



CSR Analysis of Fall 2022 In-Person and Virtual Peer Review Meetings

May 1, 2023

Executive Summary

The National Institutes of Health (NIH) depend on the Center for Scientific Review's (CSR) peer review process to ensure that all grant applications receive fair, independent, expert, and timely reviews that are free from inappropriate influences. When the COVID-19 pandemic hit in mid-March 2020, CSR shifted its review meetings online, using the Zoom platform. In Fall 2022, CSR reimplemented in-person meetings and held 1/3 of its standing study sections, and recurring special emphasis panels (SEPs), with the remainder of meetings held virtually. This was the first time that CSR held a substantial number of its regular review meetings virtually and in-person at the same time and created a natural experiment on the effects of meeting format on review processes and outcomes. Note however, that study section meeting formats were not assigned randomly. Rather, scientific review officers (SROs) volunteered to hold in-person meetings, taking into consideration the preferences of their reviewers. CSR collected two types of data to evaluate outcomes, participant survey data and multiple objective meeting measures. The purpose of the survey data, administered to all standing study sections, recurring fellowship or small business SEPs held in 2023/01 council round (n=238), was to assess whether there were any differences in reviewers' perceptions of the quality of the review and meeting experiences due to the meeting format (i.e., in-person and virtual). The purpose of meeting measures data was to evaluate roster characteristics and scoring practices and is confined to a subsample of 92 meetings.

Reviewer survey findings

- Reviewers overall prefer in-person meetings (56%)—with more reviewers who attended in-person meetings preferring in-person meetings compared to virtual attendees (74% vs 45% respectively).
- On average, reviewers who attended in-person meetings reported longer attention spans than those who attended virtual meetings—although the magnitude of this effect was small ($M = 8.52$, $SD = 1.35$; $M = 7.59$, $SD = 1.85$, respectively, on a scale from 1-10).
- 93% of reviewers "agree or strongly agree" that their review panels were able to prioritize applications according to their impact and scientific merit and 92% "agree or strongly agree" that the scientific discussions helped the panel evaluate the applications. These ratings were significantly more positive from reviewers who attended in-person meetings (vs. virtual meetings) although the magnitude of these effects is small (*prioritize* = virtual: 92%, in-person: 97%; *discussions* = virtual: 90%, in-person: 96%).
- 87% of reviewers rated reviewer engagement during their meeting as excellent or good. These ratings were significantly more positive from reviewers who attended in-person meetings (96%) vs. virtual meetings (82%)—although the magnitude of this effect is small.
- 50% of reviewers reported that they contributed to the discussion always or often. These ratings were significantly higher from reviewers who attended in-person meetings (60%) vs. virtual meetings (45%)—although the magnitude of this effect is small.

Quantitative Meeting Measures findings

- For meeting applications counts, roster sizes and reviewer workloads, no meaningful differences were observed between Zoom and in-person meeting formats.

- Roster composition did not differ between meeting formats with respect to distribution of reviewers according to faculty ladder rank.
- Small numerical differences in roster representation according to gender, minority status, and URM status were not statistically significant.
- Distributions of final overall impact scores were similar between meeting formats; Zoom meetings had a significantly lower mean (by about 1 point).
- Patterns of in-meeting score movements for overall impact scores among assigned reviewers were nearly identical across meeting formats.
- Rates of out-of-range scoring were very similar in the two meeting formats.

In the full report that follows we present first the survey data followed by the meeting measures data.

Survey Data and Analysis

Methods

CSR surveyed reviewers who participated in 238 CSR peer review meetings between September 26th to December 12th, 2022. The peer review meetings included chartered panels, and recurring small business and fellowship special emphasis panels (SEPs). Measures in the survey include application evaluation, peer reviewer quality, reviewer meeting experience and participation, attention span, format preference, demographic information, and open-ended responses. Details of the above items can be found in [Appendix A](#). Supplemental data can be found in [Appendix B](#). In all figures the numbers above the data bars display the percent of respondents that endorsed that option.

Results

The survey was administered to 6,753 reviewers, of which 3,087 completed the survey for an overall response rate of 46%. Of the respondents, 64% attended virtual review meetings (n = 1,976) and 36% attended in-person review meetings (n = 1,111) and the response rates for virtual and in-person attendees were 42% and 53%, respectively. See Table 1 for reviewer characteristics.

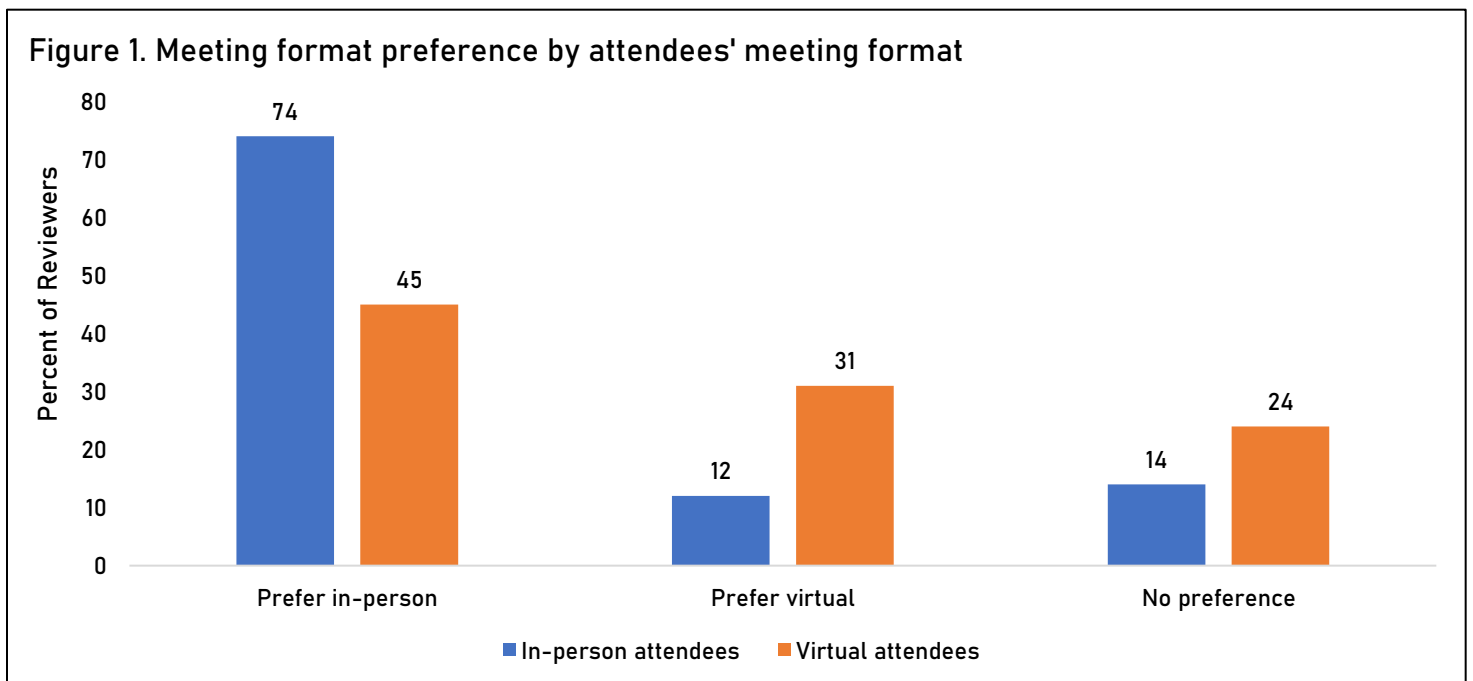
Reviewer Characteristics		% Survey Respondents (n = 3087)
Gender	Male	52
	Female	46
	Withheld	2
Race	American Indian or Alaskan	< 1
	Asian	22
	Black or African American	4
	More than one race	2
	Native Hawaiian or Pacific Islander	< 1
	White	64
	Withheld	8
Ethnicity	Hispanic/Latino	10
	Non-Hispanic	86
	Withheld	4

Under-represented minority (URM)¹		
	No	78
	Yes	15
	Withheld	7
Career Stage		
	Professor	44
	Associate Professor	37
	Assistant Professor	15
	Other	4

¹ According to OMB standards, individuals who identify with an Asian racial group other than Chinese, Filipino, Japanese, Korean, Asian Indian or Thai are considered an under-represented minority (URM). However, the current data does not allow for this level of group specificity, and therefore only Asian participants who identify as Asian and another racial group (other than white) or as Asian and Hispanic are coded as URM.

Meeting Format Preferences

- Figure 1 shows the meeting format preference of reviewers by the format in which they attended².
- Reviewers overall prefer in-person meetings, but the degree of preference varies based on the attendee's meeting format.
- In-person attendees were significantly more likely than virtual attendees to prefer in-person meetings.
- The preferences of virtual attendees were more equally distributed than preferences of in-person attendees.

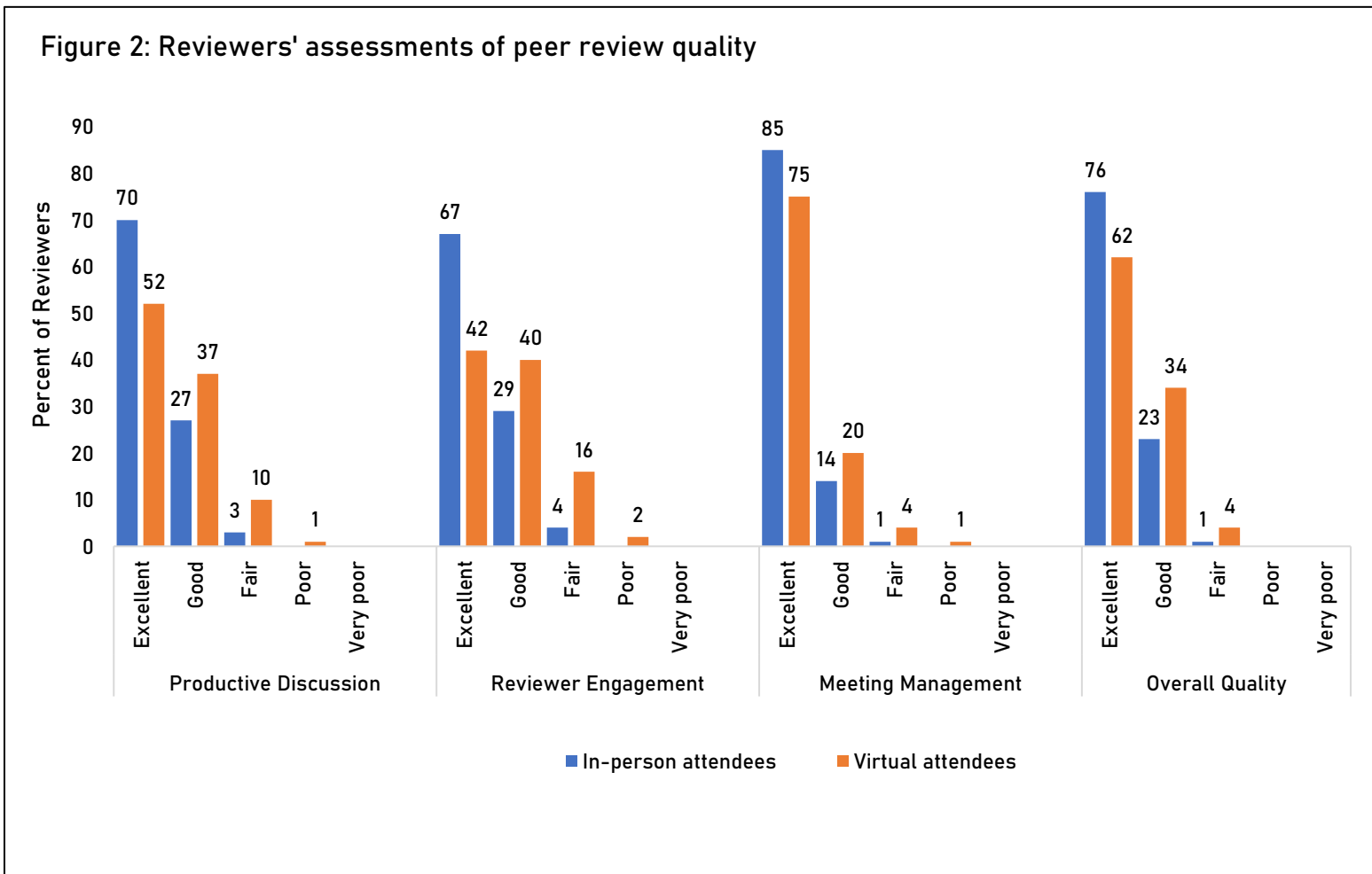


There was a significant association between the format of reviewers' meetings and their meeting format preferences $\chi^2(2) = 255.80, p < .001, \phi_c = .288$.

² Note that meeting formats were not assigned randomly. Rather, scientific review officers (SROs) volunteered to hold in-person meetings, taking into consideration the preferences of their reviewers. Thus, this finding could reflect those reviewers who, a priori, preferred in-person meetings were overrepresented on in-person panels, and that the converse was true for virtual meetings.

Quality of Review

- Figure 2 shows reviewer perceptions of the quality of the review meeting by meeting format.
- Over 60% of all reviewers thought the overall quality of the meetings were excellent.
- Reviewers who attended in-person meetings rated the meetings significantly more positive and of higher quality than those who attended virtual meetings—although the magnitude of these effects are small (see Table 2).
- The largest difference between attendee’s meeting perceptions occurred on the reviewer engagement question, with in-person attendees reporting higher reviewer engagement at their meetings.



On average, reviewers who attended in-person meetings believed that the panel had 1) more productive discussions 2) higher reviewer engagement 3) better meeting management and 4) a higher overall quality of review than reviewers who attended virtual meetings (see Table 2).

Table 2: Reviewers' Assessment of the Quality of the Review by Meeting Format

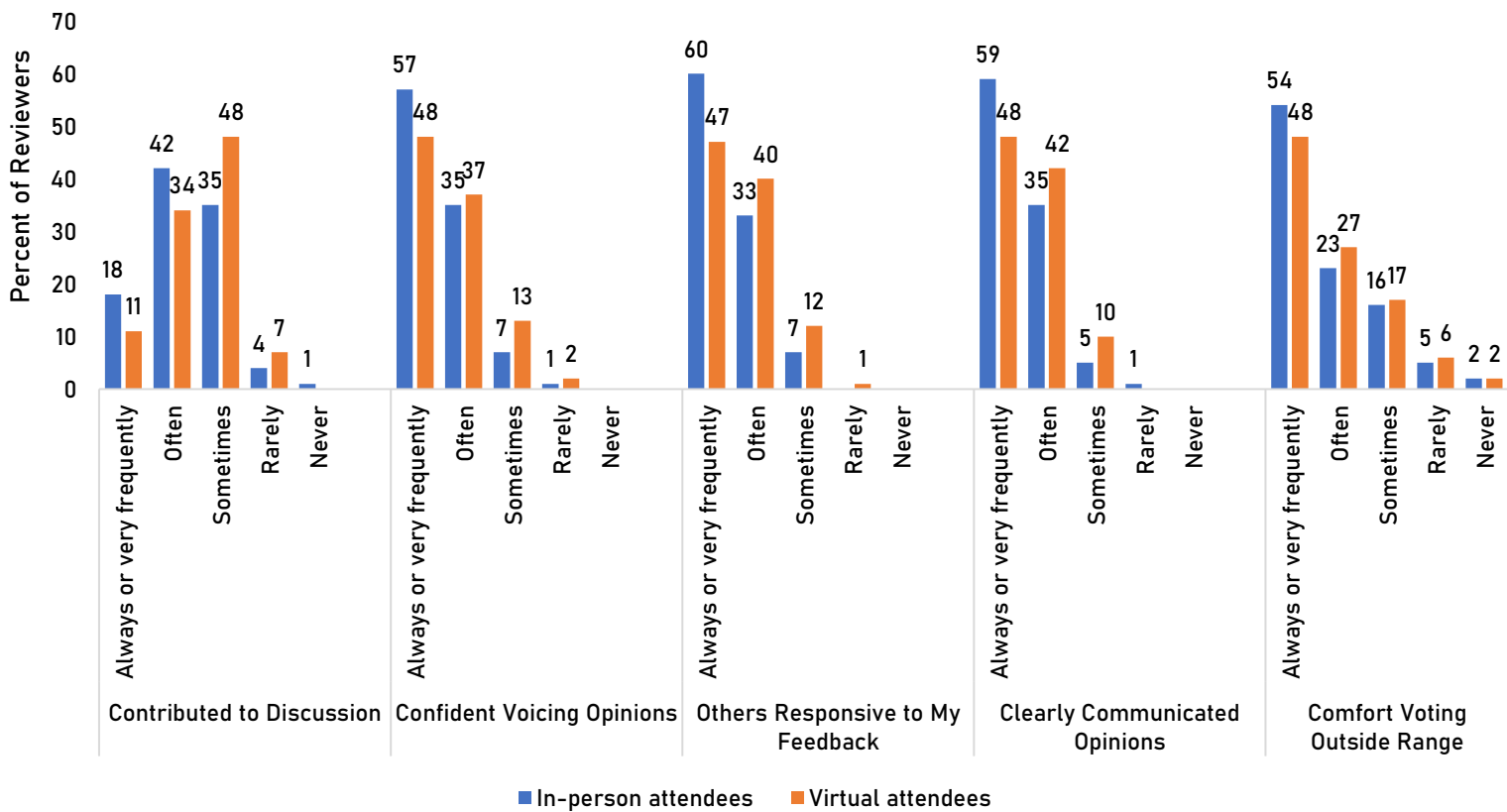
	Virtual Meetings (<i>M</i> , <i>SE</i>)	In-person Meetings (<i>M</i> , <i>SE</i>)	Independent t-test Statistic and <i>r</i>
Productive Discussions	4.39 (.02)	4.67 (.02)	$t(2837.430) = -11.889$, $p = .000$; $r = .22$
Reviewer Engagement	4.20 (.02)	4.62 (.02)	$t(2919.628) = -16.663$, $p = .000$; $r = .29$
Meeting Management	4.70 (.01)	4.81 (.01)	$t(2730.025) = -5.898$, $p = .000$; $r = .11$.
Overall Quality of Review	4.57 (.01)	4.74 (.01)	$t(2726.599) = -8.666$, $p = .000$; $r = .16$

Note: Survey questions were on a scale from 1 (very poor) to 5 (excellent), with higher scores indicating a more favorable view of the meeting.

Self-report of Meeting Experience and Participation

- Figure 3 shows reviewer ratings of their own experience and participation at the review meeting by meeting format.
- 75% or more of all reviewers rated the items listed in Figure 3 as “always” or “often”, with the exception of their contribution to the discussion.
- 60% of reviewers who attended in-person meetings and 45% of reviewers who attended virtually reported that they always or often contributed to the discussion. Reviewers’ contribution to the discussion was the largest reported difference between in-person and virtual meetings.
- Overall, reviewers who attended in-person meetings participated more and reported more favorable perceptions than those who attended virtual meetings—although the magnitude of these effects are small (see Table 3).

Figure 3. Reviewers' meeting experience and participation



On average, reviewers who attended in-person meetings stated that they 1) contributed more to the discussion 2) felt more confident voicing their opinions 3) felt others were more receptive and responsive to their feedback 4) were better able to clearly communicate their opinions and 5) felt more comfortable voting outside the range than reviewers who attended virtual meetings (see Table 3).

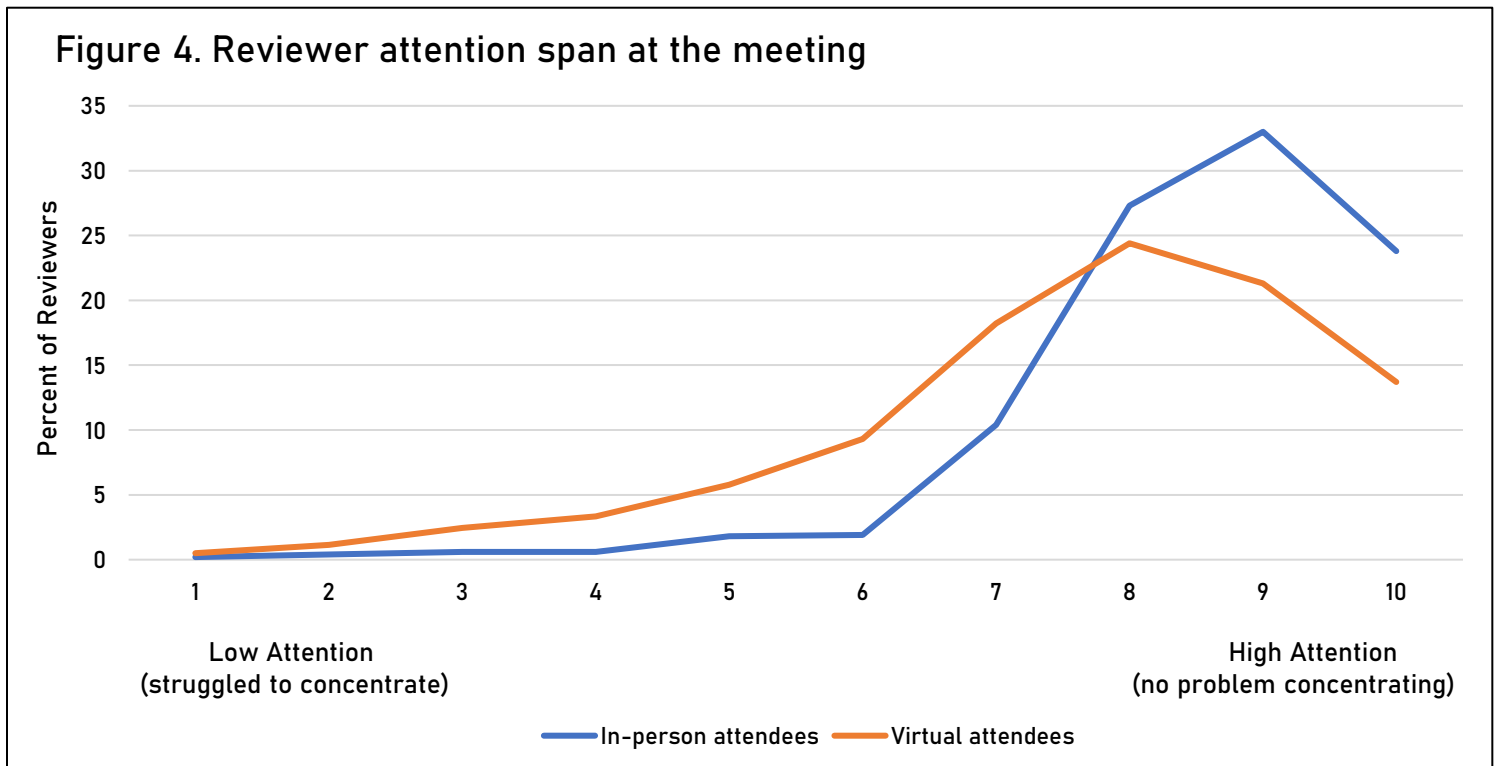
Table 3. Reviewers' Meeting Experience and Participation by Meeting Format

	Virtual Meetings (<i>M</i> , <i>SE</i>)	In-person Meetings (<i>M</i> , <i>SE</i>)	Independent t-test Statistic and <i>r</i>
Contributed to Discussion	3.48 (.02)	3.72 (.03)	$t(3080) = -8.173$, $p = .000$; $r = .14$
Confident Voicing Opinions	4.31 (.02)	4.47 (.02)	$t(2510.971) = -6.183$, $p = .000$; $r = .13$
Others Receptive and Responsive to Feedback	4.31 (.02)	4.50 (.02)	$t(2516.413) = -7.270$, $p = .000$; $r = .15$
Clearly Communicated Opinions	4.37 (.02)	4.52 (.02)	$t(2478.927) = -6.331$, $p = .000$; $r = .13$
Comfortable Voting Outside Range	4.11 (.03)	4.23(.03)	$t(2615) = -2.851$, $p = .004$; $r = .06$

Note: Survey questions were on a scale from 1 (never) to 5 (always or very frequently), with higher scores indicating more or higher quality participation at the meeting.

Reviewer Attention Span

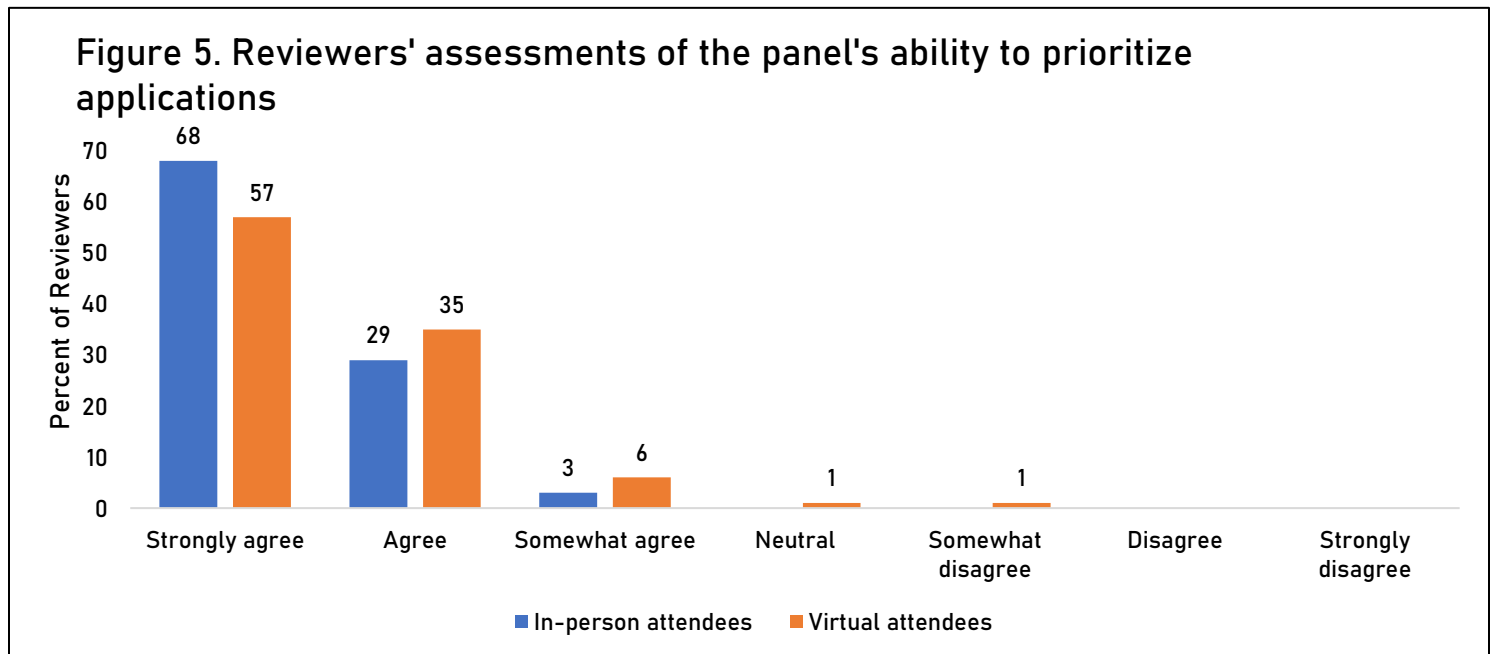
- Figure 4 shows reviewers' self-reported ability to pay attention throughout the meeting, by meeting format.
- On average, reviewers who attended in-person meetings paid attention more or had longer attention spans than those who attended virtual meetings—although the magnitude of this effect was small.



On average, reviewers who attended in-person meetings were able to concentrate better at the meeting ($M = 8.52$, $SE = .04$) than reviewers who attended virtual meetings ($M = 7.59$, $SE = .04$). This difference, $-.926$, $CI [-1.04, -.81]$, was significant $t(2881.745) = -15.916$, $p = .000$; and represented a small-sized effect, $r = .28$.

Prioritizing Applications

- Figures 5 show data capturing reviewers' perceptions of the panel's ability to prioritize applications by meeting format.
- Over 90% of all reviewers believed that the panel was able to prioritize applications according to their impact and scientific merit.
- Reviewers who attended in-person meetings were better able to prioritize applications than those who attended virtual meetings—although the magnitude of this effect was small.

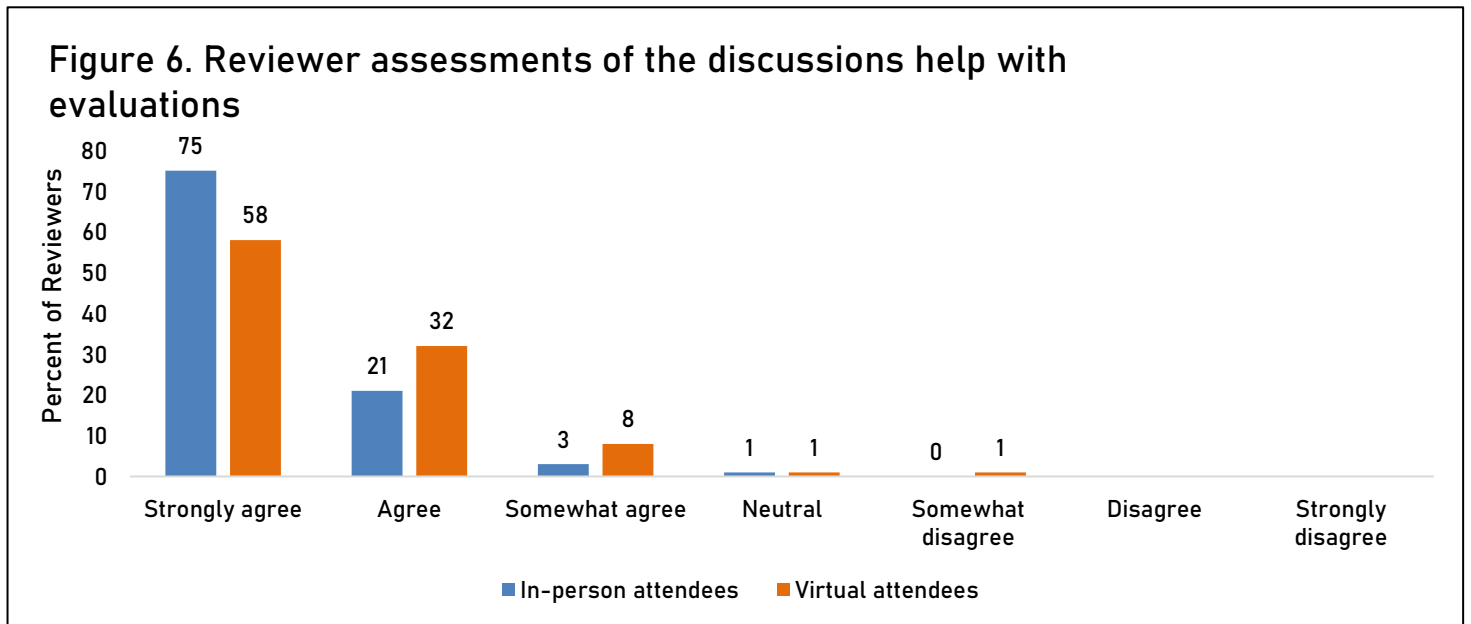


Note: Survey questions were on a scale from 1 (strongly agree) to 7 (strongly disagree), with lower scores indicating a more favorable view of the review meeting.

On average, reviewers who attended in-person meetings believed that the panel was better able to prioritize applications according to their impact and scientific merit ($M = 1.39$, $SE = .02$) than reviewers who attended virtual meetings ($M = 1.55$, $SE = .02$). This difference, $.159$, $CI [.106, .213]$, was significant $t(2667.346) = 5.874$, $p = .000$; and represented a small-sized effect, $r = .11$.

Discussions and Evaluations

- Figure 6 shows data capturing reviewers' perceptions of the panels ability to discuss and evaluate applications by meeting format.
- Around 90% of all reviewers believed that the scientific discussions helped the panel evaluate the applications being reviewed.
- Reviewers who attended in-person meetings had better discussions than those who attended virtual meetings—although the magnitude of this effect was small.



Note: Survey questions were on a scale from 1 (strongly agree) to 7 (strongly disagree), with lower scores indicating a more favorable view of the review meeting

On average, reviewers who attended in-person meetings believed that the discussions helped the panel evaluate the applications more ($M = 1.31, SE = .02$) than reviewers who attended virtual meetings ($M = 1.57, SE = .02$). This difference, $.268, CI [.216, .321]$, was significant $t(2810.331) = 9.987, p = .000$; and represented a small-sized effect, $r = .19$.

Meeting Measures Data and Analysis

Methods

This section will focus on meeting measures, outcomes, and characteristics of a subsample of the January 2023 Advisory Council review meetings, which will be referred to as the score & roster sample, while the survey results previously described in this report utilized the full sample of standing study sections, fellowship, and small business panels. The score & roster sample is derived from the methods used in a previous Zoom analyses done by CSR and, to allow comparability, was carried forward for the present analyses. The score & roster sample of 92 meetings were analyzed to compare roster composition, reviewer workload, score distributions and extent of out-of-range scoring for in-person vs. Zoom meetings held in the 2023/01 round. Detailed methods can be viewed in [Appendix C](#).

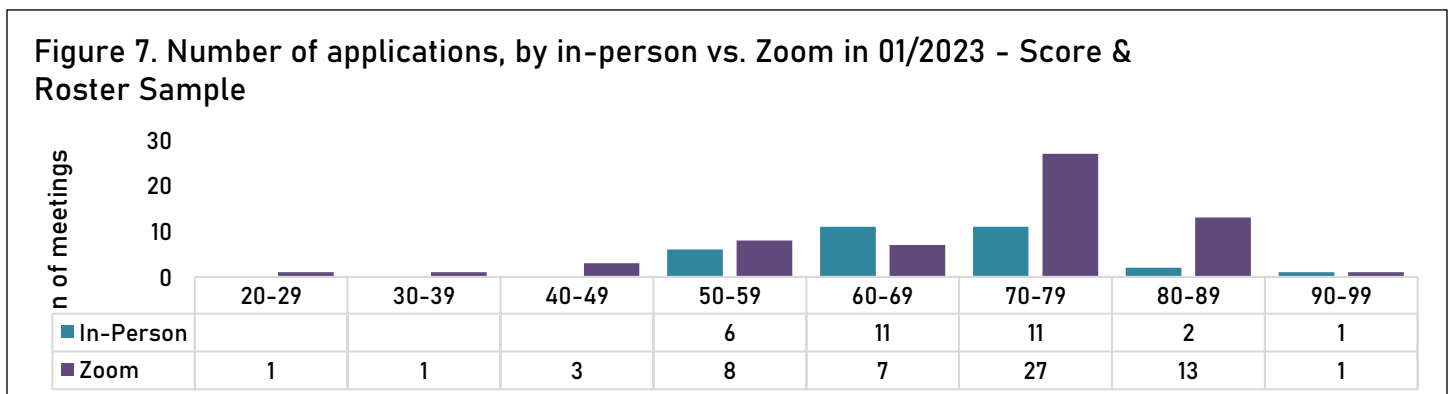
Table 4 classifies the meetings that were included in the analysis. There are two levels of analyses examined in this section: the score & roster sample (n=92, including small business and fellowship meetings), and standing study sections only (n=73). There were not enough fellowship or small business meetings in the score & roster sample to support informative, reliable analyses of the subtypes. Approximately 34% of meetings that were included in the analysis were held as in-person meetings while 66% were held as Zoom meetings.

Meeting Classification	Score & Roster Sample	Standing Study Sections Only
In-person meeting	31	28
<i>Standing Study Section</i>	28	28
<i>Fellowship</i>	2	-
<i>Small Business</i>	1	-
Zoom meeting	61	45
<i>Standing Study Section</i>	45	45
<i>Fellowship</i>	8	-
<i>Small Business</i>	8	-
Grand Total	92	73

Meeting Application Counts, Roster Sizes, and Reviewer Workload Trends

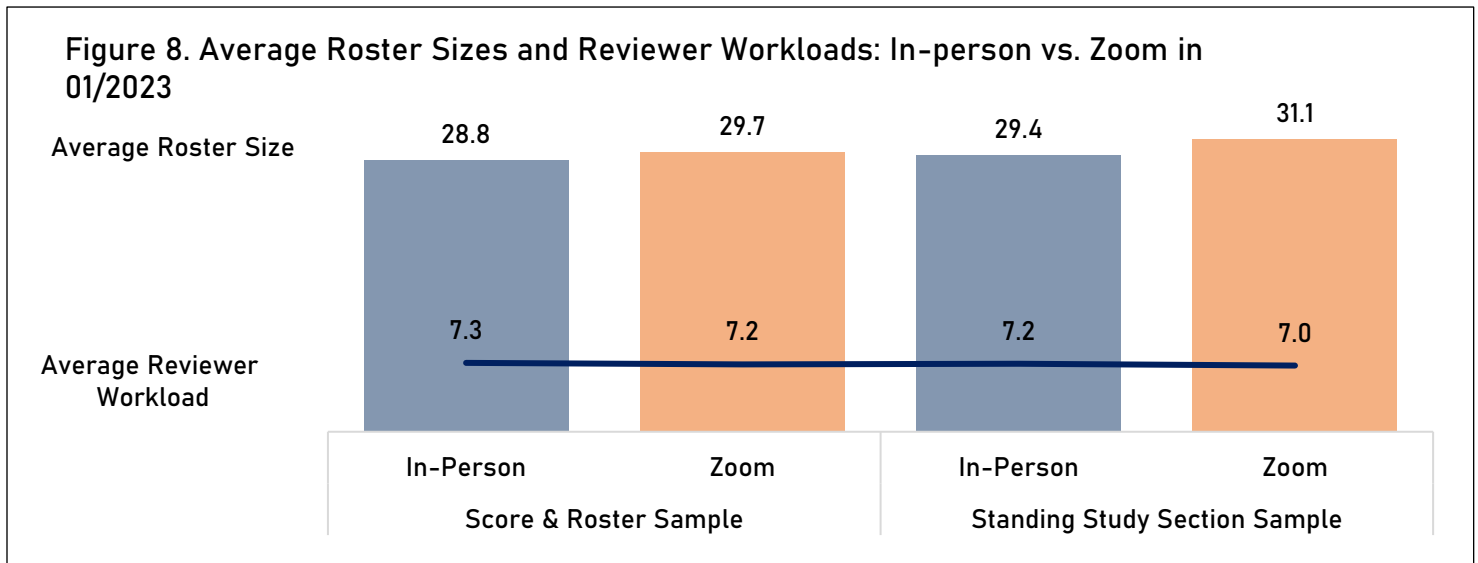
Meeting Application Counts

Figure 7 shows the distribution of meetings according to size and format – the average application counts for in-person meeting were 68.5 applications and for zoom meetings, 70.6 applications. Approximately 97% of the meetings held in-person were two-day meetings, and 94% of Zoom meetings were two-day meetings.



Roster Size and Reviewer Workloads

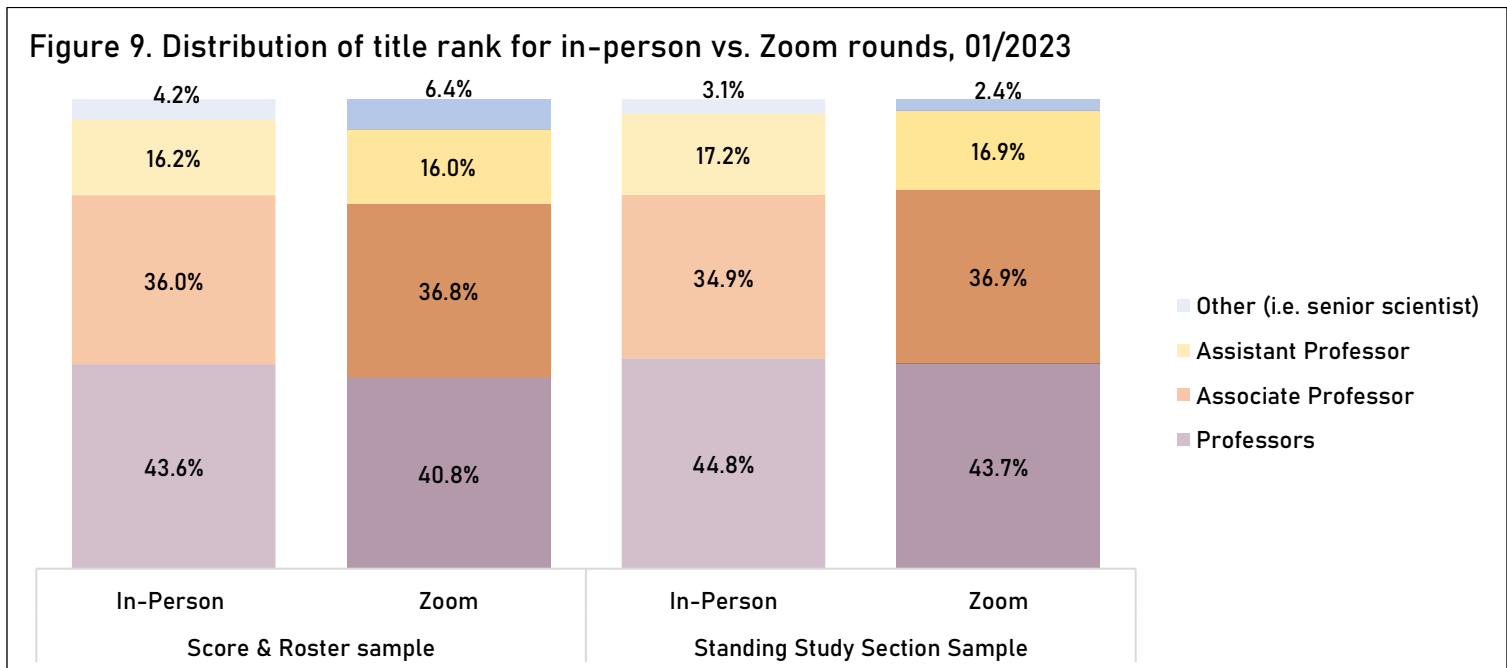
Average roster sizes and reviewer workload (average number of applications assigned per reviewer) are shown in Figure 8. Roster sizes for Zoom meetings was on average about 1-2 persons larger than those for in-person meetings while average reviewer workloads for Zoom meetings were fractionally lower than those for in-person meetings.



Roster Composition Trends

Faculty Rank Distribution

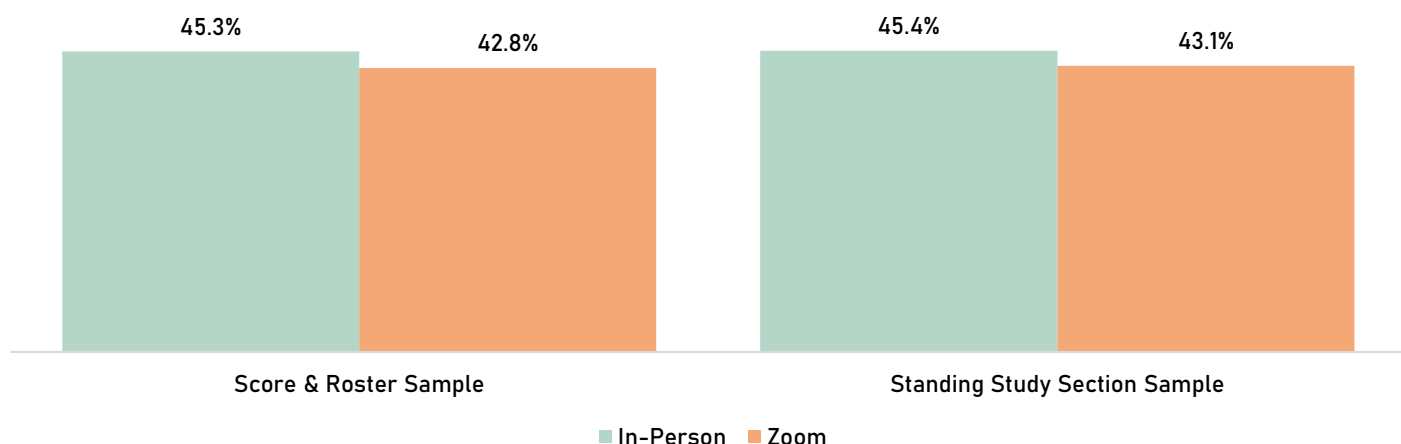
Figure 9 displays the distribution of faculty rank for in-person and Zoom meetings, by meeting type. We do not see meaningful differences in the distribution of reviewers according to faculty rank.



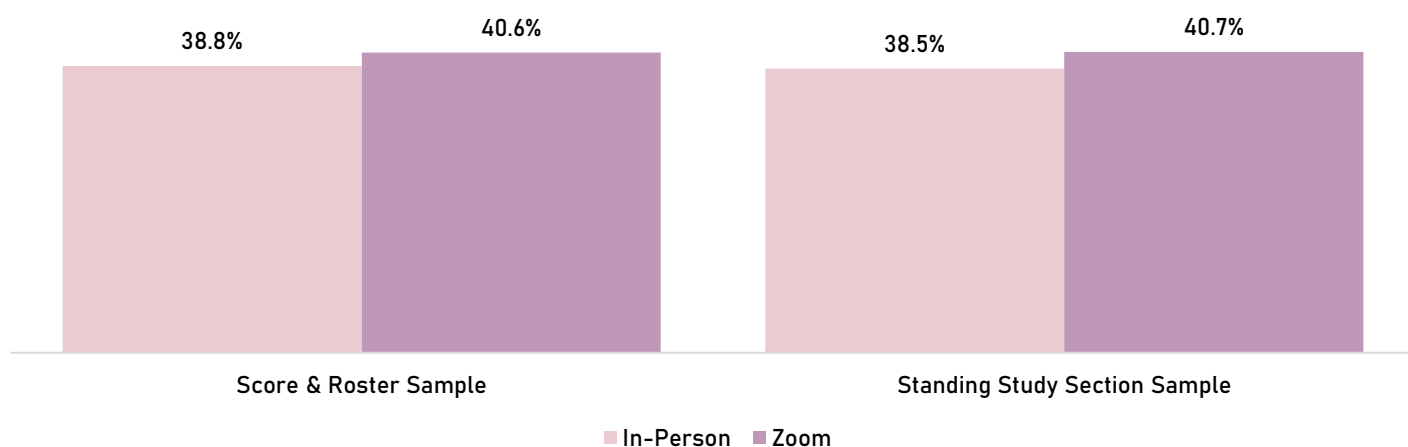
Demographic Distribution

Figures 10 and 11 display the proportion of female and minority reviewers present on meeting rosters. Across the score & roster sample, female reviewers account for approximately 43% of reviewers. Small numerical differences according to meeting format were not statistically significant (females, all meetings ($p=0.33$); females, standing study section meetings ($p=0.21$); minorities, score & roster sample ($p=0.44$); minorities, standing study section sample ($p=0.44$)).

**Figure 10. Average female representation for in-person vs. Zoom meetings
01/2023**



**Figure 11. Average minority representation for in-person vs. Zoom meetings in
01/2023**



We also assessed the representation of underrepresented minorities (URM) and women among ad hoc reviewers. Demographics of ad hoc reviewers was assessed separately because ad hocs change round to round whereas standing members do not. URMs in biomedical research at NIH are defined as individuals that identify as: Blacks or African American, Hispanic or Latino, American Indian or Alaska Native, Native Hawaiian, and other Pacific Islander. Average URM representation for all members and ad-hoc members is in Table 5 and average female representation is in Table 6.

Table 5. URM representation	Average ad-hoc URM representation	
	In-person	Zoom
Score & Roster Sample	11.8%	12.8%
Standing Study Section Sample	13.6%	15.1%

Table 6. Female representation	Average ad-hoc female representation	
	In-person	Zoom
Score & Roster Sample	46.4%	42.1%
Standing Study Section Sample	46.7%	42.1%

Numeric differences in average URM representation between in-person and Zoom meetings were not significant for either the full score & roster sample ($p= 0.29$) or standing study section subsample ($p= 0.28$).

Numeric differences in average female representation between in-person and Zoom meetings were not significant for either the full ($p= 0.42$) or standing study section subsample ($p= 0.10$).

Extent of Prior Review Service among Ad-hoc Reviewers

An independent t-test revealed significant differences in prior review service between ad hoc reviewers attending in-person and Zoom standing study section meetings in 2023/01 ($p = 0.03$), with an average of 6.7 prior reviews for in-person meetings and 5.6 prior reviews for Zoom meetings. However, the independent t-test for the full score & roster sample (including small business and fellowship meetings) was not significant ($p = 0.93$), with an average of 6.6 prior reviews for in-person meetings (-0.1 compared to standing study section sample) and 7.4 prior reviews for Zoom meetings (+1.8 compared to standing study section sample).

It should be noted that while standing study sections have standing members that are excluded from this analysis, small business panels and fellowships (in the score & roster sample) do not have standing members, so the entire panel of ad-hoc reviewers is accounted for those meetings. ECRs were included in this analysis and should be considered when interpreting data on reviewers with 0 reviews – ECRs are not allowed in small business and fellowship reviews but as per CSR policy, each standing study section meeting requires 2 ECRs. These meeting characteristics may explain discrepancies in the score & roster sample vs. standing study section sample, which includes small business panels and fellowships.

New Reviewer Recruitment

Table 7 shows the average number of ad-hoc reviewers with little to no review service that were recruited to serve on the panels. There were slightly more ad-hoc reviewers with 1-2 prior meetings on average attending Zoom meetings than in-person meetings.

Table 7. New Reviewer Recruitment	Average N of ad-hoc reviewers with 0 prior meetings in each panel		Average N of ad-hoc reviewers with 1-2 prior meetings in each panel	
	In-Person	Zoom	In-Person	Zoom
Score & Roster Sample	3.5	3.5	2.6	3.1
Standing Study Section Sample	3.5	3.7	2.2	2.6

Reviewer Excessive Service

In the score & roster sample, there were only 7 instances of reviewer excessive service, as defined by 45 or more reviews, amongst recruited ad-hoc reviewers across 5 meetings. The 7 reviewers had with a range of 46-59 previous reviews at the time of recruitment. No pattern was observed across meeting format (in-person vs. Zoom) or meeting type (standing study section, fellowship, or small business meetings).

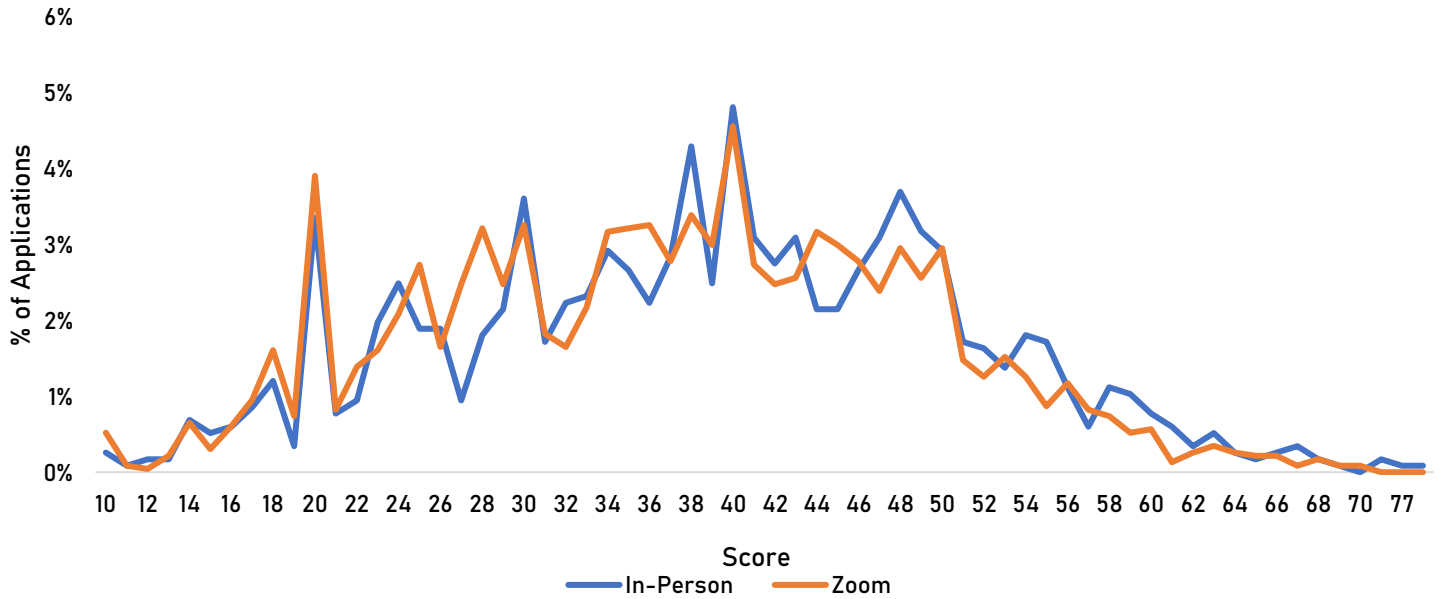
Scores Analyses

Final Overall Impact Score Distributions

Table 8 shows the descriptive statistics of final score distribution for Zoom and in-person meetings and the distribution are graphed in Figure 12. The distributions appear very similar without evidence of greater score compression or a shift of scores in either a positive or negative direction. Along with the difference in score means is statistically significant, the effect size was large with a .85 using a Hedges’s calculation due to the different sample sizes.

Table 8. Final Score Distribution Descriptive Statistics	In-Person meetings	Zoom meetings
N of Overall Impact scores	1164	2303
Mean	38.83	37.25
Median	39	38
Standard Deviation	12.03	11.6
Kurtosis	-0.41	-0.53

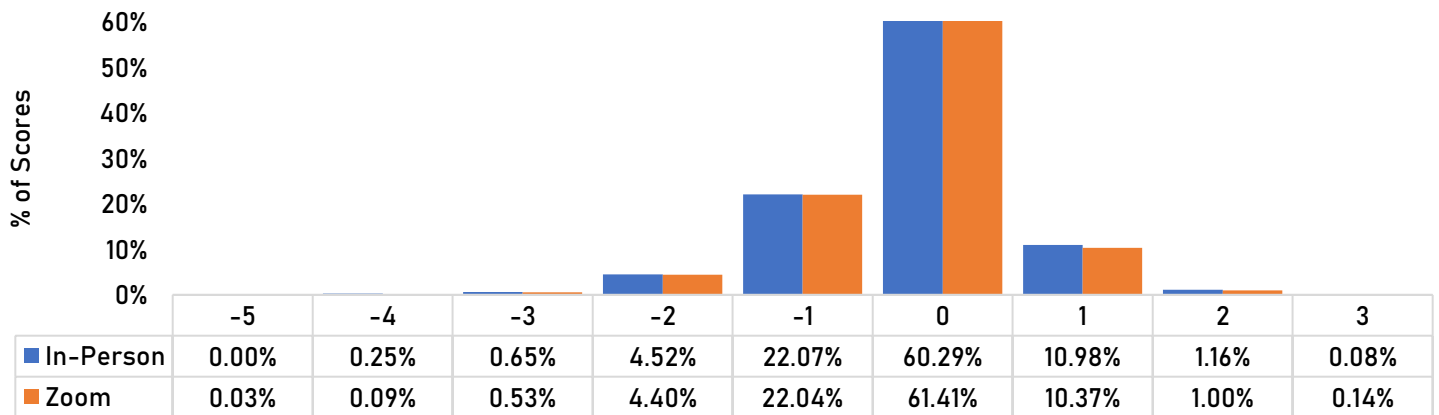
Figure 12. Score & Roster Sample: Final overall impact score distribution, in-person vs. Zoom meetings 01/2023



Changes in Scoring (Preliminary to Final Overall Impact Score)

Changes in scoring were measured in assigned reviewers by the magnitude and direction of change from the reviewer’s preliminary score to final score. Figure 13 shows the distribution of score changes by meeting format. No notable differences were observed.

Figure 13. Score & Roster Sample: Positive or Negative Movement from Preliminary Scores to Final Scores by Assigned Reviewers



Extent of Out-of-Range Scoring

Table 9 shows the percentage of final scores that were out of range for the in-person meetings and Zoom meetings. No notable differences are seen.

Table 9. Extent of out-of-range scores, In-Person and Zoom				
Inclusion Criteria	Score & Roster Sample		Standing Study Section Sample	
Meeting Type	In person	Zoom	In Person	Zoom
N of scores	28,051	57,134	25,867	44,538
% out of range scores	3.6%	3.5%	3.4%	3.3%

Appendix A: Survey methods

Participants

Reviewers who participated in 238 CSR meetings (n = 3,087) between September 26th to December 12th, 2022. The study section meetings included chartered panels, and recurring small business and fellowship special emphasis panels (SEPs).

Survey Administration

Reviewers were asked for their participation in a survey via email on the last day of the study section meeting. The email contained a weblink to the survey. Reviewers were told in the email that their responses would be kept confidential and that the survey would take less than five minutes to complete. All surveys returned by January 4th, 2023 were included for analysis.

Measures

Application Evaluation

Two items asked participants to rate on a scale from 1 (strongly agree) to 7 (strongly disagree) the panel's ability to evaluate the applications: 1) the panel was able to prioritize applications according to their impact and scientific merit, and 2) the scientific discussion helped the panel evaluate the applications being reviewed.

Peer Review Quality

Four items asked participants to rate on a scale from 1 (very poor) to 5 (excellent) the following items: 1) overall quality of review, 2) productivity of discussions, 3) level of reviewer engagement, and 4) meeting management.

Reviewer Meeting Experience and Participation

Six items asked participants to rate on a scale from 1 (never) to 5 (always or very frequently) the following items: 1) I contributed to the discussion, 2) I felt confident voicing my opinions, 3) I felt others were receptive and responsive to my feedback, 4) I was able to clearly communicate opinions, 5) I felt comfortable voting outside the range, and 6) My attention span at the meeting lasted.

Attention Span

One item asked participants to rate their attention span at the review meeting using a scale from 1-10, with 1 being really struggled to concentrate and 10 being no problem concentrating at all.

Format Preference

One question asked participants if there were no or minimal health risks from COVID-19, would they be more likely to participate in a review meeting if it was held face to face or over Zoom/video? Response options included: face-to face, Zoom/video, and no preference.

Demographic Information

Four questions were used to collect the demographic characteristics of respondents. 1) Gender: male, female, I prefer not to respond; 2) Race: American Indian or Alaskan Native, Asian, Black or African American, Native Hawaiian or Pacific Islander, White, I prefer not to respond; 3) Ethnicity: Are you Hispanic? Yes, No, I prefer not to respond; 4) Career stage: Assistant Professor, Associate Professor, Professor, Other.

Participants' race and ethnicity were used to determine whether they were an underrepresented minority or not. Non-Hispanic Asians and Non-Hispanic whites were coded as "not URM" and all other participants were coded as "URM". For participants who identified with more than one racial group, if one racial identity was not white or Asian, they were coded as "URM". Participants who identified as both white and Asian were coded as "not URM".

Open-ended Response Options

In an open-ended text box, participants were asked to please share any comments (positive or negative) about their experience or general thoughts on their recent review meeting.

Appendix B: Supplemental survey data

Supplemental Tables

Table 10. Number of Reviewers by Meeting Format (n = 3087)	
Meeting Format	n (%)
Virtual	1976 (64)
In-person	1111 (36)

*105 respondents participated in-person meetings but attended virtually. These respondents are included in the in-person meeting category.

Table 11. Number of Reviewers by Type of Committee Participation (n = 3087)	
Type of Review Committee	n (%)
Standing study section	2534 (82)
Special emphasis panels (SEP)	553 (18)

Table 12. Number of Reviewers by Type of Committee Participation (n = 3087)		
	Committee Type, n (%)	
Meeting Format	Standing Study Section	SEP
Virtual	1487 (75)	489 (25)
In-person	1047 (94)	64 (6)

*105 respondents participated in-person meetings but attended virtually. These respondents are included in the in-person meeting category.

Table 13. Geographic Residence of Reviewers and Attendee Meeting Format (n = 3087)		
Reviewers Residential Time zone	Virtual Attendee n (%)	In-person Attendee n (%)
Eastern Time (ET)	1017 (52)	599 (54)
Central Time (CT)	539 (27)	312 (28)
Mountain Time (MT)	104 (5)	41 (4)
Pacific Time (PT)	304 (15)	154 (14)
Alaska Time (AT)	1 (<1)	0 (--)
Hawaii-Aleutian Time (HAT)	3 (<1)	1 (<1)
Other U.S. time zone	6 (<1)	1(<1)
I do not live in a U.S. time zone	2 (<1)	3 (<1)

Table 14. Number of New Reviewers by Meeting Format (n = 395)	
Meeting Format	n (%)
Virtual	255 (13)
In-person	140 (13)

Table 15. Number of Reviewers who Experienced Technical Difficulties by Meeting Format (n = 563)	
Meeting Format	n (%)
Virtual	254 (13)
In-person	309 (28)

Table 16. Number and Format of NIH Review Meetings Ever Attended by Reviewers (n = 3087)		
Number of NIH Review Meetings Attended	Virtual Meetings Attended n (%)	In-person Meetings Attended n (%)
0	165 (5)	627 (20)
1	352 (11)	450 (15)
2-5	1682 (55)	777 (25)
6-15	860 (28)	802 (26)
16+	28 (1)	431 (14)

Appendix C: Detailed Methods for quantitative analyses

Overview

This document serves to document the detailed methodology for meeting selection, data collection, and analyses and reporting for the quantitative analyses in report.

Meeting Selection and Matching

This analysis utilized a matched sample of standing study sections, small business, and fellowships selected for the Summer 2020 Zoom quantitative analyses conducted by CSR. The sample has reduced from 119 meetings in the 2020 analyses to 92 meetings due to changes in study sections as a result of ENQUIRE evaluations and internal reorganization from IRG's to review branches. Meetings that did not have an equivalent for longitudinal analyses were excluded. This sample is referred to as the score & roster sample, which is a subset of meetings that were surveyed during 2023/01.

Data Collection

Information for roster and scores for the 2023/01 92 meeting sample were primarily extracted from CMM and QVR. Mail reviewers were excluded for all roster analyses, while ECRs were included but not specified. Information regarding female and minority representation were extracted from the Committee Management Module, while information regarding title rank was manually extracted from roster reports.

Extent of prior review service is measured for adhoc reviewers only participating in one of the 92 meetings in the 2023/01 sample. CSR's informatics team provided cross-sectional reviewer-level data on extent of prior review service,

adjusted for one week before the meeting to get the most accurate data at time of recruitment. Meeting counts for prior extent of service span a 12-year period and include NIH review meetings IC/CSR chartered, SEPs, and IC NACs for application funding, as well as IC meetings, telephone meetings. Non-FACA meetings, mail reviews, and CSR rump SEPs were excluded from meeting counts.

Meeting size parameters were set in the following way for comparison analyses, which were restricted to standing study sections, fellowships, and small business panels: small meetings (1-55 apps), medium meetings (56-89 apps), and large meetings (90+ apps).

Limitations

- Limited fellowships and small business panels were included in the score & roster sample, and for in-person meetings, only 1-2 meetings were included in each meeting type. Limited conclusions can be derived regarding differences in formats in these meeting types.
- ECRs were not separated out, so all roster analyses metrics will include their metrics as well, including prior review service, where they will disproportionately represent reviewers in standing study sections who have had 0 prior meetings and are “newly engaged in the review system”.