ESTIMATING PREVALENCE OF HEALTH AND BEHAVIORAL OUTCOMES AMONG SURVEY NON-RESPONDENTS IN AN EPIDEMIOLOGICAL STUDY OF CHILD PSYCHOPATHOLOGY IN RURAL MAINE.

G. V. Bobashev[1], G. Zahner[1], D. Freeman[2], and P. Biemer[1]
[1]Research Triangle Institute, Research Triangle Park, NC, 27709
[2]University of Texas, Medical Branch, TX

Key Words: Survey Non-Response, Response Propensity, Multilevel Survey, Child Psychopathology, Epidemiology

ABSTRACT

Background: An epidemiological survey of child psychopathology was conducted with children 4-11 years old residing in four rural counties in Maine, including children, parents and teachers as informants. It was difficult to obtain sensitive information (e.g., emotional problems) from this conservative, non-farm, rural population, which posed a challenge for estimating prevalence rates of the outcomes among survey non-respondents. **Methods:** In this study, we adjust for non-response in several ways using model-free and model-based response propensity weights and also estimating prevalence of outcome variables among non-respondent population. Model-based adjustment was based on a brief non-respondent survey. First, time delay from first solicitation to actual response among survey respondents was used to select candidate survey items best associated with response delay. Then, these items were incorporated in a "brief" survey of non-respondents and also were used to model the outcomes among respondents. Finally, we used the respondent model and the results of the "brief" survey of non-respondents to predict the outcomes among the non-respondents. **Results:** 2,133 eligible households were selected via a multi-level survey. Only 1,518 (71%) households have responded to the survey. About fifteen questionnaire items were selected for the non-respondents survey. Response propensity was calculated. For outcomes with high prevalence (closer to 0.5) the model-based prediction of non-response bias is a useful tool, however, for small (<0.15) the predictive power is low to draw inferences about the model-based prediction. We discuss ways to validate the non-response adjustment methods. Funded by NIMH Cooperative Agreement U01 MH51465, Gwendolyn Zahner, PhD, PI.

Introduction

An epidemiological study of child psychopathology in rural Maine was conducted to assess children's mental and behavioral status, as well as service utilization, in a population-based sample of rural families with children ages 4-11. A geographic area of four Maine counties was selected according to four criteria: 1) rurality, as defined by dimensions of population size, density and remoteness from principal metropolitan centers of New England, 2) diversity of rural occupations and range socio-economic conditions, 3) self-containment of the service system, and 4) sufficient size of target child population. socioeconomic and population density characteristics.

Table 1: 1990 U.S. Census Characteristics of the Survey Site

	County 1	County 2	County 3	County 4
Area (square miles)	1,589	3,396	3,967	2,569
Pop. Density (persons/sq. mi)	29.5	43.2	4.7	13.7
Number of Minor Civil Division (MCD's)	40	68	23	49
Density distribution – number of MCD's with persons/sq. mi. of:				
1000+	0	0	0	0
500 – 999	0	5	0	2
100 – 499	4	8	0	3
50 – 99	12	11	2	5
25 – 49	11	12	4	9
10 – 24	2	13	4	12
<10	10	17	13	18
Uninhabited	1	2	0	0
% Rural	79.9	46.6	83.5	91.0
% Non-white households	0.7	1.5	0.7	3.3
% Female headed family households	11.7	14.0	11.4	13.9

% Women in labor force	53.8	55.5	50.9	46.6
% Children 5 – 17 yrs below poverty	10.2	14.7	16.9	23.4
% Unemployed adults	6.1	6.7	9.3	10.8
% Households without full plumbing	4.0	4.1	17.7	11.2
% Households without telephones	3.9	4.1	6.2	5.5

Sampling

The baseline study cohort is based on a total of 2,086 households that were drawn through a multistage sampling procedure with over-sampling of more sparse communities. At the first stage the population was divided into two super-strata depending on the rurality of the area. One super-strata contained areas (strata) with relatively high and medium dense populations such as small city (pop. 25,000-50,000), fringe town (pop. 2,500-25,000, adjacent to small-city or fringe-town), town (pop. 2,500-25,000, not adjacent to city), and rural village (pop. 500-2,500; pop. 250-500 and density > 10 persons/sq.mi).

Sparsely populated area such as rural hamlets (pop. < 250; pop. 250-500 and density < 10 persons/sq. mi; territories), were referred to the second super-stratum. Both super-strata were divided into Primary Sampling Units (PSU). Each unit is defined as a geographically contiguous census-based area with sufficient population. Some of the PSUs were comprised of census areas with different rurality composition and were assigned a rurality code corresponding to where the highest proportion of population resided. The first and the second super-strata contained of 267 and 19 PSUs respectively and formed the PSU sampling frame. One hundred twenty PSUs were drawn with probability of selection proportional to the size of PSU (PPS sampling without replacement) from the first super-stratum and all 19 PSUs were drawn from the second super-stratum. The over-sampling of the sparsely populated areas and Native American Reservations was done to increase accuracy of estimates within these population groups.

In order to locate children 4-11 years old virtually all non-seasonal occupied households in the region were assessed and about 49,200 were listed. These households were then screened for the presence of children ages <18. Households screened with children were rostered (enumerated) for household composition and age of children. If the household contained at least one child 4-11 years of age, it was entered into a household sampling frame.

Rostering required obtaining information from children's caregivers or other adults in the household. If after several attempts at rostering the information on age eligibility was not obtained because, for example, the adults were either absent or too busy, the households were still included into the sampling frame with the goal of obtaining more information at the later steps. In some cases adults refused to provide such information and participate in any further studies. Such households were not included into the sampling frame.

The final sampling frame consisted of 13,077 households. Each fourth or second household was selected depending on whether the household belonged to the first or the second super-stratum. All households were selected from American Indian reservations. This resulted in selection of 2,257 households. Within each household only one eligible child was selected at random. Prior to any contact with potential participants, an advance letter was sent mailed to the selected household that introduced the study and described its importance. Local authorities were informed about the nature of the study. Incentive payments to participants were $30. Interviews were conducted by a team of two lay interviewers who assessed a primary caregiver, a child and a secondary caregiver where available. In order to become a completed case a household should provide at least one completed primary caregiver interview. This resulted in 1,285 completed parent interviews from non-reservation areas and 131 from households in American Indian Reservations.

Response rates:

The response rate for listing and enumeration was 90.6%, ranging from 25.1% to 99.5% in different PSU's (or ranging 85.1% to 98.5% between County/Rurality strata). Overall response rate to interview was 71%. Sources of household non-response (with raw, unadjusted percents) include "hard refusals" by primary caregivers, refusals by other adults in the household, households who were too busy or otherwise unavailable to schedule interviews, households that had moved, and households where no contact could be made. The summary of the response rates are presented in Table 2. Most of the parents (95%) who had completed the CBCL questionnaire have also provided information on child psychopathology in the DISC interview.

Table 2: Characteristics of the Sample

	4-cty	American Indian	Total	% of Total
Sampled	2,086	171	**2,257**	100
Ineligible households at interview	116	8	**124**	6
No age-eligible child	34	0	34	2
Moved outside 4-cty region	82	6	88	4
Unknown primary caregiver	0	2	2	<1
			Eligible	**% of Eligible**
Eligible households	1,970	163	**2,133**	100
Non-Respondents	582	33	**615**	29
Hard refusal, primary caregiver	330	12	342	16
Gatekeeper refusal	32	0	32	2
No contact, too busy*	220	21	241	11
Respondents	1,387	131	**1,518**	71
Mental health interview	1,285	131	1,416	66
Brief surveys	102	0	102	5

*Brief surveys attempted with 160 of 322 No Contact/ Too Busy households and were completed for 102 households.

Analysis weights

The analysis weights are calculated as inverse probabilities of selections adjusted by the response rate.

*WEIGHT= 1/ (Prob_select1*Prob_select2*Response2*Prob_ select3*Response3).* **(1)**

Where *Prob_select1* is the probability of selecting a particular PSU, *Prob_select2* corresponds to the probability of selecting a particular household from the list of eligible households. This probability generally was equal to 0.25 or 0.5 depending on the rurality super-strata and was equal 1 for the American Indian population. *Response2* is the response probability to enumeration, and *Prob_select3* with *Response3* are probabilities of selecting a particular child from a participating household and response probability of the household to interview. We have calculated the Design Effect which measures the amount of variance inflation due to such inter-cluster correlation and unequal weighting.

Handling interview non-response.

Because response rate to the interview was much lower than response to the listing and enumeration, we focus primarily on the former response rate. Traditionally it is difficult to recruit households in rural Maine in government sponsored surveys, particularly those handling sensitive topics. Many householders are reluctant to respond because of privacy concerns.

Non-response was classified by "hard" refusal when an adult explicitly refused to answer the questions and "soft" refusals such as not having time at the moment. Although for the calculation of sampling weights there was no differentiation between these two types of non-response a proportion of "soft" refusals was used in the follow-up study with attempts to be converted to respondents. Soft refusals are hypothesized by some survey methodologists to contribute more to bias than hard refusals (Groves and Couper, Proctor). Therefore, following this philosophy we assume that the prevalence among hard refusal is the same as among all respondents and focus our attention on "soft" refusals.

We also classify the respondents as Early Respondents (ER) and Late Respondents (LR). Early respondents were the subjects that have responded to the interview within the three months from initial contact. Late respondents are those who had responded later than three months. Three month period turned out to be an important threshold because it manifests a change of season and thus possibly family's lifestyle, also 90% of interviews were conducted within a three month period from the time of initial contact.

We used two approaches to account for non-response. If non-respondents are similar to the respondents in terms of their demographic and other characteristics then having non-respondent will just reduce the power of the estimates and simple weight adjustment would be sufficient. If alternatively, non-respondents on average have different characteristics than the respondents then there is a risk of obtaining biased estimates. In order to estimate the impact of non-response on the bias in the outcome and compensate for it we have developed a multi-step propensity model.

Weight adjustments will correct the prevalence estimates but will provide little information about the possible difference in outcomes between respondents and non-respondents. In order to examine the amount of possible bias we developed a prediction model. First we use the respondents and model the outcomes as functions of the key variables. Then using the model's coefficients predict the outcomes among non-respondents.

Simple weight adjustment.
We consider PSU-level non-response rate which could be also aggregated to a rurality level but we feel that PSU-level better characterizes local community beliefs and also reflects the abilities of the recruiting team to convince household respondents participate in the study. This is reflected in calculation of analytic weights.

Response propensity model.
Because little information is known about non-respondents in rural child mental health surveys, a short non-response survey was developed and "soft" non-respondents were assessed with this low-burden survey. In constructing this model we assume that the "soft" non-respondents are the main source of bias, while hard non-responce does not represent bias. Because of the assumption that non-respondents are more similar to late than to early respondents questions for the short survey were selected from the main survey based on respondents' characteristics mostly associated with delayed response. We also included few major study outcomes as a way to test for bias directly. Based on the short survey we have constructed a response propensity model that could be used to account for possible bias. The probability λ of not being a soft non-respondent for household i given the model covariates was estimated as

$$\lambda=P(Ri=1 \mid \beta^{*}X), \quad \textbf{(2)}$$

where X is a set of model covariates and β are corresponding coefficients, and hard non-respondents are considered as missing at random.
The estimates of lambda could be used instead of PSU-level non-response adjustment. If PSU indicators are added to the set of covariates then the weights would also compensate for PSU-specific effects.

Prediction models.
An alternative way to non-response weight adjustment is to estimate prevalence among non-respondents. In order to do this we can fit a model to the outcomes among respondents and then use the estimated parameters to predict the prevalence among non-respondents. We want to stress that we were estimating the prevalence among "soft" non-respondents because they are believed to be the main source of bias.

Denote an outcome as Y and measured set of covariates as X. Then the fitted model for the outcome among respondents could be written as:

$$f(Y_{resp})=\gamma^{*}X_{resp}, \quad \textbf{(3)}$$

where γ is a set of corresponding regression coefficients and f is a link function. Then the predicted prevalence $\hat{P}$ among a non-respondent could be calculated as

$$\hat{P}_{nonresp}=f^{1}(\hat{\gamma}\ {}^{*}X_{nonresp}), \quad \textbf{(4)}$$

Where $\hat{\gamma}$ is a set of regression parameters estimated from (3) and the prevalence among non-respondents could be calculated as a weighted average over all non-respondents. The confidence intervals for the fit could be calculated using Wald statistic using the standard errors for the $\hat{\gamma}$ (Collett, 1994).

Model validation.
Prediction models could be validated externally and internally. If some outcomes are included in the non-response survey the propensity model could be validated by comparing the estimate and measured prevalence of such outcomes. Using models (3) and (4) one can estimate the prevalence among non-respondents and then compare it with the observed prevalence from non-respondent survey. Although more powerful, this external way is internally contradictive, i.e., it assesses validity only for the models of observed non-respondent outcomes that are already known and no model is needed to predict them, however it explicitly validates the model.

An internal way to validate the model would be useful for an outcome that is not reported and based only on respondents. First each subject is randomly assigned to a "training" and "validating" category. Then a model is fitted to the "training" dataset and its validity is checked on a "validating" dataset. Although this approach is weakened by using only partial data it is available for all outcomes.

Results.
Variable selection for propensity model.
Some households had responded within a short period of time, for others it took a while (and multiple interviewer's contacts) before they agreed to participate. We assume that those who responded later have more similar characteristics to the non-respondents than those who responded promptly. Thus, we have considered several models that would regress the response time continuous (actual number of weeks), binary (less or more than 13 weeks), ordinal (three categories <13, 13-26 and more than 26 weeks) on the

demographic, and outcome variables such as Child Behavior Check List questionnaire and psychiatric service availability. Using backwards variable selection algorithm we have ended up with a set of variables mostly associated with the delayed response:

- Not family or relatives living in the household: Adults 18+ and children <18, number of grandparents.
- How long child lives with parents.
- Moderate child health problems: somewhat overweight, sometimes stay in the hospital, sometimes have allergies, sometimes take medication.
- Parents' education: Mother's and father's education.
- Protestant religion.
- PCG's type (mother or grandmother, based on birth year).

However the religion and parent's education questions were not included into the non-respondent survey in order to hold the integrity of the survey and not to repel the respondents.

Variables included to validate the model regardless their association with delayed response: Insurance coverage, need for psychiatric services, medication use (ritalin). We have also included an indicator of late assessment because few households were assessed late in the year and the multiple contact of non-respondents was unavailable because of the schedule problems and therefore, few households that could potentially agree to participate were treated as soft non-respondents.

Observed prevalence.

We first present the prevalence of few major outcomes measured among Early Respondents, Late Respondents and Soft non-respondents.

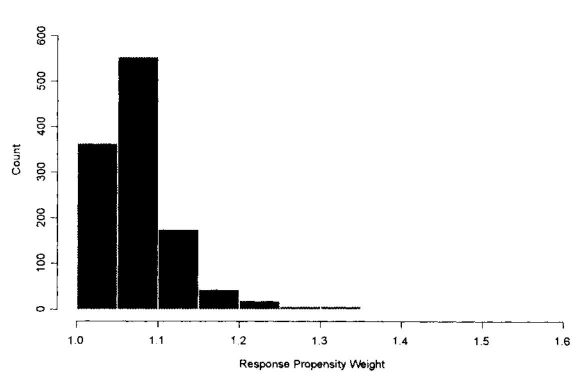
Table 3. Prevalence (%) Of Some Major Outcomes Among All Respondents, Late Respondents, And "Soft" Non-Respondents.

Variable	Early Resp.	Late Resp.	"Soft" Non-Resp
n	1,138	136	102
Insurance coverage	86.6	90.9	90.4
Ritalin use	2.4	2.9	5.8
Any Medication use	18.2	13.4	22.1
MH services needs	10.2	11.0	4.0

Response propensity model

In Figure 1. we present the distribution of the response propensities based on the model that included all the covariates. The distribution is quite smooth showing no extreme cases.

Figure1. Distribution of interview response propensity weights. These weight were calculated as inverse model-based probabilities of response.

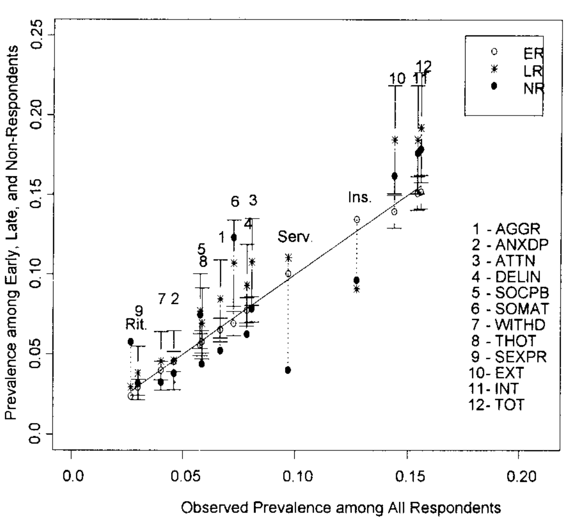


Prediction model.

Measures of Child Psychopathology, based on Achenbach's scale (Achenbach, 1991) were major outcomes of the model. These outcomes were binary variables representing clinical cases of aggression, anxiety/depression, attention deficit, delinquent behavior, social phobia, somatic disorder, withdrawal syndrome, thought problem, sex problems, as well as aggregated externalizing and internalizing problems and total problem score. For each of the outcomes we have constructed a model using the set of key variables. Because we believe that "soft" non-respondents are close to the late respondents we added an indicator of late response to the estimation model that was based on all respondents. The binary (0,1) indicator was set up such that late respondents were coded as (0) so this indicator was omitted when predicting the prevalence among non-respondents. Because the prevalence of the outcomes was quite low (few percent), having nearly fifteen variables makes the model nearly saturated and some associations could not be estimated. In order to remove such variables we used a backwards selection algorithm to remove the variables with corresponding p-values higher than 0.5. This left the modes with estimatable parameters. The summary of estimated prevalences among non-respondents is presented in

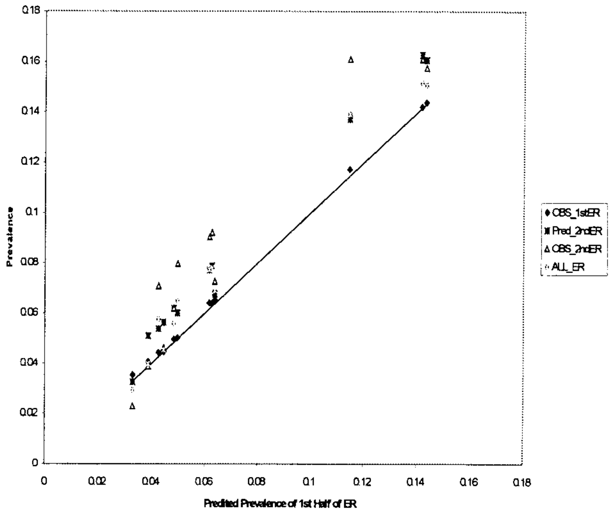
Figure2. In Figure2. we also present standard errors for the prevalence among late respondents to check if the estimated non-respondents' prevalence was significantly different. Although the model-based approach seems to be a good way to estimate the bias, the sample size and the prevalence are too small to detect significant difference between the estimated non-response and the observed prevalence. As a validity check we also plotted the observed prevalences of the variables reflecting Insurance coverage, Ritalin use, and the need for mental health services.

Figure2. Observed and model-based predicted prevalence of major study outcomes. The model was fitted to all respondents and then prediction were made for soft non-respondents. In addition we present observed prevalence of insurance coverage (Ins.), Ritalin use (Rit.) and service needs (Serv.). A solid line is added as a reference for the comparison of the observed prevalence among all respondents and predicted prevalence among soft non-respondents.



We have validated the models internally when the half of the data was used as a fitting set and another as a validation set. the results are presented in Figure3. The validation results show a good prediction capability of the models, although the predicted values still tend to be biased towards the fitted values.

Figure3. Internal validation of the prediction models. We present observed and predicted prevalence from "Training" and "Prediction" subsets (denoted 1st. and 2nd sets respectively) of Early respondents (ER). We also present a straight line corresponding to the predicted prevalence in the training set. Due to the rounding error the observed prevalence in "training" set is not exactly on the prediction line



Discussion.

The conditions that call for non-response adjustment are perhaps typical for most household surveys. Based on adjustment methodology described and used previously we have developed an approach that reflects the specificity of the studied survey. The critical assumption that was made about the studied population is that hard non-respondents are similar to early respondents and "soft" non-respondents are similar to the late respondents. It is hard to validate such assumption because no information is available on hard non-respondents and ethical considerations do not allow to reassess this population again. However, this assumption has been supported by previous research when the assessment of hard core non-respondents was possible. We have conducted a short non-response survey based on the variables that are suspected to be mostly associated with delayed response and "soft" non-response. Using the results of this small survey we have constructed models for propensity of being a soft refusal and also estimated the prevalence of the outcomes among soft non-respondents. It turned out however that when the prevalence of the outcomes is quite small and there is not enough data to base a model on, a model free approach such crude PSU-based weight adjustment could perhaps work well enough.

References

1. D. Collett, Modelling Binary Data. Chapman and Hall, 1994, London.
2. R.M. Groves and M.P. Couper. Nonresponse in Household Surveys, Wiley-Intersciense, N.Y., 1998.
3. C. Proctor, Extrapolating on day returned to adjust for mails survey non-response, JASA proceedings, 579-583, 1997