The Missing Data Problem in Randomized Control Trials

VANDERBILT School *of* Medicine

Collaborators: Sofia Ludwig[1], Jess Robb, MPH[2], Dr. Hal Sox, MD[2]

1. Vanderbilt University, 2. Patient-Centered Outcomes Research Institute (PCORI)

Background: The Patient-Centered Outcomes Research Institute (**PCORI**) is an independent nonprofit organization that is a leading funder of comparative clinical effectiveness research in the United States. As a requirement of receiving funding, researchers must provide a report that demonstrates their adherence to PCORI's standards. One area of the standards is **accounting for and addressing missing data**. This is important because In an RCT, missing outcome data can be a threat to the assumption that the randomly assigned intervention is the sole reason for differing outcomes.

Standards for Preventing and Handling Missing Data

MD-1: Describe methods to prevent and monitor missing data

MD-2: Use valid statistical methods to deal with missing data that properly account for statistical uncertainty due to missingness

MD-3: Record and report all reasons for dropout and missing data, and account for all patients in reports.

MD-4: Examine sensitivity of inferences to missing data methods and assumptions and incorporate into interpretation.

Aims

Measure outcome data missingness in completed PCORI randomized trials

Characterize how PCORI investigators address missing data: prevent it, adjust for potential bias introduced by missing outcomes, impute it.

Assess the rate of missing data (compare the rates of missing data observed vs expected)

Research Questions

What is the frequency of >10% missing outcome data in PCORI RCTs?

How often do PCORI RCTs have a plan to minimize missing data?

How often do PCORI RCTs with >10% missing data execute a plan to compare the characteristics of the study arm populations at follow-up?

What methods are used impute missing outcomes?

Methods

Selection Criteria

- A final research report (FRR*) that has completed PCORI's peer review process
- RCT design (individual or cluster randomization)
- Not "stopped early" per the study's institution, DSMB, IRB, PCORI, etc.
- Total # that met criteria: **(N = 197)**

Process

- Developed Extraction Form
- Extracted key concepts from FRRs*
- Sofia: all; Hal if >10% missingness

Emphasis on Two Part Process

- Same data collection form for **all** RCT FRRs* with **additional questions** on those with >10% missingness

Data Items

- All: includes questions about general characteristics of study and sampling.
- >10%: includes questions about analytic methods for missing data

Preliminary Results

As of 01/20/23, there are **111** that are fully or partially completed; **58 of them have >10% missing data.** That is over **52%** of the reports.

If amount of missing data is less than 10% in all arms, enter % missing and stop data collection at this point.

17.	Type of analysis of the primary study outcome.	1. Intent to treat (ITT) 2. Per protocol 3. Complete case analysis 4. Other 5. Not stated	Usually stated in Analysis section of Methods. Studies may list more than one type of analysis Intent to treat includes everyone who was randomized, regardless of getting the intervention. Per protocol means including only those who got the intervention(s). Complete case analysis means everyone who had follow-up for the primary endpoint.	No
18.	Does the Methods section describe a plan for analysis of missing data?	1. Yes 2. No	Look in the Analysis section of the Methods. Does it say anything about missing data? If so, mark Yes. Doing a text search on "missing data may help to find the text.	No
19.	Does the Report list the reasons for stopping the study intervention or withdrawing from the study and the number of participants with each reason. MD-1	1. Yes 2. No	Should be on CONSORT diagram.	No
20.	Were outcome data obtained on participants who stopped the intervention? MD-1 and MD-3	1. Yes (specifies number followed up) 2. Yes (does not specify number. followed up) 3. No or not mentioned	If yes, author should specify number (or percent) of those who stopped the intervention but gave follow-up data. Don't spend a lot of time looking; this seldom happens.	No
21.	Does FRR mention concerns about the validity of study results because of large amounts of missing data? MD-2	1. In Study Limitations section 2. in Conclusions section 3. Both 4. Other location 5. Not mentioned/discussed	Large amounts raise concern about the assumption that study arm outcome populations differ only in which intervention they receive. Look for this in the Study Limitations section of the Discussion.	No

Example of additional questions.

Discussion

This study aims to address two things: how PCORI awardees can reduce amounts of missing data in their randomized control trials, and an overall understanding of what contributes to large amounts of missing data in randomized control trials.

Moving forward, the results will inform quality improvement for PCORI research and may add to the overall literature on addressing missing data in clinical studies.

Acknowledgements

Thank you to Jess Robb and Hal Sox for this learning opportunity. I would also like to thank all the faculty and staff in the MPH program; you have all made my time at Vanderbilt truly enjoyable.