Granted project

Work Package 9

Deliverable 9.2 Policy brief on recommendations

What to do against gender bias in grant allocation?

Peter van den Besselaar, Charlie Mom

TMC Research Amsterdam

Partner	Authors and contributors	Contributions
TMC (WP leader)	Peter van den Besselaar, Charlie Mom	Authors main text, Editors
CSIC	Laura Cruz-Castro & Luis Sanz-Ménendez	Authors recommendations 1
FPS	Ulf Sandstrom	Author recommendations 2
JR	Helene Schiffbänker, Angelika Sauer, Florian Holzinger	Authors recommendations 3
ORU	Helen Peterson, Liisa Husu	Authors recommendations 3 & 4

DELIVERABLE 9.2 - OCTOBER 2023

Project no.	*824574*
Project acronym:	*GRANteD*
Project title:	*Grant Allocation Disparities from a Gender Perspective*
Instrument:	*H2020-SwafS-2018-2020/H2020-SwafS-2018-1*
Start of the project:	*01.01.2019*
Duration:	*58 months*
Work Package:	*WP9*
Due date of deliverable:	*Month 58*
Organization name of lead contractor for this deliverable:	*TMC*
Other partners involved:	*CSIC, DZHW, FPS, JR, ORU*
Document version:	*V8 – Public*

Preface

This deliverable consists of several *independent* parts. The main text tries to extract conclusions and recommendations from *all* deliverables – all research done in the GRANteD project – and from interactions with the Stakeholder Committee, with the Scientific Advisory Board, and with several members of those committees individually. We aim to keep this text as short and as clear as possible to create a basis for an exchange about the prevalence of gender bias in grant allocation, its causes, and possible strategies to reduce it. As the text is based on the research done, it is important that the reader can become acquainted with the results from the different studies. Some of the work can be accessed on the GRANteD website.[1] Another part of the research results have already been published[2] and more will be prepared for publishing in the period after the end of the EC contract (November 1, 2023). Preprints of the publications will become available too, to speed up the process of dissemination.

Integrating always requires a specific perspective, and different authors may select different perspectives. In order to have room for the necessary variety, we chose for a structure where several contributions are added to deliverable independently. These are of a different nature and formulate conclusions and recommendations from the perspective of the specific parts of the project the authors were involved in. Together the various contributions present a variety of policy recommendations, including suggestions for follow-up research.

It should therefore be emphasized that the authors are only responsible for their own text, and not for the other parts of the deliverable.

---

[1] www.granted-project.eu

[2] See next page

Some of the publications of the GRANteD Project

- Cruz-Castro, L., and Sanz-Menéndez, L. (2023). Gender bias in funding evaluation: A randomized experiment. *Quantitative Science Studies*, Advance publication.
- Cruz-Castro, L., Ginther D, and Sanz-Menéndez, L. (2023). Gender and underrepresented minorities differences in research funding. In: *Handbook of Public Funding of Research*
- Cruz-Castro, L., and Sanz-Menéndez, L. (2021). What should be rewarded? Gender and evaluation criteria for tenure and promotion. *Journal of Informetrics*
- Mom C, van den Besselaar P & Möller T (2022). Factors influencing the academic career – an event history analysis. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *26th International Conference on Science and Technology Indicators*, STI 2022 (sti22138).
- Möller T, van den Besselaar P & Mom C (2022). What is researcher independence and how can it be measured? In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *26th International Conference on Science and Technology Indicators*, STI 2022 (sti22138).
- Möller T (2023). Do female academics submit fewer grant applications than men? *Proceedings 27th International Conference on Science, Technology and Innovation Indicators (STI 2023).*
- Sandström U & Van den Besselaar P (2019). Performance of Research Teams: results from 107 European groups. In*: Proceedings 17th International Conference on Scientometrics & Informetrics ISSI 2019*, pp 2240-2252. September 2-5, 2019, Sapienza University of Rome, Italy
- Van den Besselaar P & Mom C (2023). Bibliometric indicators as Items. In: *Proc. ISSI Conference,* Bloomington.
- Van den Besselaar P & Mom C (2023). Gender and merit in awarding cum laude for the PhD thesis. In: *Proc. ISSI Conference,* Bloomington.
- Van den Besselaar, Sandström U & Mom C (2019), Recognition through performance and reputation. In: *Proceedings 17th International Conference on Scientometrics & Informetrics ISSI 2019*, pp 2065-2069, September 2-5, 2019, Sapienza University of Rome, Italy
- Van den Besselaar P & Sandström C (2020). Bibliometrically disciplined peer review; using indicators in research evaluation. In: *Scholarly Assessment Reports* **1.**

Table of contents

Introduction

This text summarized the recommendations that can be distilled from the research done in the GRANteD project, and from the interactions with the Stakeholder Committee, the Scientific Advisory Board, and at other exchanges with researchers in the field and with stakeholders in different interactions. The aim is to keep this as short and as clear as possible to create a basis for an exchange on studying and reducing gender bias in grant allocation.

After the main text was written, all partners in the project were invited to contribute with their conclusions and comments. These are added as annexes to the main text.

No effort has been made to come to a consolidated text based on consensus about the findings and consequences. It is more insightful for the reader to get a variation of inputs that can be weighted and used, than a consensus text that may hide the different and possibly contradictory options that are available.

Before formulating the recommendations, a few introductory remarks are needed. First of all, gender is one of the organizing principles at least in modern societies[3], and there are no signs that this is declining. Differentiations and differences related to gender can be expected to operate permanently. Research on gender bias in science therefore cannot be expected to 'solve' the problem of gender bias, but it can do at least two things:

- Investigating when and where gender differences are in fact forms of gender discrimination, gender bias, and

- Helping to detect the social and psychological mechanisms through which gender bias is produced.

Through this second task, one can become less ignorant about the problem and try to find effective measures to counteract gender (and other forms of) bias. Furthermore, when one finds positive developments showing a decline of gender bias - as we do in the project - that does not mean that gender bias is over. It shows that in the domain under study the gender bias producing mechanisms have become weaker and/or that enough countervailing pressure has been organized to counteract the gender bias producing mechanisms. But as long as gender remains an organizing principle in society, the main task is to continuously monitor the way gender influences social processes such as in our case grant evaluation and allocation and trying to act accordingly, based on the knowledge we have about the relevant processes and mechanisms.

We cannot proceed without making clear conceptual, empirical and measurement differences between *gender disparities* and *gender bias* when referring to research funding evaluation. The two concepts are not identical as existing gender disparities could exist and be the outcome of individual preferences, social structures, or other factors, while gender bias, as defined here[4], is caused by the influence of the gender of the applicant on the evaluation of the reviewer or the decision of the

---

[3] One could argue that this is a recent phenomenon and that over a very long period in history this was not the case (Van Schaik & Michel, *Die Wahrheit über Eva. Die Erfindung der Ungleicheit von Frauen und Männer*. 2020 Rowohlt).

[4] See Cruz-Castro & Sanz-Ménendez Deliverable D1.1.

panel, whatever causal mechanisms (reviewers' preferences, statistical bias, stereotyping, etc.) is considered. Bias in the evaluation is deviation from merit.

What do we learn from the project: Research-based recommendations.

Peter van den Besselaar & Charlie Mom, TMC Research, Amsterdam

1. In a series of case studies of research funding agencies in various countries in western Europe[5], the project analyzed the <u>evaluation of grant applications and the grant decisions</u>. The case studies cover different grant types, using different explanatory variables, and different analytical strategies, but the findings show a development over time: In the panel scores and in the grant decisions we find a trend that the (sometimes significant) advantage for men is increasingly replaced by a more neutral outcome and sometimes even an advantage for women. In a few studies, the <u>individual panel</u> level was addressed, the lowest organizational level where the evaluation takes place. There it becomes visible that an overall neutral outcome is often the result of aggregating mixed findings of neutral panels, of panels in favor of men, and of panels in favor of women. Given the relative low number of panels that were investigated, it is still an open question what the causes of these differences are.

2. The project also found that the <u>application activity</u> of men and women seems more or less equal.[6] Although there are quite some questions unanswered (see below), these findings suggest that factors like self confidence, risk avoidance and avoiding competitive situations do not seem to be different between women and men in the context of grant applications. If this is the case, the question comes up whether and why this has changed over time, or whether that has always been the case and the often-heard argument that women apply less than men seems wrong. It could also be related to the choice of the cases, which reflect the layer of applicants from research universities and not from other parts of the segmented science system.

3. Thirdly, the <u>context in which to apply for grants</u> has become much more equal, and in most aspects no gender effect or bias was found in the application survey (D7.1). For example, men and women spend an equal amount of time on the different tasks like teaching, research, supervision and so on. The main differences were in the role in the <u>caregiving</u> (more women than men report that they do the main share) and in <u>career support</u> received from a mentor (women report that they get less support than male colleagues).

> Positive developments towards equality in
>
> - grant application conditions
>
> - grant application behavior
>
> - grant evaluation and allocation

---

[5] Austria, Germany, Ireland, Netherlands, Spain, Sweden. The Polish and the Slovakian cases could not be integrated in the quantitative analyses, due to a lack of data.

[6] The applicant survey for five cases (FWF - Austria, NCN - Poland, SRCI - Slovakia, SFI - Ireland, SRC - Sweden), and the Wissenschaftsbefragung (DZHW, Germany).

4. The findings suggests that cultural (generational) changes, <u>policies</u> or other actions undertaken in the various countries[7] under study have been successful in equalizing the chances of women and men, independently of the dynamics in the panel level. However, only some of the cases have implemented reranking instruments to guarantee equality whereas others not, but the equal outcomes are a more general pattern. The GRANteD project has investigated the wide set of gender bias related risks and of gender policies as a possible answer to those risks[8], but whether policies and instruments have played a role in the observed trend remains uncertain. This was also emphasized by various researchers and stakeholders: Although it is clear that in the six countries included in the quantitative studies many gender equality policies were formulated, the question which of those were effective has not been answered yet.

5. The qualitative case studies show the tension between the formulation of <u>policies</u> and the <u>implementation</u>. With regards to the implementation, the interviews suggest that panel members and others are not always willing to implement it because they do not see advantages, or because that they are lacking the skills to do it. Often the possibility of training panel members and reviewers comes up, but the effectiveness of that is uncertain. So further research in trying to link policy interventions to changes in grant allocation outcomes is still needed.

> - More research is needed about what policies have influenced the observed developments, and what is behind the panel differences underneath the aggregated equality.

6. Interviews with panel members and other reviewers show that they find it often difficult to apply the rather general selection criteria in concrete cases. If that is the main problem, the solution is not more training for reviewers (who may not want to invest time in it), but a stronger effort <u>in defining and operationalizing the relevant criteria for evaluation</u> and in making sure that applicants provide the information that enables reviewers to use the criteria.

7. The positive developments mentioned above do not mean that the problems are resolved for ever, as it would be too optimistic to conclude that mechanisms that produce gender bias have been disappeared. We use the Netherlands as an example: The case study on gender bias in awarding cum laude for the doctoral thesis[9] shows that in PhD-thesis evaluation with a regulated process but without formalized criteria, men have a huge advantage above women to receive this highly selective recognition. Controlling for several (performance)

---

[7] And possibly the EU in the context of the ERA.

[8] Deliverables D5.1 and D6.1

[9] Peter van den Besselaar & Charlie Mom (2023). Gender and merit in awarding cum laude for the PhD thesis. In: *Proc. ISSI Conference,* Bloomington. Also: Deliverable D4.1

variables did not influence this. In cum laude decisions, all subjectivity is allowed ("we recognize quality"), and that leads to an persistent gender bias detrimental to women.

8. The Dutch career study[10] shows that women leave the science system more often and earlier than men, and women become later and less often full professor, suggesting that also gender bias in the appointment procedures persists.

> - An important domain for research on gender differences in science are the dynamics of academic careers in the research performing organizations.

9. Both examples suggest that mechanisms like gender stereotyping still exist, and therefore monitoring of the application environment and of the grant selection process seems crucial. How are success rates by gender developing, after controlling for relevant evaluation criteria, in relation to procedural, organizational, and policy changes.

   Research has a clear role in monitoring the increase or decline of bias, and research funding organizations should take up their core role in collecting the data that can be made available for monitoring and research. One should make use of the special provisions in the GDPR for statistical and scientific research. Otherwise monitoring and research will be based on small samples that are biased because of self-selection. The research in GRANteD suffered from these problems.

> - Permanent monitoring is needed, by research funding organizations and by researchers.
>
> - Collection of data is a core part of this, and RFOs should create the open access conditions so these data can be used for research on gender disparities in research funding.
>
> - The quality and size of the data collection should enable the use of analytical techniques that can result in testing causal hypotheses.

10. Funding programs are designed to advance many different objectives, and the design and definition of assessment criteria should be an integral part of developing fair instruments for

---

[10] Mom, C., van den Besselaar, P., & Möller, T. (2022). Factors influencing the academic career – an event history analysis. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *26th International Conference on Science and Technology Indicators*, STI 2022 (sti22138). https://doi.org/10.5281/zenodo.6975566. See also Deliverable D3.3. In the Swedish case study in D3.3, gender bias in the career was found too.

funding that provide all applicant equal opportunities based in merit (which includes past performance, and the quality of the research proposal) and other relevant criteria.

Adequately defining and operationalizing the dimensions to be used in grant evaluation is crucial. Firstly, because gender bias is defined as gender influencing the scores and the decision after controlling for the relevant selection criteria. Only if these are clearly operationalized these criteria can be included in selection processes. This does not only hold for the indicators for scientific performance and contributions to science, but also for other academic tasks than research and for general qualities like managing a research team, coaching team members, and so on.

11. Over the last decade, a debate emerged on new ways for evaluating research, such as evaluation in research context[11], recognizing and rewarding[12] the diverse tasks academics and academic groups have, and not anymore focus on evaluating research performance only, and especially not on bibliometric indicators only. Recent initiatives are DORA[13] and CoARA[14]. This has two aspects. Firstly, it is about including more evaluation dimensions representing the non-research tasks, and secondly it is about the relation between peer assessment and the use of indicators.

12. Although efforts exist to be more specific about the broader set of evaluation criteria, it is not clear to what extent these are implemented. One of the cases in the GRANteD project made an effort to specify evaluation criteria in terms of a series of questions to be answered in the review.[15] But the specification of evaluation questions is one step, the translation of those into a useful operationalization is the crucial next one. Without that, there is not only a *risk* that grant selection processes get biases; extensive psychological research has shown that it is unavoidable that bias and noise enter human decision making and group decision making.[16]

13. This holds for those evaluation dimensions where (bibliometric) indicators have been available (and disputed) such as for productivity and impact indicators, and for other quality and evaluation dimensions that should be included to broaden evaluation and to finetune it to the context of the work scientists are expected to do. Herein also lies a clear role for RFOs in choosing *what* it is that they consider to be worthwhile scientific output, so that both reviewers and evaluators of RFO funding processes know what to try to measure (and review in the case of reviewers).

---

[11] http://www.siampi.eu/Content/ERiC%20Guide%202010.pdf

[12] https://recognitionrewards.nl/

[13] https://sfdora.org/

[14] https://coara.eu/

[15] "How significant is the applicant's scientific productivity, impact and other merits in a national and international perspective, in relation to the research area, and the applicant's career age?" or
"To what extent does the project contribute to the applicant's ability to develop new competences and their research network, and thereby enhance their independence?"

[16] Daniel Kahneman, Olivier Sibony, Cass R. Sunstein (2021). *Noise, a flaw in human judgement.* Harper Collins Publishers.

14. Several recent examples can be given where the request for valid and reliable indicators is relevant. (i) The *narrative CV* has been embraced by research funding organizations, but it is rather unclear how those CVs can be used in a reliable and equal application evaluation process. It seems a big risk that self-presentation skills strongly influence the evaluation of the grant application: If writing style (besides scientific content) influences the evaluation outcome, the noise and bias in the grant allocation may even become larger. There should be an evaluation of experiments with narrative CVs before rolling those out in the blind.
(ii) The case studies show that RFOs are increasingly using *gender relevant research* as criterion (GiRI). To recognize this, it has been suggested to involve gender experts in evaluation panels[17], but how does one recognize 'gender experts'? And definitions and operationalizations for *gender relevant* are lacking.

15. The argument that one cannot measure everything that is relevant opens the door to subjective evaluations and consequently for bias. Also evaluation practice needs to have adequately defined and operationalized indicators. Within the GRANteD project, work has been done to further enlarge the indicators toolbox with a few that were considered important: Indicators for independence, which are increasingly important now teamwork has become the golden standard for research and indicators for cognitive mobility, which is the ability to increase the scope of the research agenda by taking up new topics. In those activities, it is crucial to be clear about validity and reliability issues.

> - There should be a substantial investment into research to develop valid and reliable measures for the relevant evaluation criteria.
>
> - Research on gender differences in grant allocation should use in the models those factors that are expected to wrongly influence the grant selection, and the criteria that should be used in grant evaluation. Operationalization is crucial.
>
> - This is essential for research on gender issues, but equally essential for a reliable and equal selection process.

16. More generally, the DORA declaration and the Leiden Manifesto are often used as arguments to reduce the role of indicators, an opinion also found in the interviews. However, the answer to wrong use of (wrong) indicators is a better use of better indicators. The GRANteD project has contributed to the development of better indicators.

---

[17] See for an analysis of the presence of gender specialist in evaluation panels (Peter van den Besselaar & Ulf Sandstrom (2019). Panel composition as a pathway too impact: Do we need stakeholder expertise to select relevant mission-oriented projects? In: *fteval Journal* **48**, p 68-73).

The criteria for assigning a proposal as 'gender relevant' was done based on a classification of 'gender relevant topics', which was found not very adequate. The qualification of 'gender specialist' was based on self-selection by reviewers in the EC experts database (Bente Knoll, Suzanne de Cheveigné, et al. (2017). *Interim Evaluation: Gender equality as a crosscutting issue in Horizon 2020. Report of the Expert Group.* Directorate-General for Research and Innovation).

17. What about the relation between indicators and peer review? Grant selection will never be an algorithmic process based on indicators. On the other hand, a process without clearly defined and measurable criteria will always tend to become biased. A main issue to address is how reliable indicators can be optimally used in the evaluation and decision-making processes. Field experiments should be stimulated and evaluated, where and one can learn from earlier experiences, such as the one described in an earlier study: *Bibliometrically disciplined peer review*.[18]

> - Field experiments are needed with models of how to integrate quantitative indicators in the qualitative assessment and decision-making processes.

18. A main issue in the discussion about quantitative variables (indicators) is whether these are neutral quality indicators, or in fact biased. If women would have to teach more than their male colleagues or have to spend more time at home for care responsibilities, men may have more time to build up a past performance record, which may influence the scores on the relevant indicators.

These indirect gender effects should be taken into account, and they point at other social domains outside the funding ecology that may have to become the focus of gender equality research and policies. In the several of the cases, these indirect effects were considered through a mediation analysis. However, they were not found which may die to the fact that several of the case studies focus on early career grants and in that phase not so much gender differences in working life and in the private situation may yet exist.

> - Investigate what context variables affect grant application processes.

19. Potential other forms of bias should not be neglected (nepotism, organizational proximity, cognitive proximity). In the four cases that were trying to assess the effect of these three other potential bias sources, some effects were found.

20. What about the generalizability of the findings? The cases are rather different in type of funding instrument, variables available, national context, and level of analysis, but still fit in the general pattern. On the other hand, the case studies were all in developed science counties and that may differentiate these cases from other parts in the world. Nearby, even the situation in the east European countries is structurally different in terms of the social

---

[18] Peter van den Besselaar, Ulf Sandström (2020). Bibliometrically disciplined peer review; using indicators in research evaluation. In: *Scholarly Assessment Reports* **1.**

position of higher education and research institutions and in terms of the positions of women in the HE and science system. This would deserve more comprehensive analysis than could be done due to the data situation.

> - Under adequate data conditions, a study focusing on other countries would be worthwhile.

21. Lessons for future projects

   - A main lesson is that getting much larger datasets is crucial when selecting cases: big data. This would enable a better specification of the models, and the use of more advanced methods including field experiments. Larger datasets would also enable a better focus on the panel level.

   - A better coordination of qualitative and quantitative approaches. In the new cases, qualitative research was leading, which is most useful when it is about not yet researched issues. As there is a huge literature on gender differences in science, it would have been better – after formulating the research model – to start with the quantitative approach and then use the qualitative investigations to deepen the findings of the quantitative studies.

   - More emphasis on the universities and on (public and private) research organizations: the sources of inequality in academia seem to originate there, and not so much in the funding organizations.

22. Lessons for EC policies

An additional question which came up in review is what our findings might imply for the EC policy agenda with respect to gender, academic careers, and reforming research evaluation. The project did not include an analysis of national governments policies or EC policies, and therefore the following text should not be read as a policy analysis. But there are several points where our results touch on EC policies: (i) gender equality policy focusing on the position of women in the science system; (ii) gender bias in careers; (iii) the gender dimension and inclusiveness in research and innovation (GIRI); and (iv) the reform of the research assessment system.

With respect to the first issue, our research suggests that gender bias in grant allocation, which was found in earlier studies, has declined and possibly disappeared, at least in the personal career grants that were the focus of the GRANTED project. This conclusion is based on a series of about fourteen case studies in a variety of countries and over a few decades, and we think this should be considered as a success.[19] However, it remains relevant to study gender bias in grant allocation for other types of grants, such as thematic and team grants, among which those of the EC.

Unexpectedly, the case studies on grant application behavior also suggest a decline in gender differences in the recent period, which also is a positive development – and in the

---

[19] See Deliverable D4.3 and Deliverable D9.1

German case study it covers all grants and not only the career grants.[20] Replicating this type of studies on a larger scale would be important. And as with many social phenomena, for grant application behavior and for grant allocation there is no guarantee that positive developments are unidirectional: depending on contextual changes, a positive trend may reverse, and continuous monitoring would be required.

Secondly, that gender bias seems almost absent in grant allocation does not imply that this is more generally the case. In other case studies done in the GRANTED project such as the studies of gender bias in careers and a study on distributing awards and prizes, we did find considerable levels of gender bias. These forms of gender bias need more research to find the causes of it, and the possible differences between countries and fields.

Thirdly, the case studies also show that the selection criteria are only defined in rather general terms and hardly operationalized, and interviews confirm that reviewers and panel members therefore find it difficult to apply those criteria in a consistent way in two respects: in the same way for all applicants, and in the same way as other reviewers do. This is in line with research on the (interrater) reliability and validity of peer review, and with psychological literature on small group decision-making and selection processes arguing that these generally suffer strongly from noise and bias.[21]

This holds e.g. for the relatively new selection criterion of GIRI: Whether a research proposal is (in)directly relevant to gender issues is rather difficult to assess, as clear and operational definitions are lacking. Work done on identifying research relevant for the seventeen SDGs also has shown that this is indeed a difficult task.

This brings us to the fourth theme where our results are relevant for the EC policy agenda: the reform of the research assessment system, which has received much attention recently.[22] Research assessment plays a role in many selection processes, predominantly in the selection of grants and of applicants for academic positions. A lot of 'manifestos' and 'declarations' have been published[23] over the last years, and all centered around principles such as the following:

- Conditions: comply with ethics and integrity rules and practices; safeguard freedom of research; assure transparency of data, infrastructures and criteria used in research assessment.
- Quality and impact: focus on quality of the research, originality, results beyond the state of the art; reward a variety of research missions (and related output); recognize contributions that advance knowledge and the impact of those contributions.
- Diversity, inclusiveness, collaboration: recognize the broad variety of outputs, not only scholarly publications but also related to other tasks of researchers: management, mentoring and supervision, interaction with stakeholders, etc.; respect the differences between disciplines research types and career stages; acknowledge the diversity of research roles and of careers; ensure gender equality.

---

[20] See Deliverable D7.1 and Deliverable D7.2

[21] Kahnemann et al, *Noise* (2021)

[22] European Commission, *Towards a reform of the research assessment system*. Brussels, November 2021

[23] Several EC reports, and others like the Leiden Manifesto, DORA, the Hongkong principles and others. See for an overview the report mentioned in the previous footnote.

This list is based on the report mentioned in footnote 22, and most of these principles can generally be supported. However, the main problem with these manifestos and declarations is that they hardly specify how this should be made operational, apart from the opinion that it should be peer review focusing on quality with where appropriate supported by (quantitative) indicators – in opposite to the current situation that is said to be dominated by quantitative indicators that in many cases are not appropriate such as the H-index and the Impact factor. However, also this principle remains rather general and operationalizations are lacking. The main instrument mentioned to do this is the increasingly popular *narrative CV* as a "qualitative method for research assessment, at least in part due to their potential to support a research culture that emphasizes meaningful research achievements over the use of flawed proxy measures of quality" (DORA blog, June 6, 2022).

Although it is undisputed that selection and decision-making are human activities, emphasizing a central role for peer review does not reflect the large literature that is rather critical about the quality of peer review processes. Furthermore, much of what is covered under peer review is at best review by experts who are often not real peers of the grant application. Finally, the role of indicators in selection and decision-making seems strongly overestimated: In our studies we found that the bibliometric variables at best explain a modest part of the variance, implying that either there is a lot of randomness in the grant decision-making, or that other criteria (variables) we did not include in the studies are more important than the bibliometric indicators used.

Based on these considerations, we would argue that to improve the evaluation of researchers, of research, and of grant applications, one needs for each evaluation context an explicitly specified set of criteria that are operationalized, and on top of that an adequate process that guarantees that the criteria are applied equally on all applicants and applications. Without that, selection and decision-making cannot avoid bias and noise.

This leads to the most concrete instrument proposed: the narrative CV. If one wants to include many assessment criteria[24], then it is essential to have information about all these criteria in a CV. To make this information manageable it should be available in some structured format. It is hardly conceivable that this can be done in a predominantly narrative CV, as applicants may not include information about all these criteria or may present it in different ways, and this may affect the evaluation. As we found in the GRANTED project and in earlier projects, writing style – similar to presentation style - does influence the scores reviewers give. A short narrative part of a CV describing the main achievements is of course important, but the criteria would need some form of quantification to enable a reliable selection.

Finally, the criticism on indicators and their use is to some extent correct, but we would not suggest a more modest role of indicators, but for developing more valid and more reliable quantitative indicators all used evaluation dimensions. Improving the indicator set remains an important theme for further research in meta-science. And it should not be restricted to

---

[24] To mention only a few – and more can be added: contributions to science, break through results, grants, contributions to collaborations and open science, societal impact and valorization in relation to various societal stakeholders, mentoring and supervision of PhD students and other early career researchers, collaboration skills, contributions to peer review, committee work and organizing conferences, managerial tasks, etc.

bibliometric indicators but use more data sources, although bibliometric data enable the development of new indicators for researcher qualities, such as researcher's independence, interdisciplinarity, and cognitive mobility[25]. More in general, there is no dichotomy of qualitative and quantitative assessment, as a mature evaluation system should be engaged in developing ways to measure qualitative criteria, to be able to compare grant applications and applicants as objective as possible.

The EC as the by far largest research funder in Europe is in the position to take the lead in defining a broad set of selection criteria, in co-developing ways to operationalize those criteria, and in experimenting with ways to organize the evaluation processes. This would be a useful addition to the work done at the level of policy development and cooperation.

---

[25] Deliverable D4.1; Mom, Van den Besselaar, Möller (2023) Determinants of cognitive mobility. *Proceedings of ISSI 2023*

Policy recommendations

Laura Cruz-Castro & Luis Sanz Menendez (CSIC)

1. Results from the experimental approach (Randomized Field Experiment) in a Spanish funding Agency (Cruz-Castro & Sanz-Menéndez, 2023) found no significant differences in the evaluations of proposals of male and female principal investigators; moreover female and male reviewers differ in their assessments, but these effects are not in line with the matching hypothesis (or with the claim that reviewers hold same gender preferences, or make gender-role congruity associations). These findings provide some policy implications regarding the rationale for increasing the number of female reviewers on the panels (a development that is the consequence of the gender policies in many countries), as a way to increase female funding success rates or their ratings, as we found a lack of effect of the gender composition of committees on the number of successful female candidates, in line with what has been found in other research reporting evaluation procedures for research hiring and promotion. However, there might be other reasons to consider quota practices from funding agencies relevant.

2. Contrary to an often-heard argument, experimental results in Spain also show that in calls where evaluation criteria are well defined and structured, and where interpretation is associated to quantitative past performance indicators, female PI led proposals are not disadvantaged in the evaluation scores.

3. Lack of significant effect of gender in evaluation could be a signal of long lasting changes in the science system or indirect effects of the visibility and salience of the gender equality policies in research, as the policy instrument analyzed explicitly mentioned in the call the commitment to gender equality in science and academia of the Funding agency, and this may have had a moderating impact on a potential gender effect, possibly linked to socially desirable behaviour or rational adaptation to a changing policy environment; from a policy perspective, this could imply that expressing commitment to gender equality policies in the calls can have a positive influence on evaluators .

Recommendations

Ulf Sandstrom (FPS)

This reflection pertains to the functioning of council work, incorporating both remote and in-person panel configurations.[26]

Serving as an expert in a specific research field necessitates the development of knowledge concerning the disparities among various publication venues. Throughout the evaluation process, which can often be time-consuming and represents an additional effort within the scientific community, the primary challenge lies in assessing the research proposal, where careful scrutiny is consistently required.

Nonetheless, for the evaluation of merits, there are shortcuts that enable a more concise and expedited assessment. This involves using an estimate of journals or conference proceedings to identify the few that exhibit superior publications compared to the others. This can be accomplished by simply perusing a list of publications, and experts do not necessarily need to rely on databases to gauge an applicant's merit based on their publications. Consequently, a bibliometric indicator such as the sum of normalized citation scores ($\sum$NJCS) is highly correlated with the merit assessment by experts.

My conclusion is that there is room for improvement in the panel peer review process, which could include the incorporation of bibliometrics for merit assessments. Furthermore, given the availability of size-dependent, field-normalized bibliometric indicators, as well as percentile models, I propose that European research funders embark on a five-year initiative to experiment with a combination of peer review and bibliometrics. One of the objectives of these experiments could be to develop a model that integrates information from each field, grounded in the overall scientific community's recognition and response.

The most advantageous aspect of such experiments is the potential to envision a future for science where researchers are not compelled to make false promises about their results in advance. They should no longer be coerced into initiating and working on projects they do not genuinely believe have the potential to endure. Instead their funding would rely on their former performances.

---

[26] It draws from my involvement in the Granted project, encompassing three key contributions: 1) a Swedish career study spanning over 40 years with about 3,000 cases; 2) an analysis of approximately 800 applicants for regular project grants to the Swedish Medical Research Council (MRC) in 2020; and 3) the examination of three funding programs financed by European funders in 2020 and 2021, including FWF, SFI, and SRC with about 600 applicants and 500 reviewers.

Conclusions and recommendations from WP6.1 and WP6.2: Practices in Panels

Helene Schiffbänker, Angelika Sauer, Florian Holzinger, JR;

Helen Peterson, Liisa Husu, Örebro University

Studying[27] work in panels (also called boards, expert teams or councils) qualitatively enabled us to understand what is going on in panels, how and when gender is or is not addressed and how this relates to gender bias. The main conclusions are:

1) ***Panel have different roles in the assessment process***: *Panels are not panels* – they differ between RFOs in respect to the role they have within the whole assessment process including decision-making. We need to be aware that RFOs have very different assessment procedures in place. Some panels assess proposals (rating, or rating and ranking), other panels assess the quality of remote review(er)s, and also intermediate forms exist. We also learned that some RFOs are continuously (slightly) adopting the role of their panels in order to improve the assessment process. E.g. in SFI, the remote reviewers are now the ones who are assessing excellence and are supposed to apply newly implemented formal policies – such as the 'holistic' concept of excellence or the gender dimension in research – in their practice. Overall, remote reviewers become more relevant in the assessment process, often they are the ones assessing new policies. Especially, if remote reviewers have greater decision-making power and panels serve more as control bodies for them, RFOs should place a greater focus on remote reviewers and their assessment work regarding discussions and analysis of gender bias.

2) **Implementation in panel practices**: Our data indicate that formal assessment policies are not necessarily applied in practice. More precisely, panellists and reviewers only partially apply the general and gender equality policies of the RFO in place for assessing applications. The challenge to bring policies into practice, the need for more reflexivity and awareness on policies, has been raised by emerging as well as by advanced RFOs. To mitigate bias, the RFO might have a strong interest in changing this behaviour.

3) **Excellence**: Excellence is still difficult to capture and to assess. To guarantee (gender) fairness, accountability is needed in applying the same standards to all applications, making the panel chair or observers (from the RFO or external) responsible for this in panel meetings and obliging the RFO to provide some support structures. It could be observed that funders increasingly ask applicants to integrate a gender approach in their research design, research questions and analysis. This means that remote reviewers and / or panel members are asked to assess how gender is addressed to increase research excellence. Our data illustrate that most panel members and remote reviewers did not understand this policy and mostly link it to the representation of women in research teams. Thus they need a better understanding of how gender is inscribed in various research topics and how gender aspects contribute to more excellent research. Furthermore, they need more gender awareness to be able to recognise and assess gender

---

[27] Our conclusion build on the following data:
- 104 Interviews conducted with different actors of the assessment process of 5 selected RFOs: RFO management, RFO staff members, remote reviewers, panel chairpersons, vice panel chairpersons and panel members, and
- Observation of 5 panels in 2 RFOs, 1 onsite, 4 online

relevance in the applications they read. Building these capacities is suggested for remote reviewers (as well as for panel members). There are international resources for this purpose, which could be made more known to RFOs, such as the Gendered Innovations Portal and the GenPORT portal. RFOs are recommended to collaborate in offering capacity building activities and in aligning formal policies including assessment indicators. Further activities to broaden excellence are related to DORA and CoARA, yet it will take time and awareness raising to gain more support for narrative CV formats, qualitative assessment in stead of h-index. To align ongoing developments for advancing assessment standards it could be an idea to establish an international forum where RFOs and further stakeholders (e.g. RFO umbrella organisations, CoARA) reflect and share their mutual experiences. This forum could be established as prestigious meeting (as it is the World Economic Forum for business leaders), organised every year or every second year and providing recommendations how to enhance standards that take gender and intersectional aspects into account.

4) **Increasing the number of female applicants and female grantees**: RFO have implemented policies that explicitly strive for (more) gender-balanced funding outcomes. To achieve this, proposals are re-ranked, aiming for a defined distribution of male and female grantees, either the same number of grantees per gender or the same share of successful applicants per gender (equal success rates). The overall RFO aim is to make sure that a guaranteed number of female grantees is seen as a signal and incentive to encourage female researchers to apply and thus increase the number of female applicants. When studying the implementation of these policies in practice we learned that the focus on numbers, on a gender-balanced *outcome* might limit the focus on the *process*. Having in mind the number of women to be funded might limit attention to and awareness of gender in all other aspects of the panel meeting, like gender-neutral language, the role of care responsibilities in respect to publications, the fact that excellence criteria can be gendered or the way the gender dimension was integrated in the content of research proposals discussed. Very generally, a focus on gender-equal approval rates has the potential risk to narrow down the understanding of gender merely to the number of women grantees, so to the representation of women. Taking gender into account in the assessment process is not really relevant as the outcome at the end is corrected anyway. Here, we should be aware, from a gender fairness perspective, that the selection of applicants who have made it to the ranking list – which serves as basis for the re-ranking – could be done with limited gender awareness and gender fairness and thus would be gender-biased. One of the main conclusions is that focusing only on gender-balanced outcomes cannot balance out a gender-insensitive assessment process.

5) **Resistances**: Understanding the reasons can be a step to change. First, policies exist, but are not known by the panel members and remote reviewers. Another possibility might be that policies are not noted or they are known, but panel members and remote reviewers do not look into them in detail, because they assume that they know what they are about. The interviews revealed that due to their various obligations, time of panel members and remote reviewers is very limited and they do not read all instructions they receive from each RFO. So it easily could happen that gender equality guidelines are not checked in detail or at all. Some reviewers have resistance against instructions from RFOs and expressed that their scientific competences and experiences are undermined by such instructions. Some argued that they are experienced researchers, and also experienced evaluators as reviewing scientific work is a core element of their professional life, reviewing constantly students' work and applications in national and

international RFOs. This way they develop their 'individual understanding' of how to practice reviewing, what to look at and what to take into account. Here the questions arise to what extent reviewers are willing to become well informed or 'educated' and how existing and/or new policies can be designed so that they have an impact on the behaviour of reviewers and panels.

6) **Building capacity**: Our fieldwork shows that applying formal RFO policies in practice does not only require awareness, but also skills, competences and knowledge. When RFOs adjust their policies or implement new ones, evaluators are requested to rethink their current concepts and to adjust their (assessment) behaviour accordingly. RFOs should encourage evaluators in that and provide effective and attractive capacity building activities.

7) **Building capacities on GiRI** (Gender-in-Research-and-Innovation): RFOs now ask applicants to incorporate the sex and gender dimension into research and innovation content, following EC requirements. When studying how this is assessed by remote reviewers and panel members, we found a lack of clarity and understanding for the concept. Gender in research content and gender balance in research teams were mixed up. Moreover, most reviewers do not understand the aim of this policy and how GiRI contributes to better research and more scientific excellence. Our data confirm that reviewers sometimes did not see any gender relevance or argued that gender is not the focus of the project. Recognising the relevance of gender in the very specific research fields – for example, climate change – requires awareness about how sex/gender impacts the generation of knowledge and innovation. It is obvious that remote reviewers need capacities to assess whether and how GiRI was implemented in the various proposals, not confusing it with the representation of women. For building these capacities, formats provided so far are guidance, guidelines, trainings, personal coaching, open access software and best practices structured by thematic fields or an expert pool of gender scholars for collaboration. Further knowledge on GiRI is currently developed by EU-funded projects like INSPIRE or GENDERACTIONplus and should be made available for reviewers.

8) **Intersectionality**: So far, most RFOs only distinguish between female and male candidates, collecting only binary gender data. Yet all members of the research ecosystem need to move towards addressing further inequalities in addition to gender, as focusing only on gender does not capture all forms of exclusion and biases in science. Further relevant categories as age, ethnic background, socio-economic status, native language or host institution, disability etc. often intersect and create multiple marginalisation. For mitigate bias in research funding, an intersectional approach takes these overlapping dimensions into account and enables us to better understand structural barriers and the needs of different subgroups. More research is needed in this direction.
So far, intersectionality is only emerging slowly at the level of RFO policies. Further efforts in this direction are needed in the near future, for example well-constructed guidelines for reviewers and panel members. For attracting a broader range of (potential) applicants from other than native cultural backgrounds, from other universities than the main grant-holders, it might be supportive to address them specifically and to make clear that contributions from very different researchers are appreciated.

Conclusions and recommendations based on WP5, *Formal policies related to grant allocation*

Liisa Husu & Helen Peterson (ORU – lead beneficiary, WP5)

Mapping national policy regimes in five different countries and mapping the specific policies in the five core-RFOs qualitatively enabled us to provide a background of how gender equality policies and policies related to risks of bias are implemented in practice in the grant allocation processes in the selected RFOs (see WP6).

The main conclusions[28] are:

1) *Policies are continually developing: During the nearly five-year runtime of GRANteD we observed how all five RFOs were* continuously developing further their policies and strategies related to gender equality. *The RFOs thus demonstrate characteristics of learning organisations in terms of equality and fairness of the research funding process.* One of the RFOs for example implemented its first ever gender equality plan during the GRANteD runtime. Another RFO completely overhauled its' assessment system. Other RFOs have implemented smaller but significant changes in their policies and practices during the project. These changes meant that the details of the policy analysis performed within WP5 quickly became to some extent outdated. However, the main results produced in D5.1, concerning gender bias risk analysis and the identification of risk areas for gender bias, are of a more general concern and of relevance for also other RFOs.

2) **Important drivers for policy change and progress:** Our data indicate that there are important drivers for change towards more equality and fairness of the funding process on regional level, national level, EU level and global level which impact the RFOs. Research intensive national contexts and overall societal emphasis on gender equality are important drivers for this development and change in some RFOs. Other drivers are cross-national collaborations between funders such as the Global Research Council, the Gender-Net Plus Consortium and the German-speaking countries funder collaboration. On the EU level the European Commission has been driving change towards gender equality also impacting the national research funding landscapes across Europe – for example through the integration of a stronger emphasis of gender equality and the gender dimension in ERA, Horizon2020 and Horizon Europe.

3) **Gender equality and equality as an important quality aspect:** One of the most important drivers for change towards gender equality that we identified is the conviction and understanding within RFOs that gender equality and equality is an important quality aspect in research, with the closely related conviction that gender bias is a threat to quality in research. This is what motivates, RFOs to continually develop and update their policies in this area more generally, and to continue to give priority to these aspects. It is also a conviction that further drives the development of specific policies regarding gender in research and innovation (GiRI) focusing on the integration of the gender dimension into research content –policies that are becoming more and more widespread in the research funding ecosystem.

---

[28] The conclusions, which have already been reported on in deliverables 5.1 and 5.2 and are here summarized, build on analyses of the following data: (i) Over 80 different documents and webpages from European, governmental, and national agencies as well as of the five RFOs, including national science policies, gender equality plans, call texts, guidelines for reviewers and applicants and annual and topical reports. (ii) Interviews conducted online via zoom with staff members in the five RFOs: a total of 33 interviews.

4) **Policies on three different levels:** The policies adopted by the five RFOs can be categorized as belonging to three different approaches to gender equality. *Fixing the numbers*: policies with a typical quantitative focus on increasing the number of women among applicants, decision-makers, reviewers, and grantees. These policies concern equal representation, gender diversity and gender balance in scientific boards and review panels. These policies are crucial for addressing unconscious gender bias in decision-making. *Fixing the organisation*: policies that primarily target the peer review process and the panels and practices during the assessment and evaluation phase. These policies, including detailed guidelines, re-ranking, and quotas, address the risk of bias appearing during panel discussions. *Fixing the knowledge*: policies that concern both reviewers' and applicants' definition and understanding of excellence. These policies emphasise the importance of including a sex and gender dimension into research design and content – in order to avoid gender bias and gender blindness in research.

5) **Advanced and emerging RFOs**: Our analyses showed that the five core RFOs in GRANteD could be characterised as belonging to three different groups with regards to gender equality policies: 1) RFOs with advanced and complex policy packages (SRC & FWF), 2) RFOs with more recent but quickly emerging commitment to gender equality (SFI) and 3) RFOs with emerging and very recently developed policy packages. The inclusion of this mix of different RFOs improved the analyses in WP5 and made a more nuanced analysis possible, contributing to a better understanding of the policy development process and the importance of the national policy regimes.

6) **Silence about gender equality**: The bias risk analysis highlighted that the most considerable risk for bias in RFOs is related to a policy silence around gender issues as that leads to a lack of some of the key features of learning organisations: when there is no strategy, no policy, no structures, no measures, no statistics, no accountability.

7) **Intersectionality and diversity**: So far, most RFOs use a binary understanding of gender, and most policies only distinguish between female and male applicants and candidates; this also means that RFOs are only collecting binary gender data. Addressing intersectionality taking into account other relevant aspects such as age, nationality, ethnicity etc. is only emerging in the policy development of the RFOs studied.

Based on these WP5 results, we propose the following recommendations, targeting RFOs policy development:

1) **Monitoring and evaluation:** RFOs need to regularly monitor and collect gender statistics on success and application rates, report them by funding instrument and scientific field, as well as produce time series on these.

2) **Transparency:** RFOs need to regularly publish gender statistics in their Annual reports and other outlets and accounts of the results of their gender equality measures.

3) **Risk bias analysis:** RFOs should perform regular bias risk analyses of their own processes and outcomes of these processes. The GRANteD bias risk checklist can be used for this purpose. Based on these risk analyses, tailor made policies should be implemented to address these risks.

4) **Funding cycle:** RFOs need to acknowledge risk areas for gender bias throughout the funding cycle and the whole funding process and address all these different areas in their policies.

5) **Knowledge production:** RFOs should develop their knowledge production on gender equality in research funding and in the scientific community more generally. The core RFOs in GRANteD showed example of this through, for example, producing GE observation reports in panels (SRC), or the Polish National Science Centre conducting the first Polish survey on gender in science.

6) **Responsibilities and accountabilities:** The responsibilities and accountabilities for implementing gender equality policy measures need to be clear within RFOs. RFOs need established Gender Equality teams responsible for the follow-up of implementation and update gender equality policies, and these need to be reporting and linked to the leadership of the RFOs.

7) **Communication and dissemination:** RFOs need to put efforts into communicating and disseminating their gender equality policies: both internally and externally to the scientific community and general public.

8) **Intersectionality and diversity:** All members of the research ecosystem need to move towards addressing further inequalities in addition to gender, as focusing only on gender does not capture all forms of exclusion and biases in science. Further relevant categories as age, ethnic background, socio-economic status, native language or host institution, disability etc often intersect and create multiple marginalisations. For mitigate bias in research funding, an intersectional approach takes these overlapping dimensions into account and enables us to better understand structural barriers and the needs of different subgroups. More research is needed in this direction, as well as dialogues between funding organisations on relevant policies and measures. So far, intersectionality is only emerging slowly at the level of RFO policies. Further efforts in this direction are needed in the near future, for example well-constructed guidelines for reviewers and panel members. For attracting a broader range of (potential) applicants from other than native cultural backgrounds, from other universities than the main grant-holders, it might be supportive to address them specifically and to make clear that contributions from very different researchers are appreciated.

9) **Further research**

In addition to policy mapping, analysis of potential gender bias risks and analysis of policy implementation conducted by GRANteD, more knowledge is needed in several areas:

a) Further research is needed on the effects of diverse policy measures. Implementation of policies is complex and always context dependent. Further research is needed to explore the effects of diverse gender equality measures and their implementation, using time-series and both qualitative and quantitative methods, that are sensitive to the dynamics of the implementation environments. This kind of research should preferably be conducted in a comparative framework.

b) Further research is needed to support RFO policy development drawing on intersectional approaches. This kind of research should be conducted as co-creation with RFOs.

c) Further research is needed on challenges and opportunities of policy development in research landscapes where gender equality policies are only emerging or even opposed as part of the growing anti-gender movements.