CONTENT ***In science, peer review is the oldest and best established method of assessing manuscripts, applications for research fellowships and research grants. However, the fairness of peer review, its reliability and whether it achieves its aim to select the best science and scientists has often been questioned. Here we present the first comprehensive study on committee peer review for the selection of doctoral (Ph.D.) and post-doctoral research fellowships. We analysed the selection process of the Boehringer Ingelheim Fonds, a foundation for the promotion of basic research in biomedicine, with regard to its reliability, fairness and predictive validity – the three quality criteria for professional evaluations. We analysed a total of 2,697 applications, 1,954 for doctoral and 743 for post-doctoral fellowships. In 76% of the cases, the decision whether to award a scholarship or not was characterized by agreement between reviewers. Similar figures for reliability were reported for the grant selection processes of other major funding agencies. With regard to fairness, we analysed whether potential sources of bias, i.e. gender, nationality, discipline and institutional affiliation, could have influenced the decisions. For post-doctoral fellowships, no statistically significant influence of any of these variables could be observed. In applications for a doctoral fellowship, evidence of a gender, discipline and institutional bias, but not of a nationality bias, was found. We therefore present some proposals for optimizing committee peer review.***

The most important aspect of our study was to investigate the predictive validity of the process, i.e. whether the Foundation achieves its aim to select the best young scientists. Our bibliometric analysis showed that this is indeed the case and that the selection process is thus highly valid: research articles from Boehringer Ingelheim Fonds fellows are cited considerably more often than the »average« publication in the Institute for Scientific Information (ISI, Philadelphia, PA, USA) journal sets »Multidisciplinary«, »Molecular Biology & Genetics«, and »Biology & Biochemistry«. These sets include journals covering the research fields in which most of the fellows publish.

Reliability, fairness and predictive validity of committee peer review

Evaluation of the selection of post-graduate fellowship holders by the Boehringer Ingelheim Fonds

Lutz Bornmann[1], Hans-Dieter Daniel[1,2]

[1]Professorship for Social Psychology and Research on Higher Education, Swiss Federal Institute of Technology (ETH) Zurich;
[2]Evaluation Office, University of Zurich, Zurich, Switzerland

• In science, peer review is one of the oldest methods for the assessment of grant applications, manuscripts submitted for publication in journals and applicants for a research fellowship. As »gatekeepers« of science, peers or colleagues requested to evaluate applications or manuscripts should recommend only those that meet the highest of scientific standards. Polanyi[(1)] regards peer review as the embodyment of the principle of mutual control, fostering judgements with respect to the novelty, accuracy and relevance of research results. Proponents of the system argue that it is more effective than any other known instrument for self-regulation in promoting the critical selection that is so crucial to the evolution of scientific knowledge.

The main points of criticism made by opponents of peer review are that (i) reviewers rarely agree whether or not to recommend that a manuscript be published or a research fellowship be awarded. Thus, reliability of the peer review process is said to be poor. (ii) Reviewers' recommendations are frequently biased, i.e. judgements are not solely based on scientific merit, but are also influenced by personal attributes of the author, applicant, or reviewer himself. (iii) The process lacks predictive validity, since there is little or no relationship between the reviewers' judgements and the subsequent usefulness of the work to the scientific community, as reflected by citations of the work in later scientific papers[(2)].

The empirical research on peer review mainly dealt with the assessment of manuscripts[(3-6)] and grant applications[(7-10)]. However, the selection

Rating	Applications for a doctoral fellowship (n = 1,490)	Applications for a post-doctoral fellowship (n = 491)
Award	62	59
Maybe an award	17	19
No award	21	22
Total	100	100

TAB. 1: Ratings given by the external reviewers for applications for a doctoral and post-doctoral fellowship (in per cent)

of post-graduate research fellowship holders, i.e. doctoral (Ph.D.) students and post-doctoral fellows, by committee peer review was given little attention. Some years ago, the Boehringer Ingelheim Fonds*, a foundation for the promotion of basic research in biomedicine located in Heidesheim, Germany, agreed to have an independent external evaluation of its selection procedure for doctoral and post-doctoral research fellowships[11]. This evaluation study aimed to answer two questions: (i) does the peer review system fulfil its declared objective to select the best young scientists? (ii) Is the main criticism raised against peer review as outlined above justified? Here we present the most important results of our study, which – analysing as it does a large data set – is the most comprehensive study on post-graduate fellowships published to date. On the basis of our results, we also propose measures for optimizing committee peer review.

The data set on which the evaluation is based

• The archive of the administrative office of the Boehringer Ingelheim Fonds contains the majority of the applications for fellowships comprising curriculum vitae, reviews, references, appraisals, protocols of the decision-making Board meetings etc. All in all, 2,697 applications received by the Foundation between 1985 and 2000 were available for analysis: 1,954 applications for a doctoral (72%), and 743 applications for a post-doctoral research fellowship (28%). The number of applications for a post-doctoral fellowship is much lower, since the Foundation ceased to promote post-doctoral scientists in 1995.

The selection process of the Boehringer Ingelheim Fonds

• Young scientists send their application to the administrative office (secretariat) of the Foundation, which ensures that the applicant and proposed project fulfil the formal conditions and that all required papers have been submitted**. Once the formal criteria have been met, the office forwards each application to an independent external reviewer. On the basis of predetermined criteria, the reviewer assesses the applicant, his or her proposed research project, and the laboratory in which the project is to be pursued, and recommends approval or rejection.

Table 1 shows the ratings given by the external reviewers for applications received from 1985 to 2000***. The reviewers recommended grants (from the Foundation) for 62% of the applications for a doctoral fellowship and 59% of the applications for a post-doctoral fellowship. In both groups, about 20% of the applications were rated »no award«.

In addition to the assessment by an external reviewer, a member of the Foundation's staff examines the application, interviews the applicant personally and submits a detailed report. The final overall rating of the interviewers reads as follows: (i) »definite award«, (ii) »award«, (iii) »maybe an award« and (iv) »no award«****. *Table 2* shows the ratings of all applications for doctoral and post-doctoral fellowships between 1985 and 2000. In both groups, about 10% of the applications were strongly recommended for an award and about 30% were rated »award«. 29% of the applications for a doctoral and 38% of the applications for a post-doctoral research fellowship were disapproved.

Finally, the applications, together with the external reviews and reports on the interview, are submitted to the Board of Trustees, which consists of seven internationally renowned scientists and is chaired by a representative of the donors. The Board convenes three times a year and – after a detailed discussion of each individual application – approves or rejects each application accordingly. From 1985 to 2000, 25% of the applicants for a doctoral and about 20% of the applicants for a post-doctoral research fellowship were successful. A comparison of these percentages with the recommendations of the external reviewers (*Table 1*) and those of the Foundation's staff (*Table 2*) reveals that both plead more frequently for approval than the Board of Trustees. About 65% of those applications rated »award« by the reviewers and about 50% of the applications rated »definite award« or »award« by the Foundation's staff did not receive a research fellowship in the end.

In a study related to panel peer review of the National Science Foundation (Arlington, VA, USA), Klahr[13 p. 151] presented similar results: ratings of the ad hoc reviewers, i.e. the external reviewers, »are more ›lenient‹ than the panel ratings.« Klahr[13 p. 152] refers to the following causes for the dis-

* *www.bifonds.de*

** *Fröhlich[12] describes the selection process of the Boehringer Ingelheim Fonds in detail.*

*** *Two experts of the Centre for Research on Higher Education and Work (Kassel, Germany) independently rated afterwards all reviews on the scale shown in table 1, since the reviewers themselves did not use a rating scale. The reliability of the experts' ratings is very high (weighted kappa coefficient = 0.96).*

**** *The interviewers used a rating scale.*

Rating	Applications for a doctoral fellowship (n = 1,920)	Applications for a post-doctoral fellowship (n = 704)
Definite award	10	8
Award	33	27
Maybe an award	28	27
No award	29	38
Total	100	100

TAB. 2: Ratings given by the staff of the foundation for applications for a doctoral and post-doctoral fellowship (in per cent)

First round	Second round	Third round
76% (*n* = 1,905)	16% (*n* = 394)	8% (*n* = 225)
Agreement ⟶		Disagreement

TAB. 3: Number of decisions made by the Board of Trustees in three rounds (in per cent; n *= 2,524)*

crepancies: »The ad hoc [the external] reviewers may have more technical proficiency, a better sense of what can realistically be accomplished in the area, and greater familiarity with the track record of the principal investigator. However, the ad hoc reviewers are at a disadvantage when it comes to making a quantitative rating of the proposal. First of all, they are generally unfamiliar with the ratings that get translated into decisions. Second, they do not have the same sense of scarce resources that the panelists do.« He recommends that external reviewers should primarily attend to the concrete attributes and faults of an application. They should also be informed that their »final overall rating is not as important as their substantive comments«[13 p. 153].

Reliability, fairness and predictive validity of the Boehringer Ingelheim Fonds' selection process

• The Board of Trustees of the Boehringer Ingelheim Fonds has the difficult task of assessing the scientific potential of the applicants as well as their research proposals and to select the best young scientists. Within the scope of the study, we investigated to what extent the Board was able to do justice to this challenging task between 1985 and 2000. The committee peer review process of the Foundation was examined with regard to the quality criteria for professional evaluations: reliability (i.e. is the selection of research fellows reliable or is the result purely incidental?), fairness (i.e. are certain groups of applicants favoured or at a disadvantage?) and predictive validity (i.e. does the process fulfil the objective to select the best young scientists?).

Reliability of committee peer review

• Human decisions are classified as reliable, when different persons come to the same or similar conclusions. In the analysis of reliability, the degree of agreement between committee members is determined.

The seven members of the Board of Trustees decide on applications in three rounds during each of the three annual Board meetings. »›A‹ means that the application is approved first time around; ›A¯‹ means that the application is put aside for the second round; and an application which is rated ›A-B‹ and below is dismissed. In the second and, if necessary, third round, the number of applications approved or dismissed depends on how much funding is still available for the session«[12]. The Foundation's secretariat states that the level of controversy in the Trustees' discussion whether to approve or reject an application increases with the number of rounds. Thus, the round in which the application is approved or dismissed should reflect the extent of disagreement between the Trustees: in later rounds, agreement tends to decrease, i.e. disagreement increases. *Table 3* shows that for 76% of the applications, the decisions of the trustees are characterized by agreement, since these applications are decided in the first round. 24% of the applications are decided under circumstances in which disagreement more or less prevails.

To determine the extent of agreement or disagreement in the Board of Trustees of the Boehringer Ingelheim Fonds, we compared our results with those of other studies. Hereby, we have to consider that in other studies the extent of agreement is not indirectly calculated by the decision round, but directly by the level of agreement between two or more reviewers' ratings. For the assessment of grant applications, the following agreement coefficients are reported: in the selection process of the Deutsche Forschungsgemeinschaft (Bonn, Germany), 82% of the reviewers' ratings are identical[14]. According to Cicchetti[15], 68% of applications receive the same assessment in the peer review system of the National Science Foundation. Hodgson[16] calculated an agreement rate of 73% for reviewers of the Heart and Stroke Foundation of Canada (Ottawa, Canada). Consequently, the extent of agreement between reviewers, and thus the reliability of the committee peer review process of the Boehringer Ingelheim Fonds, is similar to that of major funding organizations.

Criterion	»Typical« candidate
Age of applicant at first university degree	26 years
Final grade	1.4 (1.0 = highest grade)
Mobility during education	one or more changes of universities
Recommendations	two letters of recommendation
The external reviewer	recommends an award
The staff of the foundation	recommends an award
Gender	male
Nationality	German
Discipline	biology
Institution where the research project will be pursued	German university
Predicted probability for approval	50%
Predicted probability for rejection	50%

TAB. 4: A »typical« applicant for a doctoral fellowship. His chances are 50% to receive a fellowship.

Fairness of committee peer review

• Journal manuscripts or applications for a fellowship are supposed to be judged solely on the basis of their scientific merit, i.e. the scientific quality of the results or the applicant's academic achievements and the scientific quality of his/her project. Personal characteristics of authors or applicants, i.e. specific attributes, such as applicants' gender or nationality, should not influence the procedure; otherwise the fairness of the process is at risk. In a review of the literature, Ross[2] refers to 16 potential sources of bias*, Owen[17] reports 25. Both Wood *et al.*[10] and Pruthi *et al.*[18] list ten potential sources of bias.

Within the scope of our study, we investigated some of the most frequently discussed potential sources of bias: applicant's gender, nationality, discipline and institutional affiliation, i.e. the institution in which the research project is to be carried out. To identify the effect of every single potential source of bias, which could influence decisions of the Board of Trustees, we used multiple logistic regression models[19] and the statistical software package *Stata*[20]. Such models are appropriate for the analysis of dichotomous (or binary) responses. Dichotomous responses arise when the outcome is presence or absence of an event[21 p. 98]. In the case of the Boehringer Ingelheim Fonds, the binary response is coded 1 for approval and 0 for rejection of an application. If more than one independent variable, i.e. applicant's gender, nationality, discipline and institutional affiliation, was included in the logistic model, regression coefficients were estimated for all these variables and tested for their significance.

The Foundation had information on the applicant's scientific achievements up to the date of their application. We could therefore include not only the potential sources of bias as independent variables into the statistical analyses, but also the scientific performance of the applicants. We could thus distinguish between the influence of the latter and the potential sources of bias on the decisions of the Board.

The scientific performance indicators essentially comprise the criteria for approval and rejection of an application in the selection process of the Boehringer Ingelheim Fonds. (i) For applicants for a doctoral fellowship: the applicant's age at the time of his final degree; his final grades; the applicant's mobility during education, the number of recommendation letters as well as the votes of the external reviewers and members of the Foundation's staff. (ii) For applicants for a post-doctoral research fellowship: the applicant's age at the time of receiving his Ph. D., his grades, the applicant's mobility during education, the number of letters of recommendation, the number of publications by the time of application as well as the votes of external reviewers and the Foundation's staff.

Logistic regression analysis of the applications for a post-doctoral research fellowship showed that none of the examined potential sources of bias has a statistically significant influence on the decisions of the Board of Trustees. With regard to applications for a doctoral fellowship, the applicant's nationality did not statistically significantly affect the Board's decision. However, we detected a statistically significant influence of three variables hypothesized as potential biases: applicant's gender, discipline and intended institutional affiliation. The results on the selection process of the Foundation are therefore inconsistent: we found evidence for a gender, discipline and institutional bias in judging applications for doctoral, but not for post-doctoral fellowships. No bias with respect to nationality was found in either group.

To determine extent and direction of the influence of gender, discipline and intended institutional affiliation on the Board's decisions on doctoral fellowship allocations, we calculated the so-called predicted probabilities of approval and rejection respectively using *Stata*[20,22]. For the probability calculation, we first simulated a »typ-

* *Bias is defined as the influence of variables reflecting something other than the applicant's scientific merit. Such variables could be applicant's age, gender, institutional affiliation, or research field.*

ical« applicant, based on the average or most common features of all applicants for a doctoral fellowship. The »typical« applicant completed his university degree at the age of 26 with a final grade of 1.4 (best grade is 1.0). He attended more than one university during his education. In addition, he could submit two letters of recommendation with his application. Both the external reviewer and the Foundation's staff recommended him for an award. He is male, of German nationality and his first degree is in biology. He will pursue his research project at a German university (*Table 4*). This applicant's chances of receiving a scholarship are 50%, as determined by the probability computation (*Figure 1*).

If the »typical« applicant is not male, but female, the predicted probability of receiving a scholarship decreased from 50% to 33%. *Figure 1* shows that the impact of the applicant's discipline is still more important: if the applicant is not a biologist, but a chemist, the probability of approval declined from 50% to 25%. The opposite effect is observed for the institution in which the research project will be carried out: with regard to the decision of the Board of Trustees, it is obviously of advantage to choose an institute of the Max Planck Society (Germany) rather than of a German university. This choice increases the probability for approval by 17 percentage points. All in all, the results of the probability calculations point out that the Board of Trustees tend to rate particular applicant groups more highly than others.

To sum up, for applications for post-doctoral fellowships, we detected no statistically significant influence of the variables nationality, gender, discipline or institutional affiliation. For doctoral fellowships, we found no evidence for a nationality bias, but for a gender, discipline and institutional bias. This incongruent result reflects the inconsistent findings in other empirical studies investigating the fairness of peer review. For example, some studies examining gender bias in review processes point out that women scientists are at a disadvantage[23,24]. However, a similar number of studies merely report moderate or no gender effects[25,26]. Sonnert's[27 p. 47] experimental study even shows that women biologists received a better average evaluation than the men did (mean rating: 3.67 vs 3.27; $p = 0.0496$).

One principal problem that a survey of bias studies should take into account, and that affects bias research in general, is the lack of experimental studies in which individuals are randomly assigned to the contexts of interest, e.g. acceptance or rejection of submitted manuscripts. There are only very few attempts to study reviewer bias directly, i.e. in the natural setting of actual referee evaluations. Peters *et al.*[28], for instance, examined in a natural setting referees' evaluations of submitted manuscripts to American psychology journals. They looked for reviewer bias that could be attributed to their knowledge of the authors' institutions or names. As test materials they selected already published research articles by investigators from prestigious and highly productive American psychology departments. With fictitious names and institutions substituted for the original ones, the altered manuscripts were formally resubmitted to the journals that had originally refereed and published them. Eight of the nine altered articles were rejected. The bias study of Peters *et al.*[28] was however criticized for having violated ethical principles for research with human subjects[6,29-31].

The lack of experimentally derived findings makes it impossible to establish unambiguously whether work from a particular group of scientists receives better reviews (and thus has a higher approval rate) due to biases in the review and decision-making process, or if favourable review and greater success in the selection process is simply a consequence of the scientific merit of the corresponding group of applicants. In other words: the influence of institution, discipline and gender upon the Board's decisions could in fact be due to such factors as differences in the scientific quality of the research projects and/or the laboratories in which the projects are to be pursued.

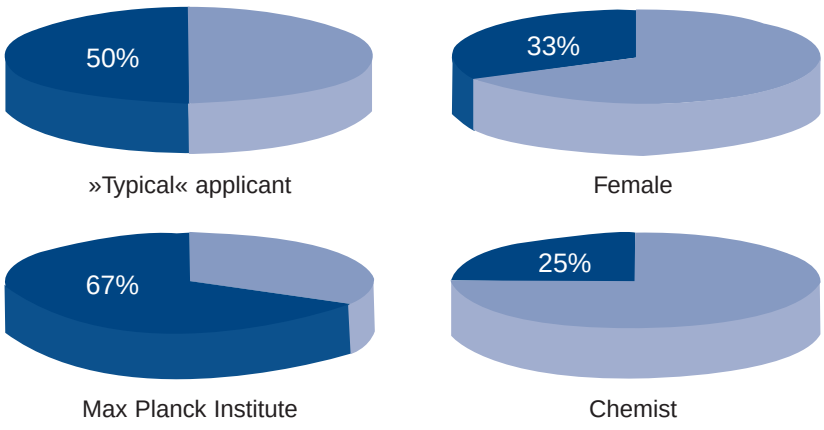

FIG. 1: Predicted probabilities for approval (dark blue segments of the circles) and rejection (light blue segments of the circles) of an application, taking applicant's gender, discipline or intended institutional affiliation into account (in per cent)

Predictive validity of committee peer review

• In the third part of our study, we examined the predictive validity of the selection process of the Foundation, i.e. whether indeed the »best« young scientists receive a fellowship. Assessing the predictive validity of decisions requires a generally accepted criterion for scientific merit. A conventional approach is to use citation counts as a proxy for research impact, since they measure the international impact of the work by individuals or groups of scientists on others[32 p. 293].

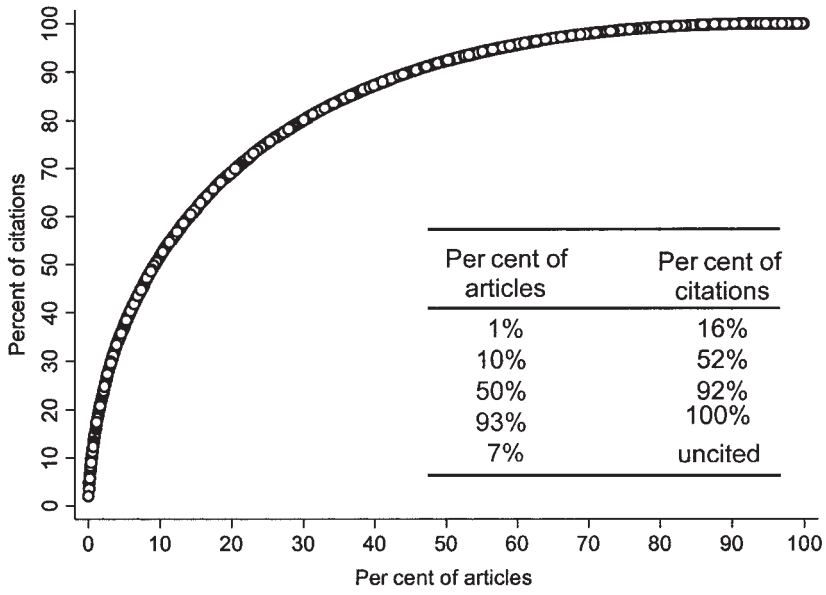

FIG. 2: Cumulative contribution of articles published by the fellowship holders of the Boehringer Ingelheim Fonds to the overall impact

»A highly cited work is one that has been found to be useful by a relatively large number of people, or in a relatively large number of experiments«[33 p. 363].

In June 2001, the Foundation asked all applicants who had been awarded a fellowship between 1985 and 1995 to send an up-to-date publication list. This list should comprise all publications since the date of approval of the fellowship to December 2000. Out of 433 Boehringer Ingelheim Fonds fellows, 225 (52%) sent their list of publications to the Foundation's secretariat. For each of the 225 fellows, the secretariat determined whether he or she worked in academia, i.e. publicly founded research, or in industry, or e.g. as a medical doctor, patent attorney or journalist. 141 (63%) of the 225 former scholarship holders had been working exclusively in academia. 84 (37%) had left academic research either immediately after completing their Ph.D. or a couple of years later. Since it can be assumed that only scientists working in academia continuously publish their results[34 p. 91], our bibliometric analyses used only the publication lists of scientists who had been working without interruption in academia.

All in all, 2,039 articles from 120 former fellowship holders were included in our analyses*. 98% of the articles were published in English and 2% in German or French. The articles were published in 508 different journals; in 36 journals, ten or more articles from fellows of the Foundation had appeared (*Table 5*). According to the Institute for Scientific Information (ISI, Philadelphia, PA, USA), the impact factors of these journals in the year 2000 varied between 32.440 (Cell) and 2.461 (Gene)**.

By the end of 2001, the 2,039 articles published between 1988 and 2000 had been cited altogether 82,099 times. The analysis of such a large number of articles always shows a highly skewed distribution of citations[35]. A large fraction of the citations is concentrated on a small fraction of the publications: the top 10% of the most frequently cited articles of the Boehringer Ingelheim Fonds fellows accounted for 52% of the citations (*Figure 2*). 7% (n = 145) of the articles were never cited; eight articles were cited more than 500 times. Due to the skewed distribution, the mean value of 40 citations (number of citations divided by the number of articles) and the median value of 17 citations (roughly half of the articles receive less and roughly half of them receive more citations) differ considerably. The citation frequency displays a distribution with a steep left flank, but a gradual slope on the right. The distribution can be approximated very well by the negative binomial distribution.

How do we know whether the citation rates for the publications of the Boehringer Ingelheim Fonds fellows are high or low? Van Raan[36] (Center for Science and Technology Studies, CWTS, Leiden, Netherlands) recommends a worldwide reference indicator for the bibliometric evaluation of research groups: »Our most important bibliometric indicator, the ›crown indicator‹, is a trend analysis over a period of, say, eight years, of the number of citations to the entire oeuvre of a research group or institute, normalized to an international field-specific reference value. In this way, we are able to demonstrate whether this group or institute is performing below or above, or even far above the international level of the research field(s) concerned«[36 p. 420]. The »crown indicator« was computed, for example, as a measure of scientific impact in an international comparative bibliometric study on the scientific performance of German medical research carried out by CWTS on behalf of the German Federal Ministry of Education and Research (BMBF, Berlin, Germany)[37].

To determine the »crown indicator« for the publications of the Boehringer Ingelheim Fonds fellows, we divided the »mean number of citations for publications from fellowship holders« by the »mean number of citations of all publications in the journal sets chosen by the fellowship

* *Of the 141 scholarship holders with an up-to-date publication list, 21 could not be included in the analysis, since the available data was incomplete.*

** *The ISI journal impact factor is a measure of the frequency with which the »average article« in a journal has been cited in a particular year. In 2000, the highest impact factor in the ISI journal's ranking list (considering all indexed journals) was achieved by the Annual Review of Immunology (50.340). In the list, Cell ranked third, Nature tenth and Science thirteenth.*

Journal	Journal impact factor in 2000	Number of articles
Proceedings of the National Academy of Sciences USA (PNAS USA)	10.789	103
Journal of Biological Chemistry	7.360	95
EMBO Journal	13.999	74
Nature	28.689	60
Development (Cambridge, England)	9.353	59
Cell	32.440	41
Science	23.872	39
Journal of Cell Biology	13.955	36
FEBS Letters	3.440	34
Nucleic Acids Research	5.396	28
Current Biology	8.393	28
Journal of Virology	5.930	26
Journal of Neuroscience	8.502	25
Journal of Molecular Biology	5.388	24
European Journal of Immunology	5.240	22
Molecular and Cellular Biology	9.669	21
Mechanisms of Development	4.154	21
Journal of Immunology	6.834	20
Genes & Development	19.676	20
Journal of Experimental Medicine	15.236	19
Gene	2.461	19
European Journal of Biochemistry	2.852	19
Biochemistry	4.221	19
Journal of Neurochemistry	4.900	17
Journal of Cell Science	5.996	17
Genomics	3.425	14
Pharmacogenetics	4.465	13
Oncogene	6.490	13
Neuron	15.081	12
Infection and Immunity	4.204	12
Trends in Biochemical Sciences	13.246	10
Neuroreport	2.696	10
Human Molecular Genetics	9.048	10
European Journal of Neuroscience	3.862	10
Biochemical and Biophysical Research Communications	3.055	10
Other journals (altogether 472 different journals, each with less than ten articles)		1,009

TAB. 5: *Journals, in which ten or more articles from fellows of the Boehringer Ingelheim Fonds had appeared after approval of their fellowships (ISI journal impact factor in 2000, n = 2,009. 30 articles published in B.I.F. FUTURA are not included since ISI does not index this journal)*

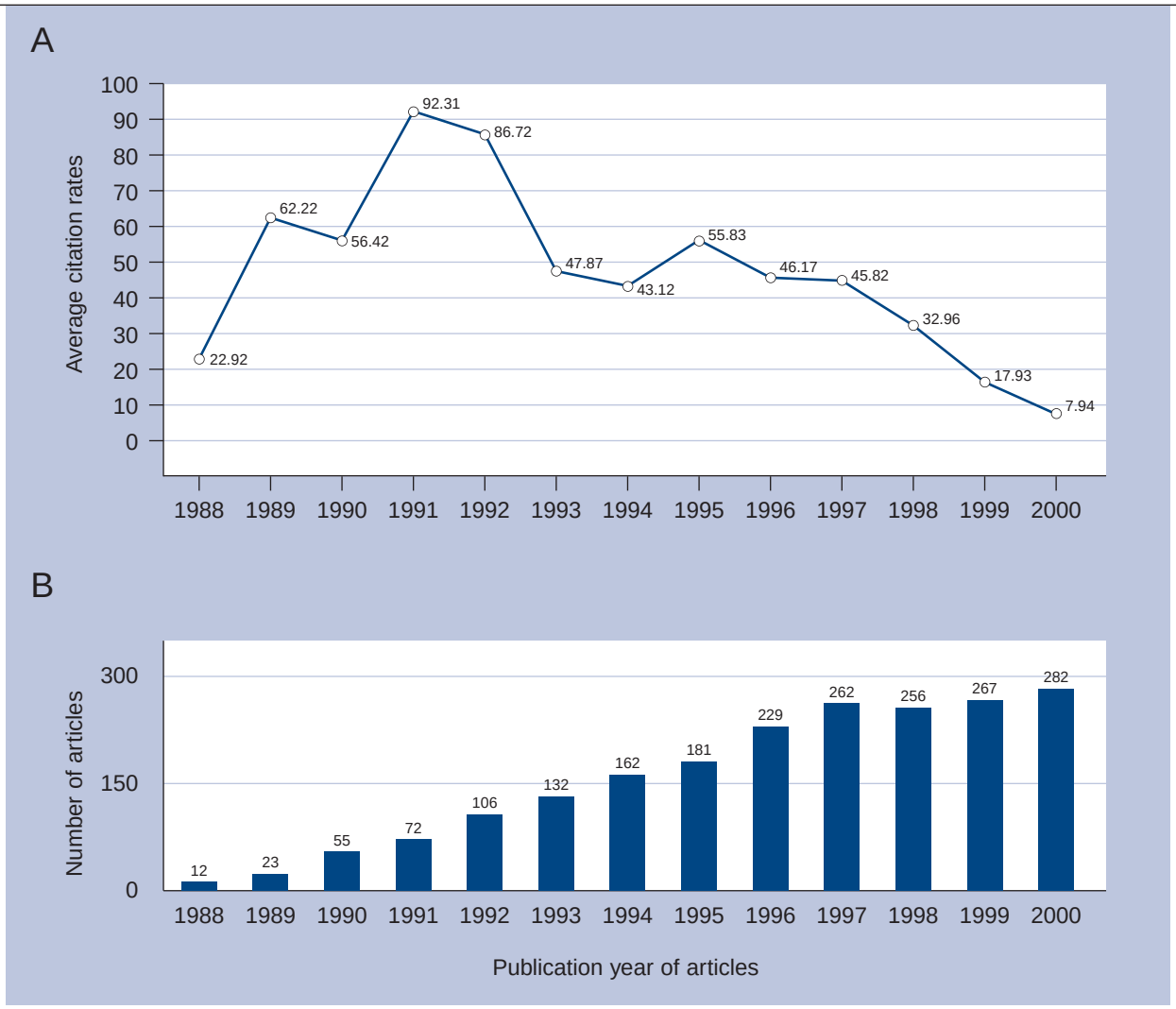

FIG. 3: (Top) Mean number of citations of the articles published by the Boehringer Ingelheim Fonds fellows by the end of 2000. For example, each of the 72 articles published in 1991 was cited on average 92.31 times by the end of 2001. (Bottom) Number of articles published in the year indicated

holders«. The quotient enables us to determine whether the citation impact of the fellowship holders is far below (indicator value < 0.5), below (indicator value 0.5 - 0.8), around (0.8 - 1.2), above (1.2 - 1.5), or far above (> 1.5) the international (western world-dominated) citation impact baseline for the chosen journal sets. With ratio values above 1.5, the probability of identifying very good to excellent researchers is very high[38].

Figure 3 (top) shows the mean number of citations of the articles published by the Boehringer Ingelheim Fonds fellows by the end of 2001. For example, each of the 72 articles published in 1991 was cited on average 92.31 times by the end of 2001, and each of the 282 articles published in 2000 was cited on average 7.94 times by the end of 2001. To calculate the »crown indicators«, we used the ISI journal sets »Multidisciplinary«*, »Molecular Biology & Genetics«** and »Biology & Biochemistry«***. Out of the 22 ISI journal sets**** we chose »Molecular Biology & Genetics« and »Biology & Biochemistry« as reference sets, since 77% of the former scholarship holders are biologists (61%) or biochemists (16%). Moreover, about a third of the research projects were in the field of molecular biology. In addition, we included the journal set »Multidisciplinary«, since a large number of papers from Boehringer Ingelheim Fonds fellows were published in the Proceedings of the National Academy of Sciences USA, Science and Nature (*Table 5, p. 13*)***** which ISI sorted into this journal set.

Table 6 (p. 17) lists the »crown indicators« of the publications classified according to journal set and year of publication. The values show that the papers of the fellowship holders were on average significantly more frequently cited than the »average« pub-

* *»Multidisciplinary« category includes journals of a broad or general character in the sciences and covers the spectrum of major scientific disciplines (e.g. Nature, Proceedings of the National Academy of Sciences USA, Science) (http://www.isinet.com/rsg/esi/).*

** *»Molecular Biology & Genetics« contains e.g. Annual Review of Cell Biology, Cell, Annual Review of Cell and Developmental Biology (http://www.isinet.com/rsg/esi/).*

*** *»Biology & Biochemistry« includes Annual Review of Biochemistry, Physiological Reviews, Endocrine Reviews (http://www.isinet.com/rsg/esi/).*

**** *Agricultural Sciences; Biology & Biochemistry; Chemistry; Clinical Medicine; Computer Science; Ecology/Environment; Economics & Business; Engineering; Geosciences; Immunology; Material Sciences; Mathematics; Microbiology; Molecular Biology & Genetics; Multidisciplinary; Neuroscience & Behavior; Pharmacology & Toxicology; Physics, Plant & Animal Science; Psychology/Psychiatry; Social Sciences, general; Space Science (http://www.isinet.com/rsg/esi/).*

***** *A comparison with other journal sets, for example »Clinical Medicine« or »Microbiology«, shows that the »average« publication in the journal sets »Multidisciplinary«, »Molecular Biology & Genetics« and »Biology & Biochemistry« has a much higher mean citation rate.*

lication in one of the three journals sets: 21 of the 30 »crown indicators«, shown in *table 6*, are above 1.5 (between 1.52 and 4.01) and seven are between 1.2 and 1.5*. Only two values (0.96 and 1.02) are in the range which van Raan[38] denotes as »average«. In the light of the mean citation rate achieved by the articles of the Boehringer Ingelheim Fonds fellows, the decisions of the Foundation's Board have a high predictive validity. Similar results were reported for the decisions of the editors of the Journal of Clinical Investigation[39], British Medical Journal[40] and Angewandte Chemie[41]: »Based on mean citation rates for accepted manuscripts and rejected manuscripts that were nevertheless published elsewhere, editorial decisions in all the existing studies reflect a high degree of predictive validity«[41 p. 56]. In addition, Chapman *et al.*[42] reported similar findings for quality ratings of graduate fellows funded by the National Science Foundation.

This high predictive validity of the selection process of the Boehringer Ingelheim Fonds is further substantiated by a second validity criterion: only 2% of the research fellowship holders, who applied for a doctoral fellowship between 1985 and 2000, returned the award while the other 98% successfully finished their research project and submitted their doctoral thesis. By comparison, the Wellcome Trust (London, UK)[43], another renowned biomedical research charity, reports that 8% of their research fellowship holders did not complete a Ph.D.

Proposals for optimizing committee peer review

• All in all, the results show that the selection process of the Boehringer Ingelheim Fonds is highly valid, that is to say, it achieves its objective to select the best young scientists. However, our study found some evidence that three potential sources of bias (gender, discipline and institutional affiliation) may influence the decisions of the Board of Trustees. It will presumably never be possible to eliminate all doubts regarding the fairness of the reviewing process. For this reason, it would surely be prudent to consider the following four proposals for optimizing committee peer review. The proposed measures are feedback, internal monitoring, programme manager and triage or prescreening. However, we do not recommend one frequently mentioned measure: the implementation of a system of quotas for certain groups of applicants. The Stereotype Threat Model[44] suggests that when people belonging to minority groups perform a difficult task in an area in which their group is considered weak, they are scared of confirming the stereotype. This psychological pressure will lead them to under-perform[45]. In the peer review process, certain groups of applicants might consider quotas as a »signal« that the funding agency expects lower scientific performance. Stereotypes could be activated in applicants, leading to applications of lower quality.

Feedback

• Many studies point out that applicants would welcome a detailed, well-founded feedback about the assessment of their application[15,46]. Albeit, the Boehringer Ingelheim Fonds, like many other funding organizations, does not give detailed feedback to applicants. It simply informs applicants briefly about approval or rejection. According to Lock[40], the reason for not providing feedback is simply that it would involve too much time, energy and money. Financial resources should be used for promotion of research and not for the selection process. However, Klahr[13] emphasizes the need for feedback, since many funding agencies – due to a limited budget and an increasing number of applications – must reject applications with favourable ratings.

Basically, there are two possibilities for including feedback in the peer review system. The first is to allow an applicant to rebut the reviewers' comments *before* the final decision about approval and rejection. »If the applicant discovers that a mistake was made in the evaluation, fairness demands that he or she have an opportunity to correct the error and to rebut the decision not to fund before the final funding decision is made«[47 p. 65]. The Royal National Institute for the Blind (London, UK), for example, send reviewers' statements to the applicant for comment. »This approach allows the applicant to identify factual errors, prioritise criticisms, and highlight what is unique about his or her application. This additional layer of discussion facilitates the working of the committee, which inevitably does not have the specific expertise required to appraise disparate external reviews. As a consequence, we have funded projects that, without this mechanism, would have been rejected, and without the delay incurred by an appeal«[48 p. 1063].

The second possibility is to inform applicants on the reasons for approval and rejection after the final decision, as practiced by, for example, the National Endowments for the Arts and Humanities (NEA, now National Endowment for the Arts, Washington, DC, USA): »NEA encourages rejected applicants to contact the program specialists to their project for explanations and suggestions for future applications. A summary of the relevant panel's deliberations is available to any applicant that requests one«[47 p. 67].

In recent years, a number of funding agencies have included feedback into the assessment process, and information is available as to whether the proposed measures have proved useful. In a postal survey, Moxham *et al.*[49] questioned applicants, successful and unsuccessful, on receiving feedback from the Wellcome Trust. The results show that over 80% found the feedback helpful. »Most interesting were the reasons given as to why this was so. Most commonly, feedback helped improve subsequent *presentations* [accentuation of the author] of proposals (82%) whilst fewer

* *The average citation rates of articles published by the fellows between 1988 und 1990, are not listed in table 6, since ISI does no longer provide the corresponding average citation rates for papers published in these years.*

claimed it helped to correct errors of thinking (24%) or reoriented their approach to research (11%) [...]. The reviewers were also enthusiastic about providing feedback; 71% were fully supportive of the idea and 98% were supportive on balance. 83% were not worried that their reports might be seen by applicants«[49 p. 12]. According to Smith[50 p. 691], feedback in the selection process of the National Institutes of Health (NIH, Bethesda, MD, USA) also improved the reviews and »comments on grants are now more detailed and careful and the whole process much more educational«. For the Association of Medical Research Charities (London, UK)[51], feedback helps to raise the standards of applications and the quality of the science. »Feedback also contributes towards openness and accountability of the peer review system and it is important that reviewers are honest and fair in their comments«[51 p. 14]. However, none of the studies published so far examined whether feedback actually did improve subsequent applications.

Internal monitoring

• The peer review process should be both monitored and examined to ensure that the funding agency's objectives are met. It is therefore important that agencies have access to comprehensive and reliable management information systems[52]. Equipped with these systems, agencies can more easily and continuously monitor the extent to which fairness and reliability are achieved. However, agencies should not only monitor the selection process itself, but also the success of the process. On the basis of the internal monitoring, funding agencies should develop, for example, guidelines for reviewers, »that warn of potential biases and suggest that reviewers try to avoid them. This may appear too simplistic, but it is a cost-effective strategy that could result in the significant reduction of unfair biases«[53 p. 167].

In its efforts to assess whether its research funds have been invested »wisely«, the Wellcome Trust has established a Research Outputs Database (ROD)[52 p. 45]. ROD was started by the Wellcome Trust in 1993 and is now operated by the City University London (UK) on contract from the Wellcome Trust. The database provides quantitative data on the published output of researchers, and links this output to the sources of research funds[54 p. 4].

Programme manager

• In the review process of journals, editors make the final decision about acceptance or rejection of manuscripts on the basis of referees' recommendations. »Where possible, peers should not make the final decisions but should advise the decision makers, who can filter peer self-interest from peers' recommendations. As a fractious horse is only as good as its rider, peer review is only as good as the program managers ... who use it, but these people are visible and can be called to account for their decisions«[55 p. 40]. The programme manager should choose the appropriate reviewers for the received application and should *calibrate* their recommendations. An application sent to equally eminent reviewers may well result in quite different reports despite mutual agreement with respect to the quality of the application. It is important to understand that the first reviewer rates the most brilliant proposal as only »very good« while the second reviewer with the same quality judgement will be off scale with praise. The assignment of a »program officer may be the single most important step in obtaining the most knowledgeable and fair review«[52 p. 145].

In recent years, some research funding agencies have adopted this procedure. In the selection process of the Engineering and Physical Sciences Research Council (EPSRC, Swindon, UK), for example, the programme manager is responsible »to the Chief Executive for meeting the objectives of his/her programme. On receiving the panel's advice, the programme manager constructs a list of successful proposals on the basis of: (i) the money made available by the EPSRC Council, (ii) the rank ordering produced by the panel, and (iii) the guidance of the panel chairman (particularly in borderline cases). The controversial point is the role of the programme managers who, despite having no direct involvement in the making of value judgements about the scientific quality of proposals, are seen as potentially exercising undue influence over the process as a whole«[56 p. 7]. In the selection process of the National Endowment for the Humanities (NEH, Washington, DC, USA), a programme officer corrects panelists who substitute reputation for merit[52 p. 102]. Even if the EPSRC and the NEH have operated with a programme manager over a long period, information on the extent to which the programme manager in fact improves the fairness of the selection process are not available.

Triage and pre-screening

• If funding agencies like the Boehringer Ingelheim Fonds have a high rejection rate in the review process, reviewers spend much unpaid time evaluating applications that will be unsuccessful. According to Gavaghan[57], frustration among reviewers and problems in recruiting them in the first place may be the consequence. Therefore, a form of triage seems desirable in the selection process »in which not all grants receive the full process and deliberations of the full committee, but are rejected at an earlier stage«[10 p. 32]. The goal is to allow peer reviewers to spend more time on top proposals and less effort reviewing – and re-reviewing – grants that are unlikely ever to get funded and to make reviewing a more satisfying experience«[58 pp. 1212-1213]. According to Marshall[58 p. 1213], applicants who are rejected using triage get the message »that this is not an application that can be moved into the fundable category simply by responding to a series of complaints«.

Triage has been used at the NIH since 1988, after a pilot study of reviewers had suggested that they are in favour of triage[58]. The study of

Vener *et al.*[59 p. 1312] with empirical data from the National Cancer Institute (NCI, Bethesda, MD, USA) shows that »the conservative model [of the NIH] is valid such that the likelihood of eliminating a highly competitive application from consideration for funding is remotely small. With the model, the process of triage is fair to applicants on the one hand and is also effective in reducing consultant workloads on the other.«

Some funding agencies have been using a two-step procedure (pre-screening) for several years: »The Wellcome Trust uses an abbreviated form for pre-screening by peers of potential applicants for a number of its fellowship schemes; only the stronger applicants are invited subsequently to submit full proposals. The Medical Research Council (MRC) has a similar screen for programme grants. The Biotechnology and Biological Sciences Research Council (BBSRC), however, prefers not to use outline proposals for pre-screening for quality (as opposed to eligibility), because they may contain insufficient information for peer review and because they might encourage larger numbers of speculative proposals, thus defeating the object of diminishing the burden on the peer review system«[56 p. 5]. The review process of the Human Frontier Science Program (HFSP, Strasbourg, France) has been modified into a two-step procedure consisting of a letter of intent phase and the review of invited full applications. »This two-step process enables the review committees to identify those applications that will be likely to succeed and greatly reduces the amount of work by applicants«[60]. The Swiss

Year of Publication	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*	*2000*
Set of »Multidisciplinary« journals										
Mean of citations for the publications of fellowship holders from the year of publication to 2001	92.31 (n=72)[1]	86.72 (n=106)	47.87 (n=132)	43.12 (n=162)	55.83 (n=181)	46.17 (n=229)	45.82 (n=262)	32.96 (n=256)	17.93 (n=267)	7.94 (n=282)
Baseline[2] for the journal set	50.00	47.31	49.86	42.17	44.40	36.71	31.09	21.67	14.75	6.33
Crown indicator (mean of citations divided by baseline)	1.85	1.83	0.96	1.02	1.26	1.26	1.47	1.52	1.22	1.25
Set of »Molecular Biology & Genetics« journals										
Mean of citations for the publications of fellowship holders from the year of publication to 2001	92.31 (n=72)[1]	86.72 (n=106)	47.87 (n=132)	43.12 (n=162)	55.83 (n=181)	46.17 (n=229)	45.82 (n=262)	32.96 (n=256)	17.93 (n=267)	7.94 (n=282)
Baseline[2] for the journal set	40.16	38.22	36.83	32.63	28.09	23.33	20.27	15.65	10.54	4.55
Crown indicator (mean of citations divided by baseline)	2.30	2.27	1.30	1.32	1.99	1.98	2.26	2.11	1.70	1.75
Set of »Biology & Biochemistry« journals										
Mean of citations for the publications of fellowship holders from the year of publication to 2001	92.31 (n=72)[1]	86.72 (n=106)	47.87 (n=132)	43.12 (n=162)	55.83 (n=181)	46.17 (n=229)	45.82 (n=262)	32.96 (n=256)	17.93 (n=267)	7.94 (n=282)
Baseline[2] for the journal set	23.04	22.30	20.79	19.24	16.44	13.89	11.85	8.88	5.77	2.56
Crown indicator (mean of citations divided by baseline)	4.01	3.89	2.30	2.24	3.40	3.32	3.87	3.71	3.11	3.10

TAB. 6: Average citation rates of papers published by fellowship holders of the Boehringer Ingelheim Fonds compared to mean citation rates of publications in the journal sets »Multidisciplinary«, »Molecular Biology & Genetics« and »Biology & Biochemistry« by publication year (1991-2000). ([1]n = number of publications; [2]baselines are measures of cumulative citation frequencies across all papers published by a journal set: an average of 50.00 for the journal set »Multidisciplinary« in 1991 means that, on average, papers in »Multidisciplinary« journals were cited 50.00 times from 1991 to the end of 2001)

National Science Foundation (SNF, Bern, Switzerland) has been using a similar two-step selection process since 1999[61 pp. 181-183].

Conclusion

• In the first comprehensive study on committee peer review for the selection of doctoral (Ph.D.) and post-doctoral research fellowships, we analysed the selection process of the Boehringer Ingelheim Fonds with regard to its reliability, fairness and predictive validity (n = 2,697). The most important aspect was to test the predictive validity, i.e. whether the Foundation achieves its aim to select the best young scientists. Our bibliometric analysis showed that this is indeed the case and that the selection process is thus highly valid.

In the analysis of reliability, the degree of agreement between reviewers was determined. In 76% of the cases, the decision whether to award a scholarship or not was characterized by agreement. With regard to fairness, we analysed whether potential sources of bias, i.e. gender, nationality, discipline, and institutional affiliation, could have influenced the decisions. For post-doctoral fellowships, no statistically significant influence of any of these variables could be observed. In applications for a doctoral fellowship, evidence of a gender, discipline and institutional bias, but not of a nationality bias, was found. We therefore present four proposals for optimizing committee peer review which could improve the fairness of the process (feedback, internal monitoring, programme manager as well as triage or pre-screening).

A number of research funding agencies have already implemented some of these proposals into their review process. In addition, information on the degree to which some of the proposed measures have proved useful are available. An internal monitoring system and a form of triage or pre-selection could be included into the selection process of foundations at no great expense. In view of the technical possibilities of modern management information systems, continuous analysis of electronically available archive data and computer-aided supervision of the selection process is nowadays indispensable. A triage or pre-selection in the selection process of applicants should be introduced, since we can assume that the number of applicants for scholarships will continue to rise in the years to come. In particular, since the Wissenschaftsrat (German Scientific Council, Köln, Germany)[62 p. 76] considers scholarships to be a better instrument for supporting Ph.D. students than regular employment at a university. Medium-term, the Wissenschaftsrat recommends to increase the number of Ph.D. students sponsored by scholarships.

Feedback to applicants after the selection process can only be realized with substantial additional financial resources. By means of feedback during the selection process, however, this optimization could be achieved at little additional financial cost. If, for example, the Foundation made the external expert reports for review of mistakes and for comments accessible to applicants, the Board of Trustees could take these comments into consideration during the selection process.

A programme manager could be employed only at great financial expense to the Foundation. Even though a programme manager can be an important corrective of expert recommendations, like the editor of scientific journals, the question is whether the benefit to the selection process justifies such high costs. We therefore suggest that a programme manager should only be considered if and when the internal monitoring system indicates biases in the selection process that cannot be eliminated by notices or guidelines for the experts or trustees.

Our bibliometric analysis showed that the selection process of the Boehringer Ingelheim Fonds is highly valid. Additional studies could further substantiate the validity of the Foundation's selection process by analysing other success criteria, such as the applicants' professional career[42,43] or by statistics on third-party funds and patents[63] of former scholarship holders. This would also provide information on the interrelation between different indicators of success.

References

1. *Polanyi, M* (1966) The tacit dimension. ***New York; NY, USA: Doubleday***
2. *Ross, PF* (1980) The sciences' self-management: manuscript refereeing, peer review, and goals in science. ***Massachusetts, MA, USA: The Ross Company, Todd Pond***
3. *Campanario, JM* (1998) Peer review for journals as it stands today – part 1. ***Sci. Commun. 19, 181-211***
4. *Campanario, JM* (1998) Peer review for journals as it stands today – part 2. ***Sci. Commun. 19, 277-306***
5. *Overbeke, J, Wager, E* (2003) The state of the evidence: what we know and what we don't know about journal peer review. In: Godlee, F, Jefferson, T (eds.) Peer review in health sciences. ***London, UK: BMJ Books, 45-61***
6. *Weller, AC* (2001) Editorial peer review: its strengths and weaknesses. ***Medford, USA: Information Today***
7. *Bornmann, L, Daniel, HD* (2003) Begutachtung durch Fachkollegen in der Wissenschaft. Stand der Forschung zur Reliabilität, Fairness und Validität des Peer-Review-Verfahrens. In: Schwarz, S, Teichler, U (eds.) Universität auf dem Prüfstand. Konzepte und Befunde der Hochschulforschung. ***Frankfurt am Main, Germany: Campus, 211-230***
8. *Demicheli, V, Di Pietrantonj, C* (2004) Peer review for improving the quality of grant applications (Cochrane methodology review). In: The Cochrane library, Issue 1. ***Chichester, UK: John Wiley & Sons, Ltd.***
9. *Wessely, S* (1998) Peer review of grant applications – what do we know? ***Lancet 352, 301-305***
10. *Wood, FQ, Wessely, S* (2003) Peer review of grant applications: a systematic review. In: Godlee, F, Jefferson, T (eds.) Peer review in health sciences. ***London, UK: BMJ Books, 14-44***
11. *Bornmann, L* (2004) Stiftungspropheten in der Wissenschaft. Zuverlässigkeit, Fairness und Erfolg des Peer-Review. ***Münster, Germany: Waxmann***
12. *Fröhlich, H* (2001) It all depends on the individuals. Research promotion – a balanced system of control. ***B.I.F. FUTURA 16, 69-77***
13. *Klahr, D* (1985) Insiders, outsiders and efficiency in a national science foundation panel. ***Amer. Psychol. 40, 148-154***

14. *Hartmann, I, Neidhardt, F* (1990) Peer review at the Deutsche Forschungsgemeinschaft. ***Scientometrics 19, 419-425***

15. *Cicchetti, D* (1991) The reliability of the peer review for manuscript and grant submissions: a cross-disciplinary investigation. ***Behav. Brain Sci. 14, 119-135***

16. *Hodgson, C* (1997) How reliable is peer review? An examination of operating grant proposals simultaneously submitted to two similar peer review systems. ***J. Clin. Epidemiol. 50, 1189-1195***

17. *Owen, R* (1982) Reader bias. ***JAMA 247, 2533-2534***

18. *Pruthi, S, Jain, A, Wahid, A, Mehra, K, Nabi, S* (1997) Scientific community and peer review system – a case study of a central government funding scheme in India. ***J. Sci. Ind. Res. India 56, 398-407***

19. *Hosmer, DW, Lemeshow, S* (2000) Applied logistic regression. ***New York, NY, USA: John Wiley & Sons, Inc.***

20. *StataCorp.* (2003) Stata statistical software: release 8. ***College Station, Texas, TX, USA: Stata Corporation***

21. *Rabe-Hesketh, S, Everitt, B* (2004) A handbook of statistical analyses using Stata. ***Boca Raton, UK: Chapman & Hall/CRC***

22. *Long, JS, Freese, J* (2003) Regression models for categorical dependent variables using Stata. ***College Station, Texas, TX, USA: Stata Corporation***

23. *Brouns, M* (2000) The gendered nature of assessment procedures in scientific research funding: the Dutch case. ***Higher Education in Europe 25, 193-199***

24. *Wennerås, C, Wold, A* (1997) Nepotism and sexism in peer-review. ***Nature 387, 341-343***

25. *Cole, S* (1992) Making science. Between nature and society. ***Cambridge, MA, USA: Harvard University Press***

26. *Ward, JE, Donnelly, N* (1998) Is there gender bias in research fellowships awarded by the NHMRC? ***Med. J. Australia 169, 623-624***

27. *Sonnert, G* (1995) What makes a good scientist? Determinants of peer evaluation among biologists. ***Soc. Stud. Sci. 25, 35-55***

28. *Peters, DP, Ceci, SJ* (1982) Peer-review practices of psychological journals: the fate of published articles, submitted again. ***Behav. Brain Sci. 5, 187-195***

29. *Chubin, DE* (1982) Reforming peer review: from recycling to reflexivity. ***Behav. Brain Sci. 5, 204***

30. *Honig, WM* (1982) Peer review in the physical sciences: An editor's view. ***Behav. Brain Sci. 5, 216-217***

31. *Fleis, JL* (1982) Deception in the study of the peer-review process. ***Behav. Brain Sci. 5, 210-211***

32. *Cole, JR* (2000) A short history of the use of citations as a measure of the impact of scientific and scholarly work. In: Cronin, B, Barsky Atkins, H (eds.) The web of knowledge. A festschrift in honor of Eugene Garfield. ***Medford, USA: Information Today, 281-300***

33. *Garfield, E* (1979) Is citation analysis a legitimate evaluation tool? ***Scientometrics 1, 359-375***

34. *Weingart, P* (2001) Die Stunde der Wahrheit? Zum Verhältnis der Wissenschaft zu Politik, Wirtschaft und Medien in der Wissensgesellschaft. ***Weilerswist, Germany: Velbrück***

35. *Lotka, AJ* (1926) The frequency distribution of scientific productivity. ***J. Washington Acad. Sci. 16, 317-323***

36. *van Raan, AFJ* (1999) Advanced bibliometric methods for the evaluation of universities. ***Scientometrics 45, 417-423***

37. *Tijssen, RJW, van Leeuwen, TN, van Raan, AFJ* (2002) Mapping the scientific performance of German medical research. An international comparative bibliometric study. ***Stuttgart, Germany: Schattauer***

38. *van Raan, AFJ* (2003) The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. ***Technikfolgenabschätzung 12, 20-29***

39. *Wilson, JD* (1978) Peer review and publication. ***J. Clin. Invest. 61, 1697-1701***

40. *Lock, S* (1985) A difficult balance: editorial peer review in medicine. ***Philadelphia, PA, USA: ISI Press***

41. *Daniel, HD* (1993) Guardians of science. Fairness and reliability of peer review. ***Chichester, UK: John Wiley & Sons, Ltd.***

42. *Chapman, GB, McCauley, C* (1994) Predictive validity of quality ratings of national science foundation graduate fellows. ***Educ. Psychol. Meas. 54, 428-438***

43. *Wellcome Trust (ed.)* (2001) Review of Wellcome Trust PhD research training. Career paths of a 1988-1990 prize student cohort. ***London, UK: Wellcome Trust***

44. *Steele, CM, Aronson, J* (1995) Stereotype threat and the intellectual test performance of African-Americans. ***J. Personal. Soc. Psychol. 69, 797-811***

45. *Cadinu, M, Maass, A, Frigerio, S, Impagliazzo, L, Latinotti, S* (2003) Stereotype threat: the effect of expectancy on performance. ***Eur. J. Soc. Psychol. 33, 267-285***

46. *Over, R* (1996) Perceptions of the Australian Research Council large grants scheme: differences between successful and unsuccessful applicants. ***Austr. Edu. Res. 23, 17-36***

47. *McGarity, T* (1994) Peer Review in awarding federal grants in the arts and sciences. ***High Technol. Law J. 9, 1-92***

48. *Fielder, A, Vinyard, H* (1998) Peer review of grant applications. ***Lancet 352, 1063***

49. *Moxham, H, Anderson, J* (1992) Peer review: a view from the inside. ***Sci. Tec. Pol. 12, 7-15***

50. *Smith, R* (1988) Glimpses of the National Institutes of Health II. Review systems and evaluation. ***Brit. Med. J. 296, 691-695***

51. *Association of Medical Research Charities (eds.)* (1995) First report of the working party on the implementation of peer review. ***London, UK: Association of Medical Research Charities***

52. *Wood, FQ* (1997) The peer review process (Commissioned Report No. 54). ***Canberra: Australian Government Publishing Service***

53. *Zentall, TR* (1991) What to do about peer review: is the cure worse than the disease? ***Behav. Brain Sci. 14, 166-167***

54. *Dawson, G, Lucocq, B, Cottrell, R, Lewison, G* (1998) Mapping the landscape. National biomedical research outputs 1988-95. ***London, UK: The Wellcome Trust***

55. *McCutchen, C* (1991) Peer review: treacherous servant, disastrous master. ***Technol. Rev. 94, 30-40***

56. *Royal Society (ed.)* (1995) Peer review. An assessment of recent developments. ***London, UK: Royal Society***

57. *Gavaghan, H* (1994) Cautious welcome to NIH peer review reforms. ***Nature 369, 269***

58. *Marshall, E* (1994). NIH tunes up peer review. ***Science 263, 1212-1213***

59. *Vener, KJ, Feuer, EJ, Gorelic, LA* (1993) A statistical model validating triage for the peer review process: keeping the competitive applications in the review pipeline. ***FASEB Journal 7, 1312-1319***

60. *Krotoski, D* (2003) The Human Frontier Science Program (HFSP): how to recognize and fund excellence in the young. In: Max Planck Society (ed.) Science between evaluation and innovation: a conference on peer review. ***Munich, Germany: Max Planck Society, 175-178***

61. *Diggelmann, H* (2003) The promotion of young scientists by the Swiss National Science Foundation (SNF). In: Max Planck Society (ed.) Science between evaluation and innovation: a conference on peer review. ***Munich, Germany: Max Planck Society, 179-184***

62. *Wissenschaftsrat* (2002) Empfehlungen zur Doktorandenausbildung. ***Cologne, Germany: Geschäftsstelle des Wissenschaftsrates***

63. *Hornbostel, S* (1991) Drittmittel im Fach Physik – ein Indikator für Forschungsleistung? ***Phys. Bl. 47, 123-125***