Minimal Standards in the Prevention and Handling of Missing Data in Observational and Experimental Patient Centered Outcomes Research

Final Report

March 15, 2012

Prepared by:

Tianjing Li, MD, MHS, PhD1; Susan Hutfless, SM, PhD; Kay Dickersin, MA, PhD; Daniel Scharfstein, ScD; James Neaton, PhD; Joseph Hogan, ScD; Roderick Little, PhD; Michael Daniels, ScD; Jason Roy, PhD; Vincent Mor, PhD, MED; Andrew Law, MHS.

1 Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street, Mail room W5010
Baltimore, MD 21205 USA
Phone: 410-502-4630
Fax: 410-502-4621
Email: tli@jhsph.edu

DISCLAIMER
All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee. PCORI has not peer-reviewed or edited this content, which was developed through a contract to support the Methodology Committee’s development of a report to outline existing methodologies for conducting patient-centered outcomes research, propose appropriate methodological standards, and identify important methodological gaps that need to be addressed. The report is being made available free of charge for the information of the scientific community and general public as part of PCORI’s ongoing research programs. Questions or comments about this report may be sent to PCORI at info@pcori.org or by mail to 1828 L St., NW, Washington, DC 20036.
Background

Consumers and stakeholders need reliable and evidence-based information for making health care choices. To meet this objective, data from patient-centered outcomes research (PCOR) must be valid (i.e. unbiased) and inferences drawn must be based on sensible and robust statistical methodology. Missing data in PCOR can seriously undermine this objective because missingness may be related to both exposure and outcomes, and thereby introduce bias. Missing data have been defined as unrecorded values that, if recorded, would be meaningful for analysis [1].

Both experimental and observational study designs can be used for PCOR. The threat to validity from missing data is similar for randomized controlled trials (RCTs) and observational studies, although the threat is potentially greater for observational data (including studies utilizing electronic health records and administrative datasets) [1]. In observational studies, data can be missing for exposures, confounders, and outcomes compared to only outcomes for well-conducted RCTs. Several recent reports have laid out the general principles and techniques for preventing and handling the problems created by missing data [1-3], most of which are highly relevant and applicable to PCOR.

In this project, we aimed to (1) identify existing guidance documents on the prevention and handling of missing data; (2) propose minimum methodological standards for the prevention and handling of missing data for PCOR; and (3) identify challenges and research gaps. We developed our methods based on the key steps suggested in the Request for Proposal instruction (see Box 1). We used “standards” to refer to our final recommended methodological guidance, following the language in the Request for Proposal, although some experts on our team feel that the term “standards” suggests an unattainable level of certainty.

Methods

1. Identify and select relevant guidance documents

1.1. Eligibility criteria: We included existing guidance documents such as research practice guidelines, recommendations, reports, principles, requirements, regulations, standards, books, archives, statements, and relevant literature on the prevention and handling of missing data for PCOR and Comparative Effectiveness Research (CER). We used PCORI’s definition for PCOR and CER (http://www.pcori.org/patient-centered-outcomes-research/; accessed March 7, 2012). Guidance documents generally represent an agency, institute, organization, or panel’s current thinking on a particular subject. We included published guidance documents as well as guidance documents in draft form. We included the most recently published guideline if multiple versions of the same guideline existed.

1.2. Search strategy: We worked with an information specialist at the William H. Welch Medical Library, Johns Hopkins University School of Medicine and developed a search strategy that combined guidance document terms with terms on missing data. Our search strategy was peer reviewed by two independent librarians. We searched the National Library of Medicine (NLM) Bookshelf on January 9, 2012 and NLM Catalog on January 14, 2012 without any language or publication date restrictions (see Appendix 1 for search strategy).
To supplement the electronic searches, we searched major regulatory agencies and organizations’ websites, including the US Food and Drug Administration, the Institute of Medicine, the Agency for Healthcare Research and Quality, the National Research Council, the National Institutes of Health, the Canadian Agency for Drugs and Technologies in Health, the German Institute for Quality and Efficiency in Health Care, the European Medicines Agency, the UK Medical Research Council, and the International Society for Pharmacoeconomics and Outcomes Research.

We hand-searched the reference lists of key guidance documents identified. We also contacted experts in the field, including groups doing registry and health records analysis.

1.3. Select relevant guidance documents: Two people independently reviewed the titles and abstracts from our searches. We retrieved the full text of all guidance documents classified as possibly eligible by at least one person. At least two people independently reviewed full text reports for final eligibility, resolving discrepancies through discussion. We documented reasons for exclusion.

2. Abstract data
We extracted the following items from included full text guidance documents (Table 1): bibliographic information, including title, publication year, organization or authors, and country or region of origin; selected characteristics of guidance development, including the purpose of the guidance, method used to generate recommendations, stakeholders involvement, inclusion of patient’s views and preferences, report of conflicts of interest, and report of independent external review.

One person extracted recommendations and noted down the pagination. This was verified by a second person. For each standard extracted, we evaluated whether the recommendation was clearly and unambiguously presented (i.e., not summarized by the data abstractor) and whether there was an explicit link between the rationale for and the recommendation (e.g., by citing supporting literature and evidence). We also determined the relevance and applicability of the guidance recommendations to PCOR by extracting the phase of research (i.e., design, conduct, analysis, or reporting) as well as the type of study design (i.e., randomized controlled trials, observational studies, studies utilizing electronic health records, patient registries, administrative data, or systematic reviews) that each guidance statement was created for. We categorized statements from multiple guidance documents if they agreed on the main point or topic domain and generated a preliminary list of potential standards.

3. Recommend minimum standards
Four people independently reviewed, refined, and shortened the preliminary list of potential standards for consensus consultation. We then used a modified Delphi approach for consultation with a consensus panel to recommend minimum standards. The Delphi approach is a structured process of obtaining opinion and information from a group of experts by means of a series of consultation and questionnaires, each one refined based on the feedback from previous responses [4,5]. This approach has been recommended by the EQUATOR (Enhancing the QUAlity and Transparency of health Research) Network in developing reporting guidelines, such as CONSORT [6].

Between February 17 and February 28, 2012, we conducted a two-round consensus survey of 10 experts on the consensus panel. Our consensus panel is multidisciplinary and includes experts in epidemiology, clinical trials, biostatistics, patient registries, electronic medical records,
and administrative data. Many members of the panel are leading experts in missing data and causal inference research.

To ensure clarity and consistency of understanding among the panel members, we communicated the context of the project to our panel before administering the survey. We emphasized that the recommended standards will “inform investigators requesting Patient-Centered Outcomes Research Institute (PCORI) funding and assist grant reviewers in evaluating research proposal so as to ensure methodological rigor in PCOR (http://www.pcori.org/funding-opportunities/past-funding-opportunities/methods-review/; accessed March 11, 2012).”

We provided descriptive information for each potential standard, including the research phase and study design the potential standard is applicable to, and the number and title of guidance documents that has described the potential standard. We provided space for comments, questions, and nomination of items not included in the list. The consensus survey was designed, administered, and analyzed using Survey Monkey.

In Round One, we asked our panel members to rate whether each potential standard is applicable to one or more study designs, and if applicable, rate each potential standard as (a) “mandatory” (i.e., must do’s), meaning that new PCOR should not be funded and initiated if the standard is not adhered to in application; (b) “highly desirable” (i.e., should do’s), meaning that the standard should generally be adhered to in an application for PCOR, but that there are justifiable exceptions; or (c) “other - not mandatory or highly desirable”, meaning that it is unimportant whether the standard is considered in application for PCOR. There could be multiple reasons for selecting “other”, for example, the recommendation may not be applicable to PCOR, is out-of-date, needs refinement, there are not enough data to support adoption of the item currently, or the recommendation is flawed. This rating schema has been successfully applied by the Cochrane Collaboration in their extensive consultation and development of the Methodological Expectations of Cochrane Intervention Reviews (http://www.editorial-unit.cochrane.org/mecir; accessed November 19, 2011).

In Round One, we also asked panel members to rate whether each standard would contribute to improving objectivity, minimizing bias, improving reproducibility, leading to more complete reporting, and enhancing transparency. These are key elements that will be considered by PCORI Methodology Committee when deciding whether to adopt or reject a standard.

We summarized Round One results, and based on a pre-specified threshold, we classified each potential standard as “mandatory” – if 7 or more experts out 10 marked mandatory for at least one study design; “not mandatory” – if 3 or fewer experts marked mandatory for any study design; or “standards that require a second round for consensus” – if between 4 to 6 experts marked mandatory for any study design. We considered that consensus was reached for the “mandatory” and “not mandatory” standards.

In Round Two, we provided summary responses and anonymized comments for those potential standards that needed a second round of voting, and asked the panel members to re-consider each item in light of others’ opinions.

After conclusion of Round Two, we conducted a structured discussion via teleconference in which results and comments from both rounds were presented. We discussed and agreed upon the final set of “mandatory” standards as well as refined wording for them. We used the National
Research Council’s report *The Prevention And Treatment Of Missing Data In Clinical Trials* [1] as the primary reference in choosing wording and terminology for standards. When appropriate, we combined standards to form a coherent and succinct presentation. We also discussed the degree to which each proposed standard contributes to patient centeredness.

4. **Identify gaps in existing guidance documents**
   We identified areas where no guidance documents exist or the guidance documents are incomplete. We summarized challenges described in the included guidance documents.

**Results**

1. **Searches and study selection**
   We identified 1790 records: 787 from our electronic search of NLM Bookshelf, 969 from NLM Catalog, 13 from searching regulatory agencies websites, and 21 from hand searching the reference list of relevant documents. We did not find any additional reports for inclusion after contacting 47 experts in the field. We retrieved and reviewed 190 full text documents and included 30 [1-3, 7-33] (Figure).

2. **Characteristics of included guidance documents**
   The characteristics of the 30 included guidance documents are described in Table 1. The included guidance documents were published between 1996 and 2011 (5 still in draft form), with more than half published after 2008. A large proportion of the guidance documents (9/30, 30%) were prepared by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), followed by 5 (16.7%) prepared by the US Food and Drug and Administration (FDA), and 3 (10%) by the Expert Working Group (Efficacy) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Almost all included guidance documents (28/30, 93.3%) were developed either in the US or have involved US stakeholders. The process for generating recommendations was described in a little over one half of the documents (16/30, 53.3%). However, fewer than half (12/30, 40%) described involving stakeholders, and only 3 (10%) reported seeking patients’ views and preferences. Two (6.7%) recent guidance documents explicitly described their applicability to PCOR. Potential conflicts of interests and independence from the funding body were reported explicitly in only 4 (13.3%) and 6 (20%) guidance documents respectively. Thirteen (43.3%) guidance documents underwent external peer review.

3. **Recommending minimum methodological standards for the prevention and handling of missing data**
   We extracted 39 potential standards from the 30 included guidance documents (Table 2). We narrowed down the number of potential standards from 39 to 24 on the basis of independent assessment by four members of the team. We asked the consensus panel to rate the 24 standards using two rounds of a modified Delphi consensus survey.

   All 10 experts on the consensus panel responded to the Round One survey, with no missing data. Based on the results of Round One, we identified 9 “mandatory standards”, 4 standards that were “not mandatory”, and 11 standards that required a second round for consensus (see Appendix 2 for summary of Round One results). Eight of the 10 experts responded to the Round Two survey within the time frame we allowed (2 working days plus a weekend) (see Appendix 3 for summary of Round Two results). We handled missing data from the 2 non-responding experts by (1) carrying out a sensitivity analysis assuming they would rate “mandatory” or not for each of the 11 standards; and (2) provided an opportunity for them and others to express their opinions during the teleconference discussed below.
After a consensus building teleconference, we included 10 mandatory standards: 3 on study design, 2 on conduct, 3 on analysis, and 2 on reporting (Box 2). We also made recommendations on the type(s) of study design and research question (e.g., intervention effectiveness, etiology, prognosis) that each mandatory standard is applicable to. We attempted to word the standards so that they were applicable to all study designs (as most of them are) and indicate when they were applicable to only some designs.

**Box 2. Minimum standards in the prevention and handling of missing data in patient centered outcomes research**

**Standards on study design**

1. The study protocol should explicitly define (1) the objective(s) of the study; (2) the intervention or interventions of interest; (3) the associated primary outcome or outcomes; (4) how, when, and on whom the outcome or outcomes will be measured; (5) potential confounders if relevant, and (6) the measures of intervention effects. The measures of intervention effects should be meaningful for all study participants, and estimable with minimal assumptions. This standard applies to all study designs that aim to assess intervention effectiveness.

2. Investigators should explicitly anticipate potential problems of missing data. The study protocol should contain a section that addresses missing data issues and steps taken in study design and conduct to monitor and limit the impact of missing data. As relevant, the protocol should include the anticipated amount of and reasons for missing data, and plans to follow up participants. This standard applies to all study designs for any type of research question.

3. Statistical methods for handling missing data should be pre-specified in study protocols, and their associated assumptions stated in a way that can be understood by stakeholders. The reasons for missing data should be considered in the analysis. This standard applies to all study designs for any type of research question.

**Standards on study conduct**

4. Whenever a participant discontinues some or all types of participation in a research study, the investigator should document the following: (1) the reason for discontinuation; (2) who decided that the participant would discontinue; (3) whether the discontinuation involves some or all types of participation. Investigators should continue to collect information on key outcomes for participants who discontinue their protocol specified intervention. This standard applies to prospective study designs that aim to assess intervention effectiveness.

5. A data and safety monitoring board should review plans for and the implementation of the prevention and handling of missing data. The committee should review completeness and timeliness of data and recommend modifications as appropriate. This standard applies to studies that include a data and safety monitoring committee.

**Standards on analysis**

6. Statistical inference of intervention effects or measures of association should account for statistical uncertainty attributable to missing data. This means that methods used for imputing missing data should have valid type I error rates and that confidence intervals have the nominal
coverage properties. This standard applies to all study designs for any type of research question.

7. Single imputation methods like last observation carried forward and baseline observation carried forward generally should not be used as the primary approach for handling missing data in the analysis. This standard applies to all study designs for any type of research question.

8. Examining sensitivity to the assumptions about the missing data mechanism (i.e., sensitivity analysis) should be a mandatory component of the study protocol, analysis, and reporting. This standard applies to all study designs for any type of research question.

**Standards on reporting**

9. All participants who enter the study should be accounted for in the report, whether or not they are included in the analysis. Describe and justify any planned reasons for excluding participants from analysis. This standard applies to all study designs for any type of research question.

10. Report on data completeness and how missing data were handled in the analysis to facilitate interpretation of study results. The potential influence of missing data on the study results should be described. This standard applies to all study designs for any type of research question.

Box 3 provides information requested by PCORI, including detailed explanation, elaboration, rationale, examples, empirical and theoretical support for each standard. We commented on the degree of implementation issues.

**4. Challenges and research gaps identified**

We found more guidance documents written for RCTs (n=14) than for nonrandomized studies (n=8). In addition, the recommendations were more likely to be explicit for the former than the latter. Through contacting experts working on nonrandomized studies, for example, key personnel at the Surveillance Epidemiology and End Results (SEER), National Cancer Institute, senior epidemiologists at Geisinger Health System, Kaiser Permanente Center for Health Research, investigators running large HIV AIDS cohorts, and experts working on claims data and electronic medical records, we did not find any guidelines or best practices explicitly written on missing data for nonrandomized study designs.

Challenges revealed from the literature and from contacting experts for nonrandomized studies include, for example, the difficulty for interpreting missing data (e.g., does absence of a specific symptom in patient registers and electronic health records indicate the symptom was not present or that the physician did not actively inquire about or document this symptom?), missing data on confounders not associated with common billing or tracking codes (e.g., smoking, over-the-counter drugs, weight), failures to follow up (due to turnover in healthcare plans), nonstandard follow-up time points, and complicity added to analysis when relevant data are not collected [11, 30, 34].

One recent guidance document, developed for RCTs, explicitly described the following research gaps [1]. These include:

(1) *Designs for the follow-up of participants in clinical trials who have dropped out of the study and who have not withdrawn their consent,*
(2) Collecting the typical rates and likely causes of missing data in various kinds of clinical trials,
(3) The effect of missing data on the power of clinical trials,
(4) How to set useful target rates and acceptable rates of missing data in clinical trials,
(5) The robustness of missing data methods such as inverse probability weighting methods and multiple imputation methods to assumptions,
(6) The assessment of goodness-of-fit for the parametric models used to analyze data from clinical trials (when there is missing data),
(7) The performance of double-robust procedures in comparison to more commonly used procedures,
(8) The impact of missingness in auxiliary variables on the various current methods, and ways of reducing the associated bias,
(9) Methods of sensitivity analysis in clinical trials, particularly for nonmonotone patterns in longitudinal data,
(10) Methods for assessing and limiting the impact of informative censoring for time-to-event outcomes,
(11) How to develop effective decision rules based on the input from sensitivity analyses, and
(12) The development of software that supports coherent missing data.

Discussion

We recommend 10 minimum standards for the prevention and handling of missing data for PCOR. These standards were identified by a panel of experts using a pre-specified systematic approach. It is worth noting that the recommended minimum standards are current “best practices”, and not the last word: “all of the recommended standards must be considered provisional, pending better empirical evidence about their scientific validity, feasibility, efficiency, and ultimate usefulness in healthcare decision making [35].”

The strengths of our approach include (1) a systematic and comprehensive search for guidance documents, supplemented with searches of regulatory agency websites and contacting a large number of experts in the field, (2) double review and abstraction of data, (3) close collaboration with a multidisciplinary panel of experts, and (4) multiple modalities of input sought from the panel, including surveys (2 rounds), regular face-to-face meetings (1 meeting/week), teleconference (9, including 3 with PCORI), and email comments (over 100).

We did not search the primary literature for standards, which may have led to omission of some important documents. Some disciplines may be more likely to report standards in manuscript form rather than in guidance documents, particularly rapidly developing areas of research, such as large database research. To counteract this limitation we searched websites and queried experts, which yielded no additional guidance document for inclusion. If we had more time, we would have requested feedback from a larger group of stakeholders (those who may submit proposals to PCORI for funding) to comment on the implementation issues. Despite our desire for inclusivity, we feel that the existing team did represent a variety of viewpoints and we frequently discussed how other investigators might interpret our standards when applying to PCORI to pursue their research.
Conflict of interests disclosure:

All authors completed and submitted the PCORI Conflict of Interest Disclosure Statement. All authors reported having no conflict of interest that may have the potential to bias their obligations under the contract.

Funding:

This study and the report were supported by the PCORI contract to review guidance documents for selected methods in PCOR (http://www.pcori.org/funding-opportunities/past-funding-opportunities/methods-review/contracts-review-of-selected-methods-in-pcor/; accessed March 11, 2012).
**Principal Investigator**

Tianjing Li, MD, MHS, PhD; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

**Co-Investigators**

Susan Hutfless, SM, PhD; Division of Gastroenterology, Johns Hopkins School of Medicine

Daniel Scharfstein, ScD; Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Kay Dickersin, MA, PhD; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

**Consensus Panel Members**

Tianjing Li, MD, MHS, PhD; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

Susan Hutfless, SM, PhD; Division of Gastroenterology, Johns Hopkins School of Medicine

Daniel Scharfstein, ScD; Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Kay Dickersin, MA, PhD; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

Roderick Little, PhD; Department of Biostatistics, University of Michigan

Joseph Hogan, ScD; Department of Biostatistics, Brown University

James Neaton, PhD; Division of Biostatistics, University of Minnesota

Michael Daniels, ScD; Department of Statistics, University of Florida

Jason Roy, PhD; Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine

Vincent Mor, PhD, MED; Department of Community Health, Brown University

**Information Specialist**

Claire Twose, MLIS; Welch Medical Library, Johns Hopkins Medical Institution

**Staff Members**

Andrew H. Law, MHS; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
Rose Iyirhiaro, MS; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

**Medical Writer for the Non-Technical Report**

Eric Vohr, MA, Johns Hopkins Hospital

**Graduate Research Assistant**

Meng Li; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
References


47. Data monitoring in committees in clinical trials : a practical perspective Ellenberg, Susan Smith; Fleming, Thomas R; DeMets, David L, 1944-. Chichester : John Wiley & Sons, c2002. NLM ID: 101149773 [Book]


