Missing Patient Reported Outcome Data in Clinical Trials: An Overview and Simulation Study

MSR19

J. S. McGinley[1], A. Savord[1], E. K. H. Chan[2], A. Larbalestier[3], Y. Liu*[2], P. S. Delong[4], RJ Wirth[1]

[1] Vector Psychometric Group, LLC, Chapel Hill, NC, USA; [2] Janssen Global Services, LLC, Raritan, NJ, USA; [3] Janssen Research and Development, Allschwil, Switzerland; [4] Janssen Global Services, LLC, Horsham, PA, USA; *At the time of this work, employee of Janssen Global Services, LLC, Raritan, NJ, USA.

ABSTRACT

Objectives Missing data occur in clinical trials and can have a negative impact on the results. In this study, we (1) outlined concepts related to missing patient reported outcomes (PRO) data, including patterns of missingness, missing data types (missing completely at random [MCAR], missing at random [MAR], missing not at random [MNAR]), and commonly used analytic strategies that address missingness (multiple imputation [MI], mixed model repeated measures [MMRM]), and (2) investigated the potential impact of missing data on PRO results in clinical trial settings.

Methods A simulation study was conducted to empirically evaluate the impact of missing PRO data in clinical trials. The simulation aligned with a hypothetical clinical trial where a PRO instrument was captured for N=500 subjects (n=250 per treatment group) at 3 visits (baseline, 3 months, 6 months) and explored various types (MCAR, MAR, MNAR) and rates (0%, 20%, 40%) of missingness. Both MI and MMRM were used.

Results Simulation results showed that when the missing PRO data were MCAR, within-group means and treatment group differences were unbiased. However, when the PRO data were MAR, unbiased within-group means were only obtained using MI, while unbiased treatment group differences were observed when using MI and MMRM. Findings also demonstrated that, regardless of the analytic strategy used (MI, MMRM), biases arose when missing data were MNAR. Supplemental simulation analyses suggested that capturing the reason(s) for missingness and integrating this information into MMRM as a covariate may help to reduce bias.

Conclusions Using MMRM and MI, researchers can handle missing data that are MCAR or MAR, but these approaches may not adequately address missing data that are MNAR. Future work could focus on designing a method for capturing information regarding reasons for missing data and developing analytic strategies that can leverage these insights to accurately characterize treatment effects.

RATIONALE

Missing data are prevalent in clinical trials, but there are limited pedagogical and methodological resources available in the context of PROs and health-related quality of life (HRQoL).

BACKGROUND

- Missing data occur in clinical trial research and have the potential to meaningfully impact the accuracy and validity of results.
- Methodologists have characterized several underlying mechanisms of missing data and research has investigated their potential implications for clinical trial research.

Patterns of Missing Data

- Two patterns of missing data are especially relevant for PRO data (Enders, 2010) (Fig 1).
  1. General: Sporadic missingness. May appear to be random, but not always the case.
  2. Monotonic: Also known as attrition or drop-out. Patients drop out or stop participating and do not provide any further observations.

Figure 1. Depiction of missingness patterns over time

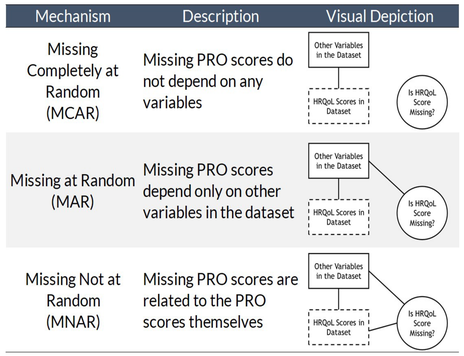
	Baseline	3 Months	6 Months	9 Months
General Missingness				
Subject 1	✓	✓		✓
Subject 2		✓	✓	✓
Subject 3	✓	✓	✓	
Subject 4	✓	✓		✓
Subject 5	✓		✓	✓
Monotonic Missingness				
Subject 1	✓	✓	✓	
Subject 2	✓	✓	✓	✓
Subject 3	✓	✓	✓	✓
Subject 4	✓	✓		
Subject 5	✓			

Notes: Check mark denotes that subject has a HRQoL score at the specific visit.

Missing Data Types

- Three core missing data mechanisms (Rubin, 1976) (Fig 2)
  1. Missing Completely at Random (MCAR)
  2. Missing at Random (MAR)
  3. Missing Not at Random (MNAR)
- MCAR and MAR can be addressed using MMRM and MI.
- MNAR is more difficult to deal with using accessible analytic methods.

Figure 2. Depiction of missing data mechanisms



Methods

- A simulation study was undertaken to:
  - Empirically evaluate the impact of missing data on PRO analyses in clinical trials.
  - Assess the potential utility of a missing data system (MDS) that captures reason(s) for missingness.
- Designed to reflect analysis of clinical trial PRO data (Table 1).

Table 1. Simulation characteristics

Area	Details
Study Design	Endpoints were mean change from baseline in HRQoL score
Simulation	R=1000 completed (non-missing) datasets N=500 total subjects (n=250 per treatment group) 2 covariates (1 binary, 1 continuous) 3 timepoints (baseline, 3 months, 6 months) At baseline: No difference between treatment groups At 3 and 6 months: Small treatment effects
Missing Data	MCAR, MAR, MNAR BL: 0% missing; 3m: 20% missing; 6m: 40% missing
Modeling	MMRM estimated with MLE (with and without MI) MMRM model-based LSM difference endpoints: Within treatment groups Between treatment groups

Results

Descriptive Statistics

- Simulation results related to descriptive statistics showed that when missing data were MCAR, descriptive statistics for HRQoL were unbiased.
- When missing data were MAR, descriptive statistics were biased unless MI was utilized.
- Biases in the descriptive statistics arose regardless of MI when missing data were MNAR.
- Findings suggested that when accurate descriptive statistics are of importance, MI should be considered when missing data are potentially problematic.

Modeling Results

- When missing data were MCAR and MAR, both MMRM using available data and MMRM with MI provided unbiased estimates of change within and between treatment groups (Table 2) (Fig 3).

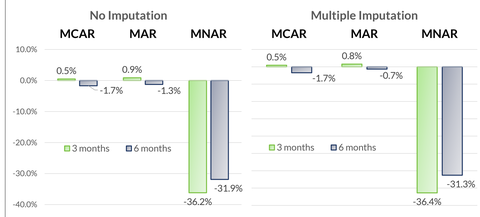
Table 2. Simulation results: Within-group change from baseline

	Within-Treatment Groups							
	No Missing		MCAR		MAR		MNAR	
	Active aEst	Control aEst	Active aEst	Control aEst	Active aEst	Control aEst	Active aEst	Control aEst
No Imputation								
LSMchg 3m	3.40	1.51	3.41	1.51	3.40	1.50	3.55	2.35
LSMchg 6m	3.39	1.49	3.37	1.50	3.38	1.50	3.54	2.25
Multiple Imputation								
LSMchg 3m	-	-	3.41	1.51	3.40	1.50	3.40	2.20
LSMchg 6m	-	-	3.37	1.51	3.38	1.49	3.39	2.09

Note: aEst = average estimates over the replications.

- The non-imputed and multiple imputed MMRMs had similar statistical power estimates (Fig 4).
- Statistical power was significantly reduced when missing data were MNAR.

Figure 3. Relative bias in treatment group difference



Note: Relative bias is computed compared to the no missing condition.

Figure 4. Empirical power to detect treatment group differences

Supplemental Simulation Results

- Results showed that not accounting for the reason for missingness (e.g., as captured via MDS), led to biased estimates.
- Bias was reduced by 24.5% to 33.6% when this information (i.e., reason for missingness) was included as a covariate in the MMRM.

Conclusions

- Missing PRO data can have a negative impact on the accuracy of clinical trial results.
- Using MMRM and MI, researchers can handle missing data that are MCAR or MAR, but these approaches will not address missing data that are MNAR.
- Preliminary simulation results suggested capturing information regarding reasons for missing data and developing analytic strategies that can leverage these insights to accurately characterize treatment effects.

References

Enders, C. K. (2010). Applied missing data analysis. Guilford press.

Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3), 581-592.

Funding

This work was funded by Janssen Global Services, LLC.