REACH US Plan for Monitoring, Analyzing, and Calculating Unit Nonresponse

Deliverable 6: Plan for Describing Nonresponse Patterns

Prepared for:
The Centers for Disease Control and Prevention (CDC)
National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP)

Prepared by:
National Opinion Research Center (NORC)
55 East Monroe Street, Suite 3000
Chicago, IL 60603

December 18, 2008

Contract: 200-2008-28054

NORC Project Number: 6590

NORC at the UNIVERSITY OF CHICAGO
Innovative solutions in research and technology

Table of Contents

Appendix

Appendix A: REACH 2010 Key Indicator Variables

REACH US Plan for Monitoring, Analyzing, and Calculating Unit Nonresponse

Nonresponse can be classified into two different forms. *Unit nonresponse* occurs when there is a failure to obtain a questionnaire or data collection form from a member of the sample. *Item nonresponse* occurs when a specific piece of information is not obtained from a responding member of the sample. This document details NORC's plan to monitor, analyze and calculate *unit nonresponse* for REACH US. At the implementation stage, modifications of this plan may be introduced in consultation with the CDC project officer.

1. Monitoring Unit Nonresponse Pattern

Unit nonresponse pattern will be closely monitored throughout data collection. NORC plans to adapt a response rate prediction method that was developed from other NORC studies (e.g., Making Connections, General Social Survey) to REACH US. This method makes use of one source of paradata, the call history dataset, to predict response rate after only a few weeks of data collection, allowing for any necessary adjustment to be made early in the field period. The prediction model uses the call history dataset from a completed study to predict the yield from a study currently underway. Specifically, the call history records in the completed survey are first grouped into cells that are formed by the age of a case (i.e., the number of weeks since the start of data collection), outcome measures (e.g., no contact, refusal, complete, etc.) and possibly other call history variables. Then the percent of cases in each cell that eventually completed the survey (or yield rate) is calculated. Next, the cases in the study underway are grouped in the same way as in the completed study and the yield rate from that study is applied to each cell to project the total number of completes from the released sample. The rationale underlying this method is that cases with similar histories have the same likelihood to complete the survey. This method performed very well on several NORC studies and we feel that it has considerable potential for REACH US. We realize, however, that the complex sequence of stages in the REACH US address-based sampling (ABS) process introduces new complexities. Appropriate modification to the existing model will be necessary as we develop and refine the system.

2. Analyzing Unit Nonresponse Bias

Unit nonresponse has two negative consequences for the quality of the estimates derived from the data. First, nonresponse reduces the sample size; when the number of responses decreases, the variability of survey estimates increases. This consequence can be counteracted by selecting a large enough initial sample so that the achieved sample size satisfies the target requirement. Even then, variability in the achieved response rate adds an additional element of uncertainty to the sample size calculation. Second, and more importantly, nonresponse has the potential to cause bias in the estimates. For means/proportions, the bias depends on two

factors: the response rate, and the difference in the means/proportions between the respondents and nonrespondents. Therefore, bias may be expressed as follows:

$$BIAS = (1 - RR) \times (\overline{X}_R - \overline{X}_{NR})$$

$where\ RR = the\ unit\ response\ rate,$

$\quad\overline{X}_R = the\ mean\ proportion\ for\ respondents, and$

$\quad\overline{X}_{NR} = the\ mean\ proportion\ for\ nonrespondents.$

Thus, bias increases as the difference in means/proportions increases, or as the unit nonresponse rate increases. While the response rate can be calculated, unfortunately we do not know the mean/proportion for the nonrespondents.

Three methods are typically used to gauge the potential impact of unit nonresponse bias in sample surveys:

(1) Comparing survey estimates with external sources of information with known accuracy;

(2) Comparing results for subsets of survey respondents who varied significantly with respect to the difficulty of persuading them to complete the interview; and

(3) Obtaining reliable data on background characteristics for both survey respondents and non-respondents and using these data as covariates in the estimation of parameters.

The first method is generally not feasible because authoritative information sources are not usually available for the populations and communities served by REACH US programs though this situation may be improved with the release of tract-level American Community Survey (ACS) data in two years' time. We believe the second and third methods present potentially promising approaches for assessing the likely magnitude of nonresponse bias and its impact on survey estimates.

Specifically, under the second approach, we plan to compare the four types of respondents that follow different paths through the ABS system: phone respondents, mail respondents after first mailing, mail respondents after second mailing, and in-person respondents. These represent sample groups with different manifest level of response propensities. We can further divide the respondents into 5 to 15 categories based on paradata of the survey process (number and outcomes of contact attempts, occurrence of an initial refusal, different modes, etc.). We can then estimate to what extent the key indicators change as we make more and more effort to contact the respondents. Particularly, the field follow-up cases (i.e., non-responders that are sub-sampled to complete face-to-face interviews) represent different categories of nonresponse to the other modes and levels of effort; for those that do complete the interview, their characteristics on the other dimensions can be identified and used to generate estimates for those that do not respond. Comparing information collected from these difficult respondents with information collected from respondents who are more cooperative may provide a valuable source of evaluating non-response bias. Mean Squared Error (MSE) comparison can

be made to see what influence these field follow-up cases have and to assess the contribution of these cases to the overall quality of the estimator.

Where there are no field follow-up cases however, we will construct an estimate for the nonresponding cases using the characteristics of responders in the response groups identified above through an appropriate weighting scheme. We will agree on a list of key indicators for estimation with the project officer later. As a reference, Appendix A presents the list of key indicators from REACH 2010.

The third approach uses frame-level information to construct nonresponse adjustment cells. Background characteristics that are available for all cases (respondents as well as nonrespondents) are used to form the basis of a model that produces weights to adjust for nonresponse. Such adjustments depend for their usefulness on the relationship between the frame variables and the target variables for the survey. Frame variables with this property are typically difficult to find, but we will endeavor to do so.

3.   Calculating Unit Response Rate

In this final section, we present our plan for calculating the unit response rates. Unit response rates are an important quality indicator for the surveys and can provide a basis for judging the potential nonresponse bias for the survey.

Weighted response rates are the only appropriate response rates for the complex designs we use for REACH US. Weighted response rates capture the fraction of the target population represented in the sample without introducing biased due to differential probabilities of selection.

Applying NORC response rate standards based on AAPOR standards, NORC classifies telephone number/housing unit disposition codes into D, ES, SI, and SE and persons within selected households into IR, ER, and C groups. Table 1 describes the resulting groups.

Table 1.   Disposition Code Categories

Category	Description
D	Sum of base weights for non-occupied or non-residential cases
E_S	Sum of base weights for cases eligible for the screener that did not respond
S_I	Sum of base weights for screened households with no eligible members
S_E	Sum of base weights for screened households with one or more eligible members
I_R	Sum of the product of household and respondent base weights for persons who were selected but then determined to be ineligible
E_R	Sum of the product of household and respondent base weights for persons who were selected but did not complete interview

C	Sum of the product of household and respondent base weights for completed member interviews

The unit response rate is the product of the screener response rate and the interview completion rate. In calculating the screener response rate, NORC applied the following definition:

$$RR_{SCR} = \frac{S_I + S_E}{S_I + S_E + E_S} \, ,$$

The interview completion rate is defined as:

$$RR_{COMP} = \frac{C}{C + E_R} \, .$$

The unit response rate is the product of the completion rate and the screener response rate:

$$RR_{INTV} = RR_{SCR} \times RR_{COMP} \, .$$

As part of our examination of unit response rates, we will calculate unit response rates for each community for the overall sample, by geographic stratification (where appropriate), by sample type and by demographic subgroups. The results of this analysis could be used in determining the cell structure for the nonresponse weight adjustment. By adjusting the weights by the inverse of the response rate within certain subgroups, bias caused by different response rates among these subgroups will be minimized (See REACH US Weighting Plan for more details).

Appendix A: REACH US Key Indicator Variables

Variable Name	Variable Description
FLU65	% of adults 65+ immunized for influenza in the past year
PNEUM65	% of adults 65+ immunized for pneumococcal pneumonia
A1CYR	% of diabetics who had HbA1C measured in the past year
FEETYR	% of diabetics who had their feet checked at least once in the past year
EYEYR	% of diabetics who had a dilated eye exam in the past year
_SMOKER2	% of population currently smoking
_FRTINDX	%of population eating 5+ fruits/vegetables per day
HIBPMEDS	% of aware hypertensives regularly taking medication
HAALL	% of population who know the signs and symptoms of myocardial infarction
STRALL	% of population who know the signs and symptoms of stroke
MAMM2YR	% of women 40+ who had a mammogram in the past 2 years
PAP3YR	% of women who had a pap smear in the past 3 years