Explaining gender bias in ERC grant selection – Life Sciences case

Besselaar, P. van den; Schiffbaenker, H.; Sandström, U.; Mom, M.

STI 2018 Leiden

23rd International Conference on Science and Technology Indicators
"Science, Technology and Innovation Indicators in Transition"

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at https://hdl.handle.net/1887/64521

ISBN: 978-90-9031204-0

23rd International Conference on Science and Technology Indicators (STI 2018)
"Science, Technology and Innovation indicators in transition"
12 - 14 September 2018 | Leiden, The Netherlands
#STI18LDN

Explaining gender bias in ERC grant selection – Life Sciences case[‡]

Peter van den Besselaar[*,#], Helene Schiffbaenker[**], Ulf Sandström[##], Charlie Mom[#]

[*] Network Institute, Vrije Universiteit, Boelelaan 1105, 1081 HV, Amsterdam (Netherlands)
p.a.a.vanden.besselaar@vu.nl

[#] Teresa Mom Consultancy bv, Amsterdam (The Netherlands)
charliesannemom@gmail.com

[**] Joanneum Research, Vienna (Austria)
helene.schiffbaenker@joanneum.at

[##] KTH, Royal Institute of Technology, Stockholm (Sweden)
forskningspolitik@gmail.com

Abstract
To explain lower success rates of female applicants in ERC (life sciences) starting grants, we collected data about past performance of the applicants, and we interviewed panel members about how selection criteria are practiced in general and specifically for female vs. male applicants. The analysis of the interviews provides empirical evidence that current evaluation practices indeed lead to gender-biased practices and outcomes. The statistical analysis shows – after controlling for several past performance variables – the prevalence of gender bias, more often in favor of men than of women.

Keywords: Gender bias; peer review; panel review; research grants; ERC; European research Council; funding.

Introduction
There is a longstanding discussion on whether gender bias influences grant selection processes, and the literature shows contradicting results [1,2,3,4,5,6,7,8,9]. However, there are three main problems with most research: (i) Most studies explain in fact only differences between success rates of men and women. However, these success rates are only meaningful when taking possible quality differences of male and female researchers into account. If these exist [10,11], gendered differences in success rate could partly or fully an effect of those quality differences and not of gender bias. To solve this problem, we have collected data to measure various dimensions of past performance, which are included in the analysis. (ii) Most studies depend on information only about the successful applicants, but not on the rejected – as the latter data are generally accessible for investigators. However, in this study we do have the data about successful and rejected applications. (iii) Bias emerges from the decision-making process, and this is often done at the level of review panels. In contrast, most studies focus on a higher level of aggregation, such as the funding instrument, or at the level of the discipline. We include here an initial analysis at panel level. We do detect gender bias, in contrast to recent reviews [6,8,9].

[‡] This work was supported by the ERC (grant 610706: GendERC project), but the funder had no influence on the design, analysis, or interpretation of the results. The work was also supported by the EC (grant 2654319: RISIS project).

We investigate the 2014 ERC Starting Grant scheme, and have access to the relevant data about the 3,030 applicants (about 95 %) that gave informed consent. We selected this case, as it is the most prestigious grant that exist in Europe for early career researchers (up to seven years after the PhD), and it is expected to strongly contribute to career opportunities of those getting the grant [12].

Starting point of the study is that overall female applicants have lower success rates (applicants/grantees) than men, most obviously in the life sciences (LS) domain. Figure 1 shows the success rates in step 1 and step 2 of the evaluation process in the nine LS panels of the StG 2014. The panel level enables to locate gender differences more accurately and potential improvements can be implemented more effectively. In this case, women have a 6 % lower success rate in step 1 and a 2 % lower success rate in step 2, which makes an overall difference of 3 %. Figure 1 illustrates that gender differences success rates vary considerably between panels. In panel LS8 (Evolutionary, population & environmental biology) women do much better than men, but in panel LS6 (Immunity and infection) it is the opposite. Also, differences exist between step 1 and 2 in the procedure, indicating the large influence of the interview with the applicant, and/or a 'gender correction' at least in a few panels (LS1, LS3, LS4). Due to space limitations, this cannot be discussed here.

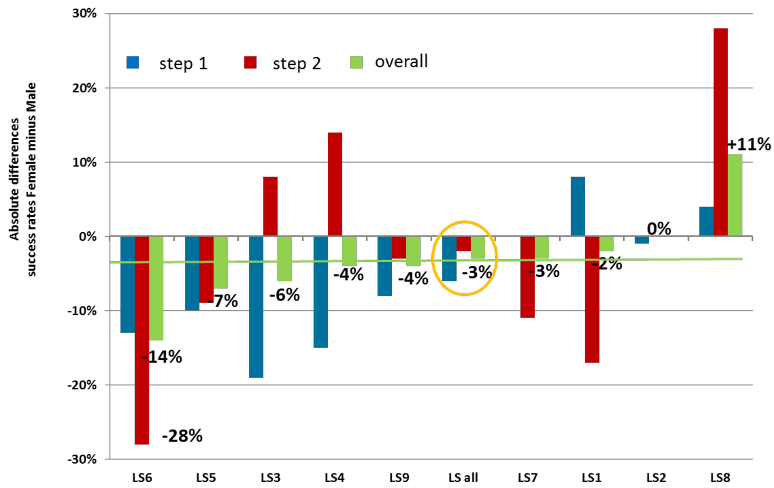

Figure 1: Success rate of female panel members, StG 2014, LS panels

As Table 1 shows, domain and field differences exist. Overall the success rate of women is higher than the success rate for men in physics and engineering (PE), possibly as part of policies to increase female participation within PE. But within PE the differences are large: women do much better in 'fundamental constituents of matter', but much worse in 'mathematical foundations'. Although in the social sciences and humanities the overall success rates are equal, between the SH fields, differences are huge: Within 'environment, space and population' (SH3: sociology, anthropology, education, communication) women do twice as well as men, and within 'markets, individuals and institutions' (SH1: economics, organization and management) it is the other way around. It would be interesting to find out why these differences occur. Is this research field specific, for example is gender stereotyping stronger in fields where 'excellence' plays a stronger role in the discourse such as philosophy, mathematics, economics [13], or is this an effect of group dynamics – so more related to personal and panel characteristics [11,14, 15]? Studying these patterns over time may answer these questions. If the pattern is stable over time, one may conclude that the field characteristics are most important, if not it may be mainly panel characteristics.

Table 1. Ratio of female success rate and male success rate

Domain	Overall	Panel with highest ratio		Panel with lowest ratio	
LS*	- 22 %	+ 116 %	Evolutionary, population & environmental bio	- 68 %	Immunity and infection
PE**	+ 13 %	+ 85 %	Fundamental constituents of matter	- 70 %	Mathematical foundations
SH***	0 %	+ 105 %	Environment, space and population	- 83 %	Markets, individuals and institutions
All	- 5 %				

* Life sciences; ** Physics & Engineering; *** Social Sciences & Humanities. Panel names from 2014.

We used a series of interviews to investigate the grant selection process and the possibility of bias entering into it. The panel processes are only weakly formalized, as are the criteria deployed by the panelists. The council has two principles implemented: (i) the only criterion that should count is excellence of the project and the investigator; (ii) panels consist of excellent researchers in their respective fields and should therefore decide among themselves what excellent applications are. But what is 'independence' and how can a panel member or a reviewer see this? And what is 'ability to do groundbreaking research'? Is that having published in *Nature*, having published a very highly cited paper, or something else? The interviews show that this in fact results in quite uncertainty and differences. For example, reviewers doubt about criteria deployed and express the need for clearer and operational criteria for 'excellence'. As a panelist tells:

> "*They give you very general guidelines like the scientific quality, the quality of the researcher, the originality of the proposal, and so on, typical of all projects. In those projects that are so related to your field of expertise you don't even need it because you appreciate them immediately. The problem comes when the projects are far from your field of expertise, then you have to be very objective in your criteria, so I have prepared a list of things I should not be forgetting.*"

This uncertainty and the well-known group dynamics that occurs in panels [11, 14] open the possibilities for bias entering in the evaluation and selection, which is strongly reinforced by the high time pressure the panels are confronted with. But, if bias is possible, does it also occur? To provisionally answer this question we use the following statistical analysis .

Approach, data & methods

Given this, we aim to predict the *applicants scores* and *application success*, using a set of independent variables related to performance (productivity, impact, previous grants, quality of the collaboration network) and to the person (age, nationality, research field, and of course gender). As decision-making on grants is done in panels, the effect of the panel is considered too – through an informal multi-level approach.

The following data were collected, and we add what variables where extracted. As the data had many formats, quite some technical work needed to be done to extract and integrate the required data (using the SMS platform - www.sms.risis.eu):

- *Age, gender, date of PhD, nationality, field of research:* from an administrative files of the ERC.
- *Earlier and current other grants*: manually extracted from the CVs.
- *Collaboration network*: semi-automatic extraction of organizations from the CVs
- *Quality of the network*: semi-automatic linking of organization names with the data in the Leiden Ranking; manual search for comparable scores of those organizations not in the Leiden Ranking.
- *Host institution*: from an administrative file of the ERC. For the quality of the host institution we use the extended 2015 Leiden Ranking scores.

- *Productivity, impact:* Downloaded from the Web of Science with a manual disambiguation. The we calculated a series of bibliometric variables, such as the number of publications, the number of fractional publications, the number of citations, the number of citations with a three years window, the share of top cited papers (1%, 5%, 10%, 25% and 50%), the number of top 10% highly cited papers (so the size dependent variant), the average number of coauthors, and the average number of international coauthors [22].
- *Organizational bias:* From the applicants' data and the panelists' data we extracted the links between applicants and panel members in terms of belonging to the same organization [9].[1]
- *Panel review scores of the applications*: from an administrative file of the ERC.
- *Decision:* from an administrative file of the ERC

We currently have a stratified dataset of about 1742 applicants, evenly distributed over the 5 scores given by the panels: A-granted, A-not-granted, B-step2, B-step1, C. We plan to collect the bibliometric data for the remaining 1288 in the future, so the results here are to some extent preliminary. The unique nature of our data is that we can combine (advanced) bibliometric indicators with a large set of other variables. These data enable several interesting analyses. For example, one may analyze whether organizational proximity (cronyism) [16], or cognitive proximity [17,18] have an effect on grant success. One may also study whether language use in review reports shows the nature of the decision-making process [19,20], and more specifically whether language use shows gender bias [21].

Analysis

Due to space limits, we restrict the analysis to the Life Sciences. Firstly, we deploy *ordinal regression* for the LS applicants in order to estimate the effect of gender on the decision after controlling for several quality (past performance) variables, the quality of the network and, and for organizational proximity. Secondly, we move to the second level, and compare the panels. We do a similar regression but on the level of the individual panels that can be compared.

Results 1: Life sciences

We used the bibliometric indicators mentioned above, the variables on the quality of the network and the host institution, and the number of grants the applicant has already acquired. We also include whether a panel member is at the host institution of the applicant, and gender. Running an ordinal regression, and after manually stepwise deleting variables that did not work, eight variables remained in the model, which resulted in a pseudo R-square (Nagelkerke) of 0.308. Table 2 shows the result.

Factors that help to get a better score are papers in high impact journals, the quality of the network, measured as the median ranking of the organizations in the network of the applicant, average number of international coauthors, and the number[2] top 10% most cited papers (fractionally counted). Negative works the average number of coauthors, as that may suggest a lower level of independence. Finally we do find effects of sexism and nepotism: women score some 0.35 points lower than male (on a five-points scale), and when the candidate has a

[1] We also started to analyze the role of *Cognitive bias* but at the moment we only have data for a few panels. We therefore do not include it here [17, 18].

[2] This is the size dependent variable, which we feel is more valid than the share of top cited-papers.

panel with a panel member that is at the proposed host institution, this gives almost a 0.6 point bonus.

Table2: Score by performance, organizational proximity and gender

	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
Number highly cited (10%) papers	0.124	0.042	8.72	1	0.003	0.042	0.207
Journal impact (NJCS)	1.133	0.11	105.386	1	0.000	0.917	1.350
Number earlier grants	0.184	0.029	39.324	1	0.000	0.126	0.241
Quality network	0.003	0.001	25.41	1	0.000	0.002	0.004
Average nr co-authors	-0.095	0.031	9.099	1	0.003	-0.156	-0.033
Average nr international co-authors	0.281	0.16	3.087	1	0.079	-0.032	0.595
Nearby panelist	0.584	0.228	6.575	1	0.010	1.031	0.138
Female versus male	-0.349	0.145	5.791	1	0.016	-0.634	-0.065

Ordinal regression; Link function: Logit
Pseudo R-square (Nagelkerke) = 0.308
Bootstrapped: 2000 samples; confidence interval 95%

This means that from a *performance* perspective, only one variable plays a role (the number of top cited papers). The other variables that influence the score are *reputation based* (journal impact related; ranking related) and network based (number of (international) co-authors). Also, the number of earlier grants has a positive effect on the score; and these grants partly can be considered as performance, but at least also partly as reputation-related. Finally we find two bias factors: after controlling for the performance and reputation variables, sexism and cronyism still have an effect on the scores the applicants get.[3]

Results 2: Life science panels

As grant decision-making tales place at the level of panels, and different social dynamics may take place in the different panels, one may expect that the levels of bias may be different in different panels. We therefore repeat the analysis for the 9 individual panels, each representing one or more specific disciplines within the life sciences. However, as at panel level the number of granted applicants is low (typically about 11 out of about 100 applicants), the number of variables that can be included is smaller, and also variables that are significant at the LS domain level, are that not anymore on panel level. Nevertheless, the variables have overall the same effect in the panel models as for the domain as a whole. In table 3 we show the sign of the variables for each of the panel-regressions. We use the same variables as for life sciences as a whole. Most have the expected effect, but some have not.

Table 3: Regression coefficients at panel level

	LS1	LS2	LS3	LS4	LS5	LS6	LS7	LS8	LS9
Number highly cited (10%) papers	+	+	+	+	+	+	+	+	+
Journal impact (NJCS)	+	+	+	+	+	+	+	+	+
Number earlier grants	+	+	-	+	+	+	+	+	+
Quality network	+	+	-	+	+	+	+	+	+
Average nr co-authors	-	-	-	-	-	-	-	+	-
Average nr international co-authors	-	+	+	-	+	+	+	+	-
Nearby panelist	-	-	-	-	-	-	+	-	+
Female versus male	+	-	-	-	-	-	-	+	+
Success rate F minus M	-	0	-	-	-	-	-	+	-

[3] Results concerning cronyism (or nepotism) confirms the follow-up study concerning the Swedish MRC reported in [23].

Interestingly, gender bias in favor of men is in 6 of the 9 panels, covering 78 % of the female applicants in the life sciences. One panel is neutral (8 %) and two are biased in favor of women (15 % of the female applicants). This needs further analysis, but gender bias may be related to the share of women in a field.

Conclusions and further work

We have shown that gender bias occurs in the life sciences, but not in all parts of the field in the same way. In most panels we find bias against women, but in two panels it is the opposite and one panel is neutral in this respect. However, the first set of panels include almost 80 % of all female LS applicants. We also found that the gender bias and different success rates are not the same, as in one third of the panels, the sign of gender bias is different from the sign of the success rates: For example, in panel 9, the success rate of women is higher than of men, but there is still gender bias in favor of men, after controlling for the performance of applicants. This means that in fact the positive success rate without gender bias would have been higher.

This analysis covers only life sciences, but we are also analyzing the other domains: social sciences and humanities, and physics and engineering. These fields are not only different in terms of gender success rates, but we also expect differences in gender bias.

Panels play an important role, therefore we will also include characteristics of the panel to the model. What panel characteristics do lead to gender bias? For example, we found a negative correlation between the number of female panel members and the female success rate (not discussed in this paper).

Finally, if one understands the dynamics of gender bias, the next question is how to reduce it. That is crucial, as the type of grants we study here have strong career implications [15, 16].

References

[1] Ahlqvist, V., Andersson, J., Söderqvist, L., Tumpane J. (2015*) A gender neutral process? – A qualitative study of the evaluation of research grant applications 2014*, Swedish Research Council, Stockholm.

[2] Beck, R, Halloin, V. (2017) Gender and research funding success: Case of the Belgian F.R.S.-FNRS. *Research Evaluation* **26**(2), 115–123.

[3] Böhmer, S., Hornbostel, S., Meuser, M. (2008) *Postdocs in Deutschland: Evaluation des Emmy Noether-Programms*, iFQ-Working Paper No. 3, Bonn.

[4] Bornmann, L., Mutz, R., Daniel, H.D. (2007) Gender differences in grant peer review: a meta-analysis, *Journal of Informetrics* **1**, No. 3, pp. 226–238.

[5] Bornmann, L., Mutz, R., Daniel, H.D. (2008) How to detect indications of potential sources of bias in peer review: A generalized latent variable modeling approach exemplified by a gender study, *Journal of Informetrics* **2**, 280– 287.

[6] Ceci, S.J., Ginther, D.k., Kahn, S., Williams, W.M. (2014) Women in academic science: a changing landscape. *Psychological Science in the Public Interest* 15 (3): 75-141.

[7] Van den Lee R, Ellemers N (2015) Gender contributes to personal research funding success in The Netherlands. *PNAS* **112**, 12349-12353

[8] Marsh, H. W., Jayasinghe, U. W., Bond, N. W. (2011) Gender differences in peer reviews of grant applications: A substantive-methodological synergy in support of the null hypothesis model, *Journal of Informetrics* **5**, 167–180.

[9] Williams, W.M., Ceci, S.J. (2011) Understanding current causes of women's underrepresentation in science. *PNAS* **108**, 8, 3157-3162.

[10] Van den Besselaar P, Sandström U (2018) Vicious circles of gender bias, lower positions and lower impact: gender differences in scholarly productivity and impact. *PlosOne* **12** (2018) 8: e0183301.
[11] Lamont M (2009) *How professors think*. Harvard University Press
[12] Van den Besselaar P, Sandström U (2015) Early career grants, performance and careers; a study of predictive validity in grant decisions. *Journal of Informetrics* **9** 826-838
[13] Leslie SJ, Cimpian A, Meyer M, Freeland E (2015) Expectations of brilliance underlie gender distributions across academic disciplines. *Science* **347**, 6216, January 6.
[14] Van Arensbergen P, Van der Weijden I, Van den Besselaar P (2014)The selection of talent as a group process; a literature review on the dynamics of decision-making in grant panels. *Research Evaluation* **23** 4:298-311
[15] Olbrecht M, Bornmann L. (2010) Panel peer review of grant applications: What do we know from research in social psychology on judgment and decision-making in groups? *Research Evaluation* **19** 293–304.
[16] Mom C, Van den Besselaar P. *Does institutional proximity affect grant application success?* Paper presented at the PEERE Conference, Rome, March 2018
[17] Van den Besselaar P, Sandström U (2017) Influence of cognitive distance on grant decisions, *Proceedings STI conference* 2017, Paris.
[18] Sandström U & Van den Besselaar P, *The effect of cognitive distance on gender bias in grant decisions*. Paper presented at the PEERE Conference, Rome, March 2018
[19] Van den Besselaar P, Stout L, Gou X (2016) Predicting panel scores by linguistic analysis. In: Ismael Rafols et al, *Peripheries, Frontiers and Beyond; Proceedings STI 2016,* Valencia
[20] Van den Besselaar P, Sandström U, Schiffbaenker H (2018) Using linguistic analysis of peer review reports to study panel processes, *Scientometrics* (2018)
[21] Van den Besselaar P, Sandström U, Schiffbaenker H, *Gendered language in applications and reviews: a linguistic analysis (In preparation)*
[22] Sandström U. & Wold, A 2015) Centres of excellence: reward for gender or top-level research? In Thinking Ahead: research, funding and the future. *RJ Yearbook 2015/2016*, pp.69-89.
[23] Sandström U. & Hällsten M. (2008). Persistent nepotism in peer review. *Scientometrics* **74** (2) 175-189.