Paradata on the NHANES Survey

George Zipf

National Center for Health Statistics

August 7, 2012

Agenda

- Review of what Paradata is

- Paradata collection on NHANES

- Analysis of NHANES Paradata

- Conclusions and Recommendations

Paradata – Definition

- Paradata are data about the data collection process.

- Examples include:
  - Time and date of data collection
  - Interviewer
  - Wave of data collection, when samples are sequentially released

Paradata – Current Collectors

- U.S. Census Bureau

- Bureau of Labor Statistics

- National Center for Health Statistics

→ *Paradata has entered the lexicon and is entering field operations*

Paradata – Means of Collection

- Data Entry

- Digital Pen

- Teleform Scanning

- Electronic Record of Contact (eROC)

Paradata – Value

- Cost per Response

- Responsive Design

- Correlation between paradata and survey values

NHANES Overview – Picture

- Obtain national medical prevalence estimates, 1999 – present

NHANES Paradata – Teleform

NHANES Paradata – Teleform

NHANES Paradata Data Quality

Completeness:
- Missing data

Validity
- Out-of range
- Vague

Three variables will be discussed here:
- Day of Week
- Time of Day
- Disposition Code

NHANES Paradata Data Quality

Day of Week:

- Missing data: 12 obs.
- Out-of-Range: 0 obs.
- Vague: 4 obs. (1 "S" and 3 "T")

Given that there were 47,659 contact attempts in 2011, post-QC day of week data quality was extremely high.

NHANES Paradata Data Quality

Time of Day:

- Missing data: 43 obs. (0.1% of total)
- Out-of-Range: 2 obs. (both "15 PM")
- Missing AM/PM: 397 obs. (0.8% of total)
- Other: a few values recorded a contact attempt at 1:00 AM, 1:25 AM, 3:10 AM, etc.

NHANES Paradata Data Quality

Disposition Code:

- Missing data: 3 observations
- Out-of-Range: 13 observations
- Vague: 0

NHANES 2011 Paradata Results

Previously available information:
- 13,244 addresses selected for screening in 2011
- 6,555 survey participants (SP) identified
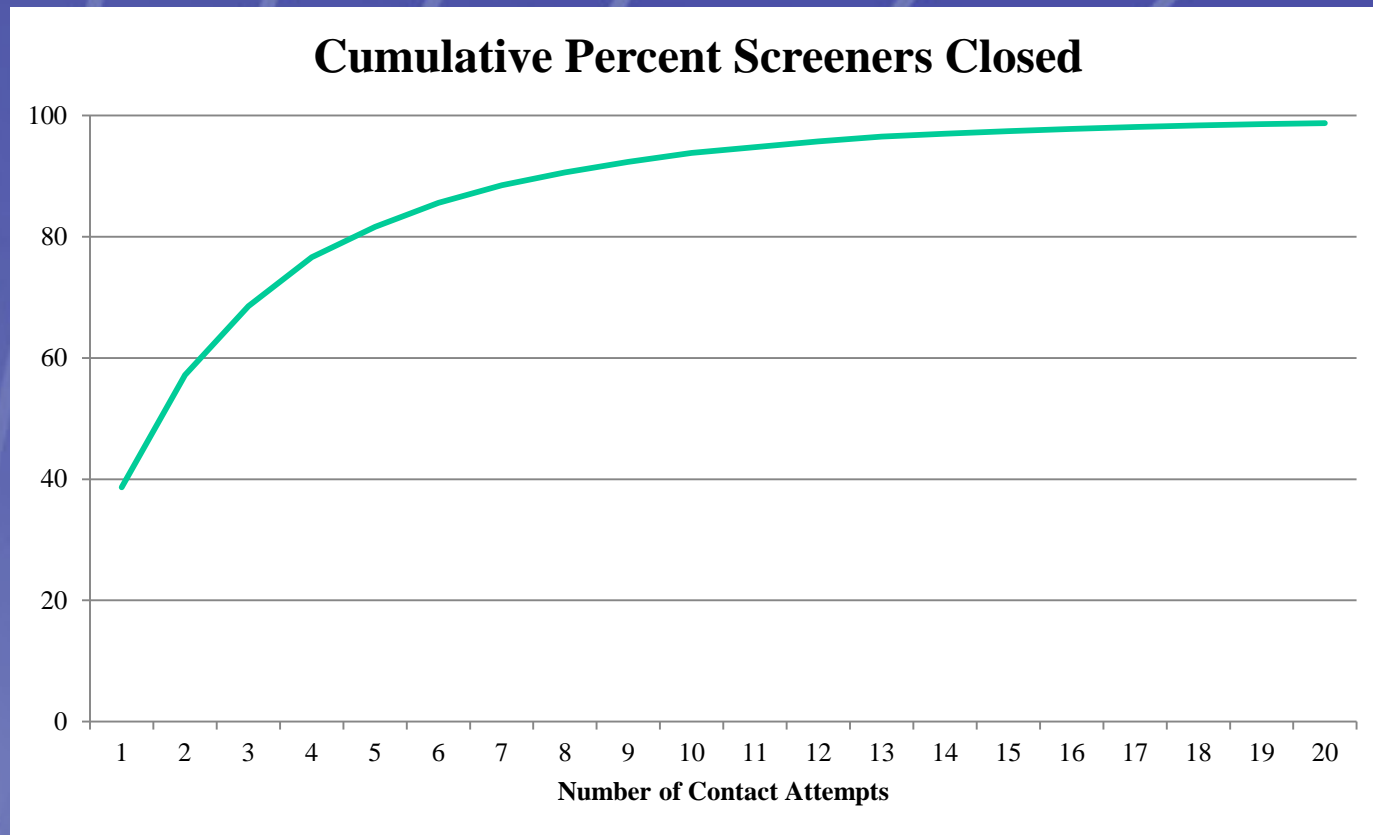- 2,975 addresses with eligible SPs

New information provided by paradata:
- 47,659 contact attempts
- 3.6 average number of contacts per address
- 7.3 average number of contacts per SP
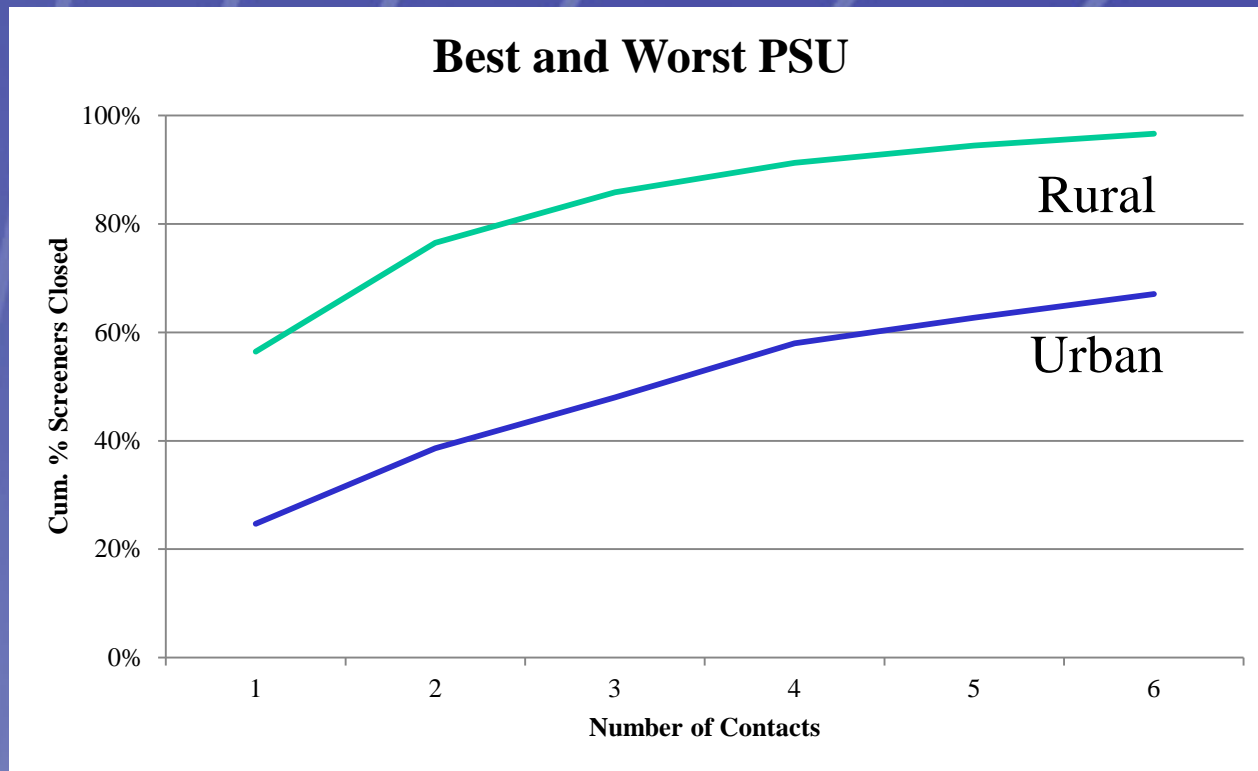- 39% Screeners closed on 1st contact

NHANES Paradata Results – 2011

- How many contacts is enough?

Cumulative Percent Screeners Closed

X-axis: Number of Contact Attempts (1–20)
Y-axis: 0 to 100

NHANES Paradata Results – PSUs

- We love small rural stands
- We do not love large, congested, urban areas

Best and Worst PSU

NHANES Paradata Results – "Not at Home"

- NHANES contacted 13,244 households in 2011
- "Not at Home" is the key factor in field operations

Stand Progress: Disposition = "Not at Home"
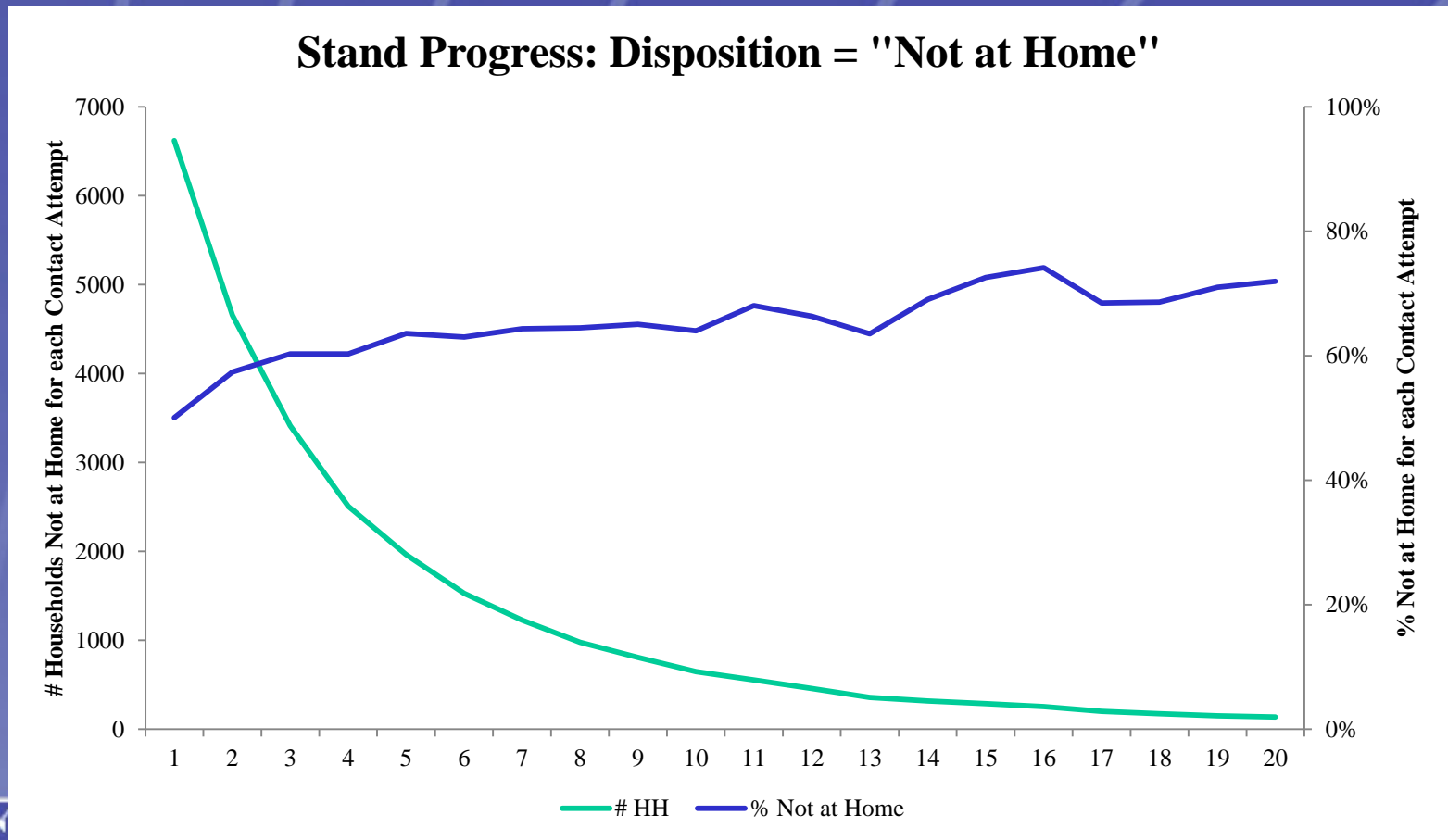
Conclusion

Teleforms:
- Variables – acceptable
- Data Quality – acceptable
- Cost – not cheap when data clean-up and training are included
- Timeliness – slow, not actionable in field

Next steps:
- Consider use paradata to set contact attempt limits (e.g. NHANES should probably be at 12)
- Refine analysis on time of day, day of week, and merge with demographic information, census data