Imputation Using the Other Pair Member

Peter Frechtel, Victoria Scott, Amy Couzens, and Andrew Moore, RTI International

Jonaki Bose, SAMHSA
Acknowledgements

- The presentation is sponsored by RTI International’s Statistics and Epidemiology Unit, with the research included in the presentation stemming from ongoing methodological work conducted under the contract for the National Survey on Drug Use and Health (NSDUH). The NSDUH is funded by the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality under contract no. 283-2008-00004C and project no. 0211838.

- The views expressed in this presentation do not necessarily reflect the official position or policies of SAMHSA or the U.S. Department of Health and Human Services; nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.
Study Background

- NSDUH provides national, state and substate data on substance use and mental health in the civilian, noninstitutionalized population age 12 and older.

- Data are collected annually on an ongoing basis from January to December.

- Approximately 700 field interviewers (FIs) staffed.

- Approximately 140,000 household screenings and 67,500 interviews completed annually.

- Conducted by RTI under contract with SAMHSA.
What is a Pair?

- From each household that is screened, either 0, 1, or 2 persons are selected.
- “Pair”: two responding household members (Chromy and Singh, 2001).
• 58.8% of respondents in pairs
• 85.7% of pairs were known to be family pairs
NSDUH Imputation

- Method: Predictive Mean Neighborhoods (PMN) (Singh, Grau, and Folsom, 2002)
- Model-based, hot-deck method
- Does not directly consider pair relationship
- Some NSDUH questions at household level or family-in-household level
- Idea: use the other pair member (OPM) as the donor when possible
When is OPM Appropriate?

- When both pair members respond, they tend to provide the same response.
- The level of agreement for imputed cases is lower than the level of agreement when both pairs respond.
- When variables are at the household level or the family-in-household level.
- When there is significant item nonresponse.
# Good Candidates for OPM

<table>
<thead>
<tr>
<th>Variable Group</th>
<th>Pair Members Tend to Agree</th>
<th>HH / Family Level Variables</th>
<th>Significant Nonresponse&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drugs</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Health Insurance</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Household Composition</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

<sup>1</sup>Nonresponse rate > 0.5%
## A Closer Look: Income

Feasibility for OPM, Income Variables, 2009 NSDUH

<table>
<thead>
<tr>
<th>Variable</th>
<th>% Agreement</th>
<th>Resp. Pairs</th>
<th>PMN</th>
<th>NR Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Income (dichotomous)</td>
<td>93.30</td>
<td>62.94</td>
<td>5.73</td>
<td></td>
</tr>
<tr>
<td>Total Income (finer)</td>
<td>66.84</td>
<td>14.29</td>
<td>9.97</td>
<td></td>
</tr>
<tr>
<td>Welfare Payments</td>
<td>97.75</td>
<td>80.66</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>Social Security</td>
<td>95.63</td>
<td>73.87</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>Supp. Social Security</td>
<td>96.57</td>
<td>79.95</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>Welfare Services</td>
<td>96.61</td>
<td>84.74</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Wages</td>
<td>93.54</td>
<td>74.07</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Food Stamps</td>
<td>95.69</td>
<td>70.40</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td># Months on Welfare</td>
<td>78.81</td>
<td>28.89</td>
<td>4.29</td>
<td></td>
</tr>
</tbody>
</table>
A Closer Look: Income (cont.)

- When to use OPM?
- Family-in-household level questions: only for family pairs? (85.7% of pairs are family pairs)
- Number of days between responses: theoretically irrelevant for income
Number of Days Between Interviews

Cumulative Mass Function, 2009 NSDUH

- 65.5% on the same day
- 86.2% within a week
- 92.2% within two weeks
- 96.7% within a month
9 regression models run: one for each of the 9 income variables

Logistic regression model:
- Dependent variable: whether pair members give the same response
- Independent variables: family pair indicator and number of intervening days
Logistic Regression Results: Income

- Family pair indicator significant for all 9 variables
- Number of days between interviews significant for 7 of 9
- “Significant but not important”
- Would reduce amount of PMN imputation by 16.9% - 40.4%
- Appears that using OPM may be a reasonable option
Logistic Regression Results: Income (cont.)

Predicted Probability of Pair Member Agreement, Receiving Income from Food Stamps, by Pair Type and # Intervening Days

- **Family Pairs**
- **Other Pairs**

Number of Intervening Days

Predicted Probability of Agreement

- 98%
- 96%
- 94%
- 92%
- 90%
- 88%
- 86%
- 84%
- 82%
- 80%
Next Steps and Other Ideas

- Household composition, health insurance, and demographic variables
- Use OPM response in PMN in imputation model (non-deterministic approach):
  \[ \ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \cdots + \beta_i \text{(pair indicator)} \\
  + \beta_{i+1} \text{(pair indicator)} \times \text{(OPM value)} + \cdots + \beta_n X_n \]
- Empirical comparison
References


More Information

Peter Frechtel
Statistician
202.974.7833
frechtel@rti.org