

The Blind Review: A Quality Assessment Measure for Review Standardization of Institutional Review Boards

Sinyoung Park

Yonsei University Health System

Cho Rong Ahn

Yonsei University Health System

Yang Hee Noh

Yonsei University Health System

Sejung Park

Yonsei University College of Medicine

Chung Mo Nam

Yonsei University College of Medicine

Myoung Soo Kim

Yonsei University Health System

Seung Min Kim

Yongin Severance Hospital

Jin Seok Kim

Yonsei University Health System

Sun Young Rha (✉ rha7655@yuhs.ac)

Yonsei University Health System

Research article

Keywords: Institutional Review Board, Standardization, Quality Assurance

Posted Date: July 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-42889/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Challenging research environment while the multinational trials are increasing, standardization of IRB review has become more important. The inconsistency is often inevitable in that opinions on the ethical part may be different, but the standardization and understanding the differences is required to assure IRB review quality. Thus, we aimed to develop and suggest quality assessment measure of IRB named “Blind review” by reviewing the same research protocols by multiple IRB panels. We further describe an analyzed result of differences to understand the mechanism of IRB standardization.

Methods: We present a description of the Blind review process. Based on the HRPP (Human research Protection Program) SOPs (Standard of Procedure), eight blind review results from January 2010 to December 2018 at a single institution with multiple panels were included using Severance Hospital HRPP database. Review results were analyzed with review scores ranged 0 to 60 points including GCP requirements and protocol issues. Panel agreement was estimated by observed multiple rater agreement. Differences between review scores according to expertise member and IRB member duration were analysed with Wilcoxon rank-sum test and Kruskal-Wallis test.

Results: Observed multiple raters’ agreement was increased from 0.444 (95% CI: 0.167-1.000) in 2010 to 0.479 (95% CI: 0.271-0.708) in 2014~2018 as IRB review experiences increases. In order to analyze the review mechanism, three GCP requirements and three protocol issues were scored (range 0 to 60). Mean values for GCP requirement and protocol issues were 19.25 ± 8.21 and 18.40 ± 9.04 , respectively. Mean score of the panels where experts participated ($n=16$, 28.13 ± 10.47) is higher than control group ($n=32$, 25.16 ± 10.96) ($p=0.93$). According to career duration of IRB members’, scores of the group whose career is less than 3 years was 25.0 ± 10.0 ($n=14$), the score of group whose career is 3-5 years was 26.3 ± 9.6 ($n=23$). The mean score of the group whose career was more than 5 years shown 27.3 ± 14.2 ($n=11$), and it is statistically significant. ($p=0.09$)

Conclusions: We suggest Blind Review as an effective measure for overseeing and ensuring IRB’s review quality and overall GCP compliance.

Background

The Institutional Review Board (IRB), also known as the Independent Ethics Committee (IEC) or Ethical Review Board (ERB), plays an important role in protecting the rights and welfare of human research subjects. [1] IRB reviews the consent process, recruitment procedure, risk and benefit analysis, and compensation arrangement, as well as scientific validity, of submitted documents. [2] In recent decades, medical knowledge and research techniques have developed dramatically. In response, laws and regulations related to clinical research ethics have become more stringent than ever. [3, 4] With increasing multinational clinical trials, both domestic laws and international regulations are becoming more important, and whenever there is any discrepancy, the stronger laws should be followed in each IRB review. [5] Such changes to the research environment create problems for both IRB members and investigators. [6] There have been discrepancies between IRB’s decisions regarding multicenter studies, even when following the same protocol. [7] These divergences may be due to different interpretations of good clinical practice (GCP), different local regulatory requirements, or the individual characteristics of IRB members.

In 1982, Veatch demonstrated the inconsistency in IRB review results for the first time. [8] Goldman and Katz [9] also analyzed IRB review results using the same protocols. The inconsistency in IRB decisions is an inveterate concern both internally and for investigators, as it becomes an additional burden. Several investigations have analyzed the inconsistencies in IRB panels' results. [10–14] Due to the characteristics of the IRB and its reviews, including its diverse members and the complex study designs, inconsistencies are inevitable. Therefore, we must analyze why such inconsistencies occur [13] based on their details. However, research on review mechanisms or its associated factors is lacking. Most data have only focused on inconsistencies using one or two protocols. Therefore, we developed “Blind review” as a quality measure that compares review results of different IRB panels following the same protocol to determine the degree of IRB standardization. We evaluated the effectiveness of this method, as well as factors affecting reviews, especially those relating to IRB members' experiences. This study pilots the Blind Review and provides information on its effectiveness based on data from a single institution.

Methods

Blind Review

We developed the Blind Review methodology as a part of quality assurance (QA) activities within human research protection programs (HRPP) to analyze consistency among IRB panels' decisions. Under this process, one protocol is selected for blind review—this protocol is submitted to every IRB panel except the one that originally made the review. Three HRPP staff members with more than two years of experience in the Human Research Protection Center (HPC) selected the protocol. Although there was no specific criteria for selection, we considered current research trends and specific issues, such as Artificial Intelligence software study, verbal consent, and germline mutation study, whose review guidelines are a work in progress. Therefore, we included various research types to compare the decisions made by different IRB panels, and did not exclude poorly designed protocols. Identical protocols were submitted to each panel simultaneously via a convened meeting, and IRB members were not notified about the Blind Review. After the Blind Review, HPC Quality Assurance (QA) staff assessed the results and shared them in the IRB member workshop. However, the investigators were not notified about the Blind Review results or that it was not obligatory to respond to every IRB panel's comments. The IRB's decision for the protocol was made by the original review panel. A detailed description of the Blind Review process is shown in Fig. 1.

Materials

As a QA measure, the HPC at Severance Hospital, Yonsei University College of Medicine, Seoul, Korea, has conducted a Blind Review annually according to the HRPP standard of procedure (SOP). We retrospectively analyzed the Blind Review records of eight protocols conducted between January 2010 and December 2018. The first three results, in 2010, were pilot tests across four panels; the other five, one for each year from 2014 and 2018, were across seven panels and streamlined with the SOP. No Blind Review was conducted between 2011 and 2013 as our institution had to develop an e-IRB system from the paper review process. We qualitatively summarized the experiences of IRB QA measures. We collected IRB member designation logs from the HPC database to assess IRB member factors such as IRB experience duration, major specialty. We

did not enroll any human research subjects, nor their personal data. Therefore, IRB approval was not required for this study.

Review criteria specified by the Severance Hospital IRB

Severance Hospital follows domestic regulations for research involving human participants, such as the Pharmaceutical Affairs Act, Bioethics and Safety Act, Korean GCP (Good Clinical Practice) guidelines, and the Medical Device Act. However, if there are issues not specified in domestic laws, and if the following regulations are stricter than the ones specified in domestic laws, the internal regulations will be provided by referring to ICH (International Council for Harmonization) GCP guidelines, regulations of DHHS (Department of Health and Human Services) of USA, federal law (Federal and local law), regulations for participants of FDA (21 CFR 50, 56, 312 and 812), etc. The criteria of IRB review and approval are as follows:

- (1) Approval: If the research activity conforms to the approval criteria defined by the related law, and if modifying the research is not recommended.
- (2) Approved with modifications: The research activity meets the criteria for approval, and the modifications required by the IRB are only minor, and would be considered of minimal risk.
- (3) Deferred: The study does not meet the criteria for approval as defined in the relevant laws and regulations, lacks sufficient information to conduct an adequate review, and/or if substantial revisions to the protocol are necessary.
- (4) Disapproved: The study is found to be unethical, without scientific or scholarly merit, and/or does not meet the criteria for approval as defined in the relevant laws and regulations.
- (5) Tabled: A study that cannot be reviewed at the meeting due to lack of time, lack of quorum, and/or extenuating circumstances.

Scoring method for IRB review results

To assess the review quality and standardization, HPC staff developed “essential review points” based on GCP and HRPP guidelines and “protocol review points” that should be detected by IRB members. These are to be discussed during review. We developed a 10-point scale, where 10 denotes both reviewers mentioning each review point, and 5 denotes the primary reviewer mentioning it but not the secondary reviewer. This applied even when a wrong comment was made. A score of 0 was allocated when none of the IRB members discussed or mentioned the issue. For GCP requirements, we identified “Risk and benefit analysis” (ICH-GCP 2.2), “Determining the continuing review interval” (ICH-GCP 3.3.4), and “Research resources” (ICH-GCP 3.1.3., 3.1.9) as mandatory review points based on GCP regulations. [15] In addition, we assessed whether the panels identified issues to be discussion according to each protocol and its detailed items, derived by the HPC staff. The consent process, vulnerability of participants, waivers for stored specimens, concerns regarding the confidentiality of participants, randomization methods, and the appropriateness of placebo procedures for the control group formed the majority of these considerations.

Statistical Analysis

Descriptive data were expressed as mean \pm standard deviations. Differences between review scores, according to member expertise and IRB experience, were analyzed using the Wilcoxon rank-sum test and Kruskal-Wallis test. [16, 17] Since Fleiss' kappa has paradoxically low values, despite the high proportion of agreement when the marginal totals of raters are imbalanced, the observed multiple rater agreement was used instead of Fleiss' kappa to assess the agreement of IRB results among panels. [18] The observed multiple rater agreement was estimated by summing all pairwise agreement tables from the panels, and its 95% confidence interval (CI) was obtained by 1,000 instances of bootstrap resampling. [19] All statistical analyses were conducted using SPSS statistical software version 23 (SPSS Inc., Chicago, IL, USA), SAS software version 9.4 (SAS Institute Inc., Cary, NC), and R software version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria). The two-sided p-value of < 0.05 was considered statistically significant.

Results

Baseline Characteristics of protocols and IRB

We executed eight different protocols of Blind Review from March 2010 to October 2018. Six protocols (75%) were reviewed in the convened meeting, while two (25%) were reviewed in an expedited review. However, during the Blind Review, all protocols were reviewed in the convened meeting. Overall, single center studies (6/8) and interventional studies (5/8) prevailed among protocols included in the blind review. The study population and notable points of discussion are described in Table 1.

Table 1
Baseline characteristics of protocols included in the blind review (n = 8)

No	Year	Center type	Study type	Study population	Major discussion issue	Panels (n)	Protocol review points
1	2010	Single Center	Non-intervention	Breast cancer	Retained tissue	4	<p>Protection of private information</p> <hr/> <p>Consent for retained tissue for FISH analysis</p> <hr/> <p>Applied revised Bioethics law in review</p>
2	2010	Multinational, Multicenter	Intervention (Phase 3)	Pulmonary hypertension	Extension study	4	<p>Requested information regarding previous main study</p> <hr/> <p>Subjects who experienced clinical relapse but enrolled in an extension study with double blind test</p> <hr/> <p>Necessity of resubmitting as a modification review of a previous study</p>
3	2010	Single Center	Non-intervention	Epilepsy	Privacy for epilepsy	4	<p>Consent process for minors</p> <hr/> <p>Concerns about information privacy and family members when conducting germline mutation test</p> <hr/> <p>Pre-screening process prior to acquiring verbal consent by telephone</p>
4	2014	Single Center	Non-intervention	Meniere's syndrome	Phone Screening	7	Proper verbal consent process

No	Year	Center type	Study type	Study population	Major discussion issue	Panels (n)	Protocol review points
							Protection of information privacy
							Proper process of pre-screening by telephone
5	2015	Single Center	Intervention	Atrial fibrillation	Proper study design	7	Informed consent form containing difficult medical terms
							Identify exact intervention type for patient/control group
							Random assignment method
6	2016	Single Center	Intervention	Sleep apnea syndrome	Sham procedure	7	Proper subject target based on investigator's affiliation
							Appropriateness of sham procedure for control group
							Discussion on additional hypertension medication
7	2017	Domestic, Multicenter	Intervention (Phase 2)	Hepatocellular Carcinoma	Rationale for study	7	Plan for human materials management and consent form
							Rationale for clinical trial
							Plan for adverse events management
8	2018	Domestic, Multicenter	Intervention	End-stage Renal Disease	Artificial Intelligence	7	Considerations for randomization, especially in the control group

No	Year	Center type	Study type	Study population	Major discussion issue	Panels (n)	Protocol review points
							Monitoring and management plan for algorithmic errors
							Discussion on whether the study should follow medical device law

Agreement among the panels on each protocol

The IRB review results of different panels were scored on a five-point scale, with “Approved” =1, “Approved with modification” = 2, “Deferred” = 3, “Disapproved” =4, and “Tabled” =5. The correspondence of IRB results among panels according to the protocol is visualized in Fig. 2. We analyzed IRB agreements across two groups—the pilot assessment with three protocols and the main assessment with five protocols. The observed multiple raters’ agreement was 0.444 (95% CI: 0.167-1.000) with protocols 1–3 in 2010, while an agreement of 0.448 (95% CI: 0.362–0.514) was observed between 2014–2018 with seven panels. In order to include recent Blind Review results, we extracted the results of panels 1 to 4 from all the reviewed protocols (pilot and main). This showed that agreement increased to 0.479 (95% CI: 0.271–0.708) from 0.444 (pilot) and 0.448 (main). (Table 2)

Table 2
Observed multiple raters’ agreement on Blind Review results

	Pilot Study (Protocol 1, 2, 3)	Main Study (Protocol 4, 5, 6, 7, 8)
Observed multiple raters’ agreement	0.444¶ (95% CI: 0.167, 1.000)	0.448† (95% CI: 0.362, 0.514)
Observed multiple raters’ agreement (Only with panels 1–4)	0.479¶ (95% CI: 0.271, 0.708)	
¶ Panels 1–4, † Panels 1–7		

Scores and associated factors regarding IRB review results

The scores ranged from 0 to 30 points for each protocol, including GCP requirements and protocol issues; thus, the total score ranges from 0 to 60 points. We analyzed the mean score and standard deviation and compared these between protocols (Table 3). The mean values of the two categories, including GCP requirements and protocol specific issues, were 19.25 ± 8.21 and 18.40 ± 9.04 , respectively. In terms of GCP requirements, the scores gradually increased from the pilot studies in 2010 through to the recent group (Protocol 1 = 15 ± 12.25 , Protocol 2 = 15 ± 9.13 vs. Protocol 7 = 22.8 ± 6.98 , Protocol 8 = 20 ± 8.16). Compared

to GCP comments, the scores for the protocol review issues were similar to the average. A comparison of the scores produced by the different protocols is shown in Fig. 3, using a graph with a standard deviation range.

To assess the impact of IRB members' characteristics on review results, we collected information on the experience of the IRB members who participated in the convened meeting. In addition, we also investigated whether there were any IRB members who had specific insights regarding the protocols that were enrolled in the Blind Review. For example, protocol 2 focused on pulmonary hypertension; thus, the IRB panel that featured a member who was familiar with cardiology was regarded as containing a member specialized in that protocol. We excluded the IRB members who had disclosed conflicts of interest regarding the protocol, as well as those panels that featured no specialized members. The differences in the review results based on the inclusion/exclusion of expert IRB members is as follows (Table 4). The mean score of the panels with experts was higher than the control group, but this result was not statistically significant ($p = 0.93$). According to the career duration of IRB members, review scores of the groups whose careers spanned less than 3 years or 3–5 years were shown to be generally low compared to the groups whose careers spanned more than 5 years ($p = 0.09$).

Table 3: Comparison of the review scores according to the protocols (mean \pm standard deviation)

	Pilot study			Main study				
	Protocol 1	Protocol 2	Protocol 3	Protocol 4	Protocol 5	Protocol 6	Protocol 7	Protocol 8
GCP	15 \pm 12.25	15 \pm 9.13	17.5 \pm 50	20 \pm 6.45	23.6 \pm 4.76	15.7 \pm 11.34	22.8 \pm 6.98	20 \pm 8.16
Protocol	18.75 \pm 10.30	15 \pm 12.90	17.5 \pm 9.57	25.71 \pm 5.34	18.57 \pm 10.69	17.14 \pm 7.56	15.71 \pm 6.07	17.14 \pm 11.12
Total	33.75 \pm 7.5	30 \pm 16.83	35 \pm 12.91	45.71 \pm 4.5	42.14 \pm 12.20	21.43 \pm 12.15	38.57 \pm 11.80	37.14 \pm 17.04

Table 4: The difference in review results depending on the presence of experts on each IRB panel.

	Review score (mean \pm SD)	p-value
Including/excluding expert members		
No expert participated	25.16 \pm 10.96	0.93*
Expert participated	28.13 \pm 10.47	
Duration of IRB members' career		
<3 years	25.0 \pm 10.0	0.09¶
3-5 years	26.3 \pm 9.6	
\geq 5 years	27.3 \pm 14.2	

*Wilcoxon rank-sum test (Mann-Whitney U test), ¶Kruskal-Wallis test, SD: Standard Deviation

Discussion

Inconsistencies in IRB decisions have been observed since the 1980s, largely due to an increase in the number of multicenter clinical trials (Goldman and Katz 1982; Hotopf, Wessely, and Noah 1995) [9, 14]; furthermore, the fact that IRB panels sometimes review the same protocol in a different way remains a major issue. IRBs aim to conform to the protocol and ensure sound research design, both scientifically and ethically. Thus, in terms of ethical considerations, dissimilar opinions are an undisputed consequence. In addition, consistency among IRBs cannot ensure a correct decision. However, major inconsistencies can affect the reliability of the IRB in terms of the wider research environment. Therefore, HRPP QA activities must verify whether IRBs identify inappropriate designs or miss essential review points.

This study analyzed the IRB minutes of panels at the same institution, following the same IRB regulations, in order to observe the consistency of review results. Several studies have shown that IRBs are relatively consistent in terms of their final decisions, but are often inconsistent regarding the reasons for their decisions. [11, 12, 20, 21] Taylor et al. [22] suggested a process for measuring the ethical quality of local IRBs. They defined seven points, including scientific value, assessment of risk, benefit, acceptable risk/benefit ratio, fair selection of subject, adequate informed consent process, and adequate mechanisms for respecting enrolled subjects, while also confirming how many points were satisfied when one protocol was reviewed by multiple IRBs. A revised common rule was also announced recently, regarding the use of a single IRB, which aims to reduce administrative costs. As of January 25th, 2018, the National Institutes of Health (NIH) requires the use of a single IRB record for most domestic NIH-funded multisite research studies. [23] Growing interest in single IRB, currently a research developed measuring process of IRB's inconsistency called "mystery shopper" and scoring system using discussion theme. [13]

Although many studies continue to find inconsistencies, it is still unclear why such inconsistencies occur in multiple IRBs. Our analysis sought to clarify this inconsistency by considering both the issue of discussion and the characteristics of IRB members. We assessed the consistency of seven IRB panel results within a single institution. Partial consistency was observed, but we have only analyzed the results in an exploratory

manner. To overcome the limitations of previous studies that had a scoring system, HRPP staff in this study comprehensively discussed and determined the review points. The total score can be considered a direct measuring tool for IRB QA activity. Based on our data, complying with the GCP requirements during an IRB review showed gradual improvement over time, while protocol-specific reviews were influenced by the characteristics of each protocol, making it relatively difficult to achieve consistency or standardization. This indicates that IRB reviews are commonly affected by specific protocols. However, the observed multiple raters' agreement and review scores both increased with an increase in the experience of panel members. Therefore, IRB members' career spans and compatibility are factors that affect IRB review processes. Thus, active and systematic training of IRB members is recommended to improve the quality of IRB reviews. IRBs are not a permanent system—rather, they should be considered as a regular interaction space for experts and non-experts. There is a significant need to support IRBs at the institutional level, because laws change continuously and often become stricter. Our data shows that the continued involvement of IRB members is essential to ensure and improve the quality of IRB reviews. Continuous monitoring of IRB reviews is necessary for agreeable standardization, excluding extreme and odd results rather than viewing them as being of equal merit. We recommend that the Blind Review process be integrated with the HRPP's activities for assessing the quality of an IRB review.

This study has several limitations. First, it was a retrospective study that was based on IRB minutes from a single tertiary hospital. This limits the generalizability of our data to the HRPP environment. However, we report our experience as a preliminary study—we provide basic information for the Blind Review and monitoring process for IRB reviews, and hope that this is explored in further studies. In addition, blindness to the reviewers can also make the results more scientific and unbiased. Second, only a small number of protocols were included in the Blind Review. However, as our center has operated HRPP since 2010 and maintains 8 years of experience in IRB review quality assessment, we believe that any degree of bias was minimal. Third, we analyzed the review scores discretionally, and the review points used for consensus among HRPP staff had no scientific criteria. Further studies should investigate and validate this as a multi-center study involving a larger number of protocols.

Conclusion

Blind Review is an effective method for overseeing and ensuring the quality of IRB reviews, as well as the overall GCP compliance. The usefulness of the Blind Review process has not been established due to insufficient evidence. However, based on our study, this process seems to be an effective monitoring method in the highly challenging environment of clinical research ethics.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets analysed during the current study are not publicly available due to the confidential data of IRB review, but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Acknowledgements

The authors would like to thank all HPC staff members for assisting in data collection and constant dedication with efforts on IRB and HRPP.

References

1. Resnik DB. Some reflections on evaluating institutional review board effectiveness. *Contemporary clinical trials*. 2015;45(Pt B):261-4.
2. Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Journal of the Indian Medical Association*. 2009;107(6):403-5.
3. Grady C. Do IRBs protect human research participants? *Jama*. 2010;304(10):1122-3.
4. Greene SM, Geiger AM. A review finds that multicenter studies face substantial challenges but strategies exist to achieve Institutional Review Board approval. *Journal of clinical epidemiology*. 2006;59(8):784-90.
5. Park S, Nam CM, Park S, Noh YH, Ahn CR, Yu WS et al. 'Screening audit' as a quality assurance tool in good clinical practice compliant research environments. *BMC medical ethics*. 2018;19(1):30.
6. Park S, Noh YH, Rha SY, Kim WH, Cheon JH. Institutional Board Review for Clinical Investigations on Inflammatory Bowel Diseases: A Single-Center Study. *Intest Res*. 2015;13(3):274-81.
7. Penn ZJ, Steer PJ. Local research ethics committees: hindrance or help? *British journal of obstetrics and gynaecology*. 1995;102(1):1-2.
8. Veatch RM. Problems with Institutional Review Board inconsistency. *Jama*. 1982;248(2):179-80.
9. Goldman J, Katz MD. Inconsistency and institutional review boards. *Jama*. 1982;248(2):197-202.
10. Angell E, Sutton AJ, Windridge K, Dixon-Woods M. Consistency in decision making by research ethics committees: a controlled comparison. *Journal of medical ethics*. 2006;32(11):662-4.
11. Edwards SJ, Stone T, Swift T. Differences between research ethics committees. *International journal of technology assessment in health care*. 2007;23(1):17-23.
12. Taljaard M, Brehaut JC, Weijer C, Boruch R, Donner A, Eccles MP et al. Variability in research ethics review of cluster randomized trials: a scenario-based survey in three countries. *Trials*. 2014;15:48.

13. Trace S, Kolstoe SE. Measuring inconsistency in research ethics committee review. *BMC medical ethics*. 2017;18(1):65.
14. Hotopf M, Wessely S, Noah N. Are ethical committees reliable? *Journal of the Royal Society of Medicine*. 1995;88(1):31-3.
15. Guideline IHJCS. Integrated addendum to ICH E6 (R1): guideline for good clinical practice E6 (R2). 2015;2:1-60.
16. Fleiss JLJPb. Measuring nominal scale agreement among many raters. 1971;76(5):378.
17. Fleiss JL, Levin B, Paik MC. *Statistical methods for rates and proportions*. John Wiley & Sons; 2013.
18. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *Journal of clinical epidemiology*. 1990;43(6):543-9.
19. de Vet HCW, Mullender MG, Eekhout I. Specific agreement on ordinal and multiple nominal outcomes can be calculated for more than two raters. *Journal of clinical epidemiology*. 2018;96:47-53.
20. Koski G. Beyond compliance...is it too much to ask? *Irb*. 2003;25(5):5-6.
21. Anderson EE, DuBois JM. IRB decision-making with imperfect knowledge: a framework for evidence-based research ethics review. *The Journal of law, medicine & ethics : a journal of the American Society of Law, Medicine & Ethics*. 2012;40(4):951-69.
22. Taylor HA. Moving beyond compliance: measuring ethical quality to enhance the oversight of human subjects research. *Irb*. 2007;29(5):9-14.
23. Taylor HA, Ehrhardt S, Ervin AM. Public Comments on the Proposed Common Rule Mandate for Single-IRB Review of Multisite Research. *Ethics & human research*. 2019;41(1):15-21.

Figures

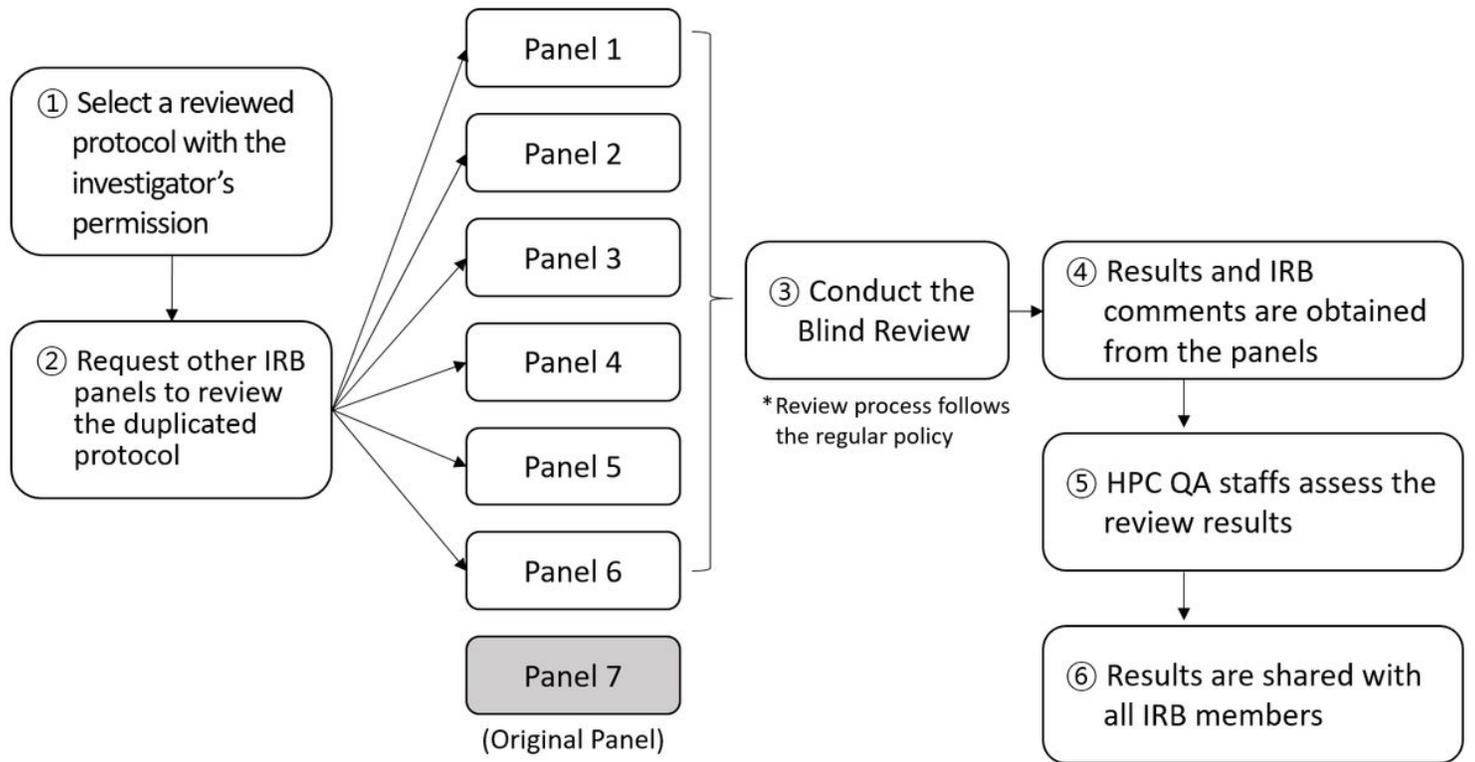


Figure 1

Flow chart of the Blind Review.

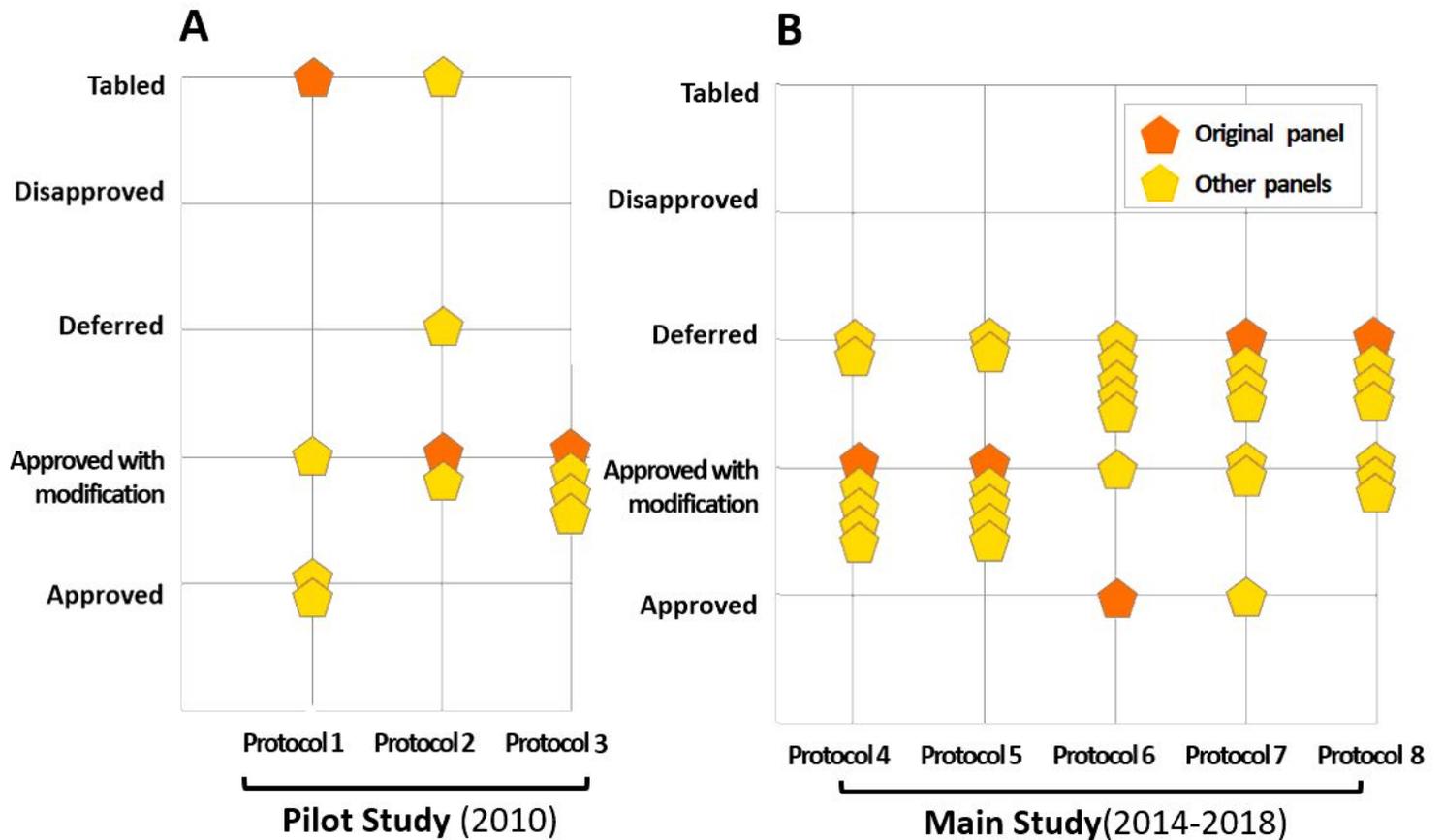


Figure 2

Blind Review results of the IRB panels according to protocols (a. Pilot study [2010], b. Main study [2014-2018]).

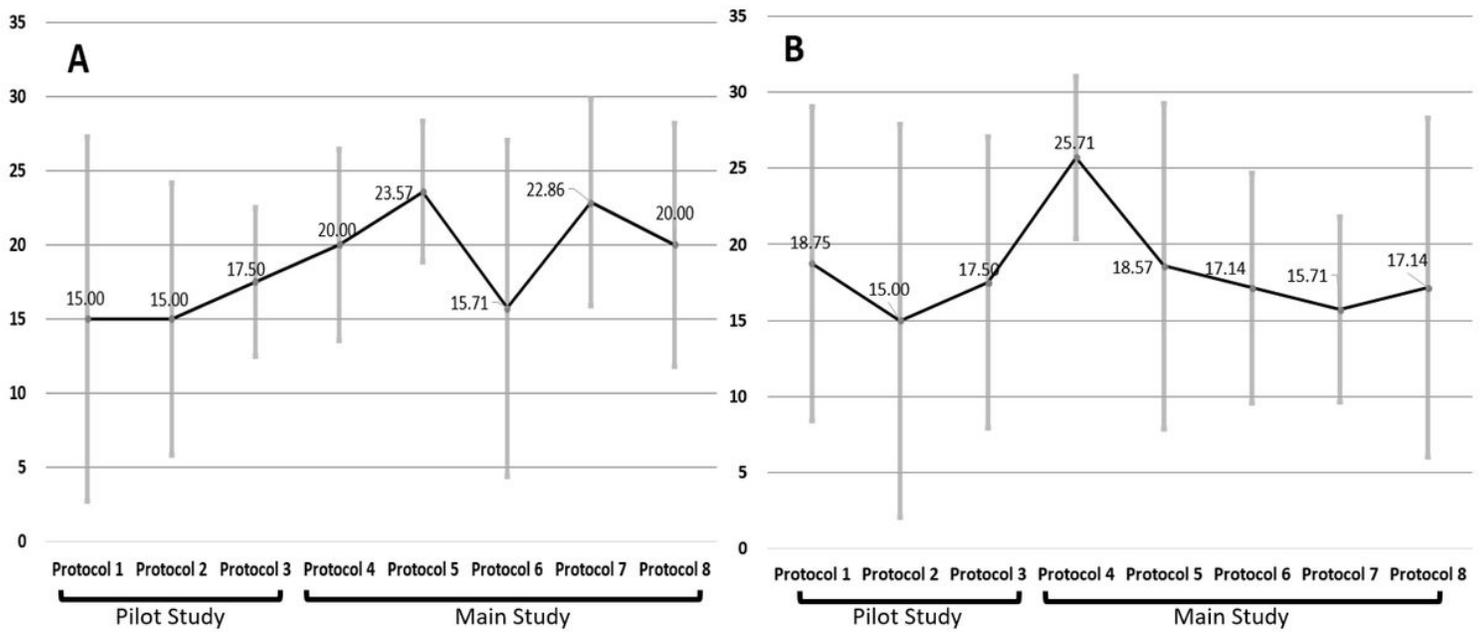


Figure 3

Changes in review scores (a. GCP requirements, b. Protocol review points).