Journal of Informetrics 2 (2008) 280–287

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

How to detect indications of potential sources of bias in peer review: A generalized latent variable modeling approach exemplified by a gender study

Lutz Bornmann[a,*], Rüdiger Mutz[a], Hans-Dieter Daniel[b]

[a] *Professorship for Social Psychology and Research on Higher Education, Swiss Federal Institute of Technology Zurich (ETH Zurich), Zähringerstr. 24, 8092 Zurich, Switzerland*
[b] *Evaluation Office, University of Zurich, Switzerland*

ARTICLE INFO

Article history:
Received 8 July 2008
Received in revised form 5 September 2008
Accepted 5 September 2008

Keywords:
Grant peer review
Gender bias
Gender effect
Gender differences

ABSTRACT

The universalism norm of the ethos of science requires that contributions to science are not excluded because of the contributors' gender, nationality, social status, or other irrelevant criteria. Here, a generalized latent variable modeling approach is presented that grant program managers at a funding organization can use in order to obtain *indications* of potential sources of bias in their peer review process (such as the applicants' gender). To implement the method, the data required are the number of approved and number of rejected applicants for grants among different groups (for example, women and men or natural and social scientists). Using the generalized latent variable modeling approach indications of potential sources of bias can be examined not only for grant peer review but also for journal peer review.

1. Introduction

According to Merton (1942), the founder of the modern sociology of science, the functional goal of science is the expansion of "true" and secure knowledge. To fulfill this function in society, a set of ideal norms became established among scientists: the ethos of science. The universalism norm requires that contributions to science are not excluded because of the nationality, gender, social status of the contributors or other irrelevant personal or social criteria (MacCoun, 1998; Ziman, 2000). Critics of peer review argue that decisions in peer review are, nevertheless, frequently biased—that is, that they are not based solely on scientific merit but are influenced also by personal attributes of the applicants (Daniel, Mittag, & Bornmann, 2007; Marsh, Jayasinghe, & Bond, 2008). But an evaluation of a peer review process that can yield reliable and valid results on the influence of potential sources of bias on the review process is as a rule very elaborate and costly. The reasons for this are: (1) The research on peer review has identified a large number of attributes of applicants that can represent potential sources of bias in the peer review process (Wessely, 1998), (2) The study design should meet the highest requirements in order to establish unambiguously that the work from a particular group of applicants has a higher rejection rate due to biases in the peer review process and not simply as a consequence of the lesser scientific merit of the group of applications, and (3) The grant peer review process is a secret activity (Tight, 2003); reviews are secured with assurance of confidentiality.

Before a research funding organization conducts an extensive evaluation study, it should therefore seek indications of the influence of potential sources of bias in the grant peer review process, (1) in order to determine the necessity for

* Corresponding author. Tel.: +41 44 632 48 25; fax: +41 44 632 12 83.
E-mail address: bornmann@gess.ethz.ch (L. Bornmann).

1751-1577/$ – see front matter
doi:10.1016/j.joi.2008.09.003

Table 1
Absolute and relative number of women among all applicants for a research grant from the SNSF and among awardees in 2004.

Disciplines (and subject areas)	Abbreviation	Number of submitted applications		Number of projects funded		Percent women among applicants	Percent women among awardees	Difference in percentages
		Women	Men	Women	Men			
Social medicine	SOMED	5	14	3	5	26	38	−11
General biology	GEBIOL	11	70	8	31	14	21	−7
Philosophy, religious studies, pedagogy, psychology	PHILO	68	126	47	71	35	40	−5
Engineering (including computer sciences)	ENGIN	17	179	15	113	9	12	−3
Experimental medicine	EXMED	19	67	12	38	22	24	−2
Social and economic sciences, jurisprudence	SOCEC	58	182	36	107	24	25	−1
Historical studies	HISTOS	11	75	8	48	13	14	−1
Basic medical sciences	BAMED	28	109	17	62	20	22	−1
Mathematics	MATHE	4	51	3	40	7	7	0
Physics	PHYSI	8	139	6	111	5	5	0
Astronomy and space research	ASTRO	4	18	3	14	18	18	1
Environmental sciences	ENVIR	12	52	8	38	19	17	1
Basic biological sciences	BABIOL	24	124	15	81	16	16	1
Chemistry	CHEMI	10	91	7	79	10	8	2
Clinical medicine	CLMED	38	134	14	55	22	20	2
Archaeology, ethnology, art history, and urban studies	ARCHO	17	62	7	33	22	18	4
Linguistics and literary studies	LINGUI	18	57	10	40	24	20	4
Earth sciences	EARTH	9	66	4	53	12	7	5
DORE (practical research)	DOREP	17	33	6	18	34	25	9
Preventative medicine (epidemiology/early detection/prevention)	PRMED	17	45	4	24	27	14	13
Total		395	1694	233	1061	19	18	1

Source: SNSF, at http://www.snf.ch/SiteCollectionDocuments/por_fac_sta_jb04_d.pdf (page 36; Retrieved: 20 November, 2007).

an evaluation study, and (2) if a necessity is found, to identify the sources of bias that should be examined more closely (Ledin, Bornmann, Gannon, & Wallon, 2007). In the following, we present a statistical method that program managers at a research funding organization can use to obtain *initial indications* of potential sources of bias in their peer review process. The method has already been used for a meta-analysis investigating gender differences in grant award decisions (Bornmann, 2007; Bornmann, Mutz, & Daniel, 2007). To demonstrate application of the method for examining the peer review process, we utilized data from the Swiss National Science Foundation (SNSF) that are published on the Internet (http://www.snf.ch/E/aboutus/facts/Pages/statistics.aspx; Retrieved: November 23, 2007). The SNSF statistics show gender-specific figures for the research projects that were approved and rejected for funding in a total of 20 disciplines and subject areas in the years 2004–2006 (see Tables 1–3).

With our statistical approach to obtaining *initial* indications of potential sources of bias in peer review processes, we are operating under the assumption that the odds of being approved among women applicants should be equal to the odds of being approved among men applicants. Unequal odds indicate a gender effect. If the effect is statistically significant, it is an evidence of bias and a detailed study of the peer review process should be conducted (see here also Women in Science & Engineering Leadership Institute, 2006).

2. Methods

For the statistical analysis we considered estimations of the odds ratio as a dependent variable. For one discipline (or subject area) j, to which the grant applications to the SNSF in a certain year were assigned, this odds ratio can be estimated as

$$o_j = \frac{d_{1j}/(n_{1j} - d_{1j})}{d_{0j}/(n_{0j} - d_{0j})}, \qquad (1)$$

where d_{1j} and n_{1j} are the number of women among approved applicants and all applicants, respectively, and d_{0j} and n_{0j} are the number of men among approved and all applicants, respectively.

The approach is to analyze the estimated gender effect across several application years at the level of different disciplines (or subject areas). As we assume that the true gender effect varies between all combinations of disciplines and application years, we estimated a *generalized linear mixed model* that explicitly allows for this variation in a multilevel framework. As

Table 2
Absolute and relative number of women among all applicants for a research grant from the SNSF and among awardees in 2005.

Disciplines (and subject areas)	Abbreviation	Number of submitted applications		Number of projects funded		Percent women among applicants	Percent women among awardees	Difference in percentages
		Women	Men	Women	Men			
Historical studies	HISTOS	13	66	12	47	16	20	−4
Engineering (including computer sciences)	ENGIN	15	180	13	120	8	10	−2
Mathematics	MATHE	3	47	3	37	6	8	−2
Astronomy and space research	ASTRO	1	18	1	14	5	7	−1
Social and economic sciences, jurisprudence	SOCEC	21	150	13	90	12	13	0
Basic medical sciences	BAMED	20	123	12	74	14	14	0
Physics	PHYSI	5	127	4	105	4	4	0
Linguistics and literary studies	LINGUI	27	37	18	25	42	42	0
Chemistry	CHEMI	9	92	7	78	9	8	1
General biology	GEBIOL	6	63	3	35	9	8	1
Earth sciences	EARTH	13	71	10	59	15	14	1
Experimental medicine	EXMED	21	71	11	40	23	22	1
Clinical medicine	CLMED	32	139	13	64	19	17	2
Basic biological sciences	BABIOL	25	137	15	97	15	13	2
Philosophy, religious studies, pedagogy, psychology	PHILO	38	111	17	57	26	23	3
Environmental sciences	ENVIR	8	59	4	42	12	9	3
DORE (practical research)	DOREP	44	55	21	31	44	40	4
Archaeology, ethnology, art history, and urban studies	ARCHO	13	38	4	21	25	16	9
Preventative medicine (epidemiology/early detection/prevention)	PRMED	4	29	0	7	12	0	12
Social medicine	SOMED	5	9	0	0	36	0	36
Total		323	1622	181	1043	17	15	2

Source: SNSF, at http://www.snf.ch/SiteCollectionDocuments/por_fac_sta_jb05_d.pdf (page 36; Retrieved: 20 November, 2007).

odds ratios violate the normality distribution, log-odds ratios were analyzed instead of odds ratios (Skrondal & Rabe-Hesketh, 2004). In our data there is a two-level structure, with applications that were submitted to the SNSF in the years 2004–2006 nested within disciplines (and subject areas) (see here also Jayasinghe, Marsh, & Bond, 2003). For the data, where *j* indicates a special combination of a discipline (or subject area) and an application year, we consider the generalized linear mixed model of the log-odds ratio (logit(o_j)):

$$\text{logit}\,(o_j) = \beta_0 + \beta_1 x_j + \zeta_{0j} + \zeta_{1j} x_j + \varepsilon_j \quad \varepsilon_j \sim N(0, \theta_j) \tag{2}$$

where

$$x_j = \begin{cases} +0.5 & \text{for women applicants} \\ -0.5 & \text{for men applicants} \end{cases},$$

and

$$(\zeta_{0j}, \zeta_{1j}) \sim N(\mathbf{0}, \Psi^2).$$

Here β_0 is a fixed intercept, and ζ_{0j} is a random intercept; β_1 is a fixed slope, and ζ_{1j} is a random slope of x_j. β_1 represents the log-odds ratio of interest (gender effect), while $\beta_1 + \zeta_{1j}$ represents the estimated 'true' log-odds ratio of the discipline (or subject area) in a particular application year *j*. The within-unit deviations $\sqrt{\theta_j}$ are simply set equal to the standard errors of the log-odds ratios, calculated from the proportions given in equation (1) (Woolf's method). The sum of the terms $\zeta_{0j} + \zeta_{1j} x_j$ (the random part) can be thought of as a total residual ξ_j. The variance of this total residual var (ξ_j) is composed of three variance and covariance components: the variance of the intercept between disciplines (and subject areas) and application years ζ_{0j}, the variances of slopes between disciplines (and subject areas) and application years $\zeta_{1j}\, x_j$, and the covariance between intercepts and slopes on the second level: cov (ζ_{0j}, $\zeta_{1j}\, x_j$). Thus, a matrix of variance/covariance components for the random effects of the procedures ($\mathbf{\Psi}^{(2)}$) can be generated.

As opposed to showing simple percentages of women among applicants and women among grant awardees in individual disciplines (and subject areas) and application years for the peer review process at the SNSF (see Tables 1–3), the analysis

Table 3
Absolute and relative number of women among all applicants for a research grant from the SNSF and among awardees in 2006.

Disciplines (and subject areas)	Abbreviation	Number of submitted applications		Number of projects funded		Percent women among applicants	Percent women among awardees	Difference in percentages
		Women	Men	Women	Men			
Mathematics	MATHE	5	40	5	30	11	14	−3
Astronomy and space research	ASTRO	3	16	3	13	16	19	−3
Linguistics and literary studies	LINGUI	23	30	21	25	43	46	−2
Clinical medicine	CLMED	34	132	16	56	20	22	−2
Environmental sciences	ENVIR	12	48	10	37	20	21	−1
DORE (practical research)	DOREP	52	66	26	32	44	45	−1
Physics	PHYSI	5	143	5	116	3	4	−1
Social and economic sciences, jurisprudence	SOCEC	34	135	25	97	20	20	0
Social medicine	SOMED	6	14	3	7	30	30	0
Chemistry	CHEMI	9	109	6	84	8	7	1
Historical studies	HISTOS	12	73	9	60	14	13	1
Preventative medicine (epidemiology/early detection/prevention)	PRMED	5	36	2	16	12	11	1
Philosophy, religious studies, pedagogy, psychology	PHILO	35	90	22	61	28	27	1
Earth sciences	EARTH	14	78	9	58	15	13	2
Engineering (including computer sciences)	ENGIN	21	222	10	159	9	6	3
Archaeology, ethnology, art history, and urban studies	ARCHO	16	41	10	31	28	24	4
Basic biological sciences	BABIOL	44	139	26	106	24	20	4
Experimental medicine	EXMED	23	66	9	37	26	20	6
Basic medical sciences	BAMED	30	115	11	80	21	12	9
General biology	GEBIOL	16	56	3	39	22	7	15
Total		399	1649	231	1144	19	17	3

Source: SNSF, at http://www.snf.ch/SiteCollectionDocuments/por_fac_sta_jb06_d.pdf (page 38; Retrieved 20 November, 2007).

using the generalized linear mixed model allows the following questions to be answered (Rabe-Hesketh & Skrondal, 2005; Raudenbush & Bryk, 2002):

- *Overall test of fixed effects*: Is there a *statistically significant* overall gender effect across all disciplines (and subject areas) and application years? How large is the *overall gender effect*?
- *Variance components*: What is the variance of the gender effect across all combinations of disciplines (and subject areas) and application years? Using the generalized linear mixed model, the individual effects (empirical Bayes prediction) can be calculated for each combination, and via a test $\left(H_0 : \sqrt{\text{var}(\zeta_{1jk})} = 0\right)$ it can be determined whether these individual effects differ statistically significantly (slope variance).

Beyond that, analysis of the SNSF data using the generalized linear mixed model has the following advantages:

- *Sample size*: Sample size is explicitly included in the estimation of the model: the gender effect for a discipline (or subject area) in an application year having a small sample is weighted less than the effect for a discipline (or subject area) in an application year having a large number of submitted applications. In the estimation, effects from small samples are reduced (shrinkage correction), as with decreasing sample size a population effect tends to be overestimated.
- *Credible intervals*: Bayes credible intervals allow us to test whether the individual gender effects for a discipline (or subject area) and application year are statistically significant. That is the case if they differ from zero—that is, if the null value does *not* lie in the credible interval.
- *Maximum Likelihood*: Maximum likelihood estimation with adaptive quadrature produces reliable parameter estimates (Rabe-Hesketh, Skrondal, & Pickles, 2004).

The generalized linear mixed model with the figures on the selection decisions of the SNSF was estimated using the gllamm procedure (Generalized Linear Latent and Mixed Models) (Rabe-Hesketh et al., 2004) in the statistical software package STATA 10.0 (StataCorp., 2007). The commands that we used to analyze the data with STATA gllamm are available online (Rabe-Hesketh, 2004).

Table 4
Maximum likelihood estimates for the generalized linear mixed model.

Effect	Parameter	Estimate	Standard error	*z*-value
Fixed part				
Intercept	β_0	0.54	0.09	−6.18*
Gender (men = −0.5; women = +0.5)	β_1	−0.23	0.09	−2.66*
Random part				
Intercept	$\sqrt{\text{var}(\zeta_{0jk})}$	0.59	0.07	8.29*
Gender (men = −0.5; women = +0.5)	$\sqrt{\text{var}(\zeta_{1jk})}$	0.32	0.11	3.03*

Note: The cov $(\zeta_{0jk}, \zeta_{1jk})$ of intercept and slope is fixed at 0.
*$p < .05$.

3. Results

Tables 1–3 show the absolute and relative number of women among all applicants and among applicants approved for awards from the SNSF, classified according to disciplines (and subject areas). The figures, published by the SNSF on the Internet in three different reports (see URL above), are for the years 2004 (Table 1) to 2006 (Table 3). The percentages in the tables (percent women among applicants and percent women among awardees) do not allow a clear assessment of possible biases for the 3 years. Across all of the disciplines (and subject areas), the percentage of women among applicants and percentage of women among awardees is approximately the same (in the tables, see the 'Total' column). However, there are clear differences between the disciplines, as for example the figures in Table 1 for the year 2004 show: in some disciplines, women appear to have a clear advantage (such as in general biology), and in some other disciplines (for example, in DORE, practical research), there appears to be a clear advantage for men. Moreover, the percentages vary in the individual disciplines across the three application years. For instance, the percentage of women awardees in basic medical sciences was 22% in 2004 (see Table 1), but it was 12% in 2006 (see Table 3) (even though the percentage of women applicants in the 2 years was approximately the same, about 20%).

3.1. Overall gender effect

Table 4 shows the maximum likelihood estimates for the generalized linear mixed model that was calculated with the data in Tables 1–3. The fixed part represents the overall regression on gender (men = −0.5; women = + 0.5); the random part represents the variability of regressions between the combinations of application years and disciplines (and subject areas), as well as the heterogeneity of effects (expressed as the variance and covariance of the intercept and slope). The estimate of β_1 (−0.23) in Table 4 (fixed part) gives the effect of the applicant's gender on approval of grant applications in the years 2004–2006 across all disciplines (and subject areas). The results show that the odds of being approved for a grant are smaller for women than for men, with an estimated odds ratio of $\exp[\beta_1\ (-0.5)–\beta_1\ (0.5)] = \exp(0.23) = 1.26$. That means that for the peer review process at the SNSF, men applying for grants have on average statistically significantly greater odds of approval than women applying for grants by about 26%.

Even if the result indicates that men have a clear advantage over women in the peer review process at the SNSF, the percentage of 26% does not completely reveal the extent of this advantage. For this reason, in the following we would like to illustrate the size of the gender effect taking a fictitious example: assume for a moment that in a 1-year time period, decisions are made at the SNSF to approve and reject 1000 applications for research grants (500 submitted by women and 500 submitted by men). Half of these applications are approved, and half are rejected. If there is no gender effect, the odds ratio of the approval rate for men and women is $p/(1-p) = 1.0$ with $p = 0.50$. In the case of gender effects, the odds ratio (or) is greater or smaller than 1.0: $or = (p+d)/(1-(p+d))$, where d is the deviation from the mean approval rate ($p = 0.50$). To solve the equation for $(p+d)$, the approval rate for men is or/(1 + or) and for women $1-(or/(1+\text{or}))$. Based on the results of the generalized linear mixed model (or = 1.26, $p+d = 1.26/2.26 = 0.558$), in our fictitious example we can therefore expect an approval rate of 55.8% (279 approvals) for grant applications submitted by men. For grant applications submitted by women, we can expect an approval rate of 44.2% (221 approvals). This makes – *ceteris paribus* – a difference of 58 approvals that is due to the gender of the applicants.

3.2. Variance of gender effects

The random part of the generalized linear mixed model in Table 4 suggests that there is a statistically significant variability in the gender effect between application years and disciplines (and subject areas). The *z*-value of the square root of the variance components $\sqrt{\text{var}(\zeta_{0jk})}$, $\sqrt{\text{var}(\zeta_{1jk})}$ is statistically significant. Fig. 1 shows the variability of the gender effect between different disciplines (and subject areas) and application years using empirical Bayes estimates as effect sizes and credible intervals for each discipline (or subject area). For a 95% credible interval, the posterior probability that the gender effect parameter for a special discipline (or subject area) and application year lies in the interval is 95%. In Fig. 1, negative mean log-odds ratios point to more favorable outcomes for men in the peer review process, and positive mean log-odds

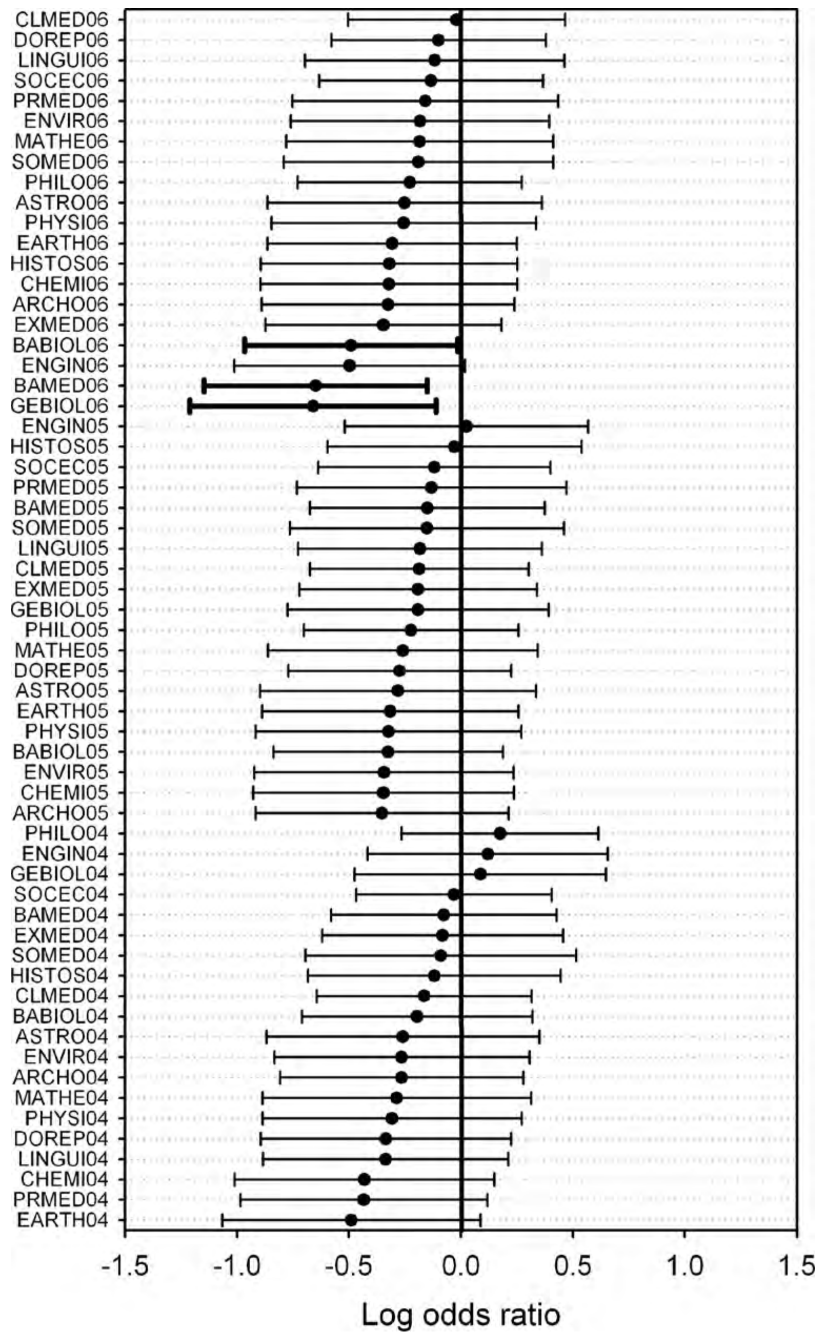

Fig. 1. Mean log-odds ratios and credible intervals for each discipline (or subject area) for the application years 2004, 2005, and 2006. Negative mean log-odds ratios point to a preference for men, and positive mean log-odds ratios point to a preference for women. Significant effects are shown by error bars in bold face.

ratios point to more favorable outcomes for women. For the SNSF disciplines (and subject areas) the estimated individual gender effects vary between −0.66 (GEBIOL, year 2006) and 0.18 (PHILO, year 2004). In other words, the gender effects (odds ratios) vary across the disciplines (and subject areas) and application years (2004–2006) within the range of about 93% (exp[(−0.66) −0.5–(−0.66) 0.5] = 1.93) in favor of *men* and about 20% (exp[(0.18) 0.5–(0.18) −0.5] = 1.20) in favor of *women* applying for grants.

For the individual log-odds ratios in Fig. 1, if the zero (no effect in the population) does not lie within a credible intervals, as is the case for basic biological sciences (BABIOL, year 2006), general biology (GEBIOL, year 2006), and basic medical sciences (BAMED, year 2006), then these effects are statistically significant. At the level of individual disciplines (or subject areas), therefore, in general biology, basic biological sciences, and basic medical sciences men have on average statistically significantly greater odds of approval than women (for 2006). For the other disciplines (or subject areas), due to too small effect sizes and/or due to too small sample size (which is also possible), the statistical power is not sufficient to obtain statistically significant effects.

3.3. Relation between the overall gender effect and statistically significant individual effects in three disciplines (or subject areas)

In order to test the extent to which the statistically significant overall gender effect of the generalized linear mixed model is mainly due to the statistically significant effects in the three disciplines (or subject areas) general biology, basic biological sciences, and basic medical sciences in the year 2006, we estimated a further model without the figures for these three disciplines in 2006. The result shows that the overall gender effect β_1 of originally −0.23 (see Table 2) drops to −0.12 and is no longer statistically significant. This means that the gender effects in the three disciplines in 2006 make up a decisive portion of the overall gender effect.

In the light of this result, we would like in the following to illustrate, similar to the illustration for the overall gender effect above, the size of the gender effect in the three disciplines using fictitious figures: let us assume for one discipline (or subject area) that in a 1-year time period, the SNSF makes decisions to approve and reject 100 research grant applications (50 submitted by women and 50 submitted by men). Half of these applications are approved, and half are rejected. Based on the results of the generalized linear mixed model, for grant applications submitted by men we can expect approval rates of 62.0% (31 approvals in basic biological sciences), 65.9% (33 approvals in general biology), and 65.7% (33 approvals in basic medical sciences). For grant applications submitted by women, we can expect approval rates of 38.0% (19 approvals in basic biological sciences), 34.1% (17 approvals in general biology), and 34.3% (17 approvals in basic medical sciences). According to this, in the three disciplines (or subject areas) approximately two-thirds of applications submitted by men will be approved, but only about one-third of applications submitted by women will be approved.

4. Discussion

In this study, we used data from the SNSF on the gender of grant applicants to present a generalized latent variable modeling approach that can be used by research funding organizations to determine whether a certain group of applicants is *possibly* disadvantaged in the peer review process. The fixed part of the model estimation for the SNSF data shows a statistically significant gender effect across all disciplines (and subject areas) in the three application years from 2004–2006. According to that, for women and men submitting applications there is a significant unequal odds ratio (the odds of being approved among women applicants divided by the odds of being approved among men applicants). Moreover, the analysis of the gender effect at the level of the individual disciplines (and subject areas) makes it clear that the overall gender effect is mainly due to the awarding of grants in three disciplines (and subject areas) within 1 year.

The analysis of the SNSF selection procedure using the generalized latent variable modeling approach thus indicates a gender bias in the SNSF selection process. As a next step it would therefore be appropriate to conduct an in-depth evaluation study—such as was done by Bornmann and Daniel (2005a, b, 2006, 2007) for the selection process at the Boehringer Ingelheim Fonds (Heidesheim, Germany) and by Ledin et al. (2007) for peer review in the fellowship program at the European Molecular Biology Organization (EMBO, Heidelberg, Germany). As both institutions had information on the applicants' scientific achievements up to the date of their fellowship applications, not only the potential sources of bias, but also the scientific performance (track record) of the applicants could be included in the statistical analyses. In this way it can be distinguished between the influence of the applicants' achievements up to the date of application and the potential sources of bias on the selection decisions (Cole & Fiorentine, 1991). This control variable approach would be expected to produce reliable and valid results regarding the influence of applicants' gender on the SNSF peer review process. The evaluation study of the SNSF selection process should examine mainly the awarding of grants in the disciplines (and subject areas) general biology, basic biological sciences, and basic medical sciences.

Although the generalized linear model to detect indications of potential sources of bias in grant peer review has numerous advantages (see above), it still has some limitations: (1) multilevel analysis requires big samples. Using Monte Carlo studies, Browne and Draper (2000) found that for accurate variance components, a lot of groups ($n = 48–50$) are needed (here: combinations of year and disciplines (and subject areas)). According to a Monte Carlo analysis by Maas and Hox (2005), with 50 and more groups the results are "in practice probably acceptable" (p. 91). In contrast to the variance components, however, the standard errors to test fixed effects show only a small downward bias, even if the assumptions of large samples and normality are not met. And even if the Monte Carlo studies available up to now examined continuous outcome variables and not odds ratios (as in the present study), we can still assume that a large number of groups is important for testing both fixed effects and variance components (this is more important than a large number of individuals per group). In the present study, we reached a sufficiently large number of groups ($n = 60$) for testing a multilevel model only through the combination of year and disciplines (and subject areas). (2) For the generalized latent variable modeling approach, there is a lack of power analyses that establish how large a sample must be in order to give proof that with a certain statistical power and alpha level an assumed effect size is statistically significant.

Finally, we would like to point out that – using the generalized linear mixed model to detect unequal odds ratios – indications of potential sources of bias (such as gender, nationality, social status) can be examined not only for grant peer review but also for journal peer review. Editors of a journal require only data on specific person/manuscript groups for a number of years and/or a number of research fields, as shown in Table 1. The use of model estimation for the evaluation of a selection process can not only give the research funding organization or the publisher indications of potential sources of biases but can also – if the results are published – increase the transparency and thus improve the perception of fairness of the peer review process (Keith-Spiegel, Koocher, & Tabachnick, 2006).

References

Bornmann, L. (2007). Bias cut. Women, it seems, often get a raw deal in science—So how can discrimination be tackled? *Nature*, *445*(7127), 566.

Bornmann, L., & Daniel, H.-D. (2005a). Committee peer review at an international research foundation: Predictive validity and fairness of selection decisions on post-graduate fellowship applications. *Research Evaluation*, *14*(1), 15–20.

Bornmann, L., & Daniel, H.-D. (2005b). Selection of research fellowship recipients by committee peer review. Analysis of reliability, fairness and predictive validity of Board of Trustees' decisions. *Scientometrics*, *63*(2), 297–320.

Bornmann, L., & Daniel, H.-D. (2006). Potential sources of bias in research fellowship assessments. Effects of university prestige and field of study on approval and rejection of fellowship applications. *Research Evaluation*, *15*(3), 209–219.

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

Bornmann, L., Mutz, R., & Daniel, H.-D. (2007). Gender differences in grant peer review—A meta-analysis. *Journal of Informetrics*, *1*(3), 226–238.

Browne, W. J., & Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics*, *15*(3), 391–420.

Cole, S., & Fiorentine, R. (1991). Discrimination against women in science: The confusion of outcome with process. In H. Zuckerman, J. R. Cole, & J. T. Bruer (Eds.), *The outer circle. Women in the scientific community* (pp. 205–226). London, UK: W W Norton & Company.

Daniel, H.-D., Mittag, S., & Bornmann, L. (2007). The potential and problems of peer evaluation in higher education and research. In A. Cavalli (Ed.), *Quality assessment for higher education in Europe* (pp. 71–82). London, UK: Portland Press.

Jayasinghe, U. W., Marsh, H. W., & Bond, N. (2003). A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings. *Journal of the Royal Statistical Society Series a—Statistics in Society*, *166*, 279–300.

Keith-Spiegel, P., Koocher, G. P., & Tabachnick, B. (2006). What scientists want from their research ethics committee. *Journal of Empirical Research on Human Research Ethics*, *1*(1), 67–82.

Ledin, A., Bornmann, L., Gannon, F., & Wallon, G. (2007). A persistent problem. Traditional gender roles hold back female scientists. *EMBO Reports*, *8*(11), 982–987.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92.

MacCoun, R. J. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology*, *49*, 259–287.

Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications—reliability, validity, bias, and generalizability. *American Psychologist*, *63*(3), 160–168.

Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, *1*, 115–126.

Rabe-Hesketh, S. (2004). Generalized latent variable modeling by Skrondal and Rabe-Hesketh. Section 9.5: Nicotine gum and smoking cessation. Meta-analysis. Retrieved 2 May, 2007, from http://www.gllamm.org/books/gum.html.

Rabe-Hesketh, S., & Skrondal, A. (2005). *Multilevel and longitudinal modeling using Stata*. College Station, TX, USA: Stata Press.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). *GLLAMM Manual*. Berkeley, CA, USA: University of California.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models applications and data analysis methods* (2nd ed.). Thousand Oaks CA, USA: Sage.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modelling. Multilevel, longitudinal, and structural equation models*. London, UK: Chapman & Hall-CRC.

StataCorp. (2007). *Stata statistical software: release 10*. College Station, TX, USA: Stata Corporation.

Tight, M. (2003). Reviewing the reviewers. *Quality in Higher Education*, *9*(3), 295–303.

Wessely, S. (1998). Peer review of grant applications: what do we know? *Lancet*, *352*(9124), 301–305.

Women in Science & Engineering Leadership Institute. (2006). *Reviewing applicants. Research on bias and assumptions*. Madison, WI, USA: University of Wisconsin-Madison.

Ziman, J. (2000). *Real science. What it is, and what it means*. Cambridge, UK: Cambridge University Press.