
Per 10 year increase	0.05 (0.03 – 0.07)	
Affiliation of the applicant		<0.001
ETH Domain	0.11 (0.07 – 0.16)	
Other	-0.19 (-0.25 – -0.14)	
Universities	0	
Nationality of the applicant		0.218
Other than Swiss	0.47 (0.44 – 0.50)	
Swiss	0	
Field of research		<0.001
Medicine	0	
Architecture	0.15 (0.05 – 0.25)	
Biology	0.27 (0.21 – 0.33)	
Chemistry	0.24 (0.17 – 0.31)	
Economics	-0.01 (-0.09 – 0.06)	
Engineering	0.07 (0.00 – 0.13)	
Geology	0.25 (0.14 – 0.35)	
History	0.32 (0.24 – 0.40)	
Linguistics	0.26 (0.18 – 0.34)	
Mathematics / Physics	0.45 (0.39 – 0.52)	
Psychology	-0.08 (-0.15 – 0.00)	
Sociology	0.01(-0.07 – 0.08)	
Call-cut		<0.001
Before introduction of guidelines	0.43 (0.40 – 0.46)	
After introduction of guidelines	0	

Notes. Results from linear mixed effects models. Marginal R²: 0.13, Conditional R²: 0.526, Intercept 4.45 (95% CI 4.34-4.57). Random effect variances: Between reviewers: 0.37, Between applicants: 0.15, Between projects: 0.08, Residual: 0.72.

In the adjusted model, substantial differences of about 0.5 points are observed for *source of reviewer nomination* and *country of affiliation of the reviewer*. Small differences (less than 0.1 point) are evident for *gender of the principle applicant* and *gender of the reviewer*. Substantial differences are also observed across disciplines. For example, scores are on average 0.45 points higher in mathematics and physics than in medicine, but about 0.1 point lower in psychology (Table 1).

Compared to crude differences, adjusted differences tended to be attenuated. For example, the crude difference between male and female applicants was 0.19 points, compared to 0.09 points in the adjusted analysis (Table 2, supplementary files).

Discussion

We retrospectively examined whether the scores given by external reviewers to grant applications submitted to the SNSF were influenced by the following factors: (1) *the source of nomination of the reviewer*; (2) *the country of affiliation of the reviewer*; (3) *the gender of the principle applicant* and *the gender of the reviewer*. We analysed data on 38,250 reviews of 12,294 unique grant applications across all disciplines between 2006 and 2016 using linear mixed effects regression.

Our analyses showed that the source of nomination of a reviewer has a statistically significant effect upon evaluation scores in the external peer review of grant applications at the SNSF. Reviewers who were nominated by applicants via the “positive list” on average tended to award higher evaluation scores than reviewers nominated by SNSF administrative offices, referees or other reviewers. This effect can be interpreted in several ways. First, applicant-nominated reviewers may award more favorable evaluation scores because they know the applicants personally and/or have received positive evaluations from the applicant in the past (Schroter 2006). This would mean a conflict of interest. Second, applicants may nominate reviewers who are experts within their field and therefore might be particularly familiar with their research and will recognize the impact and importance of their grant application. The SNSF decided to discontinue the use of the “positive list”, thereby abolishing the possibility for grant applicants to suggest their own reviewers. Of note, applicants can still submit negative lists of reviewers that should not be used because of conflicts of interest.

Our analyses further showed that the country of affiliation of the reviewer affects evaluation scores. Reviewers affiliated with a research institution outside of Switzerland tend to give higher evaluation scores than reviewers affiliated with a Swiss institution. It is possible that reviewers affiliated with institutions in countries known for high scientific productivity generally are less favorable in their evaluation of grant applications than reviewers from countries lagging behind in terms of their scientific productivity (Fischer and Reckling 2010). Switzerland consistently

has been shown to belong to the most productive countries in terms of its research outputs (Bonaccorsi and Cinzia Daraio 2007) and this might explain why we found that reviewers affiliated with Swiss research institutions award lower evaluation scores than reviewers affiliated with foreign universities. It should be noted, however, that this interpretation is based on the assumption that a country's scientific productivity is influenced by standards of scientific quality and excellence of its national research system. A further interpretation for this bias is that Swiss-based reviewers might be downgrading their competitors, in order for there to be more funds for their own projects to be funded. As the Swiss research community is small, it is basically impossible to rule out conflicts of interest as a potential reason for biases stemming from country of affiliation of the reviewer. Another possible explanation for the effect we observed is that there are other factors that contribute to the stringency of reviewers affiliated with Swiss institutions. For example, it could be that Swiss reviewers are commonly asked to review on topics related to the social sciences, law and humanities as the focus of topics within these fields is often more national than international, which would require domestic experts. As we have seen in our analyses, reviewers within these fields award lower evaluation scores than reviewers in the natural and technical sciences.

Finally, our study has shown that both the reviewer's gender and the applicant's gender influence evaluation scores. In the external evaluation of grant applications submitted to the SNSF, male reviewers tend to award higher evaluation scores than female reviewers and male applicants tend to receive higher evaluation scores than female applicants. Both of these effects were small. These findings are in line with previous research that found robust evidence of small gender differences in grant award procedures (Bornmann, Mutz, and Daniel 2007). It should be mentioned that when we adjusted for research topic, applicant's age, nationality and affiliation in the regression model, the effects related to gender decreased significantly. This indicates that observed gender differences in the external evaluation of grant applications at the SNSF can, in a large part, be explained by other, confounding factors. The SNSF started to monitor gender differences on a regular basis in 2013, overseen by an independent commission on gender and equality. The annual analyses provide the opportunity to detect and address gender effects in a timely manner and to take precautionary measures. Remaining gender effects could be explained by factors that we could not include in our model, or by the fact that external reviewers indeed might have a small bias against female applicants.

Our study has several limitations. First, this is an observational study and our inferences are from data retrospectively gathered on evaluation scores as the outcome variable and reviewer and applicant characteristics as the independent variables. The latter were not under our control and it is difficult to infer causality. Chance, bias, and confounding variables must always be considered as possible explanations for associations between reviewer and applicant characteristics on the one hand and evaluation scores on the other hand. We tried to mitigate the influence of confounding variables by adjusting for these in our regression model. Second, there were low

proportions of female reviewers and applicants, which reduces the resolution power of our observational study in terms of gender effects. From a more general point of view, it should be noted that our study covers SNSF project funding only, and does not relate to other SNSF research funding schemes including career funding for early career researchers, programme funding, awards and prizes as well as infrastructure and science communication funding. In addition to this, our results pertain to the external review of grant applications submitted to the SNSF. Our study does not cover evaluation scores awarded by members of the NRC, who serve as (co-)referees, or the final funding decisions of the evaluation body. This restricts the generalizability of our results concerning SNSF evaluation procedures as a whole.

We encourage funding bodies to monitor their evaluation processes in order to detect and adequately address potential biases in evaluation scores and final funding decisions. Further research is needed that aims to disentangle the underlying mechanisms of biases in grant funding.

Acknowledgements

This publication presents a sequel of a conference abstract presented at the 5th International Congress on Peer Review and Scientific Publication, Chicago, Illinois, USA; September 10-12, 2017. We are grateful to Andreas Limacher (Clinical Trials Unit of the Faculty of Medicine of the University of Bern) and to Rachel Heyard (Swiss National Science Foundation) for guiding us with the methodology and statistical analyses.

References

- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software* 67 (1). <https://doi.org/10.18637/jss.v067.i01>.
- Bornmann, L, R Mutz, and H Daniel. 2007. "Gender Differences in Grant Peer Review: A Meta-Analysis." *Journal of Informetrics* 1 (3): 226–38. <https://doi.org/10.1016/j.joi.2007.03.001>.
- Bornmann, Lutz, and Hans-Dieter Daniel. 2007. "Gatekeepers of Science—Effects of External Reviewers' Attributes on the Assessments of Fellowship Applications." *Journal of Informetrics* 1: 83–91.
- Demicheli, Vittorio, and Carlo Di Pietrantonj. 2007. "Peer Review for Improving the Quality of Grant Applications." Edited by Cochrane Methodology Review Group. *Cochrane Database of Systematic Reviews*, April. <https://doi.org/10.1002/14651858.MR000003.pub2>.
- Fischer, Christian, and Falk Reckling. 2010. "Factors Influencing Approval Probability in FWF Decision-Making Procedures," December. <http://dx.doi.org/10.2139/ssrn.1725985>.
- Harman, Grant. 1998. "The Management of Quality Assurance: A Review of International Practice." *Higher Education Quarterly* 52 (4): 345–64. <https://doi.org/10.1111/1468-2273.00104>.
- Harrison, Xavier A, Lynda Donaldson, Maria Eugenia Correa-Cano, Julian Evans, David N Fisher, Cecily ED Goodwin, Beth S Robinson, David J Hodgson, and Richard Inger. 2018. "A Brief Introduction to Mixed Effects Modelling and Multi-Model Inference in Ecology." Edited by Andrew Gray. *PeerJ* 6: e4794. <https://doi.org/10.7717/peerj.4794>.
- Jayasinghe, Upali W., Herbert W. Marsh, and Nigel Bond. 2003. "A Multilevel Cross-Classified Modelling Approach to Peer Review of Grant Proposals: The Effects of Assessor and Researcher Attributes on Assessor Ratings." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 166 (3): 279–300. <https://doi.org/10.1111/1467-985X.00278>.
- Marsh, Herbert W., Nigel W. Bond, and Upali W. Jayasinghe. 2007. "Peer Review Process: Assessments by Applicant-Nominated Referees Are Biased, Inflated, Unreliable and Invalid." *Australian Psychologist* 42 (1): 33–38. <https://doi.org/10.1080/00050060600823275>.
- Mutz, Rüdiger, Lutz Bornmann, and Hans-Dieter Daniel. 2012. "Does Gender Matter in Grant Peer Review?: An Empirical Investigation Using the Example of the Austrian Science Fund." *Zeitschrift Für Psychologie* 220 (2): 121–29. <https://doi.org/10.1027/2151-2604/a000103>.
- Schroter, Sara. 2006. "Differences in Review Quality and Recommendations for Publication Between Peer Reviewers Suggested by Authors or by Editors." *JAMA* 295 (3): 314. <https://doi.org/10.1001/jama.295.3.314>.
- Swiss National Science Foundation. 2016. "SNSF Funding Regulations. Funding Regulations. Regulations of the Swiss National Science Foundation on Research Grants." <https://bit.ly/2yUk0fG>.
- Bonaccorsi, Andrea, and Sapienza Cinzia Daraio, eds. 2007. *Universities and Strategic Knowledge Creation: Specialization and Performance in Europe*. PRIME Series on Research and Innovation Policy in Europe. Northampton, MA: Edward Elgar.

460 Wood, Fiona Q., and Australian Research Council. 1997. "The Peer Review Process." Issue 54
461 of Commissioned report. Australian Government Pub. Service.
462 <https://trove.nla.gov.au/version/45600880>.
463