Use of Paradata to Evaluate Medical Expenditure Panel Survey Data and Operations

Lisa B. Mirel
Steven R. Machlin
May 18, 2013
AAPOR Conference
How can paradata help assess variations in reporting of healthcare use in the Medical Expenditure Panel Survey?
Subsample of responding households from the National Health Interview Survey (NHIS).

- Annual since 1996.
- National probability sample of the US civilian non-institutionalized population.
- MEPS includes about 13,000 families and about 35,000 sample persons annually.
## MEPS Overlapping Panel Design

<table>
<thead>
<tr>
<th>MEPS</th>
<th>Panel</th>
<th>Year</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td></td>
<td></td>
<td>R1</td>
<td>R2</td>
<td>R3</td>
<td>R4</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>2012</td>
<td></td>
<td>R1</td>
<td>R2</td>
<td>R3</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td>R1</td>
<td>R2</td>
<td></td>
</tr>
</tbody>
</table>

The highlighted rows indicate the overlapping panel design for the specified years.
MEPS household respondent asked to report all health care use for family members during reference period.

Respondents asked to use notes or records to respond to questions about healthcare use during the reference period.

The reference period length depends on the date of interview.
  - An average of 2 months in panel 17 round 1
  - Range 1 week to 6 months
Zuvekas and Olin showed accurate reporting of inpatient hospital stays but some propensity for under reporting of use for Medicare beneficiaries:

- Office visits by about 19% 
- ER visits by about one-third.

MEPS uses a variety of strategies to enhance recall.

Paradata may help point to areas of under reporting.
Paradata definition

- Data about data collection process
  - Not survey questionnaire responses.

- Uses of paradata:
  - Assess various sources of survey error (e.g. non-response, measurement, and processing errors).
  - Improve quality of survey estimates.
Research question

- Are paradata variables associated with differential reporting of healthcare utilization?
Methods

- **Outcome variable:** Annualized estimate of mean office-based events per person per person

\[ Y = \left( \frac{\text{# events}}{\text{# people in family}} \right) \div \left( \frac{\text{# days in reference period}}{365} \right) \]

- **n=7,273 families**
- **Average number of people in family=3**
Main analytic paradata variables

- Reference period length
  - Length of reference period can affect reporting
  - Inverse association between reference period length and annualized health care use

- Memory aid use
  - Calendar
  - Explanation of benefits
  - Other insurance records

- Data collection path
  - Calendar path less burdensome for respondent
  - Not calendar path includes more probing, may take more time
Other paradata variables

- Number of contacts
- Language of interview
- In person/phone interview
- Reluctant respondent
- Supervisor ID
- Transferred case
- Interviewer experience
- Multi/ single session interview
- Partial/ incomplete NHIS interview
“Non-paradata” variables in model

- Income as reported in NHIS
- MSA status based on NHIS
- Number of people in the family
- % of people in family age 0-15
- % of people in family over 64
- Race/ethnicity
- Marital status
- Insurance status of family members
- Number of priority health conditions
Model

- Assess if any of the analytic paradata variables are associated with office-based utilization after adjusting for other covariates.
- Use multivariable linear regression to predict mean number of events per person standardized to one year.
- Report adjusted means.
Results: Reference period length

Distribution
- 0-<3 months: 12%
- 3-<4 months: 7%
- 4-<5 months: 8%
- >=5 months: 73%

Adjusted means
- 0-<3 months: 4.67* (p-value <0.05)
- 3-<4 months: 4.02* (p-value <0.05)
- 4-<5 months: 3.1
- >=5 months: 3.01

*p-value <0.05 compared to >=5 months
Results: Memory aid use

Distribution
- Calendar/other: 46%
- Calendar only: 23%
- Other only: 21%
- No records: 10%

Adjusted means

- Calendar/other: 6.38* (p-value <0.05 compared to no records)
- Calendar only: 5.33*
- Other only: 4.54*
- No records: 2.9

* p-value <0.05 compared to no records
Results: Data collection path

**Distribution**

- Calendar path: 54%
- Not calendar path: 46%

**Adjusted means**

- Calendar path: 4.61* (p-value < 0.05 compared to not calendar path)
- Not calendar path: 4.05

*Annualized visits per person*
The paradata variables have helped to identify potential areas for improving survey operations.
Current research

- Use paradata to assess unexpected variations in MEPS utilization data.

- Explore using some paradata variables to inform statistical adjustments for measurement error.
Thank you!

Questions?
References

