Comparing Alternatives to the Three-Form Planned Missing Data Design

Alexander M. Schoemann\textsuperscript{1}  E. Whitney Moore\textsuperscript{1}  Emily M. Meier\textsuperscript{1}  Kelly L. Reburn\textsuperscript{2}  Mark C. Bowler\textsuperscript{1}

\textsuperscript{1}East Carolina University

\textsuperscript{2}IO Psych Group

M3 2024
Outline

- Planned missing data designs
  - 3 forms design
- Complete data vs. planned missing
  - Real data example
- Random missing vs. planned missing
  - Simulation study
Missing data does not have to be a problem!

Two types of planned missing data designs:
- Time-based planned missing data designs
  - Control participant entry into the study (e.g., cohort sequential design)
- Participant based planned missing data designs
  - Randomly assign participants to receive only a subset of items
Three-form planned missing design

- Item based planned missing design
- Items are divided into 4 “sets”
  - Set X: items administered to all participants
  - Sets A, B, and C: Items administered to 2/3 of participants
    - Participants are randomly assigned to receive 2 of the 3 sets (e.g. AC)
Three-form planned missing design

<table>
<thead>
<tr>
<th>Form</th>
<th>X</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Three-form planned missing design
Three-form planned missing design

- **Advantages**
  - More items per participant!
  - Or... less fatigue per participant!
  - Less unplanned missing data (Harel, Stratton, & Aseltine, 2011)
  - Reduced practice effects (Jorgensen, et al., 2014)

- **Disadvantages**
  - Less power than a complete data design
    - Latent variable models alleviate this
  - Requires a “large” sample size
    - 100+ participants (Jia, et al., 2014)
Alternative designs

- Complete data design
  - Assign all participants to receive all items
- Random planned missing
  - Assign each participant to receive a random subset of all items.
Complete data design

- Greater power and less unplanned missing than complete data designs (Harel et al., 2015)
Complete data design

- Greater power and less unplanned missing than complete data designs (Harel et al., 2015)
- Increased fatigue for participants
Complete data design

- Greater power and less unplanned missing than complete data designs (Harel et al., 2015)
- Increased fatigue for participants
- How do parameter estimates compare between complete data and planned missing designs?
Survey of 892 real-estate agents
Survey had a total of 163 items including demographics and various work based constructs
Participants randomly assigned to complete all items ($n = 131$) or complete a subset ($n = 872$)
- Subset of items were 110 total items based on a 3-forms design
- Planned missing had ~33% missing
Compare factor model with two constructs
  - Construct 1 - Work Engagement: 9 items
  - Construct 2 - Turnover Intentions: 4 items

Engagement assessed at the start of the study, turnover intentions assessed at the end of the study

Items for constructs were split across the X, A, B, and C sets
Complete data design: Example

- Use multiple group CFA to compare:
  - Factor structure
  - Factor loadings
  - Item intercepts
  - Item residual variances
  - Latent means variance and covariances
Established configural, and weak invariance
\[ \chi^2(11) = 8.57, \ p = .661, \Delta CFI = .000 \]

Established strong invariance?
\[ \chi^2(11) = 53.78, \ p < .001, \Delta CFI = .007 \]
- Driven by two intercepts in engagement. Small differences in intercepts (\(d < .3\))

Established strict invariance?
\[ \chi^2(13) = 30.22, \ p = .004, \Delta CFI = .003 \]
- Driven by one variance in turnover intentions.
- No systematic differences in residual variances
Complete data design: Results

- Difference in latent means $\chi^2(2) = 97.77, p < .001$
  - No significant difference in turnover intention means
  - Mean of engagement is lower in the complete data group ($d = 1.10, p < .001$)
- No difference in latent variances $\chi^2(2) = 1.87, p = .394$
- Difference in latent covariance $\chi^2(1) = 3.99, p = .046$
  - $r = -0.37$ for missing data and $r = -0.56$ for complete data
Complete data design: Discussion

- No major differences in parameters between planned missing and complete data designs
- No evidence of fatigue from participants in parameters
  - Survey may be too short (~15 minutes) to observe fatigue effects
  - Almost no unplanned missing (unplanned missing <1% in both conditions)
  - Survey was (relatively) “high stakes” with strong motivation to respond
  - Small n with complete data
Random planned missing

- Easily implemented in survey software (e.g. Qualtics)
- Can include all variables, or a subset of variables
  - e.g., collect complete data on demographics and planned missing on other variables
- Increased patterns of missing data compared to 3 forms design
Simulation study comparing 3 forms design with random planned missing data

CFA model: 4 latent variables, 6 indicators each
- 24 total items
- Factor loadings between .5 and .7 within each factor
- Latent correlations between .2 and .4
Random planned missing: Simulation

- 2 missing data conditions
  - 3 forms missing data have 6 items in each set
    - Distributed across each factor
    - 25% missing data
  - Random planned missing: 25% missing for each participant
- 4 sample sizes (100, 200, 400, 700)
- All missing data handled with FIML
Random planned missing: Simulation

- Convergence 100% in all conditions
- Random planned missing replications too 2-3 times longer to fit
- No differences in parameter estimates, standard errors, or bias across 3-forms or random missing data designs
  - No differences in power for parameters
Random planned missing: Discussion

- 3-forms planned missing and random planned missing missing perform similarly in the simulation study.
- Random planned missing designs may be easier to program in survey software.
- Random planned missing designs may be harder to fit due to larger numbers of missing patterns.
  - Potential issues with coverage when not all items in a survey are used in a model.
- 3-forms designs may work better in longitudinal designs.
  - Especially with practice effects.
Conclusion

- Complete data, 3-form planned missing, and random planned missing designs perform similarly
  - With cross-sectional latent variable models
- The choice of design depends on survey length, anticipated modeling strategy, and ease of implementation
Thank you!

- Questions?
- email: schoemanna@ecu.edu