Gender and Other Potential Biases in Peer Review: Analysis of 38,250 External Peer Review Reports

Anna Severin[1,2], João Martins[1,3], Rachel Heyard[1], François Delavy[1], Anne Jorstad[1], Matthias Egger[1,2]

[1]Swiss 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 Author:

Professor Matthias Egger MD MSc FFPH

Swiss National Science Foundation

Wildhainweg 3

3001 Bern, Switzerland

matthias.egger@ispm.unibe.ch

Abstract 350 words, main text 3684 words, 3 figures, 2 tables, 29 references, supplementary materials with 1 table, 5 figures and 1 text.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

ABSTRACT

Background

The Swiss National Science Foundation (SNSF) supports fundamental and use-inspired research in all disciplines. Peer reviewers assess the proposals submitted to the SNSF. We examined whether the gender of applicants and reviewers and other factors influenced the summary scores awarded.

Methods

We analysed 38,250 reports on 12,294 grant applications across all disciplines 2006 to 2016. Proposals were rated on a scale from 1 (=worst) to 6 (=best) by 26,836 reviewers. We used linear mixed effects regression models adjusted for research topic, applicant's age, nationality, affiliation and calendar period to examine associations, and interactions between gender of the applicant and other variables.

Results

In univariable analysis, male applicants received more favourable evaluation scores than female applicants (+0.19 points; 95% CI 0.14-0.23), and male reviewers awarded higher scores than female reviewers (+0.12; 95% CI 0.08-0.15). Applicant-nominated reviewers awarded higher scores than reviewers nominated by the SNSF (+0.53; 95% CI 0.50-0.56), and reviewers affiliated with research institutions outside of Switzerland more favourable scores than reviewers affiliated with Swiss institutions (+0.53; 95% CI 0.49-0.56). In multivariable analysis, differences between male and female applicants were attenuated (to +0.08; 95% CI 0.04-0.13) whereas results changed little for source of nomination and affiliation of reviewers. There was an interaction between gender of applicant and reviewer, and between gender of applicant and calendar period. Male reviewers gave higher scores than female reviewers, with a greater difference for male than for female applicants (P=0.037 from test of interaction). The gender difference increased after September 2011, when new evaluation forms were introduced (P=0.033 from test of interaction).

Conclusions

Our study showed that peer review of grant applications at SNSF might be prone to biases stemming from different applicant and reviewer characteristics. The SNSF abandoned the nomination of peer reviewers by applicants, and made members of panels aware of the other systematic differences in scores. The new form introduced in 2011 may inadvertently have given more emphasis to the applicant’s track record, and a revision is now under discussion. We encourage other funders to conduct similar studies, in order to improve the evidence base for rational and fair research funding.

Keywords: peer review, bias, gender matching hypothesis, confounding, mixed effects models

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

BACKGROUND

In public research funding, peer review of proposals by suitable experts 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 (1). 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 bias in grant evaluations resulting from the influence of factors that are unrelated to the actual quality of the grant applications (2).

Mandated by the Swiss Confederation, the Swiss National Science Foundation (SNSF) supports basic research and use-inspired basic research in all academic disciplines. 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 (3). The proposals submitted to the SNSF are peer reviewed by at least two external experts.

Empirical studies suggest that the evaluation of applications is prone to biases that may relate to both applicant and reviewer characteristics (2)(4). Potential discrimination against women is the most frequently investigated bias in the context of grant peer review (5). In a natural experiment, a recent study of the Canadian Institutes of Health Research compared grant programmes with and without an explicit review of the track record of the principal investigator (6). The results showed that the gender gap in grant funding was due to less positive assessments of women as principal investigators, whereas the quality of the proposed research was similar for women and men (6). Of note, the SNSF introduced new evaluation forms and guidelines for peer reviewers in September 2011.

The source of nomination of reviewers was also of interest in the context of potential biases: the foundation allowed grant applicants to suggest reviewers to evaluate submissions via a “positive list”. The names put forward on the list were then considered as potential reviewers, after a careful check for conflicts of interest. A study of the Australian Research Council found that applicant-nominated reviewers tended to give better ratings than panel-nominated reviewers (7). Finally, the SNSF frequently invites reviewers from abroad to review grant applications. An analysis of data from the Austrian Science Fund suggested that

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

international peer reviewers affiliated with research institutions located in countries known for high scientific productivity were generally more stringent than national reviewers (8).

We analysed the database of the SNSF to examine the determinants of summary scores from external peer reviewers in project funding.

METHODS

Evaluation of Grant Applications at the SNSF

The evaluation of grant applications at the SNSF consists of four steps (3). After submission, the administrative office first checks eligibility and assigns grant applications to two members of the National Research Council (referee and co-referee) based on their field of expertise. In a second step, eligible proposals are peer-reviewed by external experts. External reviewers were identified in several ways: (i) grant applicants suggested experts via a "positive list", (ii) the referee of the National Research Council suggested reviewers, (iii) the SNSF administrative offices proposed experts, and (iv) experts may have declined to review but suggested other reviewers (3). For each application, at least two external independent reviews were required. The final choice of reviewers was made by the SNSF. Reviewers from the positive list were chosen only if they had the required expertise and there were no conflicts of interest. Applicants could also submit a "negative list" of reviewers who, because of possible conflicts of interest, should not be contacted.

The peer review forms and assessment scale were changed in September 2011 in an attempt to simplify the review, and to achieve a more equal distribution of scores, with fewer proposals in the top category. Up to September 2011, peer reviewers were asked to score six criteria: (i) current scientific interest and impact of the project; (ii) originality of the work; (iii) suitability of the methods; (iv) work plan, feasibility, cost; (v) experience and past performance of the applicants; (vi) specific abilities of the investigators for the proposed project. Reviewers were asked to "give a rating and provide explanatory comments" for each of the six criteria. In September 2011, new evaluation forms were introduced (3)(9), which asked experts to review proposals according to three criteria: i) the applicants' scientific track-record and expertise; ii) the scientific relevance, originality and topicality of the proposed research and, in the case of use-inspired research, the research's broader impact

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

and iii) the suitability of the methods and feasibility. Furthermore, peer reviewers were asked to declare any conflicts of interest, and given the opportunity to submit confidential comments, which would not be seen by the applicants. Up to September 2011, reviewers scored each criterion and the proposal overall on a scale from 1 to 6: (1) poor, (2) satisfactory, (3) average, (4) good, (5) very good, and (6) excellent. In September 2011 the scale was changed to (1) poor, (2) average, (3) average, (4) good, (5) excellent, and (6) outstanding. The two versions of the peer review form are reproduced in supplementary Text S1.

In the third step of the evaluation, the two members of the Council (referee and co-referee) assessed the usefulness of the peer review reports and considered them when ranking the application relative to other proposals. In the fourth and final step, referee and co-referee presented their assessment at the meeting of the corresponding section of Council. Each application was then voted on and approved or rejected (3).

Data and Variables

The outcome variable of interest was the overall evaluation score of a grant application ranging from 1 (worst) to 6 (best). Explanatory variables included meta-data on principle applicants and external peer reviewers, including source of reviewer (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 categorized as “SNSF-nominated” in the analysis. We also considered meta-data regarding the research topic of a grant application, type of institutional affiliation and age of the applicant. Finally, we introduced a dummy variable to group applications submitted before and after September 2011.

Statistical Analysis

We used a linear mixed effects model to examine the effect of explanatory variables on the overall peer-review scores (10). This model was chosen because the data are clustered and hierarchical (11). 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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

clustered at the levels of research projects, reviewers and applicants. We therefore introduced random intercepts for the identifiers of the reviewer, the applicant and the project in the model, thus taking into account the non-independence between clustered scores (12).

We ran crude and adjusted models. The latter were adjusted for gender of the applicant and reviewer, source of reviewers, country of affiliation of the reviewer, the applicant's age (per 10 year increase), affiliation (Swiss Federal Institutes of Technology and associated institutions [ETH domain], Cantonal university, other) and nationality (Swiss vs. other), the field of research (12 categories), and the period of submission of the proposal (before or after the change in peer review forms and scale). To make adjusted and crude estimates comparable, we performed a complete case analysis by deleting peer review reports with missing values for any of the relevant variables. In further analyses, we examined interactions between the gender of the applicant and the gender of the reviewer, and other variables, by including interaction terms in the linear mixed models. We thus examined the 'gender matching hypothesis', which stipulates that female peer reviewers give higher scores to female researchers and that male reviewers do the same for male applicants (13). We used likelihood ratio tests to assess the strength of the evidence for interactions.

We present crude and adjusted regression coefficients, which reflect differences in peer review scores with their 95% confidence interval (CI). The notebook of the analysis, including a summary of the different statistical models, is available online at www.git.io/fhaJx.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

RESULTS

Descriptive Analyses

We analysed the summary 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. The complete case mixed effects regression analyses were based on 37,989 reviews (99.3%).

The 12,294 proposals were submitted by 5,824 applicants: 4,516 (77.5%) men and 1,308 (22.5%) women. Female applicants were younger than men and more likely to be affiliated with other institutions (for example universities of applied sciences, the arts or teacher education) than with the Federal ETH domain or the Cantonal universities (Table 1). Women were also more likely to work in disciplines of the social sciences and humanities (psychology, sociology, linguistics) than in STEM disciplines (Science, Technology, Engineering, and Mathematics) or in biology and medicine (Table 1).

In a first step, we examined the distributions of the overall scores submitted by external reviewers (Figure 1 and Figure 2). Distributions were skewed for all variables, with grant applications more frequently being awarded high evaluation scores than low scores. The distribution of evaluation scores by gender of the principle applicant shows that male principle applicants received higher evaluation scores than female principle applicants. Similarly, the analysis of evaluation scores by gender of the reviewer showed that male reviewers tended to award higher scores than female reviewers (Figure 1). Applicant-nominated reviewers awarded higher scores than SNSF-nominated reviewers, and reviewers affiliated with institutions outside Switzerland awarded higher evaluation scores than reviewers affiliated with Swiss institutions (Figure 2).

To further explore gender differences in applicant scores, we stratified analyses by research topic (supplementary Figure S1), applicant age (supplementary Figure S2) and applicant affiliation (supplementary Figure S3). 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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

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 supplementary Figure S1).

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. Female applicants were under represented across all age groups, except for the youngest age group, and representation was particularly low in older age groups (lower panel of supplementary Figure S2). Applications submitted by applicants affiliated with the 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 under represented for all affiliations (Figure S3).

Analysis of the nationality of the applicant showed that grant applications submitted by Swiss applicants received slightly lower scores than those submitted by applicants with other nationalities, with a similar gap between genders (supplementary Figure S4). Finally, supplementary Figure S5 shows that applications submitted before the new forms were introduced received higher average scores than applications evaluated later.

Linear Mixed-Effects Models

Table 2 shows crude and adjusted differences in peer review scores by characteristics of applicants, reviewers and research proposals. In the crude model, the difference between male and female applicants was 0.19 points favouring men. More substantial differences of 0.53 points were observed for source of reviewer (0.53 points higher if the reviewer was nominated by the applicants) and country of affiliation of the reviewer (0.53 higher for reviewers from outside Switzerland). Substantial differences were also observed across disciplines. For example, scores were on average 0.68 points higher in mathematics and physics than in medicine, but 0.13 point lower in psychology than in medicine (Table 2). Compared to crude differences, most adjusted differences were smaller. For example, the adjusted difference between male and female applicants was reduced from 0.19 to 0.08 points. One exception was the difference observed between proposals evaluated before or

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

after the introduction of the peer review forms in September 2011 (0.43 points higher scores before the introduction in both analyses).

Interactions between gender of the applicants and other variables

We examined possible interactions between the genders of the applicants with the other fixed-effect variables in the model shown in Table 2. In other words, we examined whether the differences observed between female and male applicants varied across the levels of the other variables. We found that male reviewers gave higher scores both to male and female applicants than female reviewers, but this difference was considerably greater for male than for female applicants. Figure 3 shows the predicted values of the overall score from the bivariable model (P=0.011 from test of interaction). There was some evidence that the gender difference in scores became larger after the introduction of the new evaluation form (P=0.064, Figure 3). There was also strong evidence for an interaction (P<0.0001) between gender of the first applicant and his or her affiliation: the gender differences in scores were smallest for applicants based at one of the Cantonal universities, larger for the ETH domain and most pronounced for other institutions of higher education (for example universities of applied sciences, the arts or teacher education, see Figure 3). The interaction P values from the adjusted models were 0.037 (gender of peer reviewer), 0.003 (affiliation of applicant) and 0.033 (change of evaluation form). All P values from the bivariable and multivariable interaction tests are shown in supplementary Table S1.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

DISCUSSION

Research funding organizations must be concerned about possible biases in their peer review of project proposals, in order to prevent any discrimination or preference of some groups of applicants. In this study, we examined whether the scores given by external reviewers to project grant applications submitted to the SNSF were influenced by the gender of the principle applicant and the gender of the reviewer, the source of the reviewer and the country of affiliation of the reviewer. We were also interested in other factors, such as the applicant's age, affiliation and the effect of new guidelines for peer reviewers introduced in 2011. We analysed summary scores from 38,250 reports on 12,294 grant applications across all disciplines, which were submitted to the foundation between 2006 and 2016 by 5,824 applicants. To the best of our knowledge, this is one of the largest studies of peer review reports of research proposals ever conducted.

We found that female applicants received lower scores than male applicants. This gender difference was attenuated in multivariable analysis: it was partly explained by the fact that women were under represented among applicants in the fields and institutions whose proposals were rated highly, for example mathematics and physics, and institutions of the ETH domain. Although statistically these factors "explained" a substantial proportion of the gender gap, they are also a reflection of the leaky pipeline, i.e. "the phenomenon of women dropping out of research and academic careers at a faster rate than men" (14), which is well documented for Switzerland (15,16). The academic pipeline in Switzerland is particularly leaky in the social sciences, humanities, and in the life sciences. In STEM the rate of dropout of women is less pronounced, but they are a minority from the start: among PhD students only about 20% are women, whereas in the social sciences, humanities, and the life sciences the majority of doctoral students are women (16).

Ceci and Williams (17) have argued that several factors are responsible for the under representation of women, including fertility choices and work-home balance issues, which affect women in all fields, not just STEM, whereas other factors such as career preferences and gender differences in mathematics achievement and attitudes impact particularly women in math-based fields. The latter may in turn be influenced by cultural stereotypes and gender roles that lead to socialization processes that shape performance (18,19). In Switzerland, men are assigned the role of 'main breadwinner', resulting in uneven

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

distribution of housework, unfavourable fiscal policies for households with two earners, and a lack of affordable child care (15,20). At the same time, the post-doc "bubble" in Switzerland is taking place in a situation of full employment and relative shortage of skilled labour (21).

A noteworthy finding of our study was the interaction between the gender of applicants and peer reviewers. In contrast to Jayasinghe and colleagues (13), who analysed 7153 reviewer ratings at the Australian Research Council large grant programme and other smaller studies (22)(23), we found evidence supporting the 'gender matching hypothesis'. Male reviewers gave systematically higher ratings to male applicants than to female applicants, whereas the same phenomenon could not be observed for female reviewers. If such matching bias was present, male reviewers will have favoured male applicants, despite the fact that the proposals from male and female applicants were of similar quality. Alternatively, assuming proposals from male applicants were in fact stronger, female reviewers could have been biased against men and downgraded their proposals. Of note, the evidence for an interaction became stronger when adjusting for other variables, and ratings from female reviewers were generally lower than those from male experts.

Male reviewers may have given more weight to the track record of applicants than female reviewers. In this context, it is interesting that the gender gap became wider after September 2011, when new evaluation forms for external peer review were introduced. The new guidelines and form separated the criteria related to the applicants, and the criteria related to the proposed project. On the new form, the applicant's track record was the first criterion out of a total of three, whereas it was the fifth out of six criteria on the old form. Although this was not intended, the reform may have led to reviewers giving more weight to the track record of applicants, due to its prominence on the new form. Based on a Canadian Institutes of Health Research study, which showed that the gender gap in grant funding was due to less positive assessments of women as principal investigators whereas the quality of the proposed research was similar for women and men (6), Raymond and Goodman asked funders to "evaluate projects, not people" (24). We are planning additional analyses to examine whether at the SNSF the same phenomenon is at play, i.e. whether the gender gap is driven by the assessments of the track record. Furthermore, we are discussing changes to the peer review form.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Our results confirm those from the Australian Research Council, which showed that applicant-nominated reviewers tended to give substantially higher ratings than panel-nominated reviewers (7). A study of peer review in biomedical journals also found that author-nominated reviewers submitted more favourable recommendations than editor-nominated reviewers (25). This difference may be interpreted in several ways. First, nominated reviewers may have a conflict of interest because they know the applicants personally, and strive to support them. They may even have been contacted by the applicants, and asked to submit a favourable review. Alternatively, applicants may nominate reviewers who are more familiar with their field than reviewers nominated by the SNSF, and thus more able to recognize the impact and importance of the proposed research. Like the Australian Research Council, the SNSF felt that bias was the most likely explanation and decided to discontinue the use of the "positive list" in 2016. Of note, applicants can still submit a "negative list" of reviewers that should not be used because of perceived conflicts of interest.

Peer reviewers affiliated with a Swiss research institution gave lower scores than reviewers from outside Switzerland. A study of the Austrian Science Fund suggested that reviewers from countries with high scientific productivity were more stringent than national reviewers (8). Switzerland belongs to the most productive countries in terms of research output (26) and this might explain why reviewers affiliated with Swiss research institutions award lower evaluation scores than reviewers from abroad. In contrast to the Austrian Science Fund study (8), the Australian data showed that reviewers affiliated with an institution in the United States of America (USA) were more lenient than reviewers affiliated with institutions located in the United Kingdom, Germany or Australia (27), despite the fact that the USA is the country with the highest research output globally (26). Other explanations for the lower scores awarded by Swiss reviewers include greater knowledge of the local research capacity and expertise, or bias, if reviewers based in Switzerland downgraded the proposals of their competitors.

Our study has several limitations. First and most importantly, we did not examine the determinants of the final funding decision or the level of funding. It is therefore unclear whether the differences in scores analysed in the present study influenced funding decisions. Such analyses are planned for the near future. Second, this is an observational study and it is therefore difficult to infer causality from the associations observed. Chance,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

bias, and confounding variables must be considered as possible explanations for associations between reviewer and applicant characteristics and summary evaluation scores (28). We tried to control the influence of confounding variables by adjusting for these in regression models. Third, our results are relevant to the Swiss context, but may not be applicable to other countries. Finally, we examined project funding only, but not other funding schemes, such as career funding or programme funding.

CONCLUSIONS

In conclusion, our results had important implications for the evaluation of project grant proposals: we abandoned the nomination of peer reviewers by applicants, and make members of evaluation panels aware of the other factors, including the gender and affiliation of reviewers, that can influence review scores. We encourage all funding bodies to contribute to research on potential biases in research funding, and ways of preventing them (29).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Declarations

Acknowledgements

Earlier results from this analysis were presented at the 5th International Congress on Peer Review and Scientific Publication, Chicago, Illinois, USA; September 10-12, 2017. We are grateful to Angelika Kalt, Benjamin Rindlisbacher and Barbara Curdy-Korrodi for helpful comments on previous versions of this paper, and to Andreas Limacher and Lukas Bütikofer (Clinical Trials Unit of the Faculty of Medicine of the University of Bern) for advice on the statistical analyses.

Ethics approval and consent to participate

Under Swiss law, not ethics approval is required for this type of study. Peer reviewers did not provide consent. No peer reviewer, applicant or proposal can be identified from this report.

Consent for publication

Consent for publication was provided by the management board of the SNSF.

Availability of data and materials

The data analysed in this study are available to others on request for an approved research project, after signing a data sharing agreement.

Competing interests

The authors are employees of the SNSF; they declare that they have no other competing interests.

Funding

This study was funded by the SNSF (internal funds and grant number 174281).

Authors' contributions

AS, JM and ME conceived the study. JM and RH performed statistical analyses. FD and AJ contributed to data management and statistical analyses. AS and JM wrote the first draft of the paper, which was revised by ME. All authors contributed to and approved the final version.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

References

1. Harman G. The Management of Quality Assurance: A Review of International Practice. Higher Education Quarterly 52(4):345 - 364. 2002;52(4):345–64.

2. Bornmann L, Daniel H-D. Gatekeepers of science - Effects of external reviewers' attributes on the assessments of fellowship applications. J Informetr. 2007 Jan;1(1):83–91.

3. Swiss National Science Foundation. Funding Regulations. Regulations of the Swiss National Science Foundation on research grants. Version 1.1.2016 [Internet]. 2016. Available from: http://www.snf.ch/SiteCollectionDocuments/allg_reglement_16_e.pdf

4. Demicheli V, C DP. Peer review for improving the quality of grant applications (Review). 2008;(2).

5. Mutz R, Bornmann L, Daniel H-D. Does Gender Matter in Grant Peer Review? An Empirical Investigation Using the Example of the Austrian Science Fund. Z Psychol-J Psychol. 2012;220(2):121–9.

6. Witteman HO, Hendricks M, Straus S, Tannenbaum C. Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. The Lancet. 2019 Feb;393(10171):531–40.

7. Marsh HW, Bonds NW, Jayasinghe UW. Peer review process: Assessments by applicant-nominated referees are biased, inflated, unreliable and invalid. Aust Psychol. 2007 Mar;42(1):33–8.

8. Fischer C, Reckling FJ. Factors Influencing Approval Probability in Austrian Science Fund (FWF) Decision-Making Procedures - FWF Stand-Alone Projects Programme, 1999 to 2008 [Internet]. Rochester, NY: Social Science Research Network; 2010 Dec [cited 2019 May 5]. Report No.: ID 1725985. Available from: https://papers.ssrn.com/abstract=1725985

9. Project funding - SNF [Internet]. [cited 2019 May 25]. Available from: http://www.snf.ch/en/theSNSF/evaluation-procedures/project-funding/Pages/default.aspx

10. Bates D, Maechler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 2015 Oct;67(1):1–48.

11. Jayasinghe UW, Marsh HW, Bond N. 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). 2003 Oktober;166(3):279–300.

12. Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, et al. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ. 2018;6:e4794.

13. Jayasinghe UW, Marsh HW, Bond N. A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings. J Royal Statistical Soc A. 2003 Oct;166(3):279–300.

14. Directorate-General for Research and Innovation. She Figures 2015 [Internet]. 2015. 224 p. Available from: https://ec.europa.eu/research/swafs/index.cfm?pg=library&lib=gender_equality

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

15. Bataille P, Le Feuvre N, Morales SK. Should I stay or should I go? The effects of precariousness on the gendered career aspirations of postdocs in Switzerland. Eur Educ Res J. 2017 May;16(2–3):313–31.

16. Schubert F, Engelage S. Wie undicht ist die Pipeline? Wissenschaftskarrieren von promovierten Frauen. Köln Z Soziol. 2011 Sep;63(3):431–57.

17. Ceci SJ, Williams WM. Understanding current causes of women's underrepresentation in science. Proc Natl Acad Sci U S A. 2011 Feb 22;108(8):3157–62.

18. Else-Quest NM, Hyde JS, Linn MC. Cross-national patterns of gender differences in mathematics: A meta-analysis. Psychological Bulletin. 2010 Jan;136(1):103–27.

19. Eagly AH, Steffen VJ. Gender Stereotypes Stem From the Distribution of Women and Men Into Social Roles. :20.

20. Domestic activities (time used in) - UNECE Statistical Glossary - UNECE Statswiki [Internet]. [cited 2019 May 12]. Available from: https://statswiki.unece.org/pages/viewpage.action?pageId=92211215

21. Wanner P, Zufferey J, Fioretta J. The impact of migratory flows on the Swiss labour market. A comparison between in- and outflows. Migr Lett. 2016 Sep;13(3):411–26.

22. Bornmann L, Daniel H-D. Gatekeepers of science—Effects of external reviewers' attributes on the assessments of fellowship applications. Journal of Informetrics. 2007 Jan;1(1):83–91.

23. Sonnert G. What Makes a Good Scientist?: Determinants of Peer Evaluation among Biologists. Social Studies of Science. 1995;25(1):35–55.

24. Raymond JL, Goodman MB. Funders should evaluate projects, not people. The Lancet. 2019 Feb;393(10171):494–5.

25. Schroter S, Tite L, Hutchings A, Black N. Differences in review quality and recommendations for publication between peer reviewers suggested by authors or by editors. JAMA-J Am Med Assoc. 2006 Jan 18;295(3):314–7.

26. Country outputs | Nature Index [Internet]. [cited 2019 May 14]. Available from: https://www.natureindex.com/country-outputs

27. Wood F. The peer review process [Internet]. Canberra: Australian Govt. Pub. Service; 1997 [cited 2019 May 5] p. 189. Available from: https://trove.nla.gov.au/version/45600880

28. Smith GD, Ebrahim S. Data dredging, bias, or confounding - They can all get you into the BMJ and the Friday papers. Br Med J. 2002 Dec 21;325(7378):1437-+.

29. Tricco AC, Thomas SM, Antony J, Rios P, Robson R, Pattani R, et al. Strategies to Prevent or Reduce Gender Bias in Peer Review of Research Grants: A Rapid Scoping Review. PLoS One. 2017 Jan 6;12(1):e0169718.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Table 1. Characteristics of applicants who submitted grant applications to the Swiss National Science Foundation, 2006 to 2016, by gender.

The characteristics refer to the first submission of a project grant proposal during the study period.

	Male applicants (n = 4516)	Female applicants (n = 1308)
Age (mean, SD)	48.24 (8.64)	46.22 (8.27)
Affiliation		
ETH Domain	1197 (84.5%)	219 (15.5%)
Other	481 (68.2%)	224 (31.8%)
Universities	2838 (76.6%)	865 (23.4%)
Nationality		
Other than Swiss	1914 (76.9%)	575 (23.1%)
Swiss	2600 (78.1%)	731 (21.9%)
Field of research		
Medicine	1029 (76.4%)	318 (23.6%)
Architecture	145 (72.1%)	56 (27.9%)
Biology	612 (82.6%)	129 (17.4%)
Chemistry	380 (83.3%)	76 (16.7%)
Economics	290 (77.5%)	84 (22.5%)
Engineering	527 (87.5%)	75 (12.5%)
Geology	145 (85.8%)	24 (14.2%)
History	211 (75.6%)	68 (24.4%)
Linguistics	203 (66.6%)	102 (33.4%)
Mathematics / Physics	491 (89.8%)	56 (10.2%)
Psychology	223 (57.6%)	164 (42.4%)
Sociology	260 (62.5%)	156 (37.5%)
Year of submission (median, IQR)	2014 (3)	2014 (3)

Numbers (%) are shown unless otherwise indicated. Analysis based on 5824 applicants.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

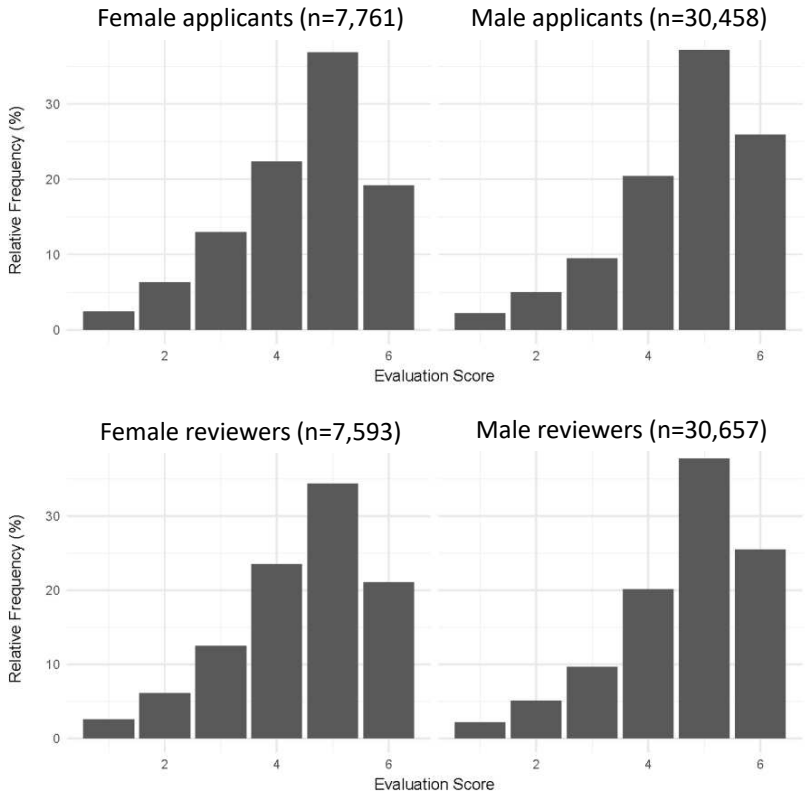
Table 2: Crude and adjusted differences in external peer review evaluation scores by characteristics of applicants, reviewers and research proposals.

Variable	Number of reviews analysed	Unadjusted difference (95% CI)	P	Adjusted difference (95% CI)	P
Gender of the applicant			<0.001		<0.001
Male	30,274	0.19 (0.14 – 0.23)		0.08 (0.04 – 0.13)	
Female	7,715	0		0	
Gender of the reviewer			<0.001		<0.001
Male	30,449	0.12 (0.08 – 0.15)		0.08 (0.05 – 0.11)	
Female	7,540	0		0	
Source of nomination of reviewer			<0.001		<0.001
Applicant	8,691	0.53 (0.50 – 0.56)		0.49 (0.46 – 0.51)	
Office	29,298	0		0	
Country of affiliation of reviewer			<0.001		<0.001
Outside Switzerland	29,396	0.53 (0.49 – 0.56)		0.47 (0.44 – 0.50)	
Switzerland	8,593	0		0	
Age of the applicant	37,989				
Per 10 year increase		0.06 (0.04 – 0.08)	<0.001	0.05 (0.03 – 0.07)	<0.001
Affiliation of the applicant					<0.001
ETH Domain	9,963	0.30 (0.26 – 0.34)	<0.001	0.11 (0.07 – 0.16)	
Other	4,075	-0.24 (-0.30 - -0.19)		-0.19 (-0.25 – -0.14)	
Universities	23,951	0		0	
Nationality of the applicant					
Other than Swiss	16,677	0.03 (-0.01 – 0.07)	0.093	-0.02 (-0.05 – 0.01)	0.218
Swiss	21,312	0		0	
Field of research			<0.001		<0.001
Medicine	7,541	0		0	
Architecture	1,391	0.13 (0.03 – 0.23)		0.14 (0.05 – 0.24)	
Biology	3,874	0.30 (0.24 – 0.36)		0.27 (0.21 – 0.33)	
Chemistry	3,244	0.46 (0.39 – 0.53)		0.24 (0.17 – 0.31)	
Economics	2,171	-0.09 (-0.17 – -0.01)		-0.01 (-0.09 – 0.06)	
Engineering	4,881	0.32 (0.25 – 0.38)		0.07 (0.00 – 0.13)	
Geology	1,168	0.49 (0.39 – 0.60)		0.25 (0.14 – 0.35)	
History	2,052	0.35 (0.27 – 0.44)		0.32 (0.24 – 0.40)	
Linguistics	2,244	0.30 (0.22 – 0.38)		0.26 (0.18 – 0.34)	
Mathematics / Physics	3,982	0.68 (0.62 – 0.75)		0.45 (0.39 – 0.52)	
Psychology	2,461	-0.13 (-0.20 – -0.05)		-0.08 (-0.15 – 0.00)	
Sociology	2,980	-0.06 (-0.13 – 0.02)		0.01(-0.07 – 0.08)	
Introduction of reviewer guidelines					
Before introduction	11,158	0.43 (0.40 – 0.47)		0.43 (0.40 – 0.46)	
After introduction	26,831	0	<0.001	0	<0.001

95% CI, 95% confidence interval. Results from linear mixed effects models based on 37,998 peer review reports.

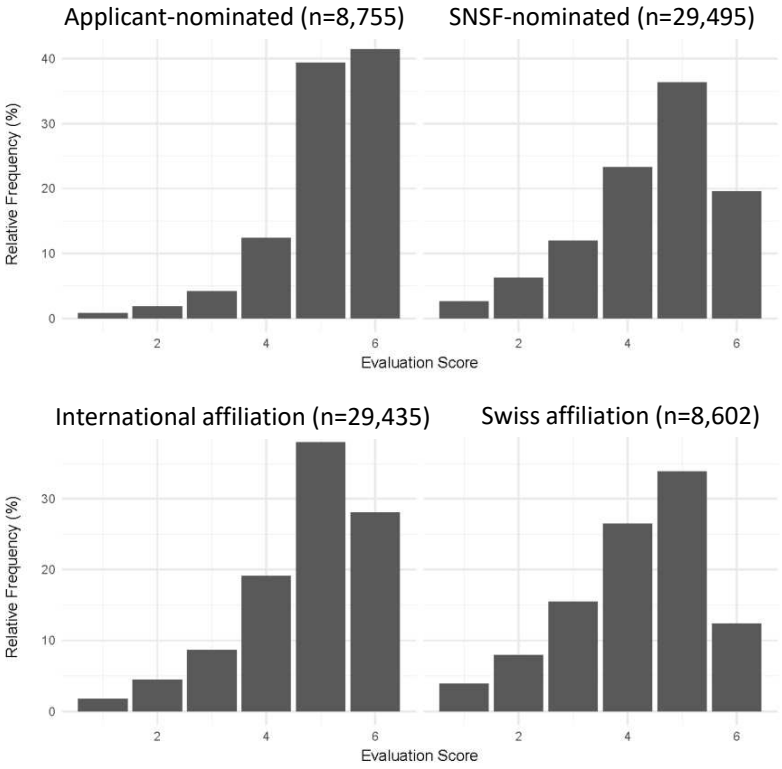
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Figure 1: Frequency distributions of external evaluation scores by gender of the applicant (upper panel) and gender of the reviewer (lower panel).
Scores range from 1 (worst) to 6 (best).



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

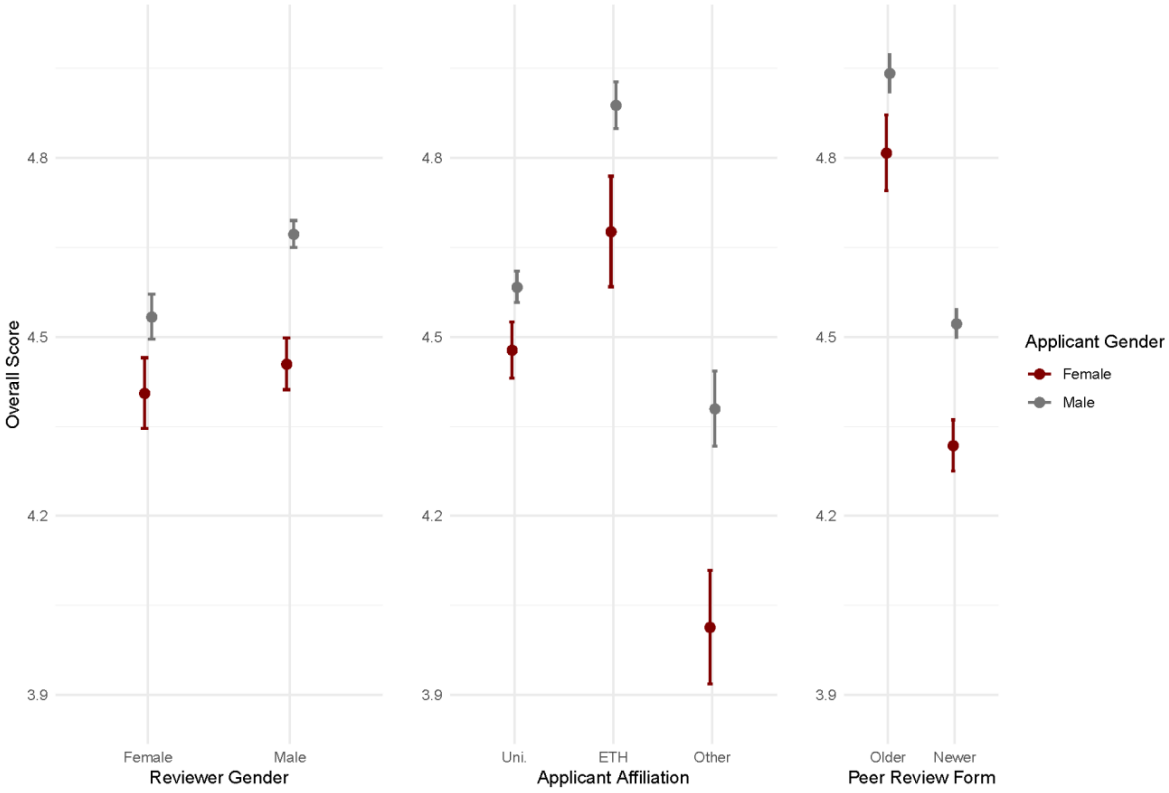
Figure 2: Frequency distributions of external evaluation scores by source of nomination of the reviewer (upper panel) and by country of affiliation of the reviewer (lower panel). Scores range from 1 (worst) to 6 (best).



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Figure 3: Gender differences in external evaluation scores by gender of the expert reviewer, affiliation and period of submission of the proposal.

Predicted values from bivariable model are shown. Scores range from 1 (=worst) to 6 (=best).



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Supplementary materials

Table S1: P values from interaction tests of gender of the applicant with other variables, based on bivariable and multivariable models.

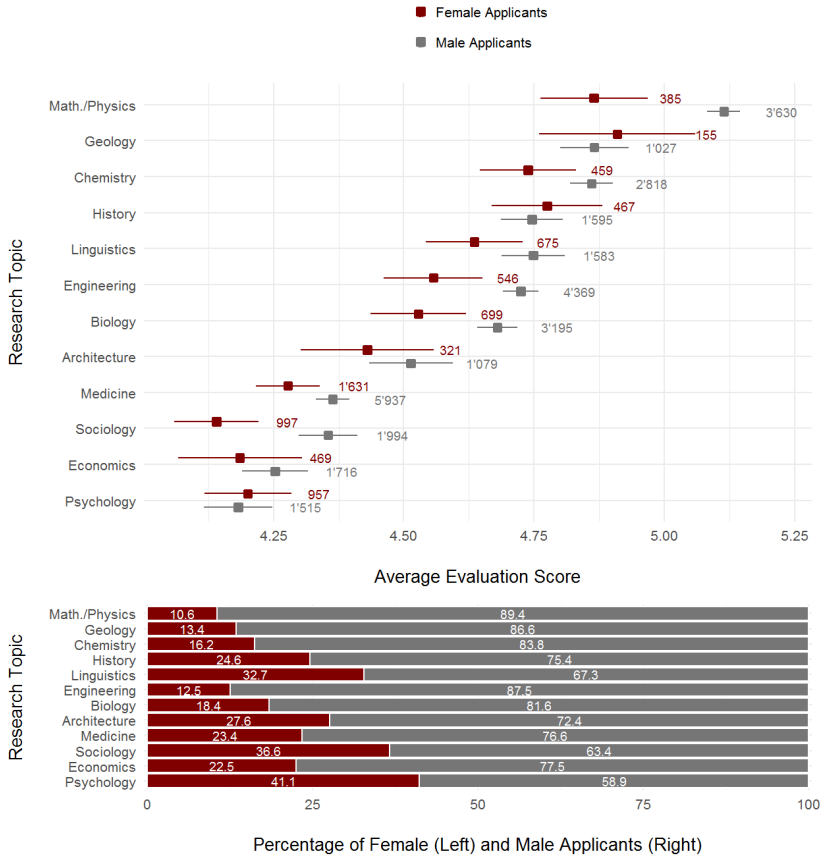
	Bivariable	Multivariable*
Gender of reviewer	0.011	0.037
Source of nomination of reviewer	0.71	0.17
Country of affiliation of reviewer	0.62	0.29
Age of applicant	0.76	0.69
Affiliation of the applicant	<0.0001	0.003
Nationality of the applicant	0.57	0.96
Research topic	0.36	0.29
Change of guidelines	0.064	0.033

Adjusted for all variables listed in Table 2 of the main paper.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Figure S1: Average evaluation scores by research topic for female and male applicants and proportions of female and male applicants by research topic

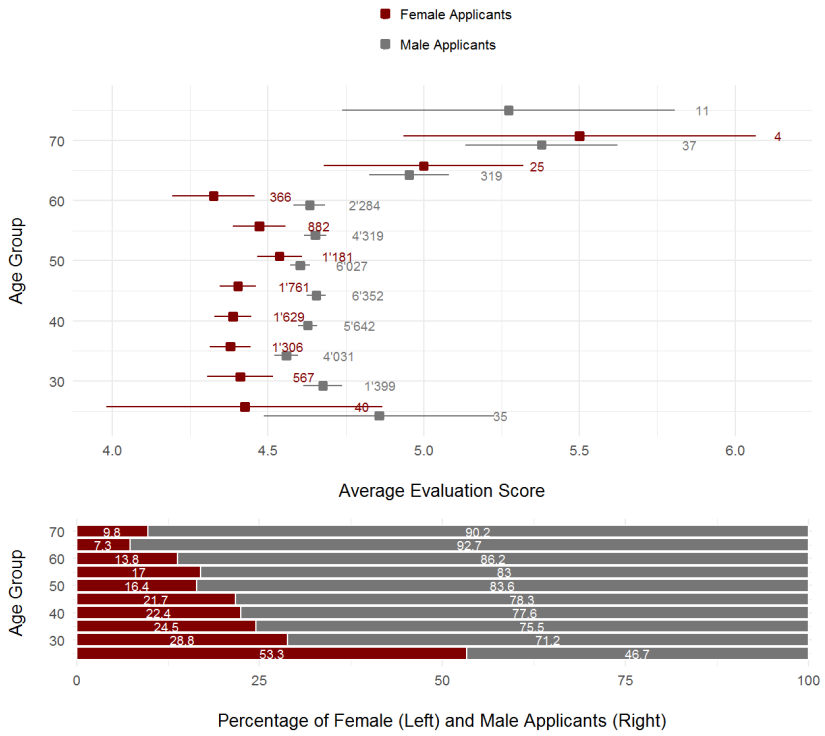
Upper panel: Horizontal lines indicate Wald confidence levels, with the number of peer review reports analysed.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Figure S2: Average evaluation scores by age group for female and male applicants and proportions of female and male applicants per age group.

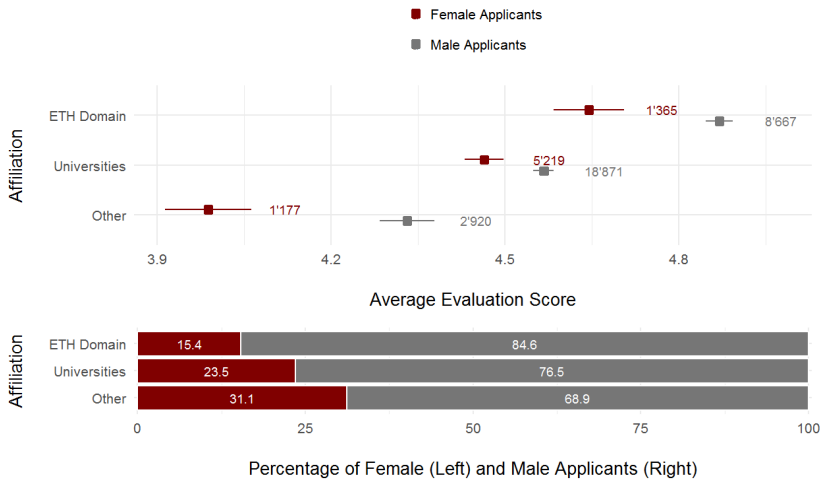
Upper panel: Horizontal lines indicate Wald confidence levels, with the number of peer review reports analysed.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Figure S3: Average evaluation scores by type of affiliation for female and male applicants and proportions of female and male applicants by affiliation

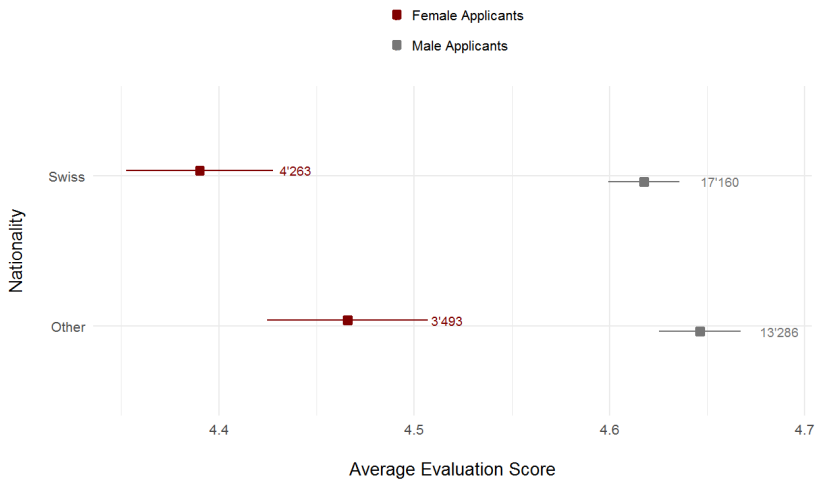
Upper panel: Horizontal lines indicate Wald confidence levels, with the number of peer review reports analysed.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Figure S4: Average evaluation scores by nationality for female and male applicant.

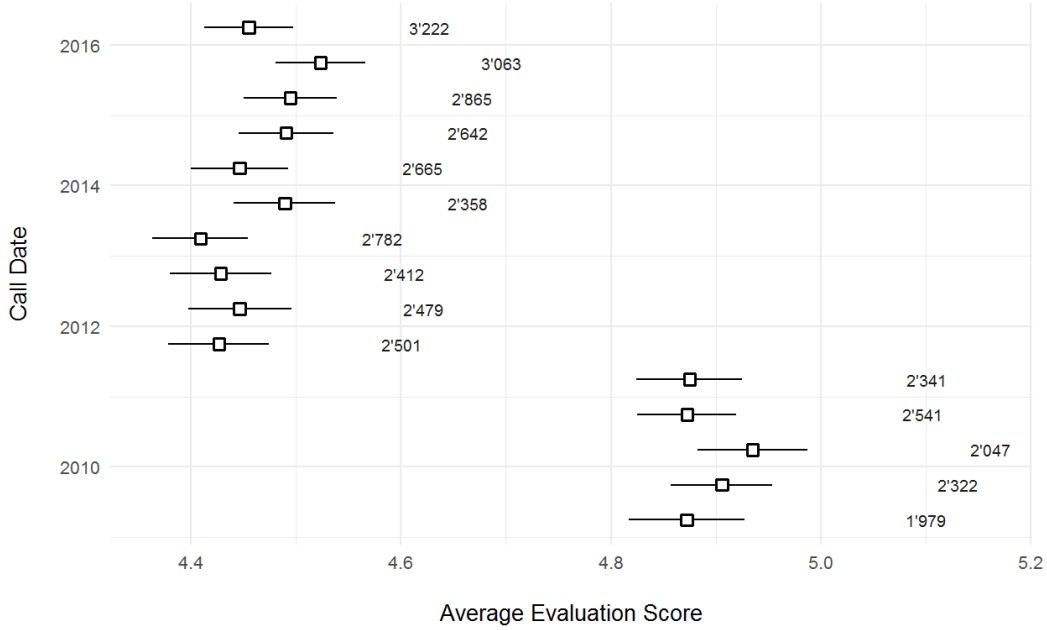
Horizontal lines indicate Wald confidence levels, with the number of peer review reports analysed.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Figure S5: Average evaluation scores by application call deadline

Horizontal lines indicate Wald confidence levels, with the number of peer review reports analysed.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

Text S1: Old and new evaluation forms.

The new forms were implemented from the 1. October 2011 call onwards.

OLD FORM

1. Synopsis

	Excellent	Very Good	Good	Average	Satisfactory	Poor		Not considered
Current scientific interest and impact of the project								
Originality of the work								
Suitability and originality of the methods to be used								
Feasibility of the project								
Experience and past performance of the applicant								
Specific abilities of the applicants for the proposed project								

Overall assessment								

Comments regarding the overall assessment

2. Detailed evaluation

Current scientific interest and impact of the project

Originality of the work

Suitability and originality of the methods to be used

Feasibility of the project

Experience and past performance of the applicant

Specific abilities of the applicants for the proposed project

Other comments

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

NEW FORM

1. Synopsis

	outstanding	excellent	very good	good	average	poor		Not considered
Applicants' scientific track record and expertise								
Scientific relevance, originality and topicality								
Suitability of methods and feasibility								

Overall assessment								

Comments regarding the overall assessment

2. Detailed evaluation

Applicants' scientific track record and expertise

Scientific relevance, originality and topicality

Suitability of methods and feasibility

3. Further comments & declaration concerning conflicts of interests (will not be forwarded to applicants)

Confidential messages

The topic of the proposed project

Declaration concerning conflict of interests (comments, if applicable

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27587v3 | CC BY 4.0 Open Access | rec: 19 Jun 2019, publ: 19 Jun 2019