Using Contact Data to Gain Operational Efficiencies

Rachel Horwitz, U.S. Census Bureau
Renee Reeves, U.S. Census Bureau
Beth Newman, U.S. Census Bureau
Demographic Statistical Methods Division – Survey Methodology

*Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.
Contact History Paradata

- Available for:
  - Computer Assisted Telephone Interviews (CATI)
  - Computer Assisted Personal Interviews (CAPI)

- Contact History Instrument Provides:
  - Information about each contact associated with a case including:
    - Time
    - Type of contact
    - Contact outcome
  - Interviewer notes
## Example of CATI/CAPI Paradata

Data are made up examples of what are in the Contact History Instrument (CHI)

<table>
<thead>
<tr>
<th>CASEID</th>
<th>PHONE</th>
<th>EMPLOYEEID</th>
<th>LOG_DATE</th>
<th>TC_TIMEZONE</th>
<th>RESPONDENT_TIMEZONE</th>
<th>ACTION_TYPE</th>
<th>DIAL_SCREEN_TIME</th>
<th>HANG_UP_TIME</th>
<th>ACCESS_TYPE</th>
<th>OUTCOME</th>
<th>FINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>200001</td>
<td>SYSTEM</td>
<td>7/10/2017</td>
<td></td>
<td></td>
<td></td>
<td>013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200001</td>
<td>111-111-1112</td>
<td>5789</td>
<td>8/30/2017 EST</td>
<td>CST</td>
<td></td>
<td>001</td>
<td>18:58:55</td>
<td>18:59:57</td>
<td>outgoing</td>
<td>035</td>
<td>195</td>
</tr>
<tr>
<td>200002</td>
<td>222-222-2222</td>
<td>4231</td>
<td>9/1/2017 CST</td>
<td>MST</td>
<td></td>
<td>001</td>
<td>14:13:51</td>
<td>14:15:18</td>
<td>outgoing</td>
<td>086</td>
<td>195</td>
</tr>
<tr>
<td>200002</td>
<td>222-222-2222</td>
<td>4325</td>
<td>9/2/2017 MST</td>
<td>MST</td>
<td></td>
<td>001</td>
<td>17:39:13</td>
<td>17:40:12</td>
<td>outgoing</td>
<td>086</td>
<td>195</td>
</tr>
<tr>
<td>200002</td>
<td>222-222-2222</td>
<td>5263</td>
<td>5/18/2017 EST</td>
<td>MST</td>
<td></td>
<td>001</td>
<td>21:31:34</td>
<td>21:34:14</td>
<td>outgoing</td>
<td>121</td>
<td>194</td>
</tr>
<tr>
<td>200003</td>
<td>333-333-3333</td>
<td>5436</td>
<td>5/14/2017 EST</td>
<td>EST</td>
<td></td>
<td>001</td>
<td>21:53:30</td>
<td>21:55:15</td>
<td>outgoing</td>
<td>087</td>
<td>194</td>
</tr>
<tr>
<td>200003</td>
<td>333-333-3333</td>
<td>5843</td>
<td>5/28/2017 EST</td>
<td>EST</td>
<td></td>
<td>001</td>
<td>16:17:15</td>
<td>16:18:43</td>
<td>outgoing</td>
<td>087</td>
<td>194</td>
</tr>
<tr>
<td>200003</td>
<td>333-333-3333</td>
<td>5943</td>
<td>6/3/2017 EST</td>
<td>EST</td>
<td></td>
<td>001</td>
<td>19:45:30</td>
<td>19:46:40</td>
<td>outgoing</td>
<td>087</td>
<td>194</td>
</tr>
</tbody>
</table>
How have contact history data been used?

- Build models to predict cooperation and refusals (Sangster and Meekins, 2004)
- Determine the optimal number of contacts before additional contacts become unproductive (Horngren, Lundquist, and Westling, 2009)
- Identify optimal contact strategies for different subgroups (Horngren, Lundquist, and Westling, 2009)
- Detect which mode is most productive for different cases (Meekins and Phipps, 2016; Tolliver, 2016)
- Limit the number of specific outcome types to reduce overall calls (Griffin and Hughes, 2013)
Recent Contact History Research in the Census Bureau’s Demographic Directorate

- Current Population Survey (CPS)
  - Call efficiencies
  - Contact limits
  - Field Test

- National Survey of College Graduates (NSCG)
  - Overall attempt limit in CATI operation
  - Outcome-based limits in CATI operation
CPS – CATI Historical Data

- **Motivation**
  - Increase in CAPI workload due to cases being recycled from CATI
  - Identifying efficiencies in the CATI operation could help control this

- **Methods**
  - Identify most effective times of day to call
  - Identify call outcomes and household characteristics that affect likelihood to complete a CATI interview using survival analysis
  - Ways call parameters could be reduced
CPS – CATI Historical Results

- Best call times
  - Early in data collection
  - Early in the morning

![Graph showing call completion rates by day and time intervals.](chart)
CPS – CATI Historical Results

- Contact attempt characteristics
  - Survival probability of completion static between 11 and 17 calls
  - 50 percent of completed interviews occurred by call 3 or 4
CPS – CATI Historical Results

- Most likely to respond:
  - Message left in first 4 calls
  - Appointment made in first 4 calls

- Least likely to respond:
  - Immediate hang up/refusal in first 4 calls
  - Cases with more than one number on file
CPS – CAPI Historical Data

- Motivation
  - Managing the rise in field data collection costs

- Methods
  - Used historical contact history data to identify:
    - Optimal number of contacts per case
      - Where response rates leveled off and estimates didn’t fluctuate
      - Reprocessed data treating cases over the identified limit as nonrespondents
    - Impact of reducing the data collection period by one day
      - Reprocessed data making respondents on the last day of data collection nonrespondents

- Analysis Measures
  - Response Rates
  - Key Estimates
  - Respondent Characteristics
CPS – CAPI Historical Results

- Optimal number of contacts per case – 11

<table>
<thead>
<tr>
<th>Metric</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Rates</td>
<td>None</td>
</tr>
<tr>
<td>Key Estimates</td>
<td>None (differences and SE's generally less than 0.01 percentage points)</td>
</tr>
<tr>
<td>Respondent Characteristics</td>
<td>None (region, MIS, tenure, age, sex, Hispanic origin, race, labor force status)</td>
</tr>
</tbody>
</table>

- Cutoff resulted in a 1.6 percent reduction in attempts (29,166)

- Impact of reducing data collection period by one day

<table>
<thead>
<tr>
<th>Metric</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Rates</td>
<td>Significantly Lower</td>
</tr>
<tr>
<td>Key Estimates</td>
<td>Percent not in labor force significantly different</td>
</tr>
<tr>
<td>Respondent Characteristics</td>
<td>Over-Represented - white persons, 70+, not in labor force</td>
</tr>
<tr>
<td></td>
<td>Under-Represented - missing tenure, age, race, age by sex</td>
</tr>
</tbody>
</table>
CPS – Field Test

- **Motivation**
  - Experimentally test the limit identified in the historical data research

- **Methods**
  - Shift interviewers from current mileage and hour constraints to a contact limit of 12
  - Compared demographically similar counties

- **Results**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Rates</td>
<td>None</td>
</tr>
<tr>
<td>Respondent Characteristics</td>
<td>None</td>
</tr>
<tr>
<td>Cost</td>
<td>None</td>
</tr>
<tr>
<td>Key Estimates</td>
<td>Significant difference in unemployment rate</td>
</tr>
</tbody>
</table>
NSCG – Historical Data Analysis

- **Motivation**
  - Part of an overall assessment of NSCG’s contact strategy, aimed at reducing costs and burden while maintaining or improving response rates

- **Methods**
  - Simulated what would happen to response rates, sample representativeness, and key estimates had CATI calls been stopped at various call cutoffs (across all phone numbers)

- **Results**
  - Determined that response rates and representativeness stabilized around 10 calls
NSCG – Field Test

- Motivation
  - Test the post hoc analysis recommendation of limiting CATI contacts to 10 calls per case

- Methods
  - Tested in the 2017 NSCG cycle as part of a contact strategy experiment
  - Random sample of sample cases received a call limit of 10; others had no limit
NSCG – Field Test Results

- Call limit reduced the average number of calls by ~4/case
- Was ~$9 cheaper/case
- However, significantly impacted key estimates and depressed response (although not significantly)
NSCG – Outcome Based Limits

- Motivation
  - Noticed runaway call attempts for different call outcomes (e.g. fax machines)
  - Imposing limits on specific outcomes may be more effective than an across the board limit

- Methods – post hoc analysis using data from the 2017 cycle
  - Identified most common call outcomes
  - Calculated the percent of CATI cases that would not have completed in 2017 under various cutoffs for each outcome
  - Selected cutoffs where the percent of lost cases is less than 0.01, 0.05, and 0.10 percent
NSCG – Outcome Based Call Limits Results

- Produced limits for each outcome and for all outcomes combined
- If the moderate cutoff were implemented:
  - 229 completes lost
  - 27,122 calls saved
- At the outcome level:
  - Lowest cutoff of 3 (number not in service):
    - 13 completes lost
    - 1,116 calls saved
  - Highest cutoff of 19 (busy signal):
    - 26 completes lost
    - 3,832 calls saved
Limitations to Using Contact History Data

- Cannot account for last ditch push at end of data collection
- Noncompliance among interviewers
- Accurate outcome coding
Conclusions and Additional Uses

- Using CATI and CAPI paradata can provide:
  - Cost savings
  - Reduction in respondent burden
  - A more complete understanding of how these operations are working and where efficiencies can be found

- Other uses:
  - Case prioritization
  - Response propensity modeling
  - Auxiliary frame data
Thank you!

Contact: Rachel.T.Horwitz@census.gov