Finding the Right Auxiliary Information for Nonresponse Adjustment Models

In Search of Zs with Desirable Properties

Andy Peytchev, RTI International
What we are (usually) doing wrong

What we should aim to be doing

What we can be doing
1. Nonresponse adjustment models are generally bad
• Few studies report model fit for nonresponse weight adjustment models
  – Generally poor model fit in general population surveys, relying on aggregate information (e.g., Census Block Group level)

  – Exceptions to this can be undesirable. Variables endogenous to nonresponse are strong “predictors” of nonresponse and can do more harm than good (more on this in a moment)

• Effort to use advanced statistical methods (e.g., machine learning algorithms) to identify complex interactions are limited by the lack of auxiliary information with desirable properties
2. Focus on explaining nonresponse is misplaced
Strongest predictors of nonresponse generally do not help with bias reduction
- Refusal on a prior call/contact attempt, number of calls/contact attempts (Wagner, Valliant, Hubbard, and Jiang, 2014)

Associations with nonresponse but not with survey variables of interest unduly increase the variance estimates (Little and Vartivarian, 2005)

<table>
<thead>
<tr>
<th>Association with nonresponse</th>
<th>Association with outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Low</td>
<td>Bias: ---</td>
</tr>
<tr>
<td></td>
<td>Var: ---</td>
</tr>
<tr>
<td>High</td>
<td>Bias: ---</td>
</tr>
<tr>
<td></td>
<td>Var: ↑</td>
</tr>
</tbody>
</table>
3. Commonly used auxiliary information lacks desirable properties
The ideal “Z” variable: an indicator for a common cause of both likelihood to respond and the survey outcome of interest.

Note: Partial figure.

The objective is seldom met in practice. Auxiliary variables tend to be associated with EITHER nonresponse OR survey variables.

4. Promising avenues for improvement
Potential for Improving Nonresponse Adjustments

- Designed paradata
  - Interviewer observations
  - Proxy reports

- Embed survey content for weight calibration
  - Consider calibration beyond traditional demographic characteristics

- Administrative data
  - E.g., student information

- Statistical methods
  - E.g., tree-based methods (limited utility depending on auxiliary data)
  - Multiple imputation for unit nonresponse, particularly for swiss-cheese pattern of missing auxiliary data and improved efficiency
- National Survey of Family Growth’s interviewer observations of sexual activity (used in nonresponse weighting adjustments)

### Table 3. Accuracy Rates for the Interviewer Judgments of Current Sexual Activity in the Last Four Quarters of the NSFG (2006–2010).

<table>
<thead>
<tr>
<th></th>
<th>Quarter 13</th>
<th>Quarter 14</th>
<th>Quarter 15 (I)</th>
<th>Quarter 16 (I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy rate</td>
<td>0.7709</td>
<td>0.7700</td>
<td>0.7809</td>
<td>0.7915</td>
</tr>
<tr>
<td>False positive rate</td>
<td>0.5732</td>
<td>0.5989</td>
<td>0.4514</td>
<td>0.5380</td>
</tr>
<tr>
<td>False negative rate</td>
<td>0.1328</td>
<td>0.1249</td>
<td>0.1488</td>
<td>0.1124</td>
</tr>
</tbody>
</table>

Designed Paradata

- Interviewer variance in interviewer observations

2015 California Health Interview Survey
- Two-stage design with screener and main interview
- Ask about health conditions for each selected household member (often a proxy report)

- Found substantial measurement error in the screener reports
- Measurement error was correlated with nonresponse
- Underreporting was correlated with nonresponse

What behaviors are related to processes generating nonresponse decisions, and are likely associated with many survey variables?

- Highly correlated with nonresponse and substantive survey variables

Two example studies using General Social Survey data (Peytchev, Presser, and Zhang, 2018)

- Voting
- Volunteering

Calibrate survey weights using benchmark estimates from a source that is not subjected to high nonresponse