



Risk evaluation in peer review of grant applications

Stephen Gallo¹ · Lisa Thompson¹ · Karen Schmalting² · Scott Glisson¹

Published online: 24 February 2018
© The Author(s) 2018. This article is an open access publication

Abstract

The process of peer review is used to identify the most scientifically meritorious research projects for funding. Impact and innovation are among the criteria used to determine overall merit. A criticism of peer review has been the perception that reviewers are biased against innovation, such as one study that found reviewers to systematically assign poorer scores to highly novel work. Moreover, reviewers' definitions for excellent research and paradigm-shifting research are different; innovative research may not always be considered excellent. Therefore, it is clear more needs to be done to understand the decision-making processes of reviewers in evaluating risk and innovation in research. In an effort to address this gap, the American Institute of Biological Sciences developed a comprehensive peer review survey that examined, in part, the differences in applicant and reviewer perceptions of review outcomes. The survey was disseminated to 13,091 reviewers and applicants, of whom 9.4% responded. Only 24% of respondent applicants indicated that innovation was addressed in their review feedback, while 81% of respondent reviewers indicated they factored innovation into selecting the best science and 73% viewed innovation as an essential component of scientific excellence. Similarly, while only 27% of respondent applicants reported receiving comments on the riskiness of their grant applications, 58% of respondent reviewers indicated that the risks associated with innovative research impacted the scores they assigned to the grant applications. These results indicate a potential source of bias in how innovation and risk are evaluated in grant applications.

Keywords Peer review · Innovation · Research funding · Grant applications · Novelty · Survey

1 Introduction

Peer review serves a gatekeeping function for the scientific community. It is the process through which research is selected for funding by subject matter expert review panels and communicated through publication in refereed journal articles. Peer reviewers judge proposals on the basis of multiple criteria. For example, when performing reviews for US federal scientific research organizations such as the National Institutes of Health (NIH 2016a) and the National Science Foundation (NSF 2017), reviewers typically determine an overall score, as well as scores for component criteria that usually include significance, innovation, methodology,

investigators and the research environment. In a study of NIH proposals, Eblen et al. (2016) found that reviewers' component scores for approach and methodology were most strongly related to the overall impact scores, followed by component scores for significance and innovation. In assessing the approach, it has been emphasized that reviewers focus on the "overall strategy, methodology, and analyses" to be "well-reasoned and appropriate to accomplish the specific aims of the project," that "potential problems, alternative strategies, and benchmarks for success" are presented and that the strategy "establishes feasibility" and manages "particularly risky aspects" (NIH 2016a). Innovation, on the other hand, has been defined as the extent to which "the proposed activities suggest and explore creative, original or potentially transformative concepts" (NSF 2017) and to which the application "challenges and seeks to shift current research or clinical practice paradigms by utilizing novel theoretical concepts, approaches or methodologies, instrumentation, or interventions," as well as whether the application is "a refinement, improvement, or new application of theoretical concepts, approaches or methodologies,

✉ Stephen Gallo
sgallo@aibs.org

¹ American Institute of Biological Sciences, Scientific Peer Advisory and Review Services, 1313 Dolley Madison Blvd. Suite 402, Mclean, VA 22101, USA

² Washington State University, 14204 NE Salmon Creek Avenue, Vancouver, WA 98686, USA

instrumentation, or interventions proposed” (NIH 2016a). These sample definitions of innovation and approach/methodology suggest tension and perhaps competing, somewhat opposing values between these review criteria, particularly with regard to the tolerance of the riskiness of the project, i.e., whether the goals will likely be accomplished based on what has been proposed. This potential competition is reflected in the research community, e.g., between emphasizing innovative research which may have inherent risk of failure but potentially high scientific payoff (e.g., Sorlie et al. 2012) and the needs for replication, reproducibility, and smaller, incremental advances (Ioannidis 2005). Moreover, reviewers associate the overall score with approach more than they do innovation (Eblen et al. 2016), which may also result in the overall scores assigned to proposals correlating poorly with the citation impact of the funded projects, presuming highly innovative work yields greater citation success (Danthi et al. 2014; Gallo et al. 2014; Li and Agha 2015).

In addition to the theoretical tension between the definitions and potentially opposing values of an application’s level of innovation and the quality and riskiness of its approach/methodology, it may be that there is considerable variation in the interpretation of these two criteria by reviewers. Some literature has suggested significant variance in reviewer preference and even biases for different types of research (Lamont 2009; Lee et al. 2013; Travis and Collins 1991). Indeed, there is evidence to suggest a large degree of subjectivity in reviewer decision-making, as inter-reviewer reliability of the same grant has been low (Cole et al. 1981, Cicchetti 1991; Bornmann and Daniel 2008). While many scientists would agree that the most exciting, innovative research should be identified through peer review and subsequently funded, it is not at all clear that there is a consensus on how, in practice, reviewers decide what that innovation should look like. Some have suggested that reviewers may judge innovative research as being less rigorous due to its novelty; therefore, reviewers are less inclined to tolerate risks and evaluate innovative proposals favorably, which is supported by recent criteria score analyses (Luukkonen 2012; Eblen et al. 2016). While reviewers likely use their subject matter expertise to identify innovative projects, some explorations have suggested that even when expertise is controlled, highly novel proposals are routinely penalized in score (Boudreau et al. 2016). Thus, if one of the goals of funding agencies is to “foster fundamental creative discoveries, innovative research strategies, and their applications” (NIH 2013), this may not be achieved, potentially due to reviewer risk aversion, although it is unclear if reviewers perceive themselves as avoiding risk.

There is a paucity of data on how reviewers evaluate risk in peer review (or how they perceive their evaluations) as well as a dearth of general descriptions of the

decision-making processes used by reviewers; what data does exist is largely on journal peer review, which is different in scope and purpose than grant peer review. Moreover, while criteria like innovation are standard in the review of most NIH and NSF grants, many applicants report bias against innovative projects, suggesting there is a discrepancy between reviewer and applicant perceptions, yet factors like reviewer and applicant attitudes toward risk have not been explored relative to peer review (Lee et al. 2013). In fact, only recently has there been a widespread call for a more robust and complete evidence base upon which to drive policy regarding the peer review of grant applications (Lauer and Nakamura 2015; Bohannon 2013; Rennie 2016). While there has been a steady increase in empirical literature on grant peer review in recent years, reviewer decision-making has been minimally explored and characterized.

The examination of reviewers’ perceptions of their evaluative practices would be particularly valuable for informing grant review processes, especially compared to applicants’ perceptions of reviewer feedback. The most obvious way to assess these perceptions is through a survey of scientists’ experience with peer review; however, surveys that currently exist in the literature center around the motivations and levels of participation of scientists in journal peer review, and not on their perceptions of their evaluations (Sense About Science 2009; Taylor and Francis 2016; Ware 2016; Ware and Monkman 2008). While some surveys have been conducted by funding agencies, they have very generally examined reviewer/applicant differences in perceptions of review outcomes, there has not yet been a more in-depth exploration of perceptions of review content (NIH 2012; NSF 2014), particularly in terms of the evaluation of specific criteria. These data would be important to shed light on decision-making processes occurring within peer review and possibly provide a basis for enhanced reviewer training to align reviewer decisions with funding agency goals and to provide feedback to applicants that would enable improved future submissions.

In an effort to better understand reviewer motivations and experiences with grant peer review, a comprehensive peer review survey was developed to address these areas. The survey was sent to individuals in the American Institute of Biological Sciences’ (AIBS) database who had either participated on an AIBS peer review panel convened by the institute or served as a PI, investigator, collaborator, or consultant on an application that had been reviewed by AIBS. The goals of the survey were to gather much needed data on several elements related to applicant and reviewer perceptions about peer review in order to inform and refine future peer review practices.

2 Methodology

2.1 Survey

A 60-item questionnaire was developed for the peer review survey. The survey was divided into five sections, three of which are relevant for this manuscript: (1) demographics; (2) investigator attitudes toward grant review; and (3) reviewer attitudes toward grant review. The other sections largely dealt with levels and types of participation and motivations for reviewing, and thus were not analyzed as part of this study.

The questions were associated with dichotomous (yes/no), multiple choice, or rating scale (1–5) answer choices. Respondents were given the option to skip any question. In addition, text boxes were provided at the end of each section to allow respondents to elaborate on their responses. Based on beta testing, it was estimated that the survey would take 15 min to complete. A full copy of the peer review survey is available in “[Appendix](#).”

One of the key areas addressed in this questionnaire was asking former applicants to identify the key areas of focus in the feedback they received from reviewers. Applicants were asked to indicate presence/absence from a list of standard criteria that were derived from those routinely present in NIH and NSF reviews as well as reviews previous reviews AIBS has conducted, including potential impact of research, the quality of the hypothesis, the research methodology, the innovation potential, the quality of the preliminary data, the responsiveness to the funding mechanism, statistical issues, the qualifications of the research team, and the appropriateness of the budget (NIH 2016a; NSF 2017).

2.2 Participants and procedures

The survey was administered through Limesurvey, a commercial, web-based survey package, and was disseminated to 16,875 scientists in AIBS’s proprietary database. AIBS has cultivated this database over many years to use as a crucial resource for recruiting for reviews it conducts on biomedical and other life sciences research proposals for a wide variety of federal funding agencies, private research institutes and non-profit research funders. Most of these reviews are ad-hoc, not recurring; a new set of scientists is recruited each peer review round to match the specific expertise needed for the given proposal set. AIBS review topic areas (and thus respondent expertise) vary from basic biological and biophysical explorations of mechanisms of disease and injury to more translational and clinical studies, including but not limited to psychology, neuroscience,

microbiology, pharmacology, engineering, physical rehabilitation, social work, and clinical trials.

To assure the anonymity of the respondents, Limesurvey uses two separate databases that prevent identifying responses being linked back to participants. The survey was open for 2 months: the initial invitation to participate in the survey was sent on September 7, 2016, a reminder was sent a month and a half later (October 24/25, 2016), and the survey was closed on November 7, 2016. Once closed, the survey was no longer accessible to invitees or respondents.

The survey responses were exported through Limesurvey and analyzed using basic statistical packages. Descriptive statistics were used to characterize the results; tests for differences in proportions were used to examine differences between applicants’ and reviewers’ perceptions for parallel questions.

As some of these data are dichotomous in nature, to measure correlations between such variables, phi correlation coefficients were calculated as was standard error and significance levels (Yule 1912) of these relationships. Chi-square tests of applicant/reviewer samples were used to compare proportions of reviewer and applicant perceived use of criteria and standard errors for these differences were calculated.

3 Results

3.1 Survey response rate

Initially, 16,875 scientists in AIBS’s proprietary database were invited via email to participate in the survey. Of those 16,875 invitations, 2737 bounced back, leaving a total of 14,138 sent invitations, which comprised 13,781 unique individuals. These individuals had either participated on a peer review panel that had been convened by the institute (36%; $N = 4986$) or had been listed as a PI/investigator/colleague/consultant/other on an application that had been reviewed through the institute (71%; $N = 9716$). About 12% ($N = 1611$) were both institute reviewers and applicants and 5% ($N = 690$) were institutional officials, administrative staff, etc. We therefore limited our base of invitees to the 13,091 individuals that received an invitation to participate in the survey and were not administrative officials. Of those individuals, 1231 responded, giving a response rate of 9.4%. As respondents were allowed to skip questions, 381 respondents had left at least one of the answers completely blank and were therefore removed from this analysis. A total of 850 had complete responses; those responses are analyzed below. It should be noted that for some questions, respondents were given the option “prefer not to answer,” or “not applicable” which was counted as a response.

3.2 Survey respondent demographics

Overall, most respondents were Caucasian males over 50 years old with PhDs, working in academia at a mid to late career stage. Most reported working over 40 h per week. The AIBS database, from which the invited population came, does not have complete demographic information; therefore, we were unable to compare the demographics of the respondents to those of the invited population. These results are summarized in Table 1.

3.3 Grant submission levels

The majority (80%; $N = 677$) of respondents had actively submitted a grant in the last 3 years while 20% ($N = 166$) had not (7 chose not applicable). These proportions are fairly representative of the total invited population (71%). Respondents reported submitting a total of 3167 grants over the last 3 years. A histogram of the number of grants submitted per respondent (Gs) yielded a fairly clear multimodal distribution of grant submission frequencies, where nearly half of actively submitting applicants (44%) reported submitting on average 2 grants a year or more (Fig. 1).

Nearly all of those who reported they submitted proposals had received feedback (91%; $N = 615$) while only 9% ($N = 58$) specified they had not (3 stated non-applicable). Of those who reported submitting proposals, 37% ($N = 250$) indicated that their last grant submission was

funded while 60% ($N = 407$) indicated that it was not (19 stated non-applicable). It should be noted that unsuccessful applicants reported submitting slightly more applications on average (4.9 ± 0.10 applications) over the 3 year period as compared to successful applicants (4.4 ± 0.13 applications; $t[655] = 2.52, p = 0.01$).

3.4 Peer review participation

The majority of respondents (75%; $N = 637$) had actively served on a peer review panel in the last 3 years while 22% ($N = 184$) had not (29 chose not applicable). This participation rate is significantly higher than that of the total invited population (36%) and is likely due to the fact that the database only contains information relative to reviews conducted by AIBS, not those of funding agencies such as NIH or NSF. Thus, a reviewer could have reviewed for NIH but not for AIBS in this timeframe.

A total of 2523 reviews from the past 3 years were reported by reviewer respondents. In terms of the number of distinct grant reviews per respondent (Gr), a multimodal distribution existed whereby 30% of active reviewers reviewed on average twice a year or more (Fig. 1). Eighty eight percent ($N = 555$) of active reviewers have submitted a grant (39% success rate) versus 65% ($N = 119$) of inactive reviewers (34% success rate). Thus, while those who review are more likely to have submitted a grant, the success rates are comparable.

Table 1 Demographics of respondents

Variables ($N = 850$)	Response					
Gender (no answer; $N = 21$)	Male	Female				
	66% ($N = 559$)	32% ($N = 270$)				
Degree type (no answer; $N = 21$)	PhD	MD	Other			
	80% ($N = 677$)	23% ($N = 192$)	6% ($N = 55$)			
Age (no answer; $N = 21$)	Under 30	30–39	40–49	50–59	60+	
	0%	2% ($N = 21$)	23% ($N = 197$)	32% ($N = 270$)	43% ($N = 362$)	
Type of work (no answer; $N = 13$)	Academia	Government	Industry	Other		
	81% ($N = 686$)	6% ($N = 52$)	4% ($N = 33$)	8% ($N = 66$)		
Race/ethnicity* (no answer, $N = 29$)	Caucasian	Asian	Latino	Black	Native American/Hawaiian	Other
	76% ($N = 648$)	11% ($N = 93$)	4% ($N = 32$)	1% ($N = 12$)	0.4% ($N = 4$)	4% ($N = 31$)
Career stage (no answer, $N = 29$)	Early	Mid	Late/Tenured	Emeritus		
	3% ($N = 26$)	28% ($N = 237$)	59% ($N = 504$)	10% ($N = 83$)		
Hours worked per week (no answer, $N = 39^{**}$)	40	40–50	50–60	60–70	70+	
	10% ($N = 85$)	25% ($N = 213$)	32% ($N = 271$)	18% ($N = 154$)	9% ($N = 76$)	

*Choose all that apply

**All Emeritus

% Respondents Versus Frequency of Gs and Gr

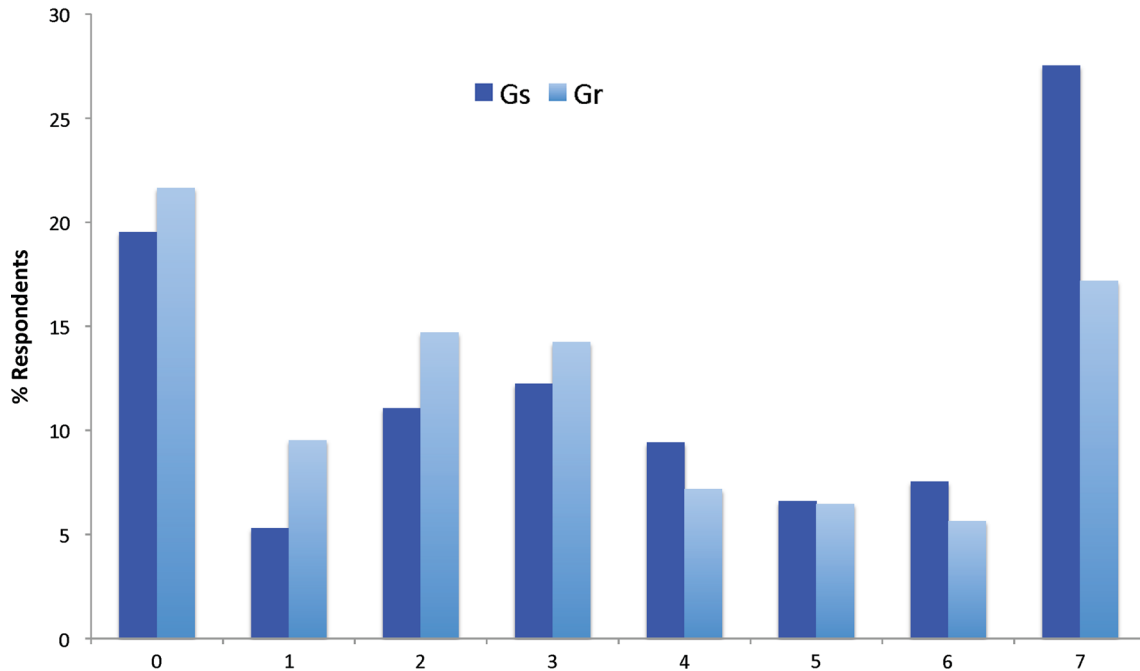


Fig. 1 % Respondents versus frequency of Gs (grants submitted) and Gr (grants reviewed). Relative proportions of respondents are represented on the y-axis while numbers of grants submitted (Gs)

or reviewed (Gr) in the last 3 years are represented in the x-axis. A total of 3167 grants were recorded to be submitted and 2523 grants reviewed in this timeframe by 850 respondents

Table 2 Areas of feedback (*N* = 677)

Variables	% Response Yes	% Response No
Methodology	53% (<i>N</i> = 359)	47% (<i>N</i> = 318)
Potential impact	33% (<i>N</i> = 224)	67% (<i>N</i> = 453)
Innovation	24% (<i>N</i> = 164)	76% (<i>N</i> = 513)
Preliminary data	17% (<i>N</i> = 112)	83% (<i>N</i> = 565)
Research team	11% (<i>N</i> = 77)	89% (<i>N</i> = 600)
Hypothesis	10% (<i>N</i> = 70)	90% (<i>N</i> = 607)
Statistical issues	6% (<i>N</i> = 40)	94% (<i>N</i> = 637)
Responsiveness to mechanism	5% (<i>N</i> = 31)	95% (<i>N</i> = 646)
Budget	4% (<i>N</i> = 29)	96% (<i>N</i> = 648)
Other	6% (<i>N</i> = 38)	94% (<i>N</i> = 639)

3.5 Applicant perceptions of review criteria and the usage of innovation and risk

Applicant respondents were asked to indicate all of the areas addressed in the critique of their last grant application. It should be noted that 16% [*N* = 109] did not indicate that any criteria were addressed and 35% [*N* = 240] indicated only one criterion, with the remaining 48% indicating more than one criteria [*N* = 328]. These results are summarized in Table 2; the three most frequently addressed criteria were methodology, impact and innovation, respectively. However, methodology was much more frequently addressed than impact or innovation, with only 24% (*N* = 164) of applicant respondents indicating that recent review feedback addressed innovation potential. Importantly, both unsuccessful and successful applicants indicated that innovation was mentioned with similar frequency, suggesting funding success did not affect commentary on this criterion (Table 3).

Table 3 Applicant perceptions of criteria (demographics)

	Age (below 50/50 and older)	Gender (M/F)	Degree (PhD/No PhD)	Recent funding success (Y/N)
Innovation	24% [45]/26% [126]	25% [107]/24% [55]	23% [129]/31% [35]	25% [<i>N</i> = 63]/25% [<i>N</i> = 100]
Risk	28% [54]/27% [131]	29% [124]/19% [37]*	27% [150]/31% [35]	28% [<i>N</i> = 116]/26% [<i>N</i> = 66]
PI track record	14% [26]/10% [51]	11% [49]/11% [24]	11% [63]/12% [14]	12% [29]/11% [45]

Table 4 Phi associations among areas of feedback

Variables	Methods	Impact	Innovation	Prelim. data	Research team	Hypothesis	Statistical issues	Funding mechanism
Methods	–							
Potential impact	0.05	–						
Innovat.	– 0.01	0.25*	–					
Prelim. data	0.12*	0.09	– 0.04	–				
Research team	0.09	0.16*	0.08	0.17*	–			
Hypoth.	0.07	0.24*	0.16*	0.14*	0.11*	–		
Stat. issues	0.05	0.03	0.03	0.14*	0.15*	0.08	–	
Funding mech.	– 0.01	0.04	0.04	0.00	0.14*	0.06	0.09	–
Budget	– 0.01	0.04	0.10*	0.08	0.11*	0.02	0.16*	0.09

Table 5 Views of review criteria (applicant vs. reviewer)

	Applicants (%)	Reviewers (%)	Test for difference
Innovation	24	81	$X^2(1) = 475$ $p < 0.001$ stderr = 2.2%
Research team	11	57	$X^2(1) = 336$ $p < 0.001$ stderr = 2.3%
Risk	27	58	$X^2(1) = 102$ $p < 0.001$ stderr = 2.8%

Similarly, age, gender, and degree did not seem to affect commentary on innovation (Table 3, proportions and N are shown, * indicates statistically significant correlation [$p < 0.01$], standard error ranged from 7.0 to 9.5%).

In addition, the majority (56%; $N = 379$) of applicants indicated that the reviewers did not comment on the riskiness of the research while only 27% [$N = 185$] stated that they did (113 stated the question was not applicable; Table 5). Again, the likelihood of comments about risk did not differ significantly by success, age, or degree (Table 3). However, a significant difference was observed by gender ($X^2[1] = 8.44$; $p = 0.004$), where female applicants were less likely to report comments on risk than their male applicants.

Phi coefficients were calculated between paired feedback areas in order to determine which areas co-occurred among

applicant respondents (Table 4; * indicates statistically significant correlation [$p < 0.01$], standard error for all correlations is 0.04). In general, correlation coefficients were small but significant between some areas. Feedback regarding innovation was significantly correlated with impact but neither impact nor innovation was correlated with methodology. Interestingly, feedback regarding the research team was correlated with receiving feedback in many other areas, including impact, preliminary data, hypothesis, statistical issues, funding mechanism and budget. Feedback regarding methodology was only correlated with feedback regarding preliminary data.

3.6 Reviewer perceptions of review criteria and the usage of innovation and risk

Active reviewers—when asked whether the evaluation criteria used at their last panel meeting were appropriate to judge the best science and move the field forward—provided an average rating of 2.3 ± 0.04 (1 = most appropriate to 5 = least appropriate) (27 indicated the question was not applicable). It should be noted that no statistically significant differences were found across reviewer gender or degree; reviewers aged 50 and older rated the review criteria as more appropriate than did reviewers less than 50 years of age ($t[609] = 2.18$, $p = 0.03$) for perceptions of criteria appropriateness (Table 6; * indicates statistically significant correlation [$p < 0.01$], standard error is reported).

Table 6 Reviewer perceptions of criteria (demographics)

	Age (below 50/50 and older)	Gender (M/F)	Degree (PhD/No PhD)
Criteria appropriateness	$2.5 \pm 0.1/2.3 \pm 0.05$	$2.3 \pm 0.05/2.5 \pm 0.07$	$2.3 \pm 0.05/2.4 \pm 0.10$
Innovation	80% [121]/82% [398]	81% [332]/85% [178]	82% [425]/81% [94]
Risk	59% [89]/58% [283]	58% [240]/60% [125]	59% [303]/58% [69]
PI track record	52% [78]/59% [287]	60% [249]/53% [111]	58% [301]/55% [64]

Eighty-one percent of active reviewers ($N = 519$) indicated that they factored innovation into selecting the best science (14% [$N = 92$] did not and 26 indicated the question was not applicable; Table 5). Moreover, 70% of reviewers viewed innovation as an essential component of scientific excellence when evaluating grant applications, while only 27% ($N = 212$) did not (22 indicated the question was not applicable). Negligible differences in reviewer inclusion of innovation were found across gender, age and degree (Table 6; proportions and N are reported, standard error ranged from 3.4 to 6.6%).

When asked whether the risks associated with innovative research impacted the scores they assigned to grant applications, 58% of reviewers ($N = 372$) indicated that they did while 35% ($N = 223$) indicated that they did not (42 indicated the question was not applicable; Table 5). Negligible differences in reviewer inclusion of risk were found across gender, age and degree (Table 6).

Fifty-seven percent ($N = 365$) of reviewers indicated that the PI's track record tempered their assessment of any detected methodological weaknesses, while 38% ($N = 243$) indicated that it did not (29 indicated the question was not applicable). Similar to innovation and risk, negligible differences were found across gender, age and degree (Table 6).

Overall, these results suggest that reviewers were significantly more likely to report that they addressed components of innovation, risk, and the mitigation of risk through the PI's track record in their reviews than applicants perceived these components to have been addressed in their critiques.

4 Discussion

4.1 Survey generalizability

Our response rate of 9.4% was low but comparable to those of similar surveys on journal peer review (7.7%: Ware and Monkman 2008; 10%: Sense About Science 2009; 5–10%: Taylor and Francis 2016) and exceeds the 2.2% rate reported in the recent PRC survey (Ware 2016), although none of these other reports were published in the peer-reviewed literature. In terms of demographics, the majority of our 1241 survey respondents was white males with PhDs who were 50 years or older, worked in academia, were largely tenured, and worked more than 50 h per week (Table 1). These demographics compare favorably to those reported in previous journal peer review surveys, such as the 2009 Sense About Science (SAS 2009) and the 2016 PRC survey (Ware 2016), in terms of gender (74% of the SAS and 70% of the PRC respondents were male) and place of work (66% of the SAS respondents worked in academia). In addition, they reflect the general gender distribution of publishing authors (Sugimoto et al. 2013). However, in terms of respondent age, the

2008 Ware and Monkman and 2016 PRC journal peer review surveys had more even distributions. The age distribution of our survey respondents is somewhat younger than the age distribution of NIH grant reviewers with NIH funding; however, it should be noted that funded reviewers are only a subsample of the NIH reviewer population (Rockey 2015).

Nevertheless, the majority of our survey respondents had submitted a grant in the last year, and 38% of those who had submitted a grant reported funding success, which is higher than the 13–23% funding success rates reported for fiscal years 2013, 2014, and 2015 by NIH (Rockey 2015). Thus, it may be that our sample is representative of more senior researchers who have success with NIH funding and of those who comprise most NIH research panels (Etcheberigaray 2014). Also, it should be noted that our sample of respondents had similar levels of grant submission to that of the total invited population and while the respondents reported higher levels of peer review than of the total invited population, it is likely that most of the reviews they refer to are non-AIBS related (e.g., NIH and NSF). Thus, we feel our respondent sample is generalizable to the larger grant reviewing population.

4.2 Attitudes toward innovation and risk

Respondents were asked to indicate the areas represented in the review feedback they received as applicants (it should be noted that reviewers were likely charged to address all of these criteria by the funding agency). Applicants indicated the three most frequent areas of reviewer feedback were methodology (53%), potential impact of research (33%), and innovation potential (24%) (Table 2). While those who received feedback on innovation were likely to also receive feedback on impact, feedback about innovation and methodology was unrelated (Table 4). These results very much align with the results from Eblen et al. (2016) as well as Rockey (2010), who also found the approach criterion to be a better predictor of overall score than the impact or innovation criteria. Yet, in spite of applicant reports, 81% of reviewers indicated that they factored innovation into selecting the best science (Table 5) and 70% viewed innovation as an essential component of scientific excellence, underscoring that applicant and reviewer respondents had significantly divergent perceptions of the use of review criteria.

Similar significant differences were also seen for the use of riskiness in review decisions, as only 27% of applicants received comments on the riskiness of their grant applications while 58% of reviewers indicated they took riskiness into account in their scores (Table 5). The focus on methods and lack of consideration of innovation and/or riskiness adds further evidence that reviewers may be (or at least be perceived as) risk-averse and also suggests that they may not be aware of this potential bias. While it is possible that

reviewers may have taken riskiness into account in their scores, they do not seem to be clearly expressing these concerns in their review comments. It is more likely that the bias is real and that reviewers are simply unaware of their risk aversion, as previous studies have documented the penalization of highly innovative work in a well-controlled peer review system (Boudreau et al. 2016).

Some have suggested that highly innovative research may be associated with more reviewer uncertainty about their judgments of the methodology, which may lead to lower scores (Luukkonen 2012; Boudreau et al. 2016). While reviewers may consider innovation in their decisions, they likely give this less weight than methodological weaknesses, which is reflected in the content of the reviewer's critique. Also, methodological weaknesses are usually more concrete, while estimations of innovativeness are more subjective and subject to individual perception, which likely influence the focus and length of the evaluation for each of these criteria. In addition, our results indicate the lack of a relationship between feedback concerning innovation and methodology (Table 4). Thus, the reviewer's written critique will appear to applicants to focus on methodological issues. This differential weighting of criteria by reviewers to synthesize an overall score for the scientific merit of an application has been identified as commensuration bias (Lee 2015) and may explain why the vast majority of reviewers indicate they factored innovation into their critiques, despite the low frequency reported by applicants (Table 5). Such a bias suggests that it may not be possible to reliably focus on both innovation and methodological rigor; some have suggested distinct funding mechanisms to address this challenge (Gewin 2012; Ioannidis 2011).

However, it may also be that concerns about innovation and riskiness influence reviewer perceptions of other criteria. For instance, grant reviews are not double blinded (reviewers know applicants' identities and are usually charged to evaluate the research team) and our results show that 57% of reviewers indicated that the PI's track record tempered their assessment of any detected methodological weaknesses (Table 5). In fact, our analysis has indicated a significant relationship between the research team and many of the other criteria factors, suggesting that PI track records strongly influence reviewers' perceptions of the value of the application (Table 4). Thus, while methodological weaknesses may be weighed more heavily than innovative ideas, reviewers' risk tolerance may be enhanced by applicants' track records of innovation and impact. Indeed, some studies have suggested that professional connections and knowledge of applicants' work affect reviewer scoring (Gallo et al. 2016; Li 2015), although it is unclear if PI familiarity results in the funding of more innovative proposals. However, it should be noted that only 11% of applicants reported comments about the

research team (Table 2), lending further support to a significant discrepancy between applicant and reviewer perceptions, which may establish a perception of prestige bias and cronyism among applicants.

In general, reviewer and applicant demographics and even funding success had little effect on respondent perceptions of review feedback, suggesting the pervasiveness of these discrepancies in the review process (Tables 3, 6). The exception seems to be with female applicants and their low level of reported comments on riskiness (Table 3). Some have suggested that a subtle gender bias exists in grant peer review where reviewers more often tend to see male applicants as leaders and score their applications more generously (Magua et al. 2017). If this is the case, it may be that reviewers temper their assessments of methodological weaknesses to a lesser degree for female applicants. More work is needed in this area to confirm whether this is true.

If reviewers are in fact biased, how might those who conduct peer review address biases? One suggestion has been to train reviewers to properly interpret and weigh the different review criteria to achieve the goals of the funding agency, journal, etc. (Lee 2015). However, when reviewers are simultaneously asked to assess feasibility and innovation, implicit risk aversion may result in resistance to achieving equitable weighing of component criteria. Also, there is likely considerable variation in risk preferences between individuals (Weber and Hsee 1999; Slovic 1999) and across different teams of peers (Gardner and Steinberg 2005). These risk preferences may be difficult to change with training, although to date, neither measurements of reviewers' and review panels' risk preferences nor the effects of training have been assessed.

Another important consideration is the lack of consistency in reporting of any of the review criteria areas in the feedback (Table 2), where despite an expectation that nearly every reviewer critique might include a focus on methodology, it is only reported 53% of the time. This variance may contribute to the considerable inter-rater reliability found in many peer reviews (Cole et al. 1981; Cicchetti 1991; Bornmann and Daniel 2008). Moreover, respondents indicated an average of 2.3 ± 0.04 with regard to the appropriateness of the criteria, with younger respondents finding the criteria less appropriate. However, it is not clear that reviewers have a consensus on which criteria are most appropriate and or should be weighted the highest. There is clearly a tension between criteria like high risk–high reward innovation and sound methodology. While groups like NIH have created special funding mechanisms to support risky innovation (NIH 2016b), the review criteria for R01 grants still include both innovation and methodology as criteria (NIH 2016a). Thus, it may be that more discussion needs to take place among the scientific community as to whether current review criteria are appropriate for most reviews, if other

criteria need to be considered, and if criteria should be prioritized to promote reviewer reliability.

One important absence in our analysis is the linkage of reviewer/applicant responses to the same proposal. This type of analysis would be very interesting but can only be accomplished through the willingness of the funding agency to allow access to the data to make such comparisons. Also, the lack of correlation between feedback concerning preliminary data and innovation is particularly surprising (Table 4), as often there is a lack of sufficient preliminary data to support highly innovative research approaches, so much so that many funding mechanisms to support innovative research lack the requirement for preliminary data. It may be that reviewer consideration of preliminary data does not weigh as heavily in reviewer's minds as other criteria or more likely it may influence feasibility concerns in the methodology comments. More research should be conducted exploring reviewer consideration of preliminary data relative to other criteria.

In conclusion, it is apparent that more studies are needed to tease out reviewer decision-making processes in grant peer review, particularly with regard to risk tolerance. The differences seen in this study between the perceptions of grant reviewers and applicants have unearthed an interesting area that warrants further investigation if we are to better understand and optimize the peer review process.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Full survey

Survey is listed below.

What is your gender?

Please choose all that apply:

Male

Female

Prefer not to answer

What is your age?

Please choose only one of the following:

Under 30

30–39

40–49

50–59

60+

Please specify your race/ethnicity

Please choose all that apply:

American Indian or Alaska Native

Asian or Asian American

Black or African American

Hawaiian or Other Pacific Islander

Hispanic or Latino

Non-Hispanic White/Caucasian

Other

Prefer not to answer

What type of degree(s) do you have?

Please choose all that apply:

PhD or other research doctorate

MD

DDS

DVM or VMD

Other

Prefer not to answer

What type of an organization do you work for?

Please choose only one of the following:

Academia

Government

Industry

Other

What stage of career have you reached?

Please choose only one of the following:

Early career

Mid career

Late career/tenured

Emeritus

On average, how many hours do you work each week?

Please choose only one of the following:

40 h

40–50 h

50–60 h

60–70 h

70 + h

Please provide any comments that justify your responses under Section 1, Demographics.

Please write your answer here:

Section 2: Grant submission and peer review experience

Have you submitted a grant for peer review in the last 3 years?

Please choose only one of the following:

Yes

No

If you answered yes to submitting a grant for peer review in the past 3 years, how many grant applications have you submitted in that time frame?

Please choose only one of the following:

1

2

3

4

5

6

7 or more

Did you receive reviewer feedback on your last grant submission?

Please choose only one of the following:

Yes

No

Was your last application successful, i.e., were you funded?

Please choose only one of the following:

Yes

No

Have you served on a peer review panel in the last 3 years?

Please choose only one of the following:

Yes

No

If you answered yes to serving on a peer review panel in the last 3 years, how many peer review panels have you served on in that time frame?

Please choose only one of the following:

1

2

3

4

5

6

7 or more

If you answered yes to serving on a peer review panel in the past 3 years, please select the mode of your last peer review panel meeting.

Please choose only one of the following:

Face-to-face

Remote (video/teleconference)

Internet-assisted

Other

How many ad-hoc reviews (usually one or two grant applications reviewed telephonically that are being evaluated in a panel meeting setting) have you performed in the past 3 years?

Please choose only one of the following:

0

1

2

3

4

5

6

7 or more

Have you reviewed for a journal in the last 3 years?

Please choose only one of the following:

Yes

No

If you answered yes to reviewing for a journal in the past 3 years, how many submissions have you reviewed in that time frame?

Please choose only one of the following:

1

2

3

4

5

6

7 or more

What is a higher personal priority: grant review or journal review?

Please choose only one of the following:

Grant review

Journal review

Both are equal priority

Neither is a priority

Please elaborate on your responses under Section 2, Grant submission and peer review experience.

Please write your answer here:

Section 3: Investigator attitudes toward grant review

If you answered yes to receiving feedback on your last grant submission, please answer Section 3 of the questionnaire. If you answered no, please proceed to Section 4.

On a scale of 1–5 (1 most useful, 5 least useful), overall how useful was the reviewer feedback you received on your last grant submission?

Please choose only one of the following:

1

2

3

4

5

On a scale of 1–5 (1 most useful, 5 least useful), how useful was the reviewer feedback in improving your grantsmanship?

Please choose only one of the following:

1

2

3

4

5

If you were not funded, on a scale of 1–5 (1 most useful, 5 least useful), how useful was the reviewer feedback in improving your future submissions?

Please choose only one of the following:

1

2

3

4

5

On a scale of 1–5 (1 most useful, 5 least useful), how useful was the reviewer feedback in informing your future scientific endeavors in the proposed research area?

Please choose only one of the following:

- 1
- 2
- 3
- 4
- 5

Did you feel the reviewer feedback was well written, cohesive, and balanced?

Please choose only one of the following:

- Yes
No

Did you feel the reviewer feedback was fair and unbiased?

Please choose only one of the following:

- Yes
No

Overall, in what area(s) did the reviewer feedback primarily focus?

Please choose all that apply:

- Potential impact of research
- Hypothesis
- Research methodology
- Innovation potential
- Preliminary data
- Responsiveness to funding mechanism
- Statistical issues
- Qualifications of research team
- Budget
- Other

Did the reviewers comment on the riskiness of the research project?

Please choose only one of the following:

- Yes
No

Based on the reviewer feedback you received, do you feel that the reviewers had the appropriate expertise to evaluate your grant application?

Please choose only one of the following:

- Yes
No

Please elaborate on your responses under Section 3, Investigator Attitudes Towards

Grant review.

Please write your answer here:

Section 4: Reviewer attitudes towards grant review

What are your reasons for accepting an invitation to serve on a peer review panel?

Please choose all that apply:

- Desire to give back to the scientific community
- Networking opportunities

Informing your own grantsmanship

Gaining exposure to new and innovative scientific areas

Enhancing your career/resume

Expectation from the funding agency

Honorarium

Other

Do you feel that serving as a reviewer on peer review panels has positively impacted your career?

Please choose only one of the following:

- Yes
No

If you feel that serving as a peer reviewer has positively impacted your career, in what ways has serving as a reviewer influenced your career?

Please choose all that apply:

- Bolstered your career
- Improved your grantsmanship
- Increased your exposure to new scientific ideas
- Improved your networking/collaboration opportunities
- Other

In general, which type of panel meeting format do you prefer?

Please choose only one of the following:

- Face-to-face
- Virtual [teleconference/videoconference]
- Internet-assisted

On a scale of 1–5, (1 most influential, 5 least influential), please rate the following factors in influencing your selection of preferred panel meeting format:

Please write your answer(s) here:

- Logistical convenience
- Level of communication among panel members
- Networking opportunities
- Likelihood to participate on panel

In the last 3 years, how many times have you declined an invitation to serve on a peer review panel?

Please choose only one of the following:

- 1
- 2
- 3
- 4
- 5
- 6
- 7 or more

What were your reasons for declining an invitation to serve on a peer review panel?

Please choose all that apply:

- Limited free time
- Poor expertise match
- Personal reasons (holiday, sickness, travel)
- Review timeline too compressed
- Conflict of interest

Issue with funding agency
 Other
What is the maximum number of peer review panels/committees you prefer to serve on per year?
 Please choose only one of the following:
 1
 2
 3
 More than 3
What is the maximum number of days you prefer to attend a peer review panel meeting?
 Please choose only one of the following:
 1
 2
 3
 More than 3
What is the maximum number of R01-type grant applications you prefer to be assigned for a peer review panel meeting?
 Please choose only one of the following:
 1–2
 3–4
 5–6
 7
 More than 7
What was the actual number of days of your last peer review panel meeting?
 Please choose only one of the following:
 1
 2
 3
 More than 3
What was the actual number of R01-type grant applications you were assigned to review at your last peer review panel meeting?
 Please choose only one of the following:
 1–2
 3–4
 5–6
 7–8
 More than 8
On average, how many hours did you spend reviewing each grant application before the panel meeting?
 Please choose only one of the following:
 1–2
 2–3
 3–4
 4–5
 5–6
 7 or more
Please elaborate on your responses under Sect. 4, Reviewer attitudes towards grant review.
 Please write your answer here:

Section 5: Peer review panel meeting proceedings
Please answer the following questions in relation to your last peer review meeting. On a scale of 1–5 (1 most definitely, 5 not at all), was your scientific expertise necessary and appropriately used in the review process?
 Please choose only one of the following:
 1
 2
 3
 4
 5
On a scale of 1–5 (1 most definitely, 5 not at all), from your perspective was the expertise of the other panel members necessary and appropriately used in the review process?
 Please choose only one of the following:
 1
 2
 3
 4
 5
Did the grant application discussions facilitate reviewer participation?
 Please choose only one of the following:
 Yes
 No
Were the grant application discussions fair and balanced?
 Please choose only one of the following:
 Yes
 No
On a scale of 1–5 (1 most useful, 5 least useful), how useful were the grant application discussions in clarifying differing reviewer opinions?
 Please choose only one of the following:
 1
 2
 3
 4
 5
On a scale of 1–5 (1 extremely effective, 5 no effect), did the grant application discussions affect the outcome?
 Please choose only one of the following:
 1
 2
 3
 4
 5
On a scale of 1–5 (1 most appropriate, 5 least appropriate), were the evaluation criteria appropriate to judge the best science and move the field forward?
 Please choose only one of the following:
 1
 2

3
4
5
[]On a scale of 1–5 (1 extremely important, 5 of no importance), how important is the PI's track record to assessing an investigator initiated (R01)-type application?

Please choose only one of the following:

- 1
2
3
4
5

[]In general, does a PI's track record temper your assessment of any detected methodological weaknesses?

Please choose only one of the following:

- Yes
No

[]On a scale of 1–5 (1 most definitely, 5 not at all), did the grant application discussions promote the best science?

Please choose only one of the following:

- 1
2
3
4
5

[]Was innovation factored into selecting the best science?

Please choose only one of the following:

- Yes
No

[]Did you view innovation as an essential component of scientific excellence when evaluating the grant applications?

Please choose only one of the following:

- Yes
No

[]Did the risks associated with innovative research impact the scores you assigned to the grant applications?

Please choose only one of the following:

- Yes
No

[]On a scale of 1–5 (1 completely, 5 not at all), how much did the seniority of your fellow panel members influence your evaluations during the panel deliberations?

Please choose only one of the following:

- 1
2
3
4
5

[]Was the format and duration of the grant application discussions sufficient to allow the non-assigned reviewers to cast well informed merit scores?

Please choose only one of the following:

- Yes

No

[]On a scale of 1–5 (1 extremely useful, 5 not useful at all), how useful was the Chair in facilitating the application discussions?

Please choose only one of the following:

- 1
2
3
4
5

[]Please elaborate on your responses under Section 5, Peer review panel meeting proceedings.

Please write your answer here:

Thank you for taking the time to fill out the survey. Have a wonderful day!

Submit your survey.

Thank you for completing this survey.

References

- Bohannon J (2013) Who's afraid of peer review. *Science* 342:6154
- Bornmann L, Daniel HD (2008) The effectiveness of the peer review process: inter-referee agreement and predictive validity of manuscript refereeing at *Angewandte Chemie*. *Angew Chem Int Ed* 47(38):7173–7178
- Boudreau KJ et al (2016) Looking across and looking beyond the knowledge frontier: intellectual distance, novelty, and resource allocation in science. *Manag Sci* 62(10):2765–2783
- Cicchetti DV (1991) The reliability of peer review for manuscript and grant submissions: a crossdisciplinary investigation. *Behav Brain Sci* 14(1):119–186
- Cole S, Cole JR, Simon GA (1981) Chance and consensus in peer review. *Science* 214(4523):881–886
- Danthi N et al (2014) Percentile ranking and citation impact of a large cohort of National Heart, Lung, and Blood Institute—funded cardiovascular R01 grants. *Circ Res* 114(4):600–606
- Eblen MK, Wagner RM, RoyChowdhury D, Patel KC, Pearson K (2016) How criterion scores predict the overall impact score and funding outcomes for National Institutes of Health peer-reviewed applications. *PLoS ONE* 11(6):e0155060
- Etcheberrigaray R (2014) CSR and NIH peer review: maintaining highest quality and fairness. <https://www.healthra.org/download-resource/?resource-url=/wp-content/uploads/2014/09/Rene-Etcheberrigaray-Durham-September-2014-Meeting-Presentation.pdf>. Accessed 22 Feb 2018
- Gallo SA, Carpenter AS et al (2014) The validation of peer review through research impact measures and the implications for funding strategies. *PLoS ONE* 9(9):e106474
- Gallo S et al (2016) The influence of peer reviewer expertise on the evaluation of research funding applications. *PLoS ONE* 11(10):e0165147
- Gardner M, Steinberg L (2005) Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Dev Psychol* 41(4):625
- Gewin V (2012) The sky's the limit. *Nature* 487(7407):395–397
- Ioannidis JPA (2005) Why most published research findings are false. *PLoS Med* 2(8):e124. <https://doi.org/10.1371/journal.pmed.0020124>

- Ioannidis JPA (2011) More time for research: fund people not projects. *Nature* 477(7366):529–531
- Lamont M (2009) Pragmatic fairness: customary rules of deliberation. How professors think: inside the curious world of academic judgment. Harvard University Press, Cambridge, pp 107–158. <http://www.cunyawards.org/wp-content/uploads/2015/07/Lamont-2009.pdf>
- Lauer MS, Nakamura R (2015) Reviewing peer review at the NIH. *N Engl J Med* 373(20):1893–1895
- Lee C (2015) Commensuration bias. *Philos Sci* 82:1272–1283
- Lee C et al (2013) Bias in peer review. *J Am Soc Inform Sci Technol* 64(1):2–17. <https://doi.org/10.1002/asi>
- Li D (2015) Expertise vs. bias in evaluation: evidence from the NIH. HBS Working Paper 16–053, Harvard Business School, Boston
- Li D, Agha L (2015) Big names or big ideas: Do peer-review panels select the best science proposals? *Science* 348(6233):434–438
- Luukkonen T (2012) Conservatism and risk-taking in peer review: emerging ERC practices. *Res Eval* 21(1):48–60
- Magua W et al (2017) Are female applicants disadvantaged in National Institutes of Health peer review? Combining algorithmic text mining and qualitative methods to detect evaluative differences in R01 reviewers' critiques. *J Women's Health* 26(5):560–570
- National Institute of Health (2012) Enhancing Peer Review Survey Results Report_2012. https://enhancing-peer-review.nih.gov/docs/Enhancing_Peer_Review_Report_2012.pdf. Accessed 22 Feb 2018
- National Institutes of Health (2013) About NIH: mission. <http://nih.gov/about/mission.htm>. Accessed 22 Feb 2018
- National Institutes of Health (2016a) Definition of criteria and considerations for research project grant (RPG/X01/R01/R03/R21/R34) critiques. https://grants.nih.gov/grants/peer/critiques/rpg_D.htm#rpg_overall. Accessed 22 Feb 2018
- National Institutes of Health (2016b) NIH Common Fund announces 2016 high-risk, high-reward research awards. <https://www.nih.gov/news-events/news-releases/nih-common-fund-announces-2016-high-risk-high-reward-research-awards>. Accessed 22 Feb 2018
- National Science Foundation (2014) NSF's merit review process fiscal year 2014. <https://www.nsf.gov/nsb/publications/2015/nsb201514.pdf>. Accessed 22 Feb 2018
- National Science Foundation (2017) National Science Foundation's merit review criteria. https://www.nsf.gov/pubs/policydocs/pappg17_1/pappg_3.jsp#IIIA. Accessed 22 Feb 2018
- Rennie D (2016) Let's make peer review scientific nature 535:31–33. <https://doi.org/10.1038/535031a>
- Rockey S (2010) <https://nexus.od.nih.gov/all/2011/03/08/overall-impact-and-criterion-scores/>. Accessed 22 Feb 2018
- Rockey S (2015) <https://nexus.od.nih.gov/all/2015/07/30/understanding-capacity-peer-review/>. Accessed 22 Feb 2018
- Sense About Science (2009) Peer Review Survey. <http://archive.senseaboutscience.org/pages/peer-review-survey-2009.html>. Accessed 22 Feb 2018
- Slovic P (1999) Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield. *Risk Anal* 19(4):689–701
- Sorlie PD, Bild DE, Lauer MS (2012) Cardiovascular epidemiology in a changing world—challenges to investigators and the national heart, lung, and blood institute. *Am J Epidemiol* 175(7):597–601
- Sugimoto CR, Lariviere V, Ni C, Gingras Y, Cronin B (2013) Global gender disparities in science. *Nature* 504(7479):211–213
- Taylor & Francis (2016) Peer review: a global view. <https://authorserv.ices.taylorandfrancis.com/peer-review-global-view/>. Accessed 22 Feb 2018
- Travis GDL, Collins HM (1991) New light on old boys: cognitive and institutional particularism in the peer review system. *Sci Technol Hum Values* 16(3):322–341
- Ware M (2016) Peer Review Survey. Publishing Research Consortium, London
- Ware M, Monkman M (2008) Peer review in scholarly journals: perspective of the scholarly community—an international study. Publishing Research Consortium, London
- Weber EU, Hsee CK (1999) Models and mosaics: investigating cross-cultural differences in risk perception and risk preference. *Psychon Bull Rev* 6(4):611–617
- Yule GU (1912) On the methods of measuring the association between two attributes. *J R Stat Soc* 75:579–652