RTI INTERNATIONAL | turning knowledge into practice

The Accuracy of Interview Paradata: Results from a Field Investigation

Kevin Wang and Paul Biemer, RTI International

Prepared for the Annual Conference of the
American Association for Public Opinion Research, Chicago, IL

May 16, 2010

Acknowledgement

- This project is funded by the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract no. 283-2004-00022 and Project no. 0209009
- The views expressed in this presentation do not necessarily reflect the official policies of the Department of Health and Human Services; nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government
- This presentation is sponsored by RTI International's Survey Research Division
- Many thanks to the interviewers and supervisors who took part in the study and RTI International project staff who helped make it possible

Overview

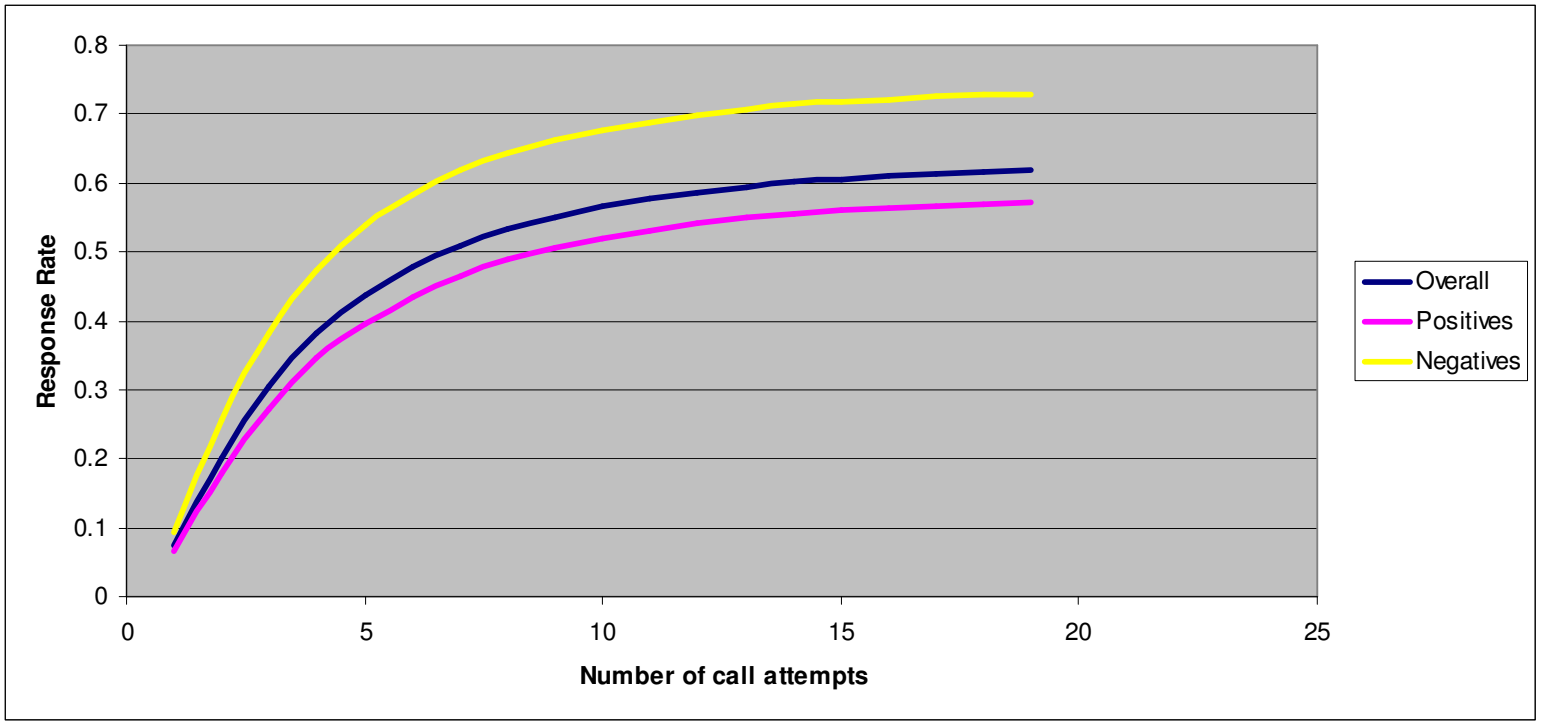
- How can callback data be used to address nonresponse bias through weighting adjustments?
- How did we examine the quality of callback data?
- What did we learn about our callback data?
- What are the implications for addressing nonresponse bias through weighting adjustments that use callback data?

How do we adjust for nonresponse bias?

- Nonresponse bias is a function of the nonresponse rate and differences between respondents and nonrespondents on survey items of interest
- Adjust for nonresponse by weighting up respondents who resemble nonrespondents on variables related to response propensity and the survey items of interest
- Data on callbacks is available for both respondents and nonrespondents; is it related to both response propensity and survey items?

What information could the number of calls tell us about response propensity?

Response Rate for Positives (e.g. “users”), Negatives (“non-users”) and Overall as a Function of Number of Call Attempts



Using data on callbacks to adjust for nonresponse: a simple callback model

$$\Pr(\text{response after } T \text{ callbacks}) =$$

$$\sum_{t=1}^{T}\left[\begin{array}{c}\Pr(\text{response at callback } t \mid \text{no response by } t-1)\\ \times\Pr(\text{no response by } t-1)\end{array}\right] =$$

$\sum_{t=1}^{T}\alpha(1-\alpha)^{t-1}$, where α is the probability of response at callback t

$\Pr(\text{response after } T \text{ callbacks})$ is called the *response propensity*

It can be used in a weighting adjustment to compensate for nonresponse

Small errors in *T* can have large implications for nonresponse adjustments

For example, suppose $\alpha = 0.2$

For $T = 3$, response propensity should be 0.49 (2.04)
If $T = 2$ is recorded, the estimate drops to 0.36 (2.78)
If $T = 4$ is recorded, the estimate increases to 0.59 (1.69)

- If T is not accurate, the estimates of response propensity will be biased and the callback model NR adjustment will not perform well
- Our task in this study is to determine if interviewer records of T are accurate

Purpose of the study

- Gather information from field staff on practices for recording callback data to assess the potential for measurement error in the callback data
  - What are factors that lead interviewers to over- or under-report callbacks?
- Gather recommendations for improving procedures/methods for gathering callback data to reflect visits

Study Background

- The National Survey on Drug Use and Health (NSDUH) is a general population survey
- Conducted by RTI under contract with SAMHSA since 1988
- The nation's leading source of information on substance use behaviors and mental health
- Data collected in all 50 states and the District of Columbia

Study Background (cont'd.)

- Data collected on a quarterly basis each year
- Approximately 700 field interviewers staffed
- Approximately 180,000 household screenings and 67,500 interviews completed annually
- In-person survey
- Screening and interview phases

Methods

- Informal survey of interviewers on reporting scenarios and practices conducted between September 8 and September 16, 2009
  - 601 responses from 653 interviewers (92 percent response rate)
  - Sample Question:

    You approach a controlled access community containing several dwelling units. While trying to enter, you are stopped by a security guard and after speaking with him briefly, he refuses to let you enter. Later in the day, you make a second visit and notice the same security guard standing outside. You do not attempt to speak to him again and decide to work in another part of the segment. Would you enter a ROC to document this second visit to the community in your iPAQ?
- Conference calls with two groups of field supervisors, regional supervisors and regional directors on September 29 and October 5, 2009

Callback data

- For each visit to a sampled dwelling unit, information is entered into Record of Calls (ROC) data through handheld computer
- Data gathered include call outcomes (interim and final result codes), time of day and date of call attempt, respondent reasons for refusal, open-ended notes to help with case management and scheduling
- Information is transmitted on a daily basis from interviewers to supervisors

More incentive to underreport visits than overreport

- Overreporting or “padding” can be caught from timesheet reviews and these interviewers are removed from the project
- Underreporting may occur because of pressures to keep a case “alive”; if too many unproductive visits are recorded, the case may get closed out
- Underreporting may also occur to avoid being perceived as not using time effectively

Non-motivational sources of underreporting – Drive-bys

- Modeling would consider drive-bys as visits since they inform contactability but they are generally not entered as a visit
- Reported if an attempt is made; not reported if no visible change in household (e.g. no car in driveway)
- More likely to be reported later during the quarter when documentation of a visit is needed
- Some summarization of visits in one ROC (e.g. made visits at 9AM, noon, 3PM and 5PM but no one home in all visits)

Other non-motivational sources of underreporting

- Up to two persons can be selected for the interview (A and B person); interviewers may enter a call attempt for just one person rather than both; or get records mixed up
  - 17 percent of interviewers responded that they have recorded callback data for the A person under the B person (or vice versa) “often” or “sometimes”
- “One and done” instructions
  - Supervisor instructs interviewer to make one last visit and if no one is home, close out
  - When asked about a hypothetical situation like this, 44 percent of interviewers said they would only record the final result code

Summary

- Underreporting seems to be more frequent than overreporting of call attempts
- The degree of underreporting can vary a lot by interviewer. This interviewer variance is very difficult to model in a NR adjustment framework
- Depending on the level of precision needed for callback modeling, potential changes range from modest (e.g. interviewer prompts when there are two interviews in a household) to extensive (e.g. new definitions of visits and procedures for recording visits)
- Difficult to obtain more accurate data on visits without additional burden to interviewers with an existing system; more feasible if designing a new system

Future Work

- Advances in technology may provide means for more accurately recording visits without additional burden to interviewers
  - GPS to show when interviewers are at a household
  - Computer Audio Recorded Interviewing (CARI) or other means of recording contact with anyone at the household
- Need a process for recording management decisions to limit callback attempts (e.g. response rate or sample size target has been achieved so close out remaining cases)
- For NR adjustment, more complex models can be developed to model the error in *T*

Questions?

- Email Kevin Wang (kwang@rti.org)
- Thank you!