ASA Comments on the Draft Update of the PCORI Methodology Standards

Prepared by the ASA Health Policy Statistics Section, Committee on Funded Research, and other ASA members

June 8, 2026

The American Statistical Association (ASA) appreciates the opportunity to comment on the draft update of the PCORI Methodology Standards. As the nation's largest professional organization of statisticians, the ASA is committed to upholding methodological rigor and the proper use of data everywhere. ASA members have long valued PCORI and the agency's desire to bring the best analytic tools and methods to patient centered data and comparative effectiveness research. The PCORI Methodology Standards are an important public facing tool that assists all health researchers provide higher quality output with better consistency and reproducibility. Guidance on advanced topics such as causal inference, adaptive trials, cluster-randomization, Bayesian methods, and artificial intelligence are just some of the areas where PCORI's inclusion in methodology standards has been a boon to statisticians working with researchers who were not aware of or appreciated these modern methods.

Now, we also thank PCORI for giving our organization and the public as a whole the opportunity to comment on the existing standards as well as some proposed changes. Overall, this feedback can be summarized by the following important points, followed by more minor suggestions:

1. In general the standards seem reasonable and represent best practice for research and we appreciate the substantial work involved in putting them together.
2. A broad comment that we encourage the committee to keep in mind is that many of the standards are still fairly vague and will require substantial judgement by researchers and

proposal reviewers. We encourage the methodology committee to create trainings in the standards, and perhaps illustrative examples of the scale/scope of the language expected, to help facilitate strong use of the standards.

3. In addition, the standards are extensive and adhering to them – and documenting adherence to them – will take substantial work. We encourage consideration of whether some, or the full details of them, could be documented only after a project is funded rather than at the proposal stage.
    - A specific place this comes up is PROT-2, which is really only relevant once the project is underway; is the idea at the proposal stage just that the applicant team will attest that they will do this in the future?
    - Another example is SD-2: Describing strategies for mitigating threats to internal and external validity is a VERY big topic and could take 10 pages to describe. (And also overlaps with some of the other standards). Further guidance on what is specifically expected here would be useful.
    - Another example along these lines is GOV-3; it is not clear what will be looked at/for in terms of a governance process.
4. As noted below with a few more specific examples, not all of the standards are relevant for all types of projects. It would be useful to give more specific guidance to researchers regarding which sorts of studies need to report on which standards.
5. There is overlap between some of the standards – e.g., the Observational Research standards overlap with some of the others about confounding, etc. A future revision may consider consolidating some of the standards or re-organizing them. (Previously there were cross-cutting standards and then more specific standards for particular types of studies and study designs, which felt like a somewhat more natural organizational structure and made it more clear which standards are relevant for which types of studies).
6. PCORI outlines the need for power calculations in submissions as part of a data analysis plan (DA-1). However, the standard is written with several lines of 'as applicable' leaving applicants to self-determine whether a power calculation is necessary. This is likely appropriate in that they do make more sense (and are more feasible) for some sorts of studies than others, but also opens the line for bias toward certain studies. We request that PCORI amend the language to advocate that everyone provide a reason for the number of individuals they plan to recruit for any aspect of the study (including qualitative and quantitative studies). Often statisticians perform power calculations on proposed research that have complex designs where a complex simulation study might be needed and has not been undertaken which adds a burden that others can self-select against with review committees left to decide whether the sample size needed justification. Regardless of whether the proposed study is experimental with a well

designed protocol, an observational study utilizing an existing registry, a focus group/qualitative assessment, or other active research; an intelligent defense and reason for the number of individuals being recruited should be given. To not do so biases evaluation of research directly involving statisticians from those who do not by adding an additional requirement to evaluation.

7. That said, we also recognize that some analyses of secondary data may have unclear or unnecessary power analyses. Some researchers, for example, argue that power analyses are unnecessary for observational data analyses: https://pubmed.ncbi.nlm.nih.gov/34461211/.
8. These last few points relate to the broader issue that the standards are somewhat "uneven" in that some types of studies/methods have more rigorous standards, and it isn't always clear why that is the case (see below for some comments on this point related to causal methods). We recognize the tension that not all standards are relevant for all types of studies (as we note in various comments) but also recommend that there be clarity about what is expected when, so that applicants can direct their energy towards the most relevant standards for their type of study..
9. Relatedly, and as one example, PCORI funds a large number of methodological projects that move statistical theory and artificial intelligence methodology forward in a patient centered way; these projects sometimes have great difficulty in adhering to principles PE-1 to PE-6 as getting actionable patient and caregiver feedback on mathematical innovations may be impossible. PCORI has done wonderful work in incorporating how to deal with some of this in reviewer training, but it is not reflected in the Methodology Standards. We ask that a specific callout be added that recognizes that proposals and projects that are submitted to the methods RFA should strive to maximize engagement but recognize that engagement with patients or caregivers might not always be possible or appropriate. It might make sense to add language like ", and as consistent with the aims of the proposed work" to the end of some of the standards, to signal that some could/should be adapted.
10. We believe the language on OOKV-2 should be strengthened: "Use patient-reported outcomes when patients or people at risk of a condition are the best source of information" can be interpreted by some to not use patient-reported outcomes (PROs) if they feel that there is a better metric. We recommend considering instead language such as "Use patient-reported outcomes when patients or people at risk of a condition unless there is a well-defined argument to not do so."
11. We recommend changing the heading 'Decisional Dilemma' to 'Evidence Gap or Decisional Dilemma' as some projects may not represent such a dilemma but would provide output that fills a necessary evidence gap. In addition, the added language throughout the standards on Decisional Dilemma does not apply broadly to all methodological studies. We thus suggest retaining the 'evidence gap or decisional

dilemma' language in DD-1 for all instances of Decisional Dilemma that are currently in the draft standards.

12. The language in the Conceptual Model (CM) section assumes a comparative effectiveness study without stating it explicitly. We suggest PCORI either rephrase or start both standards CM-1 and CM-2 off with 'If proposing a clinical comparative effectiveness study…'.
13. The requirements in OOKV for 'all' variables is burdensome for studies using large registries and recommend changing to 'key' variables.
14. Under the Real World Data (RWD) header, we suggest that PCORI add a note that says that if the goal is trial emulation then the curated data should reflect the emulated trial and reflect only the information that would be available at each time point of an emulated trial.
15. The language in PE-1 through PE-6 are somewhat similar and overlapping; if they really are meant to be distinct it will be useful to have clarified wording and/or examples of what is being meant by each one. Fundamentally there seem to be 3 aspects: 1) identification of the patient stakeholders; 2) description of how they would be partners and collaborators in the research; and 3) resources to support their involvement.
16. The Artificial Intelligence standards are unclear in terms of what is being talked about – any AI (including machine learning?), only LLMs? And is it in relation to AI tools being studied for clinical use? Or the use of AI for research purposes? An AI intervention as a comparator also raises completely different issues from AI used as part of the study workflow or as part of an analysis strategy. The potential uses of AI are so heterogeneous and raise such a wide range of issues that you may consider whether AI should be folded into other sections of the standards rather than pulled out as its own separate section (e.g., AI-4 would make more sense as part of the comparators section). And AI as part of the analysis should be covered in data analysis, etc.
17. It is not clear why there are specific standards for propensity score and instrumental variable designs but not for other non-randomized designs, such as regression discontinuity or interrupted time series/difference-in-differences.
18. OR-8: Why is preregistration mentioned only for target trial emulation and not for other types of designs?
19. Many PCORI funded studies use small numbers of clusters. We suggest that the CRD standards should include an explicit requirement for investigators to address small-sample challenges at both the design and analysis stages. (1) At the design stage, investigators should describe any restricted randomization procedures employed to improve baseline balance, including stratification, matched-pair randomization, or covariate-constrained randomization (rerandomization). (2) At the analysis stage, investigators should describe and justify the specific methods used to obtain valid

inference under small-sample conditions. These include, but are not limited to: covariate adjustment to reduce residual variance, bias-corrected sandwich variance estimators, degrees-of-freedom adjustments (e.g., Satterthwaite or Kenward-Roger corrections), and randomization-based inference procedures. When restricted randomization is used at the design stage, the analysis also must respect the randomization procedure employed. A suggested new standard could be worded as: "Describe and justify the methods used to address small-sample challenges at both the design and analysis stages. At the design stage, describe any restricted randomization procedures used to achieve covariate balance. At the analysis stage, describe and justify the use of small-sample correction methods, particularly when the number of randomized clusters is fewer than 30. Justify that the chosen analytic approach is appropriate for the number of clusters available and that the randomization procedure used at the design stage is reflected in the analysis."

20. Relatedly, given the increasing use of stepped wedge designs, but their complications, CRD-1 could possibly be expanded to require that investigators using a stepped-wedge design explicitly justify its use over a parallel-arm CRT, and describe how the design's unique methodological features will be addressed. Suggested modification to CRD-1 could be like (needs to be substantially shortened or become its distinct standard): "If a stepped-wedge design is proposed, justify its use over a parallel-arm cluster-randomized design, and describe how the analysis plan will address challenges specific to stepped-wedge trials, including secular trends, the longitudinal correlation structure, and potential heterogeneity of treatment effects across exposure duration."

More direct / minor suggestions:

1. The term ‘core elements of patient centeredness’ is vague and a more direct list or immediate definition from PCORI of these core elements would be useful.
2. IC-2 has a typo in the spelling of reproducible.For MD-3, we recommend replacing ‘sensitivity analyses’ with ‘sensitivity and other analyses’ as not all missing data methods would be considered sensitivity analyses which a researcher might misconstrue as not a valid method to PCORI.
3. MD-4: We are curious why imputation is called out in particular, and not also methods such as non-response weighting adjustments or longitudinal models that use maximum likelihood approaches for dealing with missing data? It may be more appropriate to frame the standard about saying how missing data will be handled and justification of the approach, rather than focusing only on imputation.
4. PROT-1: This is one of the standards that is really only relevant for comparative effectiveness designs (not, e.g., methods or other types of projects); clarifying that it is

only relevant for such studies would be useful. (The same issue comes up for CM-1, where the language presumes a comparative effectiveness study).

5. SD-3: The language should probably be “ensuring allocation concealment…”. Also, allocation concealment may not be relevant for all studies using a randomized design. It may be appropriate to add “, as appropriate given the study design and aims” at the end of the standard.
6. GOV-4: The details listed here may be more appropriate in a data safety and monitoring plan document.
7. RER-1: Given that most studies likely won’t use formal “sampling” we recommend the use of a term such as “selection” or “recruitment” instead of “sampling.”
8. RER-2: When outcomes vary and are measured over time it may be important to also consider approaches to maximize on-time assessment.
9. OOKV-3: The definition of “proxy variable” in this standard should be clarified; it is unclear what is meant by proxy in this context.
10. IC standards: It may be useful to request clear articulation of the estimand being targeted. (This does come up in the DA standards; they could potentially be cross-referenced).
11. DM standards: DM2-4 seem redundant with DM-1, as they would all be elements included in a data management plan. (This is also one that will be very burdensome to finalize at the proposal stage and may be more appropriate to finalize after funding).
12. RWD-3: It is unclear how to operationalize this documentation or what is meant by the broad statements about data provenance, appropriateness, and data quality. Again, examples would be useful.
13. DA-6: Similarly, it is unclear what is being asked for here that wouldn’t be covered by other standards around variable definitions, analysis plans, etc.
14. MD-2: What is meant by “sources of missing data” and how does MD-2 differ from what would be covered by MD-1?
15. Observational Standards: “Observation” can mean just observing what people do (no causal intent); we suggest using a term like “Non-experimental [or non-randomized] causal inference studies”.
16. OR-1: What is meant by this standard that wouldn’t be covered by other standards? And what is meant by “study components”?
17. OR-2: Consider having this standard discuss only covariate balance; the data quality assessments would presumably be covered by other standards.
18. OR-3: Is “parameters” the right term to use here, or should it be something like “measures”?
19. OR-9: Again what is meant to be included here that wouldn’t be covered by other standards?

20. CRD-2: It is unclear why contamination is only discussed for cluster randomized trials, when actually contamination is an issue for many studies, and may be actually less of a problem in cluster randomized designs.
21. A potential missing standard HEI-3: Power analysis tailored to hybrid study type, with explicit hypothesis testing framework for co-primary outcomes. The HEI standards do not mention power analysis, but the statistical design challenges can differ across hybrid types. This gap is most consequential for Hybrid Type 2 studies, which simultaneously evaluate effectiveness and implementation outcomes as co-primary endpoints. In this setting, investigators must explicitly specify the hypothesis testing framework — conjunctive (both effectiveness and implementation outcomes must meet the threshold for the intervention to be considered successful) versus disjunctive (success on either outcome is sufficient) — because these frameworks have fundamentally different implications for Type I error control, statistical power, sample size requirements, and ultimately for decision-making. Suggested new standard: "Describe and justify the power analysis or sample size rationale tailored to the selected hybrid study type. For Hybrid Type studies with co-primary effectiveness and implementation outcomes, explicitly specify the hypothesis testing framework and justify this choice based on stakeholder discussions about the decision-making context."
22. HEI-2 requires a causal model that distinguishes implementation strategies from clinical interventions and links them to their respective outcomes. However, the standard does not require investigators to specify the hypothesized mediation pathway by which implementation strategies are expected to influence clinical outcomes through implementation outcomes. In hybrid designs, this mediation structure can be central to the scientific rationale — if an implementation strategy improves uptake of an effective intervention, the clinical benefit flows through that uptake. Suggested addition: "If applicable, describe the hypothesized mediation pathway through which implementation strategies are expected to affect clinical outcomes via implementation outcomes, and specify whether the study is designed to evaluate this mediated pathway."
23. SR-1: Is the NASEM systematic reviews reference the report from 2011 on finding what works in healthcare? That specific reference is unclear.

The ASA is fully committed to helping PCORI bring the best science to patient centered research possible. Questions or comments may be directed to ASA Director of Science Policy Steve Pierson: spierson@amstat.org.