GRANT ALLOCATION DISPARITIES
FROM A GENDER PERSPECTIVE

Identifying gender bias and its causes and effects

Partner	Authors and contributors	Contributions
TMC (WP leader)	Peter van den Besselaar, Charlie Mom	Authors and editors of the report. Integrating and coordinating the work and contributions. Developing the analytical framework and study design. Co-authors section 6.2, 6.5, 6.6 and 6.8. Author Annex 1.
CSIC	Laura Cruz-Castro, Luis Sanz-Menéndez	Authors and editors of the report. Developing the analytical framework and study design. Co-authors section 6.5 and 6.6.
DZHW	Stefan Hornbostel Torger Moller	Developing the heuristic model. Co-author of section 6.2 and 6.6. Commented on the text.
FPS	Ulf Sandstrom	Author section 5.5. Co-author section 6.6 and 6.8.
JR	Helene Schiffbaenker, Florian Holzinger	Co-author of section 6.1, 6.5, 6.7, and Annex 2. Commented on the text.
ORU	Liisa Husu	Author section 5.4. Co-author section 6.1, 6.5, 6.7 and Annex 2. Commented on the text.

DELIVERABLE 2.1

OCTOBER 2020

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:	50 months
Work Package:	WP2
Due date of deliverable:	Month 11
Organisation name of lead contractor for this deliverable:	TMC
Other partners involved:	JR, ORU, CSIC, DZHW, FPS
Document version:	v06
Review of document version V4:	Luis Sanz-Menéndez and Helene Schiffbänker Suzanne de Cheveigné and Jacques Mairesse (SAB members)
Review of document version V5:	PO and external reviewer (Silvia Gómez Anson).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Abstract

Grant selection and decision-making generally take place in panels. Therefore, the main focus of the project is to measure the level of gender bias at the panel level, and to explain the differences in gender bias using characteristics of the panels and of the (organizational) context of the panels. A second source of gender bias addressed in the project is the application process. The project also investigates the factors influencing the decision to apply for research grants, and the possible gender differences emerging from that. Finally, the question is addressed whether possible bias in grant allocation translates into gender bias in the subsequent academic career. This report develops the conceptual framework and the study approach that should enable to clarify prevalence, causes, and (career) effects of gender bias in grant allocation.

Keywords: Research grants; panel review; peer review; gender; gender bias; inequality.

History of the elaboration of version 5:

In this deliverable, the approaches and models introduced in the proposal are updated, and the same holds for the various work packages and their relations. The first version of the deliverable was discussed in the Berlin meeting in September 2019, in the months after that, and in the meeting in Madrid at the end of January 2020. These discussions have made clear that the GRANteD project integrates a quantitative approach and a qualitative approach. The qualitative approach will collect data about the panel processes (using interviews and observations),about the attributes of panellist and applicants, and about the evaluation procedure (usingthe written materials / the evaluation forms). This may result in new gender processes not yet covered in the literature. However, in the previous version 4, these two approaches were separated, whereas in the new version it is much more integrated.

New in this version:

- The text is radically reorganized in order to integrate the treatment of the existing casesand the new cases (the old chapter 7 has been integrated in chapter 6).
- The qualitative and quantitative approaches are now integrated (in chapters 5 and 6), the main request of SAB reviewer 1.
- We now explain at various places how the analyses will be implemented, the main request of SAB reviewer 2.
- Examples emerging from the analysis of the ERC data have been removed.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Authors and contributors:

Name	Organisation	Contribution
Peter van den Besselaar **Charlie Mom**	TMC	Authors and editors of the report. Integrating and coordinating the work and contributions. Developing the analytical framework and study design. Co-author section 6.2, 6.5, 6.6 and 6.8. Author Annex 1.
Laura Cruz-Castro **Luis Sanz-Menéndez**	CSIC	Authors and editors of the report. Developing the analytical framework and study design. Co-authors section 6.5 and 6.6.
Stefan Hornbostel **Torger Möller**	DZHW DZHW	Developing the heuristic model. Co-author of section 6.2 and 6.6. Commented on the text.
Ulf Sandström	FPS	Author section 5.5. Co-author section 6.6 and 6.8.
Helene Schiffbaenker **Florian Holzinger**	JR	Co-author of section 6.1, 6.5, 6.7, and Annex 2. Commented on the text.
Liisa Husu	ORU	Author section 5.4. Co-author section 6.1, 6.5, 6.7 and Annex 2. Commented on the text.

Dissemination level	
PU	Public

Consortium: The GRANteD consortium consists, after Janury 2020, of six partners:

JOANNEUM RESEARCH Forschungsgesellschaft mbH (JR) (Austria),

OREBRO UNIVERSITY (ORU) (Sweden),

AGENCIA ESTATAL CONSEJO SUPERIOR DEINVESTIGACIONES CIENTIFICAS (CSIC) (Spain),

DEUTSCHES ZENTRUM FURHOCHSCHUL- UND WISSENSCHAFTSFORSCHUNG GMBH (DZHW) (Germany),

TERESA MOM CONSULTANCY BV (TMC) (Netherlands)

FORSKNINGSPOLITIK SVERIGE AB (FPS) (Sweden)

To address questions and comments please contact: Prof Peter van den Besselaar, TMC, Middenweg 203, 1098 AN Amsterdam. Email: peter@teresamom.com

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Table of content

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

1. Introduction

The aim of this deliverable is to further develop the research framework into an operational version, describing in more detail what will be done and how and where the parts of the project come together. The original framework that is in the GRANteD proposal remains the frame of reference and is not included or summarized here again. More precisely, this deliverable translates the framework into more concrete designs and methodologies, and this will be input for (decisions about) the planning of the work.

Why do we need an updated version of the research design? As promised in the proposal, the project design has to be updated using the results of WP1, the literature review (D1.1) and the mapping of Research Funding Organizations (D1.2). This is needed as the project should go beyond the current state of knowledge about gender bias in grant allocation and about the effects of (gender bias in) grants on differences in academic careers of men and women. Deliverable 1.1 informs comprehensively about where the current research challenges are.[1]

The overall scientific aims[2] of the project are to (i) clarify the concept of gender bias (in contrast to gender gaps) in grant allocation and careers, which requires distinguishing merit-based and policy-based selection criteria from gender stereotyping and other particularistic criteria; (ii) identify occurrence of gender differences and gender bias in the two different phases: the application phase leading to the decision to apply for funding (gendered self-selection and gendered access to support available when preparing the application); and the selection phase (gender bias in the decision-making process); (iii) determine the impacts of (gender biased) funding on careers of male and female researchers; and (iv) investigate the causes of the various forms of gender bias.

This can be broken down into several research questions, where the panel dynamics is the focus, as there the main (preparations for) the decisions are made. However, also other phases will be studied as the model presented later in this report shows: the application phase and the post-grant career.

What data need to be collected in order to answer these questions is described in detail in the Annex to this deliverable.The basic idea of the research approach is simple, but the resulting project is complex.

[1] Cruz-Castro, L. & L. Sanz-Menéndez (2019). Grant Allocation disparities from a Gender Perspective: Literature Review. Synthesis Report. GRANteD Project D.1.1 http://dx.doi.org/10.20350/digitalCSIC/10548

[2] From the Grant Agreement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

2. What do we know about Gender disparities and how to advance knowledge

Allocation of grants and potential gender bias occur in the context of social interactions between panel members, and between applicants and evaluators. The space in which the interaction occurs is institutionalized by the RFO, usually, at the panel level. Previous research has usually paid limited attention to the proper level of analysis of the potential gender bias: the organizational level. However, there is some earlier research suggesting that if gender bias occurs, it is in the decision-making processes in the *panels*, in the interaction between panelists and where attributes of the applicants and panelists play a role. Research suggests different levels and directions of gender bias between panels in the various disciplines. There are panels where gender bias is against women, but there were also gender-neutral panels, and panels where women profit from gender bias.[3] The question then comes up of why these differences occur between the panels within one organization? Is this the result of a random process? Or are there structural conditions at the panel level that determine whether bias exists or not, and in what direction? What panel and panelist characteristics make panels biased or neutral? Within an RFO, it cannot be the organizational context as that is the same for all panels, but when comparing RFOs, this may be important. It could also be an effect of different disciplinary culture. Therefore, the Granted project focuses on the *panel level* of analysis, on the space where interaction between applicants and panelists takes place; and more specifically, on the panels and their decision making processes. The larger organizational characteristics come in when comparing the various RFOs.This implies for the GRANteD project that the new case studies to be developed in five RFOs will focus on the panels, to get enough information about the interaction between applicants and evaluators, in as many panels as possible, to reduce the effects of measurement errors.

As far as the direct *grant selection phase* is concerned, the task of the project (and this deliverable) is to develop an approach for (i) measuring the prevalence of gender bias at the panel level, taking merit of the applicants into account (WP4), (ii) comparing panels in terms of processes, practices, structures, and organizational context (WP4, WP6), and (iii)comparing the RFOs in their broader national context, including the various gender equality policies at national or organizational level (WP5).

The aim of this approach is toidentify the causes of gender bias, and to identify which gender equality policies are effective. And, the model that follows from this can be applied on the existing cases, and more comprehensively on the new cases, as here data collection will be based on the approach.

With respect to gender equality policies, these may have an effect at the panel level or not, depending on several factors: The effects of policy implementation are an empirical question. A policy may be implemented in different ways, even within the same organization and national setting, which may affect the effectiveness of the policy.[4]Of specific importance is the question about which the local implementation of rules

[3]Van den Besselaar P, Schiffbaenker H, Sandström U, Mom, C, Explaining gender bias in ERC grant selection. STI 2018 Conference Proceedings, 346-352

[4]Husu and Callerstig 2018.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

are,since they could explain thedifferent effects within their local context. If the new cases show enough variety in terms of the gender policies, we may investigate which of them are effective to create a less biased and more merit-based decision making about grants.Policies, by the way, may not only focus on gender equality within the RFO, but also focus on what happens before and after selection processes: the application phase, and the academic career.

The *application phase* is here seen as the period in which potential applicants decide whether they do apply or not. A common observation in earlier studies is that compared to the gender composition of the population of *potential* grant applicants, women are underrepresented in the population of applicants.[5] This suggests that women are less likely to apply for grants than men do. Why is that the case? What work and what private life related factors do play a role here? The issue is even more important as the gender differences that result from the application phase may be much larger than those that emerge during the panel selection processes. If so, that would also have serious implications for gender equality policies, and the loci where these policies should focus on. This is addressed in WP7.

A last crucial question relates to the effect of inequality of grant allocation on the *academic careers* of men and women. Receiving a grant may influence the career, and if there is gender bias in grant decisions, this may lead to gender differences in the subsequent careers of women and men. There are many more factors that influence the direction and speed of careers, and these factors themselves may be influenced by the career. The resulting vicious circle[6] of grants, research projects and their output, and academic positions leads to challenging questions for the research design.[7] To what extent grants influence the careers is the topic of WP3.

Qualitative and quantitative data will be collected for the new cases, and these data will be used in a series of (statistical) analyses at the panel level. On top of that some form of case comparison will be done to find out what contextual factors play a role. A previous version of this deliverable was submitted in April 2020, and the current (revised) version elaborates the model and its implementation in some more detail.

[5]Dickson D (1997) Female scientists wanted.*Nature***390**, p431

[6]Van den Besselaar P, Sandström U, Vicious circles of gender bias, lower positions and lower impact: gender differences in scholarly productivity and impact. *PlosOne***12** (2017) 8: e0183301. https://doi.org/10.1371/journal.pone.0183301

[7]Mairesse J, Pezzoni M, Visentin F (2019) The impact of family characteristics on the gender publication gap: evidence for physicists in France. Interdisciplinary Science reviews 44, 204-220; Van Balen B, van Arensbergen P, van der Weijden I, van den Besselaar P, Determinants of academic careers.*Higher Education Policy***25** (2012) 313-334

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Methodological considerations

In deliverable 1.1[8], the state of the art in the literature has been summarized. This report shows that a lot of research has already been done on gender bias in grant allocation and the possible gender biased effects on academic careers. In order to move beyond the state of the art, the project has chosen for a broad, model based approach in order to find out the prevalence and causes of gender bias in grant allocation. The objectives related to this are to identify:

- what factors influence the (different) application behavior of men and women
- whether and where gender bias occurs in the grant allocation process,
- the causal mechanisms that produce gender bias in the allocation of grants (in the one or the other direction),
- the (contextual) mechanisms in which grant applications are submitted and decided on,
- the causal mechanisms that link grant allocation with different career opportunities.

In order to realize this, we will heavily rely on the state of the art in the relevant literature. The first approach is multi-theoretical, implying that mechanisms are used from a variety of theories. Firstly, the existing cases and the new casestudies are expected to enable such a strong multi-theorybased analysis at the panel level, as well as a comparison between the different funding organizations. Secondly, the applicant survey and other (existing) datasets will improve our knowledge about the application process, and possible gender bias that may emerge there. Thirdly, a longitudinal dataset will be created to study over time the effects of grants on careers, and whether this works different for men and women.

A strong multi-theorybased approach has it benefits, but may also be blind for uncovering new phenomena that have not yet studied. Therefore, an explorative approach deployed using interviews, observations and document analysis to develop *new sensitizing concepts* to understand the risk, contextual background including gender equality policies, possible forms of and dynamics of gender bias, including gender bias within the panels, among reviewers and within the RFOs. The explorative approach should lead to a *new understanding* of the processes leading to gender bias, where gender bias may exist, under what kind of conditions, and how RFOs approach the issue. This is done mainly through investigating how the actors perceive and interpret the situation.

As is well known, qualitative explorative approaches have their value in providing new and deeper understanding of phenomena, which can be used for further hypothesis development. At the same time they have limitations in terms of internal and external validity and reliability. But when the exploration is successful, this will lead to new sensitizing concepts that could enrich the theory based models. The qualitative analysis

[8]Cruz-Castro, L. & L. Sanz-Menéndez (2019). Grant Allocation disparities from a Gender Perspective: Literature Review. Synthesis Report. GRANteD Project D.1.1 http://dx.doi.org/10.20350/digitalCSIC/10548

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

is based on Acker's theory of gendered organizations (TOGO)[9], which argues that organizations are gendered, and that organizational structures and processes are not gender neutral. Acker's theory will be used as an analytical tool to explore the multiple and complex ways organisations are gendered and how gender inequalities are produced, reproduced (but also resisted) in practice within different organizational processes. The TOGO aims to identify common mechanisms which - across organizations - produce bias towards women along five different analytical dimensions which are in practice intertwined: (i) Gender segregation and divisions, (ii) gender interaction with the organization, (iii) gendered culture, (iv) gendered identities, and (v) gendered organizational structures.It should be noted, that the interviews and observations are also used for data collection within the framework of the multi-theoreticalapproach, especially where the focus is onpanel processes and practices.

How far we can get depends heavily on the data that can be collected for the new cases – especially data on the panels (and panelists). To make this all possible, the RFOs that will be studied in the new casestudies should provide the project with the data as specified in the data template in the Annex to this report: Review reports and panellist /reviewer scores; project text and CV;and panelists' characteristics[10].

As the literature review makes clear, we distinguish between an *outcome* (e.g., occupational differences, or grant success differences) and the *mechanism or process* that result in this outcome. Even if we observe gender differences, this does not need to be result of discrimination, but can be the result of neutral mechanisms in relation to relevant group differences. Therefore, gender bias in grant allocation cannot be identified by simply looking at differences in success rates (grants) and differences in occupational structure (careers). If success rates of men and women are equal, the process is not necessarily unbiased, and the other way around, if men have higher success rates than women, the process is not necessarily biased. In Granted, we will avoid using data on success' differences to conclude about the processes of creating inequality, such as bias or discrimination.

If that would be the case, the issue of gender bias would be easily solved, as one only needs to define a base line and then divide the funds proportionally to that base line among male and female applicants. For example, if one takes the gender distribution of the applicants as baseline, one can simply split the evaluation between a male and a female group, and divide the funds proportionally. However, this is only an acceptable solution if one either assumes that men and women are by definition of an equal quality, or that merit does not need to play the primary role in grant allocation.[11]

[9] Acker 1990, 1992, 1998

[10] See the data template in the annex 1 of this report for an overview.

[11] Interestingly, some RFOs seem to take this standpoint: For example, the current gender equality strategy of the Swedish Research Council (approved by the Board in 2014 and available at www.vr.se) states that "the Swedish Research Council assumes that research capacity exists to the same extent in both sexes") But it is of course very important to specify whether one talks about *capacity as potential* or about *capacity as realized.* For grant selection, capacity as potential can be important, especially for early career grants. Whether selections based on potential were correct, can be tested by looking at predictive validity.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

However, GRANteD acknowledges that fundamentally the science system should be merit-driven, but that in practice there are many deviations.[12] The main question then is whether the grant decisions and the decisions on hiring and promotion are based on some form of discrimination – that is neglecting the merit dimension of a person or group partly or fully. So, the proof of whether there is gender bias relates to the selection/decision *processes*. The *result* of those processes is something different, and may also be the result of a biased process but also of a non-biased process of occupational, institutional and field segmentation and segregation, where men and women tend to concentrate in different occupations, institutions, fields, and positions, as apart from the quality of the process, other variables influence the success rate, like gendered individual choices. How merit should be measured, and whether that is gender-neutral will be discussed later in this report.

Using the deliverable D1.1 we select which variables have to be taken into account for analyzing grant application behavior and grant decision making, and the same holds for academic careers. And these need to be integrated into a well-specified (set of) model(s) that steer the analysis.

For advancing our knowledge, the selection of the new cases should be so that they do fit in the strategy formulated in this deliverable and in the mapping of European RFOs (D1.2), to avoid convenient sampling. One issue that comes up within this context, is that, a *comprehensive* model that assesses the impact of grants on careers would need to takeinto account RPOs (Research Performing Organizations) like universities and PROs (Public Research Organizations). Decisions about employment and promotion take place at this level, also mainly through recruitment panels, and those panels value project grants, fellowships, or competitive fund raising in general differently. It might also be the case that those panels are affected by different types of bias (gender, cognitive, institutional) so, analyzing the effect of grants on careers, should, as in the rest of the project, control for past performance, as well as for the possible existence of bias in the panels evaluating careers. However, the complexity will be higher in the sense that organizations could use not only merit assessment but other forms of assessing the worth of the candidate for the organizational missions as criteria for evaluation.

The literature review and earlier experiences has resulted into the basic model described in the proposal. The data-template (included in this report) is a translation into data requirements. In the new cases we may more easily be able to measure the organizational/process/procedure/practices/policies data (which are only partially available in the existing cases), but also here we may see some principal and practical issues. (i) Not getting data on rejected applicants is the most obvious risk in the new cases. Getting the data should be an absolute requirement for accepting new cases. (ii) Another issue with the new cases are the past publication performance data of the applicants. These could be collected from Web of Science or another bibliometric

[12] Note that a strong trust in merit as an organizational principle has in some contexts shown to paradoxically lead to gender and other biases (Castilla EJ & Bernard S (2010). The Paradox of Meritocracy in Organizations. Administrative Science Quarterly, 55 (4): 543–576 ; https://gap.hks.harvard.edu/paradox-meritocracy-organizations).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

database, but that is very laborious as experience has shown. May be the RFOs do have those data in applications, or they may have (Scopus or WoS or Orcid) IDs of the applicants, but also this is uncertain. Above that, the coverage of these databases is disputed for many research fields. The solution is to addseveral items about past publication performance to the applicant survey (WP7). Some exploration on how to improve measurement of past publication performance will be done in Wp4.1. (iii) Merit is not covered by bibliometric data alone, as there are various other signs of merit that could be included. For these, the applicant's CVs are a main source. But extracting data from those CVs is equally resource consuming. So also for these other signs of merit, itemswill be added to the survey.What the relevant indicators are (e.g., awards, grants, collaborations) will be addressed in WP4.1.(iii) For the model, it would be preferable to provide more systematic data on the interaction between applicants and evaluators, e.g., reports onthe panel process made by council officers who participate in the panel.(iv) In order to have a maximum benefit from this multi-method design, it is useful that interviews (and possibly observations) provide some core structured data on panel members and panel processes. WP4,WP5 and WP 6.1/2 will interact about what that implies for the interview protocols, the coding, and the aggregation from individual panelists to panel data.

The new cases: selection criteria

- One criterion is variation between funding instruments, this in order to avoid '(self) selection bias'. The existing cases focus strongly on the 'top instruments', like career grants (Emmy Noether, VENI, Ramon y Cajal). It is important that there are enough cases for 'normal' (often team based) applications (like in H2020).
- A second criterion is that next to interviews, observations and policy documents, also the type of data we have for the existing cases should be provided by the core-RFOs – such as review scores and review reports, proposal and CV, and a variety of data about the structure, demography and membership of the panels. Details can be found in the Annex: the data template.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

3. The Level of Analysis

As emphasized above, the proper level of analysis for GRANteD is the place where the interaction between applicants (in vivo and mediated by documents) and evaluators takes place: The Panel level.

As the number of panels in the new cases is now foreseen to be in the range of a small N,[13]a statistical comparison is difficult). However, we may do a Qualitative Comparative Analysis (QCA) to solve this. QCA, a set-theoretic method developed by Charles Ragin[14] and colleagues, is better suited to analyze situations where causality is conjunctural, and equifinal (i.e. where different causes may combine to bring about an outcome of interest and where there is more than one path to an outcome).In order to proceed beyond the analyses possible in the fourexisting cases as mentioned in WP4, there is a need to have more systematic data in the five 'new cases'.

Additionally, it is a crucial methodological requirement that for assessing the effects of the "gender policies in RFO", we have panels from the new RFOs with very different level of commitment with gender policies; "self-selection" by RFOs to show their successful policies has to be avoided.

Another way to proceed could be an experimental approach based on various forms including randomized controlled trials (RCT). Within an RFO panels can be selected where in one of the two the RFO experiments with a new procedure/practice/structure, and in the other keeps the old procedure/practice/structure. Then we can experimentally test whether certain policies have the intended gender effect or not.

Finally, GRANteD studies also the application phase and the career effects, and in both cases it may be not so much the RFO but the PROs that play the main role. Especially in relation to careers, this would be a useful follow-up project.

[13]The model we decided to apply has a strong focus on the panel level, which is in line with the lessons from the literature review. However, the number of panels we will have access to remain for the moment limited. This has to be solved inT2.4 at the time of Milestone 2

[14] C.C. Ragin, *The comparative method; moving beyond qualitative and quantitative strategies*, Berkeley, University of California Press 1987; Rihoux, B. (2006). Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods: Recent Advances and Remaining Challenges for Social Science Research. *International Sociology*, *21*(5), 679–706.Ragin, C. C., and B Rihoux. 2008. *Configurational Comparative Methods: Qualitative Comparative Analysis*. 1 edition. Thousand Oaks: SAGE Publications, Inc. C.C. Ragin, *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago, University of Chicago Press 2008; C.Q. Schneider, C Wagemann, *Set-theoretic methods for the social sciences*, CUP 2012.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

4. The heuristic model

The term 'model' as usedin the proposal may be somewhat misleading, as it suggests that one overall statistical model will be developed that then is tested on the data collected in the various WPs. But the overall model in fact consists of various middle range models, each describing one or more mechanism that plays a role in the various loosely coupled processes that together form the grant and career system.

Processes related to the structure of the pool of potential applicants:

a. The effect of gender composition of the discipline on the pool of available candidates

b. The effect of the RPO hiring and promotion evaluation practices (and other organizational features) on application behavior.[15]

c. The effect of application behavior and self-selection as source of bias.

d. Do gender differences in choice of research topics lead to different success of M and F applicants?

Related to the grant selection process

e. The effect of *self-presentation* by applicants on the selection.

f. What other individual (merit and non-merit) *characteristics* explain the scores applicants get, and the outcome of the selection process? After controlling for relevant other variables, does gender (directly or indirectly) predict scores and success?

g. The effect of the distribution of women across more or less reputed institutions on gender on scores and success

h. Do other biases like cognitive proximity and nepotism[16] occur? and does this has a gender effect?

i. As gender bias may occur differently between different panels, what *panel characteristics* cause[17] these differences in gender bias. Which of these may be used for intervening?

j. Do gender-related policies of RFOs (such as extension, gender training for panels, monitoring of the panels) have an effect on the scores/decisions?

k. What contextual factors contribute to gender bias in grant allocation (e.g., panel rules and procedures, organizational gender policies)?

Related to careers:

l. Self-selection in careers which may indirectly lead to gender differences in grant application and success.

[15] Not yet in the model (fig 1 below).

[16] One may wonder whether nepotism is the right term here, as the focus is on friendships, professional networks, and organizational proximity - and not on relatives. Therefore cronyism may be the better term.

[17] From a social constructivist perspective, one may prefer "are linked to" for "cause".

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

m. Contextual factors in careers in RPOs and academia: mentoring, support, division of labor (research time available in post), research field, integration in research groups, networks and collaboration patterns, which may indirectly lead to gender differences in grant application and success.[18]

n. Do grants have an effect on the direction and speed of careers, and does that lead to gender disparities?

o. Does the national science funding system affect the value that RPO give to grants in employment and promotion decisions?

As already mentioned above, we are studying the science system and therefore it is obvious that *merit* should play a role in grant decision making. We operationally define merit as *academic performance* which is expected to influence grant and for career decisions. What counts as past performance is of course an issue and we return to that below. But here also a few issues come up:

- Are merit criteria gendered in the way that women have less opportunities for equally good performance scores that men have? And if so, why would this be the case?
- Can we identify and deploy the (possibly field dependent and the panel dependent) merit criteria that would be deployed in a merit driven world, or do we focus on the merit criteria that seem to be deployed *de facto*, even if they are considered inadequate (such as the H-index or the journal impact factor)?
- How does the deployment of different merit criteria relate to the level of gender bias? Who (women or men) would profit from the use of merit criteria, and who from the deployment of reputation criteria?

The following model illustrates the (coupled) processes we are studying. Each of the processes has to be investigated in order to identify the mechanisms that produce (if at all) gender bias. In the model, one can distinguish the following elements: application, decision making, panel processes, council processes, and contextual factors. And for the career part of the model, we find the role of grants, the role of other factors such as preferences, support, RPO decision making.

[18] For an empirical approach to these issues: Van den Besselaar P, Sandström U, Vicious circles of gender bias, lower positions and lower impact: gender differences in scholarly productivity and impact. *PlosOne***12** (2017) 8.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

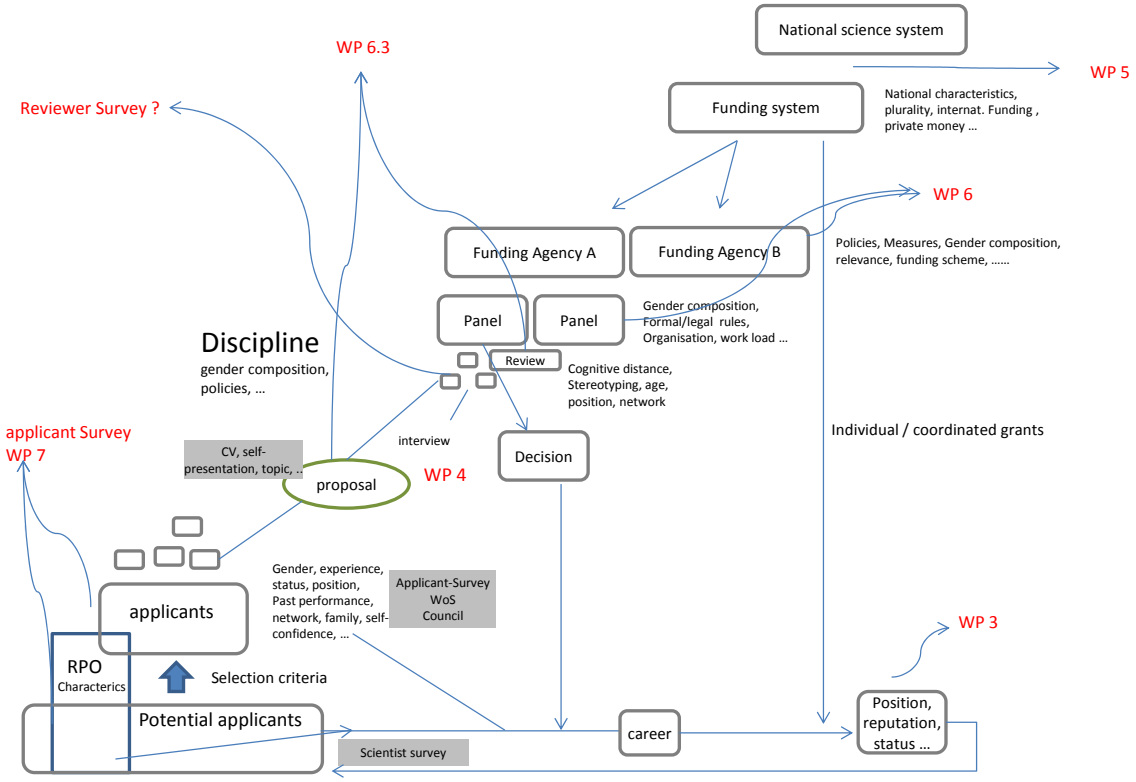


Figure 1: The GRANteD heuristic model (Stefan Hornbostel)

In the next sections, we will discuss the various parts of the heuristic model that also refer to WPs and tasks in the proposal. In section 5, we focus on the research design following from the model, and in section 6 we give details about the operationalization:

- The application phase (WP3, WP4, WP7)
- The prevalence of gender bias (WP4, WP7)
- The self-presentation and self-confidence (WP4, WP6.3, WP7.2)
- The panel practices, procedures, processes; and panelists' characteristics (WP3, WP4, WP6.1/2)
- Organizational policies (WP5)
- Panel contexts (WP5)
- Effect of grants on careers (WP3 & WP4).

The heuristic model is tuned towards the RFO and the relevant processes taking place there. However, for the effect of grants on careers and on output (post performance), we may need a different design/model, focusing on RPOs (Research Performing Organizations), to assess the impact of grants in the probability of advancing careers of men and women. This will be further developed in WP3.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

5. Foundations of the Research Design

5.1 Application phase

In the application phase, decisions are made as whether to apply and where. From an analytical point of view even if we analyze individual decisions we should we aware of the tension between two ideas: Were they pushed or did they jump? Previous research has confirmed gendered patterns of self-selection.

Are there gender differences in application behavior?, and are there gendered processes that may lead to different application behavior and to differences in the entrance situation of male and female applicants?. For example, if male applicants get more institutional support, they may produce better proposals. And if female researchers have more household and childcare responsibilities, they may have a lower past performance when they apply than their male competitors.

It might be analytically useful to distinguish the contextual (structural explanation) that women are under-represented in the type of positions from which they would typically apply,[19] from other type of explanations based on the factors affecting supply (attitudes towards competition, preferences about work-life balance) and demand (role of networks, mentoring, institutional support, gender discounting, etc.).

Several questions have to be answered here.

- *Are women and men differently selective when applying for a grant and why?*
- *Are negative feedbacks from past experiences of application/competition more influential for women's decision to write new applications?*
- *Do men and women select different research areas and topics that may give them a differential access to grants? E.g., if men would select more hot topics and mainstream topics and if women would select more often applied, peripheral, or interdisciplinary topics, this may be a cause for differential grant success – an indirect gender effect*
- *Are there differences in attitude towards competing for grants and jobs, to work-life balance, to family life?*
- *Role of family division of labor on performance and grant success.*
- *Different institutional support or mentor support.*
- *Are women over-represented in positions (part time/fix term/teaching, etc.) from which they are not eligible/expected to apply?*
- *Are women under-represented in more reputed or research-intense institutions and does this affect their self-expectations of success?*
- *Are there gender differences in motivation to apply?*

[19] See e.g., Huang et al. (2020) *PNAS*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

- *Are there gender differences in research group membership, leadership, formation of groups that are linked to application activities?*
- *Are there gender differences in task profiles (teaching, research, admin) that may be linked to application behavior?*

5.2 The prevalence of gender bias

(Gender) bias occurs when other factors than merit comes into play. In the project, as in much previous socialresearch, we take an approach that has been called *sophisticated residualism*, meaning that – as simply as possible - we assume a general model of the following type:

> Male/Female differences in grants (the rewards in this case) = M x (Differences in Merit + Differences in Human Capital + Differences in Social Capital) + Discrimination.

Where **M** stands for **mechanisms** by which the different types of inputs produce differential rewards. Merit could be operationalized as the different forms (or indicators/measures) of past performance; meaning previous differences in results. Human Capital is related to some forms of capabilities of the applicants, while Social Capital factors are much more related with the reputation and the networks. Discrimination is what remains after taking account of Human Capital, Social Capital and Merit.In this model the allocator of rewards are the Panels of the RFOs, but that is the topic of the next section.This model does not preclude that gender differences in scores on in principle correct indicators for merit can be based on discrimination and gender. These differences can be based on cumulative differences in merit accumulation and assessment, assessment of human capital and social capital accumulation, but sees discrimination as a separate "mechanism". (Cf. Cumulative disadvantages, Matheus/Mathilda effects etc.).[20]

Accordingly, in the specific space in the heuristic model, the M factor could have various formulations and specifications of the incentives and mechanisms that translate agents' preferences into rewards. Several existing and new indicators will be used to predict the panel score and application success. Does gender have an effect after taking a series of other relevant variables into account? If so, this will be interpreted as gender bias.

If merit is important, it is necessary to specify what counts as merit in practice, and what should count as merit. In the current proposal, we used the following merit indicators:

- *Publications*
- *Impact*
- *Independence (to be further developed)*
- *Innovative contributions to science (to be developed)*

[20] Bask M, Bask M (2015) Cumulative disadvantages and the Matthew Effect in life-course analysis. PLoS ONE 10 (11); Rossiter MW (1993)The Matthew/Matilda Effect in Science. *Social Studies of Science*, **23** 325–341.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

- *Quality of the project (to be developed)*

Others indicators represent reputation more than merit, but we know that they are used in practice by evaluators, and/or specified by the RFO[21]:

- *Earlier grants*
- *Awards and prizes*
- *Quality of the network*
- *Ranking of university where the PhD was obtained*
- *Impact of the journals published in*
- *Time to degree*
- *Current academic status*
- *Social background, ethnicity, nationality.*

These variables can partly be derived in a standard (but very laborious because of the large number of cases) way from Web of Science and Scopus, and they need technical work to upscale (independence) or to be developed (innovative contributions). The table 2 below shows other merit and non-merit performance dimensions that may be taken into account. This will be further discussed in WP4.

Table 4. Top 10 characteristics of talent in general and concrete talent evaluation[a]

General evaluation	Concrete grant evaluation
Being social (sc)[b]	Publication record (pc)
Acquired grants as previous recognition (pc)[c]	Elaboration research proposal (pc)
Ability to work hard (ic)	General comprehensiveness (pc)
Ambition (ic)	International experience (pc)
Publication record (pc)	Authenticity (ic)
Ability to work independently (ic)	Enthusiasm (ic)
Enthusiasm (ic)	Originality (ic)
Perseverance (ic)	Self-consciousness (ic)
Writing skills (pc)	Ambition (ic)
Being proactive (sc)	Hot topic in research proposal

[a]Characteristics within the same cell were mentioned just as often by the interviewees and are therefore ordered alphabetically.
[b]sc = social capital, pc = professional capital, ic = individual capital.
[c]Although grants can be considered professional capital, interviewees referred to them more directly as symbolic capital: previous acknowledgement of prestige.

Table 2: Selection criteria[22]

[21] Increasingly RFOs specify what should not be taken into account (numbers of publications, or the JIF and the H-index), but they remain generally vague in what should count (e.g., the sole criterion is 'excellence'). All criteria may be open to different interpretations and therefore to bias, and this holds especially for vague criteria.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Of course, merit criteria may be applied in a biased way (different for men and women) and the score on merit criteria can be biased as men and women may have different possibilities to develop the performance (e.g., Matthew and Mathilda effects). We take that into account as this would point at where the gender discrepancies are produced (and where they might be counteracted).

5.3 The panel processes, practices and procedures

If we can detect potential gender bias, the next step is to find the factors that produce this gender bias: what is behind gender bias in selection processes?Returning to the previous formulation regarding the determination of disparities in outcomes resulting from discrimination or bias in the processes, we should advance into two parallel directions.

Firstly, we consider the allocation process as an explicit social interaction process between applicants and evaluators; for that we need to include in the equation a new factor: the panel structure and other characteristics

> Male/Female success rate differences in grants (the rewards in this case) = M x [(Differences in Merit + Differences in Human Capital + Differences in Social Capital) x (Panel Structure)] + Discrimination.

Where Panel Structure stands for the various factors and characteristics of the Panel composition that need to be taken into account too, e.g., the proportion of men and women in the panel.

Secondly, we need to go further into the potential different M (mechanisms) by which the different types of inputs (merit+humancapital+social capital) produce differential rewards; As previously mentioned, the M factor could have various formulations and specifications of the incentives and drivers that translate agents' preferences into rewards.

Now we move into some of those (competing) mechanisms potentially accounting for discrimination. The literature suggests several, but the issue is then to (i) formulate the mechanism, and (ii) operationalize those in a way that we can measure those.

Past research disagrees over the mechanisms, how those mechanisms relate to scientific merit values and which are the effects on the functioning of the overall research system. However, some lines of explanation have been explored with two main underlying ideas: the "preferences" of the actors and the "information" that the actors have. In both cases, actors involved used cognitive mechanisms as shortcuts for decision-making, but in the first case there is room for considering the interest and power involved in the decision, and in the second case stereotypes resulting from "statistical" learning may operate.

[22]Van Arensbergen P, van der Weijden I, van den Besselaar P (2014) Different views on scholarly talent – what are the talents we are looking for in science? *Research Evaluation***23**, 273-284.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

What mechanisms are mentioned in the literature? Without being exhaustive:

- *Gender stereotyping (Stereotyping can be both a preference-based mechanism (taste for a particular group) and a lack of information type of mechanism. The implications for policy are very different.*
- *Gendered criteria (criteria not taking into account gender differences in life course)*
- *Gendered application of in principle neutral criteria (assessing men on their strength and women on their weaknesses)*
- *The level of competition within the panel (measured in terms of overall success rate).*
- *The work load within a panel (the number of applications to be evaluated).*
- *Nepotism and institutional proximity (both are clear examples of preference-based discrimination mechanisms)*
- *Cognitive bias which may have different effects for men and women.*
- *Gender differences in topic choice (e.g., mainstream versus ore marginal topics; new versus established topics)*
- *Sexism: opinions about the role of women in science*
- *What individual practices help us understanding gender bias?*
- *How are the rules of the formal negotiation and decision-making process (WP5) including the role of panel chair applied in practice?*
- *Have any gender-related critical incidents been observed? What hypotheses about different success-rates, performance differences, ongoing discourses, do reviewers or panelists share? Are these supported by empirical evidence from evaluating the case itself or from the literature?*

An explanation could be that in most cultures***, gender stereotypes***exist that link men and women to different occupations. In the science system this means that the position of professor is implicitly associated with men, and not with women. That would in a panel lead to assessing men of higher performance/merit than women, even when they are in fact equal. This can be expected to be an automatic process, which can be stronger under work pressure, and weaker after a training intervention, or under formalized procedures and assessments, where relative scores need to be explicitly argued.

Another explanation could be that the ***composition of the panels*** is important, and that a very low number of female panel members may not be able to counteract stereotyping, and that gender bias will be stronger in that case. An alternative story is that the female panel members had to fight very hard during their career to get where they are and that they therefore have internalized the stereotypes even stronger. This would imply that success of female applicants is negatively associated with the number of female panel members.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

A third explanation could be that the level of ***formalization***of procedures may influence the panel group dynamics, and through that whether men and women are treated the same or not. This may include things as guidelines, monitoring, and possibly mandatory training.

What characteristics of the panels (processes, practices, procedures, language use, composition and structure) can explain the different levels of gender bias we measure at the panel level? This is the core question, as the relevant characteristics may show in what directions one needs to look for strategies to reduce gender bias. Social psychological theory and research has identified a series of mechanisms that may be relevant: Group composition (such as coverage of expertise, acknowledgment of expertise, intergroup status, intragroup social relations; ingroup versus outgroup identity), Group norms, Level of cohesiveness, Motivation, Accountability, Panel membership (prototypical versus peripheral), Information availability and distribution.[23] The list of items to take into account in the interviews is underdevelopmentby WP5 and WP6.1/2 in collaboration with WP4.

As already explained, interviews with RFO staff and with panel members will firstly bring new insights about practices related to the assessment of applicants and to the decision making processes. What panel practices help to explain why panel outcomes may differ? By going into the qualitative data in sufficient depth, we aim to detect new characteristics, and to add empirical evidence to findings from previous research.[24] Secondly, the interviews will provide data for the variables that are included in the model, which contributes to testing hypotheses outlined in the conceptual framework developed in this deliverable. Thirdly, we hope to gain new insights in panel dynamics that can be transformed into additional variables for the model.

5.4 Policies and organizational context

RFOs to be studied in GRANTeD are public funding organizations, regulated not only by legislation but also by national research policies, and by their internal policies, both explicit gender equality policies and also seemingly gender-neutral policies which might have an impact on gender equality and gender bias. These national and organizational contexts of the RFOs in which funding applications and funding decisions take place, will be studied in WP5. Do research funding systems with high research intensity (measured by share of R&D expenditure of GDP, private/public research funding ratio, proportion of researchers in the labor force) show different patterns, in regulations and policies related to RFOs, including gender equality policies, accountability and monitoring, than systems with middle range or low research intensity?

[23] Olbrecht M, Bornmann L (2010) Panel peer review of grant applications: what do we know from research in social psychology on judgement and decision-making in groups? *Research Evaluation* **19** 121-129; Van Arensbergen, van der Weijden, van den Besselaar, The selection of talent as a group process; a literature review on the dynamics of decision-making in grant panels. *Research Evaluation* **23** (2014) 298-311

[24] E.g., Van Arensbergen P, van der Weijden I, van den Besselaar P (2014) Different views on scholarly talent – what are the talents we are looking for in science? *Research Evaluation* **23**, 273-284.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Do differences in gender ratios of researchers by disciplines or fields in the national setting be relevant for understanding gender bias and risk of gender bias, such as differences in the share of women among potential applicants (potential applicant pool), researchers in the system, and at the professorial level (reviewer/panelist pool)?

Do funding landscapes with strong political steering and accountability for gender equality demonstrate smaller gender bias? How do funding agencies understand and deal with the issue of gender bias? How is the political steering on gender equality translated into regulations and policies at the RFO level, including panel composition and construction? The identified contextual aspects coming out (or not) in the panel processes can be traced in the WP6 interviews of the chairs and panelists as well as the observations (e.g. argumentation using contextual aspects, including policy goals). Different policies have been implemented by RFOs to counteract gender bias in grant allocation and to provide equal opportunities for males and females, such as informing panel members about gender bias and gender stereotyping, asking for explicit reporting on the gender distribution, or extension policies in the application phases. Therefore, investigating the organisational context in which these GE policiesare implemented is equally important as analysing the national or international research landscapes. This will provide us with comprehensive data to understand the concrete organisational framework conditions that enable the occurrence or mitigation of gender bias in research funding.

The RFOs and applicants operate in a national and international research landscape and funding regime, including specific regulations and policies guiding the national RFOs, as well as gender equality policy landscape. These will be mapped and analyzed in WP5. There is interesting variation in these contexts within Europe and among the countries we plan to conduct GRANTeD research. Linking contextual analysis with other data gathered in GRANTeD it may be possible to answer questions like: Is a certain type of research landscape and funding regime more likely to show high or low gender bias? Are contexts with high degree of political engagement with gender equality and national gender equality guidelines for RFOs showing less gender bias in funding allocation than contexts more silent or inactive in this area? Is high gender equality and gender equality awareness within the national setting and the RFO linked to low gender bias?

5.5 Grants and careers

With respect to grants and careers, grants may play a role in career progress at the one hand, and career differences may affect grant success. It is also important to note that when talking about the impact of grants on career we refer at least to two dimensions: impact on production and impact on promotion. This can only be studied for existing datasets (WP3 and WP4) as only there, long-term impacts on careers can be studied.

Analyzing the effects of grants on advancing career and checking if those effects are different is another important objective of the project. The issue brings to our attention that it is also an important part of the project to qualify the predictive validity of the selection process in different Panels and RFO instruments.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Monitoring the applicants (successful and not) in an RFO could provide some first ground analysis, but in order to investigate the effect on the career, we will followa sample of applicants over time, whether they are able to enter the academic system, and how they are being promoted.

It would be needed to differentiate between more 'elitist' career grants, the individual grants shaping careers, and the 'normal' often topic-based grants in which there is an *indirect* selection of a PI. It is not the criteria from the call that selects the applicant, but the internal team dynamics and division of labor may decide who acts as applicant, as PI.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

6. Advancing the operationalization

The questions listed in section 5.1. will be addressed using a variety of data sources as summarized in the Annex. In this chapter, it will be discussed how the research can be implemented, sometimes referring to examples from the existing cases, but mostly in a more general way, showing what is possible with the data we aim to collect.

6.1 The explorative study (WP5, WP6.1, WP6.2)

The exploratory study will use interviews and observations as the main instruments. The interviews are semi-structured. A feasible number of structured questions will be added, in order to construct the variables needed for testing the models.

The structured information at the panelist level will be aggregated to the panel level, e.g., by taking the average score of the panelists belonging to the same panel. The information from the first open part of the interview may be used to get an idea how the averaging will be done, or whether some additional information at the panel level is needed to have a good estimate.

Interview data are collected by semi-structured interview guides. All interviews are recorded and transcribed, transcripts are analyzedusing MAXQDA, applying qualitative content analysis for a variety of themes.[25] The core of this method is a theoretically based, inductively developed system of categories, which determine the aspects and elements to be filtered out of the available material. The following procedure for coding the material will be followed:

- A set of categories (codes) will be developed, based on the research questions, on the results of the literature review, as well as on the short interview reports. The researchers involved will start coding the material with the same initial set of categories. This will enhance comparability and methodological rigor of our research. The code book will be discussed between all involved researchers to establish a common understanding.
- Testing the developed categories by applying them to the material: Through coding the material the developed categories will be tested and new categories/topics will emerge. These new codes will be included into the code book by each researcher.
- Comparison of code books: After having coded about 10-15% of the material, the code books will be exchanged and the experiences discussed. For instance, the frequency of codes between researchers can be compared to get a picture of which codes are very prominent and which are not. Also the concrete coded sequences can be compared to see whether everyone did apply the codes in the same way (consistency analysis – inter coder reliability). Also new emergent

[25]Mayring, P. (2002), Einführung in die qualitative Sozialforschung: Eine Anleitung zu qualitativem Denken, Beltz Studium, Weinheim. Flick, U., Kardorff, E.v.u. Steinke, I. (Eds.) (2000), Qualitative Forschung: Ein Handbuch, Rowohlt, Reinbek bei Hamburg. Gläser, J. u. Laudel, G. (2010), Experteninterviews und qualitative Inhaltsanalyse: Als Instrumente rekonstruierender Untersuchungen, Lehrbuch, Wiesbaden.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

codes should be compared between researchers: whether the same topics, categories, codes have emerged and whether they are labelled in the same way. All emergent codes will be incorporated into the code book.

- Establishment of a final code book: Based on these discussions and exchange of experiences a final code book will be developed. This will be used to code the rest of the material. If necessary because of major changes in the code book also the already coded material will be recoded.
- Reporting of emergent codes and exchange of experiences: After the final code book has been established it is still possible that new important topics might emerge, for instance for core-RFO specific questions. To cover these topics each partner is allowed to complement the code book with new codes but is also responsible for documenting any changes. We propose to exchange coding experiences between partners on a regular basis (once a week, every second week) to facilitate learning, to enhance comparability and to share emergent codes, topics.
- Since we will not ask panel members explicitly about their understanding of gender or how they assessed excellence/merit or gender bias, we look for notions that refer to them explicitly or implicitly along different questions. As an example: To understand the construction of excellence, we look for different dimensions and notions that were depicted in the data: What do reviewers think is excellence? How did they evaluate and judge it? Were there any difficulties or challenges or were they clear about how to assign this criterion to applicants? And if they did, how exactly did they proceed? We continue to constantly compare, break down, discuss and group concepts until we are able to present inductively generated dimensions of excellence. Through this open approach we are able to reconstruct how reviewers view and understand the excellence criterion.
- All transcripts are analysed on the level of an individual panel member, but as we aim to interview 25 panel members per RFO, we are able to aggregate them on panel level.

6.2 Application phase (WP7)

To test whether self-selection (= the decision to apply for a grant) is different for male and female applicants we need to collect data about attitudes and behavior that may shape application behavior, and may influence performance and application success. Among the included variables are competitive thinking, excellence/performance, mobility, work life balance and family, as well as the idea of "the ideal applicant".

The decision to apply for a grant may depend on many factors, which partly can be investigated using an applicant survey, which will be sent to all applicants of core-RFOs that will have provided consent to participate in the GRANteD research project.[26] The

[26] If a potential new case can only deliver very partial data due to among others privacy regulation, it may not be selected.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

design of the survey will be guided by the questions listed in section 5.1, but the survey will also cover other items related to performance. Data about potential applicants should be collected in a different way – see the end of this section.

- Support for applicants: Writing competitive research proposals needs a lot of time and effort. Support can be provided by the RFO, the home institution, but also by supervisors, mentors or peers. The survey will contain items to provide detailed data about what support applicants have received from who. Does receiving support increase the chances for grant success? Do men and women get the same support? In a previous project we found hardly any differences between men and women. In the GRANteD project we will include more specific and detailed questions on the support structures.

  Additional analyses could be done on the differences in the networks of male and female applicants. For this we may use the coauthor relations, or the data provided in the CV. Different networks may influence the likelihood to be invited as co-applicants.

- Earlier experiences with grant applications: The survey will collect data on the number of earlier applications to research councils where the applicants were the principal investigator and whether they were successful or not, including the amount of funding received. This can be used as an indicator for their application experience and track record. It will help to answer the question as whether there is a self-selection effect: Do women have another threshold then men for applying?: It might allow assessing whether men and women have different experiences before applying and how this affects grant success. Besides earlier grants we also include questions on awards that have been received prior to the application.

- Belonging: The survey will include items to measure *feelings of belonging* to academia. The level of belonging can influence self-selection, as belonging is expected to correlate with the intension of following an academic career. Do men and women differ in terms of belonging, and does this influence application behavior? Belongingness could be related to other factors that shape application behavior and grant success like support mechanisms.

- Family situation: Does the family situation influence the probability to apply for a grant? In order to assess this, the applicants survey will include questions about family situation, the age and number of children, and how much time is spend on domestic labor and child care and how this is divided between the applicants and an eventual partner. This information will also be used to investigate whether domestic labor andchild care influence the performance of women and the probability to win grants.

- Evaluation of the procedure: Items will cover the application and evaluation process, including the interview with the panel.

- Past performance: Firstly questions like earlier grants, awards and prizes, societal impact, the (status of the) PhD supervisor and institution, etc., but also about publications and citations. The latter would replace the (resource consuming) collection of bibliometric performance data.

- Potential applicants:A crucial aim of WP7 is to provide empirically grounded evidence on gender bias for potential applicants. Only with the old cases (7.1 German Scientist Survey,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

7.3 German Doctoral Panel) this can be done. The German Scientists Survey (conducted in 2009, 2016 and 2019) is a randomized sample of all scientists employed at German universities. The German doctoral panel based on all academics who have received their doctorates in 2014. With the old cases we analyze the gender differences between all researchers (as potential applicants) in terms of their application behavior and (non) success with various RFOs, their employment status and position, their past performance (publications), their institutional research conditions and family situation. While the advantage of the old cases is to explore gender differences between applicants and non-applicants, the advantages of the new applicant surveys are that the questionnaires will be improved to investigate aspects that were neglected in the previous surveys (e.g. support in the grant application process).

- Analysis outline: The analysis will be performed on the level of each funding program. Whether it will be possible to analyse the data on the level of panels will be decided depending on the response rates and consequently on the distribution of responses between different panels.

  The data collected through the applicant survey will be analysed in two ways. On the one hand we will look for differences in the survey data concerning different variables or sets of variables like different feelings of belongings, organisational grant application support, childcare responsibilities, attitudes towards competition or work-life balance, past performance or perceptions of the assessment process related to gender, (academic) age, position etc. On the other hand, we will relate the survey data to the scores given by panellists or reviewers, the results of the assessment process itself and the results of the linguistic analysis of the CV and application. This will enable us to investigate whether differences observed in the survey data are also having an influence on how grant applications and CVs are written and whether these differences result in different scores and consequently in different funding decisions – meaning funded or not-funded grant application.

- Descriptive statistics: Descriptive statistics will show the basic features and differences in the applicant survey data set. The identified differences between groups will be tested whether they are significant. As we expect small case numbers and not normally distributed data we will use non-parametric tests.

- Causality: A crucial question for the *operationalization and implementation* is how to find causal relations. The dependent variable is the application decision, which is a binary variable. We will use logistic regression to find out which of the selected variables do influence the application decision, and whether there is gender specific self-selection. The important question is how to define the counterfactual. As we include a series of background variables in the survey, we will be able to compare pairs of men and women, who are about identical in all other relevant variables.

- Generalization: We finally aim to compare the results of the applicant survey in the core-RFOs with the results of the German Scientist Survey (T7.1) and the German Doctoral Panel (T7.3). Therefore, we will develop the applicant survey in T7.2 in close collaboration with the colleagues from the DZHW and based on the above mentioned survey instrument. Of course, it will not be possible to replicate these instruments as

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

they serve a different purpose than the applicant survey. But as far as possible, the described procedure will allow to perform similar analysis for specific parts of the data and research questions in T7.2 as in T7.1 and T7.3.

6.3 Measuring merit (WP4)

- Merit: In most studies about gender bias in grant allocation, merit is not taken into account. We will do that, but also do not want to restrict the measurement of merit in only bibliometric terms. The following merit related factors will be considered and if possible taken into account in the analysis – as far as available data and resources allow:
  - Past academic performance
  - Awards received by applicants
  - Overall quality of the proposal[27]
  - The newness of the proposal
  - The independence of the applicant.
- Gender and merit. Although the above mentioned merit variables can be accepted as valid, they may have a gendered dimension. For example, get a grant, and is this mediated by the effect of the family situation on performance? Women that did give birth to children in that period can ask for an extension, and we expect that they will. The information about the extensions can be used as operationalization of the family situation. The following questions can be answered: (i) has the family situation an impact on output and impact of the applicants (e.g., more household tasks may lead to less research time and to a lower academic output), and (ii) has this possibly a gendered effect on grant success? One would expect that baby/child care takes time away from doing research and will lead to an average lower past performance score than those without children.Another possibility to collect data on care responsibilities of women and men applicants is the applicant survey, which would provide more complex and differentiated data on care responsibilities.
- Prizes and awards may be a relevant factor. It is to a large extent neglected, and not much literature is available. In order to assess whether it is a useful variable to include in a study on gender bias (and to include it in the data collection efforts), we have analyzed in WP4.1a Dutch case where men had a twice as high chance to be awarded with cum laude for their PhD (6% vs. 3% of all PhD recipients). We show to what extent this difference can be classified as gender bias, using logistic regression. Please note that *cum laude* is the highest grade you can get for your PhD thesis in the Netherlands, and it is only awarded in exceptional cases (in our case 5% of all PhD receivers. This is different from e.g., Germany, where cum laude is not very good. It comes only after summa cum laude andmagna cum laude. If one wants to stay in research in Germany, magna cum laude or summa cum laude are needed. These results lead to the decision to ask applicants for awards and prizes in the applicant survey.

[27] An earlier project showed a very strong correlation between the score for the CV of the applicant and the score for the project proposal. This suggests that the overall quality score is redundant, and will lead to multicollinearity problems.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

- Measuring publication performance. Although the use of bibliometric databases for measuring scholarly productivity and impact is well established, it remains disputed due to e.g., the only partial coverage of output in many research fields. We will use the survey to collect performance data, to have all fields treated equally. But a disadvantage is and a disadvantage is the that suffers of course from the self-reporting problem. In WP4 we will do some work on the question how to measure productivity and impact as good as possible in a feasible way.
- Bibliometric indicators. Some bibliometric indicators need some refinement, as existing ones are not always valid. An example is how to measure impact. For this, an improvement is the P-model, which will be explained in D4.1. Another issue is whether women get fewer citations than men do. This is a simple question, and important, as it may indicate whether impact indicators are gender neutral or not. But it is complex to answer, as most papers are authored by reams consisting of men and women. An approach to this will be presented in D4.1.
- The newness of the proposal, and the level of innovativeness. For this we need to scan the literature and develop a usable indicator. To produce an unambiguous indicator, that is not (very) sensitive for biased use is a real challenge,and we do not expect to go in WP 4.1 beyond a proof of concept.
- The independence of the applicants:The concept has been developed, but applying it on a large-scale needs further development of the indicator, within WP4.1.

6.4 The prevalence of gender biasat the panel level (WP4)

- A variety of variables may influence the panel score, such as
  - Personal variables (gender, age, academic age, nationality, academic position, (international) mobility, self-confidence, presentation style).
  - Performance variables (publications, citations, highly cited papers, earlier grants, independence).
  - Reputation variables (journal impact, awards).
  - Network variables (co-authors, collaborators, earlier affiliations, host institution, nepotism, cognitive distance).
  - Scientific interests (topic choice).
  - Panel characteristics
  - Self-presentation of applicants: we need an operational definition that is text-based so we can measure it through the language used in the CV and the proposal.
  - Presentation style will be analyzed using the same data.
  - Self-confidence may be measured through language use in the CV, performance through Web of Science, Scopus, CV or a survey, and the network through the CV.
- Mixed models will be used for a multilevel analysis of the effect of gender on the panel scores the applicants receive, and also measure the effect of the various other variables mentioned above. As we control for the theoretically important covariates, we can be confident that we can detect the impact of the gender variable. Multiple logistic regression can be done to identify the effect of gender on the decision. In an earlier

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

study, we found mixed result.[28] The innovative character of the GRANteD project is that it includes a wide variety of the relevant independent variables, which together have not been taken into account in earlier studies.

- How would gender bias work, what are the mechanisms? Panel members may weight performance, position and project quality differently for men and women; this would imply that such neutral merit criteria are deployed in a gender biased way. In a statistical terminology, there are interaction effects: the effect of high-quality performance, project and a high position may be smaller (or different) for women than men. As our earlier research[29] shows that performance and position relate to gender, one may expect that taking performance and position into account, the gender effect becomes lower. This suggests that studies without past performance probably overestimate the effect of gender on the decision – as part of the effect disappears when taking mediation into account. Also application behavior may explain a part of the gender differences. This leads to a multiple mediation through position, through performance, and application behavior:

    a. Gender -> performance -> grant decision

    b. Gender -> position -> grant decision

    c. Gender -> application behavior -> grant decision

    d. Gender -> grant decision.

  Of the total gender differences in the grant decision, a part is explained by *position* and *performance* (including quality of the proposal). The remaining part of the difference is explained by gender and can therefore be called bias (asexplained in Deliverable 1.1). Additionally, other types of gender bias may be in the application phase and in the career.[30] Where relevant, we will include mediation in the regression analysis. We intend also to explore the use of structural equation modeling, as a way of handling endogeneity.

6.5 The panel processes, practices, and procedures (WP6, WP4)

If we have measured the level of gender bias at the panel level (see previous section), the next question is what panel characteristics may influence the level of gender bias. Relevant are the structure of the panel in terms of gender and country distribution of the members, the experience of the members and the chair. We also need information about the level of (implicit) gender stereotyping at the individual panel member level, and of the opinions of the panelists about whether gender bias is an issue within the science system at large, and in the grant allocation system more specific. We also need data about the group processes that

[28] For preliminary results: Van den Besselaar P, Schiffbaenker H, Sandström U, Mom, C, Explaining gender bias in ERC grant selection. STI 2018 Conference Proceedings, 346-352

[29] Van den Besselaar P, Sandström U, Vicious circles of gender bias, lower positions and lower impact: gender differences in scholarly productivity and impact. *PlosOne***12** (2017) 8

[30] Van den Besselaar P (forthcoming) *Has the glass ceiling disappeared in science*? Presentation at the Gender Summit in Amsterdam, October 2019.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

for sure influence the decision making. We need data about the workload (the number of applicants by panel member) tooand possibly about the overall gender stereotyping in the panel. Much of this information needs to come from the interviews, and from the administrative data of the RFO.

- The group dynamics. The dynamics of small groups is known to influence the decision, enabling various forms of bias to emerge. The following types of bias may emerge during the group process, and have an influence on the level of gender bias: (i) gendered use of criteria, (ii) nepotism, (iii) cognitive bias. In wP4.1 we will address whether these will be addressed.
- Criteria deployed. We have no survey data or observation data about what criteria are deployed. We will have interview data for a selection ofpanelists.But we plan to have also review reports of all panelists, and we can try to find out whether panel differences exist in criteria mentioned or implied in those review reports. This we use as indicator for what criteria are deployed, and this can be supported by the results of the interviews. A previous project was done in this way at the instrument level (the Starting Grant), which showed that the review reports indeed enable to reconstruct several core aspects of the decision making process.[31] This analysis should be done for the new cases at the panel level. If that succeeds, these criteria differences may be used to explain differences in gender bias at panel level.
- Nepotism: The institutional affiliations of the panelists will be collected. We define institutional proximity as when in a panel a panelist from the applicant's *host institution*is involved. Previous research shows that this form of proximity influences the probability to get funded. The results suggest that female and male applicants profit from this in a different degree. If so, nepotism can function as mechanism that produces gender bias.[32]
- Cognitive bias:In an earlier project, we studied whether the field of research of applicants influenced the probability to get funded. The question focuses on the cognitive distance between applicants and panelists. Using bibliometric data, one can measure the cognitive distance between the individual applicants and panelists, in order to answer the following questions: (i) does cognitive proximity increase the probability to get funded? And (ii) Are there a gender differences in the cognitive distance between applicants and panelist? So do women and men profit differently from cognitive proximity? An analysis of a few panels shows that we can estimate cognitive distance, and some statistics was done on the result – suggesting that men profit more from cognitive proximity than women do[33].

  Funders also may prefer some research topics over others, e.g., because some are seen as scientifically more promising, and others may be seen as more relevant to society. If that is the case, different preferences of men and women for research topics may lead to gender differences in success. More specifically, this effect can have two variants, one

[31]Van den Besselaar P, Sandström U, Schiffbaenker H (2018), Using linguistic analysis of peer review reports to study panel processes. *Scientometrics***117**, 313-329

[32] Mom C, Sandström U, Van den Besselaar P (2018) Does Institutional proximity affects grant application success?*STI 2018 Conference Proceedings*, 1579-1585

[33]Van den Besselaar P, Sandström U, Influence of cognitive distance on grant decisions, *Proceedings STI conference* 2017, Paris. IS13, 7pp

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

panel related and the other RFO related. The first is *cognitive proximity*, related to the topic preferences of panel member – as discussed above. The other is a *portfolio effect*, as the topics female applicants prefer may get less funding. Whether men and women propose different research topics can be investigated by overlaying the proposal themes on a global science map. A possibility to detect the proposal-topic for such an analysis is by extracting the in-text references that are used in the proposals.

- Language use to measure gender stereotyping (WP6.3): From the literature, gender stereotyping emerges as one of the mail causes of gender bias. This should be measured as good as possible. A survey could be used for that, but there are other (additional) strategies. In a previous paper[34] we showed that linguistic analysis of review reports indicate what linguistic word categories are used for positive and negative evaluations. In the new cases we will add to that the gender dimension: can we detect through linguistic analysis whether men and women are evaluated differently, and more precisely, can we detect gender bias in the review reports that points at gender stereotypes?This analysis can be done at the level of the instrument, at the level of the disciplines and at the level of the individual panels.

  The core-RFOs should provide textual data for further analysis: proposal, CV and the review reports. As the core-RFOs are also expected to provide other data (see Annex), we can include those in the analysis. This holds especially for the scores and the links between panelists, scores and review reports. Combined with the data from the panelist interviews, this can provide a rich analysis of the gender disparities in the scores, related to the opinions, stereotypes and other characteristics of panelists. It will contribute to a detailed answer to the question whether the panel structure and composition makes a difference with respect to gender bias, and the answer will go much further than only the sex-composition of the panel.It however would be good to have also another more direct measureof gender stereotyping, e.g., through the IAT.

  For several languages, the dictionaries are available. In WP6.3, we will try to find ways what to do with other languages.

- Experimentally testing of gender bias: "Investigate experimentally whether reviewers exhibit a bias against female applicants that could contribute to the gender disparity in academic science". Task 6.4 (Experiment - Randomized control trials (RCT) in experimental settings) definition was "*With the collaboration of some of the RFOs involved in the stakeholders committee, we will design a set of experiments based on allocation of "funding applications" randomly assigned either a male or female name, to a set of reviewers (males and females) from different fields and disciplines, to assess the existence of significant differences in the review scores and in the content of the comments*". A plan B was mentioned in the proposal defining the experimental setting in a "RPOs" or, even, in a Lab experiment context.

  Implicit in this approach is the idea that gender bias, if it exists, is brought to the evaluation process by panelists, their interactions and panels' dynamics. There are different possibilities, highly dependent on the background discipline, to explain the

[34] Van den Besselaar, Sandström, Schiffbaenker, Scientometrics, 2018.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

possibility of gender bias in grant allocation processes.[35] From Economics, various competing explanations have been developed, of which two may interest us; one based on preferences and other one focused on the role of available information. The theories based on preferences can be divided into "taste-based" on the one hand, or "statistical discrimination" based on information on the other hand; in this case we would be talking about the weight of stereotypes and cognitive shortcuts, among other mechanism. From psychology, the "role congruity" theory indicates that the expected evaluation would be more positive if the characteristics of the "group" evaluated are aligned with the typical social roles attributed to such group. From Sociology, the focus has been on aspects associated with identities and interactions between actors (networks of relationships, "cognitive" and "institutional" proximity, etc.).

Task 6.4. will focus primarily on the potential processes underlying gender bias from the evaluator's perspective (rather than that of the "object" or person being evaluated). That is, from the point of view of testing the model, the focus will be on Panel member's judgments and evaluations of other males and females' scientific competence to try to measure the existence of bias in the assessment, and whether the bias relates with the gender or other attributes of panelists in the context of Panel dynamics.

Task 6.4 will analyze whether, given an equally qualified male and female grant applicant, panel members evaluate showing better appraisal of the male or the female applicant. All participants will receive the same materials (for example a CV and a summary proposal), which will be randomly assigned either the name of a female researcher, a male researcher or, even, anonymized; researchers' gender will thus be the only variable that will differ between conditions of applicants. Using previously validated scales participants will evaluate (i) the quality of the materials and (ii) make a recommendation.

The implementation will be characterized by the investigator's control of the treatment conditions, with the setup of two or three experimental groups among panelists.Participating panelists will be randomly assigned to one of three applicant gender conditions (female, male or no name).

Randomization is the core of RCT; through the randomization of the treatment confounding variables could be ruled out, and we will be able to answer the questions of a) whether or not there are significant differences in the evaluation scores that can be attributed to the treatment (gender of applicant), and b) whether the anonymization of the materials regarding the gender of the applicant has an effect.

Firstly, we will analyze the average scores awarded to the applicant by each treatment group, and the statistical analysis will test the existence of significant differences in the means of the score values across groups; this is known as average treatment effects (ATE), and could be measured using t-tests and other statistical tools.

Secondly, in order to explore causality, some of the relevant attributes of the panelists included in the model will also be tested, for instance: sex (homophily), subfield (cognitive distance), institutional affiliation (institutional proximity), position (rank), academic age,

[35] Cruz-Castro, L. & L. Sanz-Menéndez (2019). Grant Allocation disparities from a Gender Perspective: Literature Review. Synthesis Report. GRANteD Project D.1.1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

etc. The purpose will be to analyze the potential mediating effects of relevant panelist' attributes on the relationship between the gender of the applicant and the evaluation scores. Depending on the final nature of the dependent variables (and their measurement), different types of regression models could be implemented.

Potentially, the experiment could also include a series of questions to locate the respondent along various attitudinal scales about women and science or tests like IAT, to explore the operation of further mechanisms related to stereotyping.

6.6 Testing of the modelfully and partially (WP4)

The new cases should provide the data to test more of the elements of the models specified in Ch 4 and 5:

- Measuring the level of gender bias at the panel level, using various forms of regression.
- Relating the levels of gender bias with a series of panel characteristics.
- Comparing the RFOs in a systematic way, in order to explore what contextual factors, including policies, may influence the prevalence of gender bias.
- In order to generalize our results, we may do a meta-analysis, using our own findings and findings of other projects – if available.

The existing cases will be used to test several parts of the model. This will provide additional support for our main findings. Here we discuss how the existing cases will add to the empirical evidence.

- In 2005, the Swedish MRC has provided a set of proposal evaluations of all proposals. Each member of the panel, except those with a conflict of interest, have given a grade (1-5) in three dimensions (proposal idea, proposal methodology and, track record or competence). We can link the grades with the panelist who gave them. These data make it possible to investigate whether conflict of interest in relation to one of the panel members also affects the grading from other panel members. The theory[36] suggests that if a panel member who reads the application over summer notifies that the applicant has an association[37] to one of the other panel members of the panel, he/she is inclined to give a higher grading than would have been done otherwise. If there are specific patterns that establish between CoI of that type and bibliometric scores per individual, then we might be able to say that peer review is sensitive to that type of nepotism. In this case, this is established already before and while the panel member reads the proposal, and the grading is kept in the process (for the Swedish MRC).

  Taking point one into consideration, we will also investigate something that has been discerned by several investigations: Randomness in the decision-making process. Research has shown that if different panels evaluate the same application, it can get very different scores. This is probably a large component in the explanations of success or

[36] Sandström &Hällsten (2008). Persistent nepotism in peer review. *Scientometrics***74**: 175-189.

[37] As supervisor or collaborator, co-author, department colleague, etc.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

failure: If reviewers do not agree on how to rank strengths and weaknesses, a proposal is almost always turned down.[38]CoI situations may even make that problem stronger.

We might also have data from 2012 for replicating the analyses mentioned above – has something changed over time? The questions to answer is (1) Is it the case that there is a gender disparity in how the applications are rated due to differences in weights of strengths and weaknesses (probably) based on bias in the process? (2) Is there a pattern in the reviews that would indicate that the review process treats applicants that are of the same academic age, career stage and status with different standards?

- The Emmy Noether data (695 applicants, 365 granted and 330 rejected) provides an exemplary case to explore gender-specific characteristics of the evaluation procedure of an early career excellence grant. Moreover, it will be used to further studywhether the career effects are different for man and women The data is not as comprehensive as ideally would be the case, so only parts of the model canbe tested. Past performance analysis will be applied for all applicants (2000-2006). The Emmy Noether data also allows answering questions raised in WP7 (with the Emmy Noether applicant survey) and questions of the WP3 (career effects). This link may not be possible with other old and new cases.
- The existing data about the Ramon y CajalProgramme can contribute to answer the following questions:
  - Existence of gender bias. This analysis is feasible with the existing data at the panel level.
  - Gender differences in performance.
  - The program applies an extension policy; we will do a before and after analysis.

  The plan is to collect new data on the panelists and, if successful, the case could contribute to addressing the following questions:
  - Do organizational and cognitive proximity between panelists and applicants play a role? (Conditional on whether the needed data can be made available).
  - What are the career effects of bias in grant allocation? (Conditional on whether the needed data on the current employment of applicants can be made available).
- The VENI dataset will be extended in order to test the relation between the very early career difference of winning the prestigious VENI grant (within three years after the PhD) and the further career. The dataset is very useful, as the start position of the applicants is very similar, enabling a causal analysis. However, as the period between the application and the post-performance measurement and career measurement was not long, we will collect new current data on performance and career level, and replicate the analysis for a longer time span. It will create a much better picture of the effect of grants on careers.

[38] Cicchetti, D. V., The reliability of peer review for manuscript and grant submissions: a cross-disciplinary investigation. *Behavioral and Brain Sciences*, 14 [1991]: 119–135; Pier et al., Low agreement among reviewers evaluating the same NIH grant applications. *PNAS* March 20, 2018, vol 115 (2) 2952-2957.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

6.7 Panel processes: policies and context (WP5)

- National research landscapes (WP5.1.) will be analysed by R&D intensity, integration of gender equality in research policy, gender balance in Research and Innovation, and societal overall gender equality, using global indicators and national policy document data. Gender equality policies of the core-RFOs are explored through national policy and RFO documents (WP5) and interviews with RFO staff (WP6.1.), complemented by interviews with panelists (WP6.2.). The RFO policies will be analysed using thematic analysis focusing on their (1) framework and coverage, including national and legal remits, (2) responsibility and management, (3) policy tools applied, and (4) sustainability. The policies related to the specific funding instrument chosen for detailed analysis by GRANTeD will be analysed focusing on their (1) framework and aims, specifically whether gender equality aims are explicitly included, and (2) policy tools directly related or relevant to gender equality (including encouragement for women to apply in calls for funding, coaching workshops for female applicants, obligatory/voluntary gender training for panels/reviewers, parenthood-related policy tools, target for share of female grantees, gender-related criteria for evaluation). Variables to be tested in the model will be developed in collaboration with WP4. For example, policy tools used can be easily developed to binary variables.
- At the level of the organizations, documentation needs to be collected about the policies, measures and procedures and the relevant aspects of those procedures (which may differ between RFOs but also within an RFO for different funding instruments). Using interviews, and if possible also review reports, one can assess whether and how the policies, measures, and procedures are implementedin practice, and if not, what kind of deviations occur. The interest is thus both in the scope and content of these policies and measures as well as their implementation in practice. The variable used for the model should reflect more the practices than the written rules. The focus of the contextual analysis will be to better understand the practices wheredecisions are taken and where gender biasmay occur. Over the different cases, we try to have similar descriptions, which should make comparison possible, e.g., using the comparative approach.

- <u>Mapping the context</u>: The following characteristics of national context will be explored in a way that enables systematic comparison. It will to a mapping of
  - national research funding regimes (EU, OECD, national research policy documents)
  - specific regulations and policies guiding national RFOs (national policy documents and legislation, national steering documents of RFOs)
  - national gender equality policies

- <u>Gender bias risk analysis of evaluation criteria</u>:
  - what are the formal evaluation criteria, including merit criteria, and do they include potential risk for gender bias?
  - what kind of gaps and inconsistencies can be identified?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

- does the RFO organize training for the panels on the evaluation system, and does this training include discussion on potential gender bias?

- Formal regulations and policies concerning funding decisions
  - Panel composition (gender balance regulations and de facto composition)
  - Selection of reviewers (is there a reviewer database, do panel members or applicants propose reviewers, gender balance targets among reviewers, monitoring)
  - Decision making process (degree of formalization, documentation, role of panel, rating and ranking)
  - Evaluation criteria and how they are defined
  - Written guidelines to the panels and panelists
  - Written guidelines to the applicants

- Gender equality policies:There is a large number of gender equality measures in RFOs identified in the literature, by previous projects of the consortium and through stakeholder contacts of the GRANTeD team. These include:
  - Providing training to avoid gender bias and gender stereotyping
  - External observers
  - Formalized processes, strictly controlled by the panel chair
  - Formalized scoring methods
  - Roleof the RFOs offices
  - One-step versus two-step process of allocations
  - Type of scales
  - Rules about conflict of interest
  - Reducing stress in the panel by reducing work load
  - Policies compensating for gender bias produced elsewhere in society

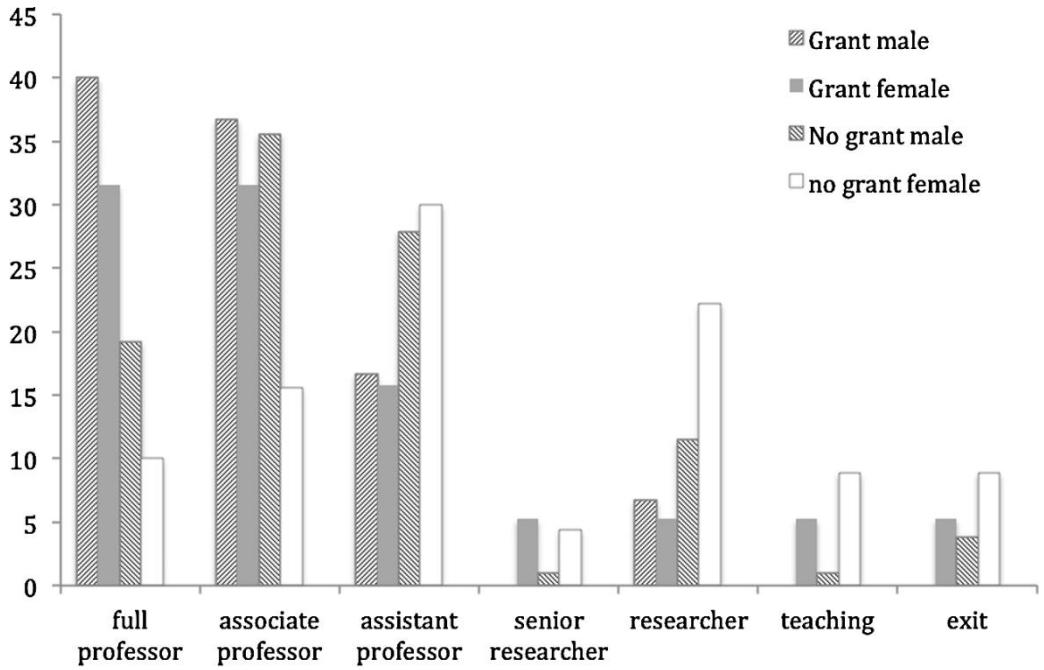
  A complete overview will be provided in one of the deliverables.

- Extension policies: An example is that mostly women take care of most of childcare and household duties. A way to counteract this is the extension-policy that most funders nowadays have: women get an extension of the application deadline (in case of the ERC, the deadline is seven years after the PhD) for each child born in that period (eighteen months per child). That should compensate for the time needed to care for the baby. The effect of these policies – related to the way they are implemented – can be tested in terms of their effects. To investigate whether the policy works, one may compare men and women without extension with women that used the extension regulation. Do users of the extension have a similar success as those that did not, thereby taking into account that they also have become older, which may negatively correlate with grant (and career) success.
- Required variety: In order to provide a robust analysis of the effect of different policies we need variation.The selected RFOs should have different levels of implementation of gender policies and instruments to reduce potential bias.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

6.8 Grants andcareers (WP3, WP4.3)

- Types of grants: We distinguish between the effect of 'elitist' grants (VENI, Emmy Noether) and the more 'regular' grants (MRC case). The former may have a much stronger effect on the careers than the latter.
- Career effects: In the VENI case, we collected data about post performance[39] and career position, about 9 years after receiving the VENI (early career) grant. Receiving the grant indeed makes a difference for careers, as grantees move up quicker than non-grantees. The effect does not seem gendered, as for grantees and non-grantees, men move up faster than women do. That seems independent of having received the grant.



- A longitudinal study of the relation between grants and careers

1. The longitudinal study will focus on Sweden, due to the availability of detailed data at the individual level, which are not easily available elsewhere.
2. We will monitor the supply to the scientific workforce by using the PhDs achieved after 1980 (data from LIBRIS database). That will give us the possibility to answer how many *potential researchers* there were in each period. By using gender-API and manual resources, we can assign gender to about 90-95% of these by the use of first names. We will be able to follow gender disparities in the supply to the workforce and to study gender differences in this respect.

[39] There is a question about whether issues of predictive validity (post/grant publications of granted and non-granted applicants) and impact of grant on careers should be included. In some paragraphs it has been mentioned that a design is needed focusing on the role of RPOs (and not RFOs) in order to assess the impact of grants in the probability of advancing careers of males and females.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

3. Using the personnel registers (personnel catalogues, internal telephone books etc.) of each university in Sweden[40], we will be able to measure how many of the PhDs do continue at the university[41] each year, how many that exit from the university system, how many that goes to other activities, and how many that jumps between the sectors.[42] These data will help to answer the questions: (i) Are there gender disparities in how men and women choose their careers? (ii) Is it changing over time?
4. The university catalogues are available per year and will be used to measure (i) years until academic positions for each man and woman in our sample that enters the system, (ii) how many years it takes to become an associate professor (docent, senior lecturer), (iii) how many years it takes to become a professor? Etc. This can be done for each of the positions at the university.
5. By following the PhDs in their publication behavior from their first publication onwards, we can show how many (men and women) are potential grant applicants. We aim to follow researchers by creating disambiguated files so that we can build bibliometric indicators per individual and rank them in each period.[43] Several different indicators should be used but we will mainly look for *size-dependent indicators*, type total impact (SNIP total), total P-Model etc. With this information we will be able to questions whether "excellence" has any role in the selection of person's that try to be granted and those that stay in the university system compared to those that exit e.g. to the company sphere. To summarize: we will in WP3 be able to contribute to the understanding of application behavior and gender differences with these historical data. Also, we will be able to add knowledge to how these figures change over time.
6. There are about 100-150 rookies (new individuals) that try to enter into the grant system (of Research Councils) each year in the period of 1980-2000. Their dissertations give information on supervisor(s), and we plan to create a network of supervisors and their connections by (a) departments and (b) publications. The association between people at universities *and how they are selected to the panels* in the research councils is up to now an underdeveloped theme in science studies. Sweden should be one of good cases for such an investigation, as it is a small country with quite some geographic distances. The association between people as supervisors on the one hand, and panel members on the other is an important aspect that relates to the conflict of interest protocol given by the research councils.[44]

[40] Focus is on universities that in Sweden cover about 90-95% of the R&D money for basic research; only a small part goes to the institute sector.

[41] Earlier research based on large-scale registers show that in Sweden more women than men leave academia, so exit cannot be the explanation to vertical segregation in the HE system (see Silander et al. *Higher Education* 2013, p. 173-188

[42] For those that choose an international career, a CV search will be done.

[43] The method is described in Sandström & Wold (2015),

[44] Historical comparison of gender inequality in scientific career across countries and disciplines. PNAS 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

7. Using the rookies and by following that sample of applicants, we can find out how characteristics like gender, age (based on year of birth), academic age (based on PhD year), network composition (distance to the panel club), cognitive distance to the panel members; relative independence to the supervisor(s), and other factors influence their *chances of getting granted* the first time they apply for a grant, and the second time etc. In all these investigations, we will have a focus on gender disparities.[45] Long [1992] suggests considering maternity leaves as possible causes of prolonged non-publishing time spells. Long and later Mairesse and Pezzoni (2015) put forward the effect of non-publishing time spells as an explanation of the productivity gap and the latter authors also points at the lack of longitudinal studies within the field of interest for gender studies hindering a deeper understanding of the factors that could explain gender disparities in science.[46]
8. We will prudently investigate which bibliometric indicators do best in mirroring the decisions of the panels. Such information will lay the ground for a discussion on peer review as the basis for success in careers. But, first, to answer that question, we have to investigate what factors can explain careers in science. Is it the case that those who follow in the footsteps of their supervisors do have a better chance to make job career in the university system? Or is it publication success, i.e. scientific influence (impact) over colleagues what explains career outcome (time to a professorship) in the long run? Or is it other factors like academic age, field competition, distribution over areas, the newness of research (novelty or breakthrough or 'high risk-high gain), etc.

[45] Huang et al. (PNAS 2020) show that career length explains quite some of the productivity differences between men and women. However, in their research, there's no information on actual age. We will test whether the effect remains, when taking age into account. Furthermore, it is unclear how reliable their data are, especially in relation to correct disambiguation. For determining career length, this is crucial.

[46] Mairesse Jacques and Michele Pezzoni. [2015]. "Does Gender Affect Scientific Productivity? A Critical Review of the Empirical Evidence and a Panel Data Econometric Analysis for French Physicists". Revue Economique, 66(1), p.65-113.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

7. Integrating the results (WP9)

The basic integration will be done on several levels:

- We will pool the panels from the several case studies to do an as large as possible analysis at the panel level. This should result in the identification of those panel characteristics that relate to gender bias.
- If we find different levels of gender bias between the studied RFOs, a case comparison becomes interesting. Using the qualitative comparative analysis approach[47], the existing and new cases will be compared, in a way that we (i) know the level and direction of gender bias at the panel level (W3, WP4, WP6.3, WP6.4) and, at the same time, have enough systematic data on the panels that we can explain the different levels of gender bias at the panel level in terms of characteristics of the panelists, the panels, and their organizational and wider context.
- Additionally, we will be able to use the applicant survey and the existing scientist survey (WP7) to better understand whether possible bias emerges in the application process due to gender differences in support in the RPO and in the private situation, and/or whether gender disparities are the effect of choices made by men and women. Here also data from the existing cases may be useful, as indicated above (section about WP3 and WP4).
- Finally, the scientists survey (WP7) and the longitudinal analysis in WP3 will lead to more insight into the effects of grants on careers, and the role that a variety of other variables play.

Together this promises a rich picture that may not only show whether and if so, how strong gender bias in grant allocation exists, but may also improve our understanding of where it is generated and what might be done about it.

[47] Ragin CC, Rihoux B (2008) *Configurational Comparative Methods: Qualitative Comparative Analysis*. 1 edition. Thousand Oaks: SAGE Publications

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

ANNEX 1: The data template

Introduction

This annex specifies the data needed to answer the research questions asked in the GRANteD project. We summarize which data are required for the *new cases*, and where these data will be used for.

Building an integrated dataset

The first data requirement is that we can link the data about applicants and the data about panelists. That requires that we start with non-anonymous data. The anonymization will be done after the datasets have been created and used by the project researchers. So, for the linking process we do need:

- Data on applicants: name, age, sex, academic age, affiliation, nationality, discipline.
- Date on panelists: name, age, sex, affiliation, nationality; university of the PhD and university of the bachelor degree; discipline.
- Administrative data of the various RFOs

Answering the research questions(complete model)

We have a series of questions about the prevalence of bias and the causes of bias. I list them here, and specify what data we need to answer those.

Q1. Is there gender bias at panel level, how strong, and in what direction?
This question aims at predicting the panel scores and the application success by a set of performance measures, a set of background variables, and gender. We would need the following data:

- performance data:
  - applicants survey.
- bias: gender and affiliation of applicants and of panelists, and cognitive distance between applicants and panelists
  - administrative data RFO
  - WoS data for applicants and panelists (cognitive distance)
- scores in as many dimensions as possible linked to the individual panelist, and the final scores and decisions; review reports.
  - administrative data RFO
- personal characteristics of applicants and panelists
  - administrative data RFO
  - applicants survey
  - CV of applicants
  - WoS data for applicants data and panelists (measuring independence).

Q2. Does application behavior (self-selection) play a role?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Do women and men apply equally often? And if not, what factors play a role? To answer that question, we need data on

- social factors influencing application behavior
    - CV of applicants
    - applicants survey
    - advertising, instructional and other texts/docs of RFO

Q3. Gender differences in performance?
Merit plays a role in grant decisions, and the question is how merit can be measured. A second question here is whether men and women show systematic performance levels, and why this is the case. To find this out, we need data on

- social factors influencing gender differences in performance, like the academic position of the applicant, resources network, family situation, etc.
    - CV of applicants
    - admin data RFO
    - applicants survey

Q4. Does self-presentation play a role?
It is often claimed that women do not 'sell' themselves as good as men do. In order to find out whether this plays a role we have to analyze the way male and female applicants present themselves.

- writing style of applicants in relation to success
    - proposal text
    - CV text

- Presentation style of applicants during interviews (if F2F interviews is part of the procedure).
    - Observations of panels
    - Proposal and CV texts

Q5. Do organizational and cognitive proximity between panelists and applicants play a role?
Apart from gender, other forms of bias may play a role, possibly interacting with gender bias. The most obvious are nepotism (a social relation between the applicant and panel members), and cognitive similarity (applicant and panel member(s) being active in the same research topics). To investigate that, we need data on

- affiliation of panelists and applicants
    - Administrative data RFO (panelists and applicants);
- specialty of panelists and applicants
    - WoS data

Q6. Do special measures (like application extension) play a role?
In order to improve gender equality in the application process and in the results, several specific policy measures have been implemented in grant allocation procedures. An example

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

is (for academic age-related grants) allowing women a longer application period if they have given birth to children. Which measures are implemented, and what is the effect?

- what regulation exists?
  - documents RFO
  - applicants survey
- which applicants used what regulation?
  - administrative data RFO

Q7. Do panel dynamics and panel structure play a role?
The decisions are made (or prepared) by panels, and if gender bias occurs it is also at the panel level. Can we identify the panel characteristics that distinguish between panels in which gender bias occurs from those where no gender bias occurs? We need data on

- Panel structure
  - administrative data RFO
- Panel dynamics
  - interviews with panelists
  - RFO internal reports from/on panels
- Level of gender bias in panelists frames, stereotypes
  - -
  - review texts of individual applicants;
  - interviews with panelists
  - applicants survey

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

Unit of analysis	Data	Specification	Source
Applicants	Personal data	Name, gender, age, year of PhD (academic age), nationality, Institutional affiliation, Orcid or other bibliometric ID, name of PhD supervisor	Council administrative data
	Opinions	On gender in science, Evaluation of selection processes and procedures	Survey applicants
	Application experiences	success with various RFOs and the specific RFO	Survey applicants
	Social situation	Status; position; career path[48]; past performance[49]; integration into a research networks; independence; institutional research conditions; family situation; life and career goals; self-confidence[50] Maybe a quality check of the assigned WoS-publications by the applicants instead of bibliometric author disambiguation.	Survey applicants
	Application behavior	self-selection and self-presentation	Survey applicants
	CV	To measure network, past performance, and self-presentation (linguistic analysis)	Application
	Proposal	To measure network self-presentation (linguistic analysis)	Application
	Review/evaluation reports	To measure gender stereotyping in the panels (linguistic analysis)	Council
	Scores from reviewers / panelists	With the reviewer/panelist info	Council admin. data
	Overall score		Council admin. data
	Grant / no-grant		Council admin. Data
	Performance data	To measure past performance	Scopus/WoS
	Use of special regulation	Extension because of giving birth, military services, etc	Council admin. Data
panelists / reviewers	Personal data	Name, gender, affiliation, age, nationality	Council admin. data
	Opinions and experiences	Measuring of stereotypes and frames, opinions about gender in science, characteristics of the deliberation and decision process Additional issues may come up in the design	–interview selection of panelists
	Bibliometric data	Measuring of cognitive distance to applicants	Scopus/WoS
panels	Panel composition	Age, gender, field, nationality	Council admin. data
	Field of the panel		Council admin. data
	Type of funding instrument		Council admin. data
	Criteria and procedures	e.g., detailed selection criteria	Council documents
other	Policy documents	How are potential applicants addressed (linguistic analysis)	Council documents
	Conflict of Interest regulation	How are conflict of interests handled at the panel level, and elsewhere (nepotism)	Council documents

Table 1: Data needs

Q8. What are the career effects of grants and of bias in grant allocation?

This is the focus of WP3 and a part of WP7. I leave those out for a moment, and add this later. In any case we have several data to study this, or will collect those.

[48] To replace or complement the CV, and when want to avoid internet CV search.

[49] Broader than only via WoS, and again as alternative for CVs.

[50] Psychological tests items can be used here.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

- Longitudinal data about applicants and their career steps and application success
  - Data on several Swedish research councils
  - Data from the German researchers survey
  - The applicants survey,
  - Web and WoS data (for applicants' careers in the existing cases).

Practicalities

Several data are collected by the GRANteD team, such as bibliometric data, Web data, and the survey data. For the latter we need collaboration from the RFO with address and a support letter for applicants and (possibly) panelists explaining that the RFO finds the study important.

The other data should relatively easily be available at the RFO. Most of the data/documents will be available as XLSX or PDF file. So, the work load should not be high for the RFO.

If needed, we can have a look at what format the data have, and how they can be retrieved as easy as possible.

Selection procedures and models

Depending on the model of selection, the data needs may be different. For example, in models where the panel is crucial, we may focus on the panel members and leave the reviewers out. However, there are also procedures where the assessment is done by external reviewers only (mainly?) and where the committee/panel only (merely?) has the task to 'count' the scores from the review reports and rank the applications. The different procedures may have different levels of bias. And that would be an important result.

For such comparison we also would need data about the reviewers, as reviewers' characteristics may also have aneffect, independently from the procedure.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

ANNEX 2. Variables for interviews and observations (WP6.1/WP6.2)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

cont ext	National science and gender equality framework conditions	Research policy	Level of Innovation performance	Summary Innovation Index Alternative: Index for attractive research systems			EIS2019
		National funding system	Budget funding instruments	Share of competitive research funding in the national research funding budget	2	Low to high (exact scale needs to be developed); is there a scale measuring this in the EIS?	
		Equality policy + Gender regime	Status quo of gender equality in each society	Global gender gap index rank position 2019 Alternative: Score value of the EIGE Gender Equality Index			Global Gender Gap Report EIGE Gender Equality Index
		integration of gender equality in R&I	National framework for gender equality in research and innovation	integration of gender equality			
Leve l	Dimension	Sub-Dimension	What will be investigated in detail	variable name	Relevan-ce for model 1= very 3 = not	Scales	Data source
RFO	governance		Organizational profile			1= researcher driven (bottom up) 2= top down 3= legally based	Policy docs reports Interviews RFO staff
	GE policies	Formalisation of GE activities		Existence of a Gender equality plan		Exists Does not exist	Analysis of Policy docs
		Experiences with GE activities		Years of GE activities		Number of years	Analysis of Policy docs Interviews RFO staff
		Relevance of Gender equality	GE as a policy objective	Commitment to GE in statutes, mission etc.		Promoting GE is a relevant policy objective of each RFO	Analysis of Policy documents
			Commitment of top level management	Top manage-ment Gender commitment		High to low	Interviews RFO staff
				Understanding of gender inequality		Individual deficits Structural reasons Combined approach	Analysis of Policy documents Interviews RFO staff
			Gender statistics	Monitoring of GE		performed regularly irregularly not performed at all	Analysis of Policy docs
	Funding instrument	Competitiven ess		Overall success rate		Percentage of successful proposals	RFO data
	Process or-ganisation	Formal procedures	Transparency of decision making process	Presence of clear assessment procedures		Procedural rules for assessment procedures are defined? Guidelines how to implement these rules are available?	Analysis of Policy documents Interviews RFO staff
				Decision making rules		Majority: which majority rule? Unanimity	Analysis of Policy docs Interviews RFO staff
				Rating scales		Fine graded (comprehensive scale) Roughly graded (narrow scale)	Analysis of Policy documents
			Selection of panellists	Clear rules& procedures for selecting panelists?		Yes partly no	Analysis of Policy docs Interviews RFO staff
				Clear criteria for selecting panelists?		Yes partly	Analysis of Policy docs

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

						no	Interviews RFO staff
			Time pressure	Average duration for assessing one application		Number of minutes available for assessment	Analysis of Policy docs Interviews RFO staff Interviews w. panellists
		application and selection process	Transparency of criteria	Openness and ambiguity of assessment criteria		Number of (general and abstract) criteria (like excellence independence) open to interpretation without clear assessment guidelines and operationalisation	Analysis of Policy documents
			Accountability	Clear structures of individual accountability		Group accountability vs. individual accountability	Analysis of Policy docs
				Presence of process observers foreseen		Yes no	Analysis of Policy docs Interviews RFO staff
				Are names of panellists published?		Yes no	
				Focus of assessment		1= focus on PI 2= focus on CV 3=CV and PI equally important	Analysis of Policy docs
			Eligibility	Extension regulation in place		Yes no	Analysis of Policy documents
				Target group of extension regulation		Women only Men only Men & women	Analysis of Policy documents
				Number of extended years		Number of years	Analysis of Policy documents
			GE measures	Gender bias training for panellist		Yes no	Analysis of Policy docs Interviews RFO staff
				Other GE measures		Which ones?	Analysis of Policy docs Interviews RFO staff
Social process and practices within panels	Panel Structure (gend ERC wp4)	Sex composition	Is	Disciplinary Heterogeneity		Panel is composed of panellist from heterogeneous sub-disciplines Panel is composed of of panellist from a narrow set of sub-disciplines	Analysis of Policy documents Interviews RFO staff core-RFO data
		Disciplinary composition		Disciplinary Heterogeneity		Panel is composed of panellist from heterogeneous sub-disciplines Panel is composed of panellist from a narrow set of sub-disciplines	Analysis of Policy documents Interviews RFO staff core-RFO data
		Geographic/institutional composition					
		Panel chair		Sex of panel chair		1= female 2= male	core-RFO data
				Experience of panel chair		Number of terms served as panel chair Number of terms served on the panel	core-RFO data Interviews RFO staff
		Panel composition		share of F/M panel members			
				Perceived share of F/M panel members			interviews
	Assessment practices	Accountability		Formal pro-cess obser-vers present		Yes no	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.

				Formal pro-cess obsevers intervening		Yes no	
		Gendered practices	Criteria applied differently to female and male researchers	Criteria dropped		1= always 2= … 3=	
				Double standards			
		Systematic application of criteria	Criteria applied systematically in each panel	Weighting of criteria		Criteria are weighted differently by panellists Criteria are weighted in the same way by panellists	Observations
				Emergence of new criteria		In panel discussions new assessment criteria emerge Only defined criteria are used in the assessment process	
	Decision making in practice		workload	Applicants per panelist			
			Focus of assessment	Focus of assessment practices		1= majority of panellists focus mainly on PI 2= majority of panellists focus mainly on CV 3= majority of panellists focus equally on CV and PI	Observations Interviews
			Power dynamics	How many panel members participate actively on average in negotiations?			Observations
	GE		Relevance of GE in RFO			1 = very relevant ….	Interviews self-assessment panellists and panel chair
pane lists							
	stereotypes				Mean panel scores go into model	Will be further developed together with Ruth van Veelen	IAT test sent to ALL panelists with IC
		Linguistic analysis				Need to be specified in data request sheet of core-RFOs (MoU)	Review reports protocols

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824574.