Power of Planned Missing Designs in Longitudinal Panel Designs

Kimberly Gibson, Alexander M. Schoemann, Fan Jia, Graham G. Rifenbark, Terrence D. Jorgensen, Mijke Rhemtulla, Wei Wu, Todd D. Little

---

**Planned Missing Data Designs**

- Participants are randomly assigned to conditions in which they only answer a subset of questions, or participate in a subset of measurement occasions.
- The planned missing data are missing completely at random (MCAR), so they can be estimated without bias.

**Why use Planned Missing Designs?**

- **Increase Data Quality (Validity)**
  When participants have fewer questions to answer or fewer repeated measurements, they are less likely to be fatigued, and more likely to offer high-quality data (Raghunathan & Grizzle, 1995).
- **Reduce Cost**
  Planned missing designs require less data collection than complete data designs, reducing the cost of measurement. For example, studying cortisol trends requires multiple sample measures per day across several days (Graham et al., 2006).
- **Reduce Unplanned Missing Data**
  Fatigued participants are more likely to skip questions or assessment occasions, or drop out entirely. Unplanned missingness can lead to biased results.
- **Quantify Practice Effects**
  Participants who skip a measurement occasion will not be contaminated on the next occasion, making it possible to disentangle growth effects from practice effects (e.g., McArdle & Woodcock, 1997).

**Missing Data Estimation**

- **Full Information Maximum Likelihood (FIML)**
  FIML estimates a model in the presence of missing data, using all available data to determine the most likely parameter estimates.
- **As long as the model is correctly specified (e.g., relevant interaction terms are included, multilevel structure is specified if it exists), and missingness is MAR or MCAR, parameter estimates and standard errors will be unbiased.**

**Simulation**

- 200 datasets per condition (N = 500) were simulated according to a 3-timepoint panel design
- MCAR missingness on 1-3 variables per timepoint was imposed according to a 3-form design (Graham et al., 2006)
- Cross-lag path values varied from .1 to .9 across conditions
- We investigated power to detect cross-lagged paths as the amount of planned missingness increased

**Three Form Design**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Common Set X</th>
<th>Variable Set A</th>
<th>Variable Set B</th>
<th>Variable Set C</th>
</tr>
</thead>
<tbody>
<tr>
<td>High miss</td>
<td>X1</td>
<td>X2, X5</td>
<td>X3, X6</td>
<td>X4</td>
</tr>
<tr>
<td>Med miss</td>
<td>X1, X4</td>
<td>X2</td>
<td>X3, X5</td>
<td>X4</td>
</tr>
<tr>
<td>Low miss</td>
<td>X1, X2, X4</td>
<td>X3</td>
<td>X5</td>
<td>X6</td>
</tr>
</tbody>
</table>

**Results**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Percent Missing</th>
<th>$\beta_4$</th>
<th>$\hat{\beta}_4$</th>
<th>$\overline{SE}_{\beta_4}$</th>
<th>$\overline{SE}_{\hat{\beta}_4}$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hi miss</td>
<td>27.78</td>
<td>.1</td>
<td>.091</td>
<td>.050</td>
<td>.054</td>
<td>.090</td>
</tr>
<tr>
<td>Med miss</td>
<td>22.22</td>
<td>.1</td>
<td>.091</td>
<td>.050</td>
<td>.054</td>
<td>.083</td>
</tr>
<tr>
<td>Low miss</td>
<td>16.67</td>
<td>.1</td>
<td>.092</td>
<td>.050</td>
<td>.053</td>
<td>.081</td>
</tr>
</tbody>
</table>

Note: Convergence is 100% for all models with N = 500. Results are for one cross-lagged path. Results for other parameters and other models are similar.

**Conclusions**

- Parameter estimates resulting from the PM design were unbiased and efficient.
- Power suffers little whether 1/6 or 3/6 variables per timepoint are missing.
- With N = 500, power to detect cross-lagged paths increases to 1.0 as the parameter value reaches .4 or higher, regardless of missingness condition.
- With lower N, convergence may become a problem.

**References**