Two Approaches to Evaluate Missing Clinical Outcome Assessment Responses: A Simulation Study

Qin S, Ma J, Nelson L

RTI Health Solutions, Research Triangle Park, NC, United States

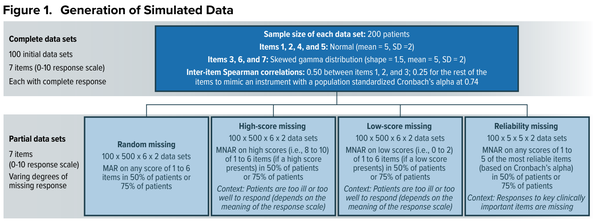
BACKGROUND

- Missing responses are common in multiple-item or daily diary clinical outcome assessments (COAs) and may be missing-at-random (MAR), such as a patient not recording responses due to device failure, or missing-not-at-random (MNAR), such as a patient skipping a response due to disease severity (e.g., a patient experiencing a high level of pain).
- Missingness may impact the reliability and measurement error of COA scores and ultimately reduce or inflate the statistical power in determining treatment efficacy. The US Food and Drug Administration (FDA) recommends that missing data scoring rules be specific to each COA with consideration of the hierarchy of item-level clinical importance and determined during the COA development process.[1,2]

OBJECTIVES

- This simulation study examines the feasibility of using COA scores' measurement error to support a missing data scoring rule for developing clinical trial endpoints.

Figure 1. Generation of Simulated Data



SD = standard deviation.

EVALUATION METHODS

- Two statistics were computed for every partial data set and the corresponding complete data set:
  - The standard error of measurement (SEM) of the 7-item mean score, where SEM = SD × $\sqrt{(1 - \textit{Cronbach's alpha})}$, yielding $SEM_{partial}$ and $SEM_{complete}$
  - Intraclass correlation coefficients (ICCs) between the 7-item mean score from partial data and the one from complete data, using a two-way analysis of variance model with mixed effects, yielding $ICC_{partial\text{-}complete}$
- For every initial data set and corresponding partial data sets, missingness is considered as not having an impact on the scores' measurement error or reliability if the following criteria are met. (Otherwise, it is flagged.)
  - At least 95% of the 500 $SEM_{partial}$ are in the range of 0.9 × $SEM_{complete}$ to 1.1 × $SEM_{complete}$ for the random and high/low-score missing
  - At least 95% of the 500 $ICC_{partial\text{-}complete}$ ≥ 0.81 for random and high/low-score missing
  - $SEM_{partial}$ is in the range of 0.9 × $SEM_{complete}$ to 1.1 × $SEM_{complete}$ for reliability missing
  - $ICC_{partial\text{-}complete}$ ≥ 0.81 for reliability missing

RESULTS

- The SEM method was more stringent and sensitive to the change in the number of missing items than the ICC method for the random missing (Table 1) and reliability missing (Table 2).
- The SEM method was more stringent than the ICC method in high-score missing (Table 1) and also in low-score missing until four items (Table 1).
- The fewer flags in the high-/low-score missing arise in part because the actual percentages of patients with missing and the actual numbers of missing items depend on response distributions. For example, when 50% of patients were given the chance of missing, the high-score missing up to one item tended to have approximately 3% of the 200 patients with missing; this percentage rose to approximately 14% when up to six items were allowed to miss (including 10% missing one item, 3% missing two items, and 1% missing more items). The low-score missing up to one item tended to have about 9% of patients with missing; this percentage rose to approximately 45% when up to six items were allowed to miss.

Table 1. Flag Rates of Random Missing and High-/Low-Score Missing

No. of Items or Days Allowed to Miss	Evaluation	Flags Across 100 Initial Data Sets: Patients With Random Missing 50%	Patients With Random Missing 75%	Patients With Chance of Missing High Scores if Present 50%	Patients With Chance of Missing High Scores if Present 75%	Patients With Chance of Missing Low Scores if Present 50%	Patients With Chance of Missing Low Scores if Present 75%
1	Method 1: 95% of $SEM_{partial}$ in range	16	45	0	0	10	38
	Method 2: 95% of ICC ≥ 0.81	0	0	0	0	0	0
2	Method 1: 95% of $SEM_{partial}$ in range	41	62	0	0	9	52
	Method 2: 95% of ICC ≥ 0.81	0	0	0	0	0	0
3	Method 1: 95% of $SEM_{partial}$ in range	74	81	0	5	12	49
	Method 2: 95% of ICC ≥ 0.81	0	0	0	0	0	0
4	Method 1: 95% of $SEM_{partial}$ in range	97	97	0	23	14	55
	Method 2: 95% of ICC ≥ 0.81	0	68	0	0	4	94
5	Method 1: 95% of $SEM_{partial}$ in range	100	99	6	56	17	63
	Method 2: 95% of ICC ≥ 0.81	100	100	0	0	76	100
6	Method 1: 95% of $SEM_{partial}$ in range	100	100	20	75	32	77
	Method 2: 95% of ICC ≥ 0.81	100	100	0	0	99	100

Table 2. Flag Rates of Reliability-Based Missing

No. of Items or Days Allowed to Miss	Evaluation	Flags Across 100 Starting Sets: 50% Patients With Reliability Missing	70% Patients With Reliability Missing
1	Method 1: $SEM_{partial}$ in range	10	30
	Method 2: ICC ≥ 0.81	0	0
2	Method 1: $SEM_{partial}$ in range	15	33
	Method 2: ICC ≥ 0.81	0	0
3	Method 1: $SEM_{partial}$ in range	34	54
	Method 2: ICC ≥ 0.81	0	0
4	Method 1: $SEM_{partial}$ in range	65	77
	Method 2: ICC ≥ 0.81	10	16
5	Method 1: $SEM_{partial}$ in range	94	94
	Method 2: ICC ≥ 0.81	47	82

SEM Method Performance

- The random missing (Figure 2) and reliability missing (not shown) results were similar.
  - There was no trend for the impact of missing on Cronbach's alpha. The impact of missing was observed on SD. Most differences between $SD_{partial}$ and $SD_{complete}$ were positive and increased when more items were randomly missed.
  - Hence, $SEM_{partial}$ increased and became greater than $SEM_{complete}$, indicating the change in $SEM_{partial}$ was responsive to the change in missing.
- The high-score and low-score missing results were unique (Figures 3 and 4).
  - A slight upward trend was observed in high-score missing where partial alpha appear higher (better) than alpha of compete data. A downward trend was observed in low-score missing (Figure 3). Most of the differences between $SD_{partial}$ and $SD_{complete}$ were negative (Figure 4).
  - Hence, for high-score missing, $SEM_{partial}$ appeared lower (better) than $SEM_{complete}$, indicating that change in $SEM_{partial}$ was responsive to the change in missing. For low-score missing, the decreased SD offset the impact of decreased alpha, which may explain the lower responsiveness of the SEM method than the ICC method at high levels of missing (Table 1).

Figure 2. Random Missing: SEM and SD

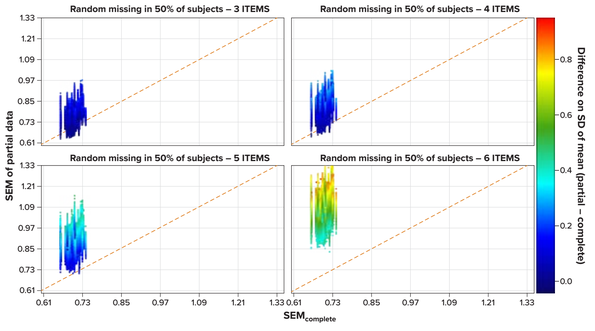


Figure 3. High-Score and Low-Score Missing: Cronbach's Alpha and SD

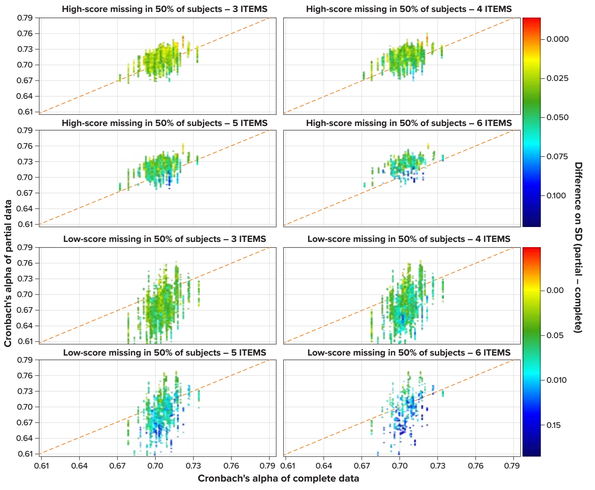
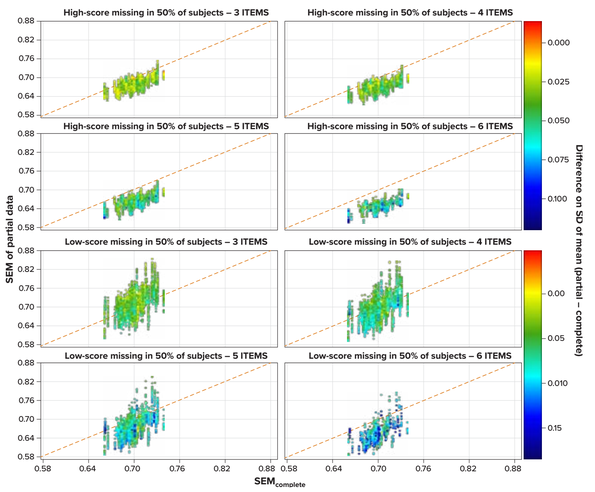


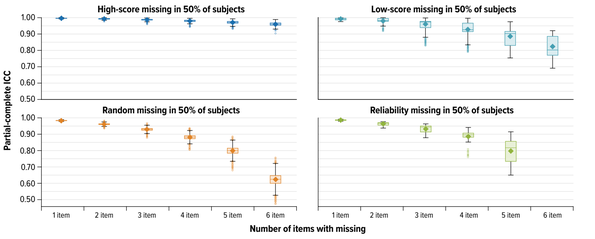
Figure 4. High-Score and Low-Score Missing: SEM and SD



ICC Performance (Figure 5)

- The impact of random missing on $ICC_{partial\text{-}complete}$ was greatest. The means of $ICC_{partial\text{-}complete}$ for reliability missing were similar to random missing, but the variability of ICCs was greater, which explains the fewer flags at high levels of missing (Table 1 and Table 2).
- Followed by low-score missing, high-score missing had the smallest impact on $ICC_{partial\text{-}complete}$.

Figure 5. All Types of Missing: Partial-Complete ICC



DISCUSSION AND CONCLUSION

- The SEM method was shown to be sensitive to the change in missing and responsive (general monotonically) in random missing and high-score missing when the majority of scores were low.
- In low-score missing when the majority of scores were low, the ICC method is more sensitive. However, overall, the ICC method is not as responsive as the SEM method; this result could be attributed to the 0.81 cutoff and/or the simulated moderately large (relative to scale) between-person variability of COA scores. Its performance may improve in a clinical population with restricted score variability.
- Random missing appeared to support the most restrictive missing rule. This, in part, may be attributed to the dominating effect of its largest actual percentages of patients with missing and the largest numbers of actual missing items. Therefore, for application on a real data set, the simulation for random missing is likely sufficient.
- Both of the proposed SEM and ICC methods are sample-specific, and the resulting missing rule may be more stringent in situations with larger percentages of patients with missing.

REFERENCES

1. Food and Drug Administration. https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf.
2. Food and Drug Administration. https://www.fda.gov/downloads/Drugs/NewsEvents/UCM620708.pdf.

CONTACT INFORMATION

Shanshan Qin, PhD
Senior Psychometrician, Patient-Centered Outcomes Assessment

RTI Health Solutions
Research Triangle Park, NC

Phone: +1.919.541.6011
E-mail: sqin@rti.org

Presented at: ISPOR 2019; May 18-22, 2019; New Orleans, LA