ADJUSTING FOR NONCOVERAGE OF NONTELEPHONE HOUSEHOLDS IN THE NATIONAL IMMUNIZATION SURVEY

Michael P. Battaglia, Abt Associates Inc.; Donald J. Malec, National-Center for Health Statistics; Bruce D. Spencer, Northwestern University; David C. Hoaglin, Abt Associates Inc. ; Joseph Sedransk, Case Western Reserve University

Contact Author: Michael P. Battaglia, Abt Associates Inc., 55 Wheeler Street, Cambridge, MA 02138

KEY WORDS: Random-digit dialing, simple poststratification modified poststratification, model-based estimation

1. Introduction

The National Immunization Program is currently engaged in an effort in 78 areas throughout the United States to raise immunization coverage rates among young children. These 78 areas consist of either whole states or sub-state areas, including the District of Columbia. In order to provide baseline estimates of immunization rates for children 19 through 35 months of age and to monitor ‘change in these rates, the National Immunization Survey (NIS) formerly known as the State and Local Area Immunization Coverage and Health Survey, is being conducted in each of the 78 Immunization Action Plan (IAP) areas. Beginning with the second quarter of 1994 and continuing through the fourth quarter of 1997, the data collection effort uses telephone sampling and interviewing methods to conduct independent quarterly surveys in each of the 78 IAP areas. The sample sires are large enough that four consecutive quarters of survey data can be combined to provide annualized estimates of the coverage rates for five antigens (DTP, Polio, MMR, HiB, and Hepatitis B) within *each* of the 78 IAP areas.

As described in more detail by Ezzati-Rice *et al.* (1995), the data collection methodology and sample design of the NIS use list-assisted randomdigitdialing (RDD) methods to sample households for computer assisted telephone interviews (CATI). The RDD approach makes it possible to screen for households containing one or more age-eligible children, and to conduct the required number of interviews for the identified age-eligible children in each quarter. A key disadvantage of the RDD sampling approach is that it gives children residing in nontelephone households a zero probability of selection. Although one can use an RDD sample to generalize to the population of age-eligible children in telephone households in a straightforward manner, the NIS aims to generalize to the entire population of age-eligible children residing in households in each IAP area.

The estimates obtained will be a function of the estimation techniques used and two other IAP-specific factors: (1) the proportion of age-eligible children residing in nontelephone households, and (2) the difference between age-eligible telephone and nontelephone children with respect to vaccination rates. In the United States as a whole, about 12% of age-eligible children reside in nontelephone households. The noncoverage rate, however, varies considerably among IAP areas, ranging from a low of about 2% to a high of about 25%.

At a national level the Immunization Supplements to the 1992 and 1993 National Health Interview Survey (NHIS) indicate that 19- to 35-month-old children in telephone versus nontelephone households have vaccination rates that differ by a considerable amount. For example, 59% of telephone children had received four or more doses of DTP vaccine, three or more doses of Polio vaccine, and one or more doses of MMR vaccine, and were therefore considered up-to-date, whereas only 41% of nontelephone children were up-to-date. More generally, nontelephone children tend to have lower rates of vaccination than telephone children.

The substantial numbers of nontelephone households in many IAP areas and the large differences between telephone aud nontelephone children's vaccination rates indicate considerable potential for noncoverage bias in several IAP areas. Any candidate estimation technique for the NIS must recognize this potentially large bias and attempt to adjust for differences between the telephone and nontelephone groups. We are studying three estimation techniques: simple poststratification, modified poststratification, and a model-based approach. Common to all three techniques is an initial assignment of base sampling weights and an adjustment for unit nonresponse.

The first step in the weighting methodology assigns to each telephone number in an IAP area a base sampling weight, equal to the reciprocal of the probability of selection. Because the probability of selecting a household in an RDD sample is proportional to the number of distinct telephone numbers associated with the household, we divide a child's base weight by the number of nonbusiness voice-use telephone numbers in the child's household (up to a maximum of three). Unit nonresponse can occur at several stages in the interviewing process, and the amount of information available about a nonresponding phone number varies with the stage. We apply a simpliied weighting-class adjustment methodology, dividing the population into subgroups and applying multiplicative factors to the weight, for three categories of noninterviews: residential status of the telephone number unknown, household reached but nothing more known, and eligible household reached but nonresponse questionnaire not completed. Thus each of the three estimation methods starts with the nonresponse-adjusted base weights.

Sections 2, 3, and 4 discuss simple poststratification tion, modified poststratification, and model-based estimation, respectively. Section 5 compares the resulting estimates of immunization coverage and considers some areas for further research.

2. Simple Poststratification

A widely-used estimation technique for an RDD sample is to poststratify the nonresponse-adjusted base sampling weight. That is, the weighted distribution of the completed interviews is brought into agreement with the population control totals for a specific set of poststratification cells. For the present survey, the population control totals come from National Center for Health Statistics natality data files. The general idea behind poststratification is to select variables that are related to unit nonresponse and/or noncoverage and are correlated with the key subject-matter variables. The natality data include variables such as date of birth, race of mother, Hispanic origin of mother, and education of mother. These variables, as demonstrated in an analysis of the 1992 and 1993 NHIS, are related to noncoverage and vaccination status.

Simple poststratification assumes that the percentage vaccinated within each poststratification cell is the same for both telephone and nontelephone children. The NIS produces several estimates of vaccination coverage. Most can be characterized as uptodate estimates; that is, a child is considered uptodate if he or she has received at least a specified number of vaccinations of a particular type. One of the primary estimates of vaccination coverage is the percentage of children that are 4:3: 1 up-to-date (i.e., 4 or more DTP, 3 or more Polio, 1 or more MMR). Our analysis of the 1992 and 1993 NHIS indicated that the percentage of 19- to 35-month-old children that are 4:3:1 up-to-date is lower for non-telephone children than for telephone children within the categories of the potential poststratifiers. This suggests that simple poststratification techniques will be only partially successful in eliminating noncoverage bias.

Simple poststratification requires population control totals that correspond to variables collected in the survey questionnaire. These variables include state and county of residence, race of child, Hispanic origin of child, age of child in months, race of mother, Hispanic origin of mother, and education of mother. The NCHS natality data are the best source of the desired population control totals.

The NIS has been used to form estimates of vaccination coverage for Q2, Q3, and Q4 of 1994 combined. The vital-statistics data for each set of vaccination coverage estimates are defined by choosing the range of dates of bii that corresponds to children 19 to 35 months of age as of the midpoint of the estimation time period.

Estimation for Q2, Q3, and Q4 of 1994 used poststratification cells based on the education of the mother, the race/ethnicity of the mother, and the age of the child. The number of cells varied among IAP areas.

Use of the natality data file to form the required population control totals has some limitations. First, the natality file provides a universe of live bii, and therefore it does not reflect a reduction in the population of children from infant mortality. To adjust for infant mortality, state-specific infant mortality rates by race group are applied to the vital-statistics data.

Second, the natality file will not reflect children born outside the United States who immigrate to this country before reaching the age of 19 to 35 months. This immigration increases the population size of children, and the effect is likely to vary considerably from IAP area to IAP area. The Public Use Microdata Samples (PUMS) from the 1990 Census are used to estimate the number of two-year-olds in each state who were born outside the United States.

Third, the natality file records state, county, and city of residence at time of birth. Children may move from one IAP area to another by the time they reach 19 to 35 months of age. The average annual interstate migration rate for children in the one- to four-year-old age group is 3.6%. The PUMS and vital-statistics data are used to estimate in-migration and out-migration in each IAP area.

Using the adjustment procedures described above, the distribution of births is tabulated for the poststratification cells in each IAP area. These counts provide the population control totals used for Q2, Q3, and Q4 of 1994.

3. Modified Poststratification

The analysis of the 1992 and 1993 NHIS indicated that the relationship between telephone ownership and the various uptodate vaccination coverage variables cannot completely be accounted for by individual-level demographic and socioeconomic variables or by county-level demographic, socioeconomic, and health-care-related variables. This implies that, within the poststratification cells developed from the natality data file, the up-to-date vaccination rates differ between telephone and nontelephone children. It is therefore likely that the simple poststratification technique will be only partially successful in reducing noncoverage bias. Is there a way to achieve additional bias reduction within the poststratification framework? At a national level, the NHIS Immunization Supplement (an in-person interview survey) can provide estimates of vaccination rates for telephone and nontelephone children for the various poststratification cells. This information can in turn be used to split each poststratification cell into two subcells: one representing up-to-date children, and the other representing children who are not up-to-date. Poststratification then can be used to adjust the weights of the children within these subcells. The resulting weights should yield some further reduction of noncoverage bias.

For a given IAP area, the natality data file yields the number of children in poststratification cell g. This control total, N_g can be split into estimated telephone and nontelephone totals by applying (from the 1990 Census) P_g, the proportion of children in poststratification cell g that reside in telephone households. From N_g and P_g one obtains the estimated number of children in nontelephone households, $N_{g0} = N_g(1 - P_g)$, and the estimated number of children in telephone households, $N_{g1} = N_g P_g$.

Using the NIS sample of telephone children and the 1992 and 1993 NHIS national samples of telephone children, the ratio of the two up-to-date rates on the 4:3: 1 vaccination series was formed:

$$r_{21_g}^{(431)} / r_{11_g}^{(431)},$$

where $r_{21_g}^{(431)}$ is the IAP-specific NIS up-to-date rate for telephone children in poststratification cell g and $r_{11_g}^{(431)}$ is the corresponding NHIS national uptodate rate.

The NHIS also yields the national up-to-date rate for nontelephone children in poststratification cell g, $r^{(431)}$. With this information, one can develop an estimate tof the number of children that are 4:3:1 up-to-date in poststratification cell g of each IAP area:

$$\hat{N}_g^{(431)} = \hat{N}_{g1} r_{21_g}^{(431)} + \hat{N}_{g0}(r_{21_g}^{(431)} / r_{11_g}^{(431)}) r_{10_g}^{(431)}.$$

Its complement, $N_g - \hat{N}_g^{(431)}$, equals the estimated number that are not up-to-date. Thus, splitting each poststratification cell into two subcells allows the children to receive a poststratified weight that is a function of whether they are up-to-date. If certain poststratification cells within an IAP area tend to have a higher than average proportion of nontelephone children and the nontelephone children have a lower 4:3: l-up-to-date rate than telephone children in that cell, the not-up-to-date children in that cell will receive a higher poststratification weight adjustment than the children who are up-to-date.

Some algebraic manipulation gives the modified poststratification adjustment factor in the up-to-date subcell,

$$P_g + (1 - P_g)(r_{10_g}^{(431)} / r_{11_g}^{(431)}),$$

and a somewhat more complicated expression in the not-uptodate subcell.

Although the modified poststratification technique offers the potential for a greater reduction of noncoverage bii, it has two major limitations. First, it relies on the ratio of national up-to-date estimates from the 1992 and 1993 NHIS for the poststratification cells in all IAP areas. These national ratios may not apply to the individual IAP areas; but the NHIS is not large enough to produce the up-to-date ratios at, say, the Census Region level. Second, the approach takes only one survey vaccination measure, 4:3:1-up-to-date, into account in forming the subcells. Other vaccinations such as HiB are not incorporated into the adjustment.

4. Model-Based Estimation

The NHIS is too sparsely spread over the IAP areas to allow for direct estimates of the relationships of the vaccination rates for nontelephone to telephone children in each individual IAP area. Instead, a statistical model can take into account the characteristics of the individual children in the NHIS and also allow for geographic variation not accotmted for by those characteristics. Ultimately the model can be applied to the data from the individual children in the NIS. In this way the model-based approach works with the data at a finer level of detail than is possible with simple or modified poststratification.

Logic of the Model

Within poststratification cell g we cross-classify the population of age-eligible children according to whether the household has a telephone and whether the child is up-to-date on a particular vaccination, v:

$N_{g1}^{(v)}$	vaccinated telephone children
$N_{g0}^{(v)}$	vaccinated nontelephone children
$\tilde{N}_{g1}^{(v)}$	unvaccinated telephone children
$\tilde{N}_{g0}^{(v)}$	unvaccinated nontelephone children

Considering vaccinated and unvaccinated children separately, we use the NHIS data on telephone and nontelephone children to develop models that allow us to estimate $N_{g0}^{(v)}$ and $\tilde{N}_{g0}^{(v)}$ from the NIS data.

We illustrate the role of the model for vaccinated children. For child i with nonresponse-adjusted base sampling weight W_i, the strategy is to adjust W_i to reflect the overall probability that a vaccinated child is selected into and participates in the NIS. From an extensive list of child-level and county-level variables that are available in the data from both the NHIS and thee NIS, we use the NHIS data to develop a logit model for T(X), the probability that a vaccinated child with characteristics x (the variables in the model) resides in a telephone household. (We assume that $\tau(\mathbf{x})$ is the same at the time of the NIS as in the NHIS.) A vaccinated child with characteristics $\mathbf{x}_i$ has (estimated) probability $\hat{\tau}(\mathbf{x}_i)$ of residing in a telephone household and probabiity $1-\hat{\tau}(\mathbf{x}_i)$ of residing in a nontelephone household. Here $\hat{\tau}(\mathbf{x}_i)$ is the result of evaluating t(x) from the NHIS logit model at $\mathbf{x}_i$, the characteristics of child i in the NIS. The probability that child i is selected into and responds to the NIS, given that he or she resides in a telephone household, is $1/W_i$. Thus the overall probability is $\hat{\tau}(\mathbf{x}_i)/W_i$, and taking the reciprocal yields the weight $W_i/\hat{\tau}(\mathbf{x}_i)$.

For unvaccinated children a parallel development produces an adjusted weight in the same form. Then, by applying a common multiplicative factor to the adjusted weight for each NIS child in poststratification cell g (vaccinated and unvaccinated), we bring the weighted sample total into agreement with the population control total for cell g. This step implicitly yields estimates of

$$N_g^{(v)} = N_{g1}^{(v)} + N_{g0}^{(v)} \text{ and } \tilde{N}_g^{(v)} = \tilde{N}_{g1}^{(v)} + \tilde{N}_{g0}^{(v)} .$$

Implementation

Because telephone status at the individual level is dichotomous, a binomial logit model is appropriate for the 1992 and 1993 NHIS data. Specifically, the random-effects logit model is

$$\ln\{\tau_i/(1-\tau_i)\} = \mathbf{x}_i\beta + \delta_{c(i)},$$

where c(i) is the county for child i, τ_i is the probability that the household of child i has a telephone, $\mathbf{x}_i$ is a (row) vector of individual-level and county-level variables (which may include indicators for demographic groups), and the δ_c are independently and identically distributed as $N(0,\sigma_\delta^2)$. For counties represented in the NHIS data, the estimated county effects, $\hat{\delta}_{c(i)}$, allow for geographic variation beyond that accounted for by the variables in $\mathbf{x}_i$. The variance σ_δ^2, when estimated, allows one to assess the error in applying the nontelephone adjustment to all IAP areas.

Building the Actual Model

In developing estimates from the NIS data for Q2, Q3, and Q4 of 1994, the modeling was based on one vaccination measure, 4:3:1-up-to-date, and on three demographic groups that combined the age of the child and the education of the mother. A single model (with some indicator variables and interactions) emerged as adequate for the full data set, combiig vaccinated and unvaccinated children in all three demographic groups.

The randomeffects model resulted from first building a fixed-effects model and then using the same explanatory variables in a random-effects model that incorporated county effects. Approximate maximum-likelihood estimates for β, the $\delta_{c(i)}$, and σ_δ^2 are obtained by identifying the $\delta_{c(i)}$ as missing values and using the EM algorithm (see Stiratelli et al., 1984). The estimate of σ_δ^2 was small; hence, the additional variation in the estimate from local effects was assumed to be negligible.

Calculating Model-Based Estimates

For each variable in the random-effects model, the definition in the NIS data is essentially the same as in the NHIS data. Then, for each child in the NIS, the estimate of $\tau_i = \tau(\mathbf{x}_i)$ is

$$\hat{\tau}_i = \exp\{\mathbf{x}_i\hat{\beta} + \hat{\delta}_{c(i)}\}/[1 + \exp\{\mathbf{x}_i\hat{\beta} + \hat{\delta}_{c(i)}\}].$$

(When the child's county, c(i), is not represented in the 1992 or 1993 NHIS data, the maximum-likelihood estimate is $\hat{\delta}_{c(i)} = 0$.) Multiplying the child's nonresponse-adjusted base sampling weight, W_i, by $1/\hat{t}_i$ yields the model-based weight $W_i/\hat{t}_i$. This step adjusts the weight of a child in a telephone household to account for similar children in nontelephone households. Within each of the 78 IAP areas, the above model-based weights are summed within each poststratification cell and then adjusted so that their sum matches the control total derived from vital statistics.

5. CONCLUSIONS

If no NHIS estimates of immunization status were available, simple poststratification would be our preferred method of adjusting for noncoverage of nontelephone households. Simple poststratification generally reduced the estimated immunization levels, with an average decreasing from 56.6% to 55.7% or .9% (see Table 5.1). Some few IAPs had their estimates increase, whereas others had varying amounts of decrease. The root-mean-square (RMS) sire of the change was 1.3% (see Table 5.2).

The availability of the NHIS estimates for both telephone and nontelephone households allows us to use modified poststratilication. The modified poststratification lowered the simple poststratified estimate for each IAP. After the modified poststratification, the average change from the unpoststratified estimate was a decrease of 2.22, and the RMS size of the change was 2.5%.

The modified poststratification would suffice for national estimates, but it fails to reflect local variability in the up-to-date ratios for telephone and nontelephone households. Use of model-based weights was intended to account for that variability. Indeed, the model-based weights changed the overall level of the immunization estimates only slightly, with the mean over IAPs a scant .1% higher than it was under the modified poststratified weighting. As expected, however, many individual IAP estimates changed by larger amounts. The standard deviation (over IAPs) of the change in estimate was .7% when the weighting was changed from simple poststratification to modified poststratification or when the weighting was changed again from modified poststratification to model-based weighting. These changes are fairly large in light of the standard deviation of the IAP estimates, which was 4.7% or 4.8% under any one of the weightings (Table 5.1).

It might seem surprising that the model-based weighting did not introduce more IAP-to-IAP variation in the estimates (compared to modified poststratification), because the model-based weighting employed additional IAP-level and county-level covariates for making the coverage adjustments. A possible explanation is that, although the model-based adjustments should increase the variability of the expected values of the IAP estimates for nontelephone households, they will not necessarily increae the variability of the overall estimates for the IAPs. Assume that the modified poststratified and model-based 4:3: 1 estimates cover the full population of telephone and nontelephone children in an IAP area, and that the simple poststratified estimates cover only the telephone population. Knowledge of the proportion of children 19 to 35 months of age in the IAP area allows one to solve for the implied nontelephone modified poststratified 4:3: 1 estimate and the implied nontelephone model-based 4:3: 1 estimate. The results shown in Table 5.3 indicate that the coefficient of variation of the nontelephone model-based 4:3:1 estimates is 53% higher than the CV of the nontelephone modified poststratified 4:3: 1 estimates.

Future research on this front will attempt to improve the logit models underlying the model-based weighting, and to extend the modeling to vaccinations other than the 4:3:1 series. Each model-based adjustment is specific to the particular vaccination for which the model was fit. It may be useful to fit the individual models for each vaccination, derive the model-based weights, and use them to form alternative marginals for the IAPs. One could then attempt to rake a single set of model-based weights, such as the 4:3:1 model-based weights discussed here, to match the marginal distributions. Alternatively, one could construct a model for the 4:3:1 series that also includes indicators for other vaccinations.

References

Ezzati-Rice, T.M., Zell, E.R., Battaglia, M.P., Ching, P.L.Y.H., and Wright, R.A. (1995) "The Design of the National Immunization Survey." *1995 Proceedings of the Section on Survey Research Methods* Alexandria, VA: American Statistical Association.

Stiratelli, R., Laird, N., and Ware, J. (1984) "Random-, Effects Models for Serial Observations with Binary Response," *Biometrics* 40, 961-971.

Table 5.1 Alternative Estimates of Vaccination Rates for 4:3:1 UTD for IAPs, Q2 to Q4 of 1994

Weighting Method	Summaries of Alternative IAP Estimates			
	Mean	**Median**	**SD**	**CV**
No poststratification	**56.6%**	**56.8%**	**4.7%**	**8.3%**
Simple poststratification	**55.7**	**55.7**	**4.8**	**8.6**
Modified poststratification	**54.4**	**54.6**	**4.7**	**8.6**
Model-based weighting	54.5	54.7	4.8	8.8

Table 5.2. Changes in Estimates of Percent 4:3:1 UTD under Different Weightings

Weighting Methods	Summaries of Differences between Alternative Estimates		
	Mean	**SD**	**RMS size**
Simple poststratification - no poststratification	-0.9%	1.0%	1.3%
Modified poststratification - no poststratification	-2.2	1.2	2.5
Model-based weighting - no poststratification	-2.1	1.2	2.4
Modified poststratification - simple poststratification	-1.3	0.7	1.5
Model-based weighting - modified poststratification	0.1	0.7	0.7

Table 5.3. Alternative Estimates of Nontelephone Children's Vaccination Rates for 4:3:1 UTD for IAPs, Q2 to Q4 of 1994

Weighting Method	Summaries of Alternative IAP Estimates		
	Mean	**SD**	**CV**
Modified poststratification	42.8%	4.6%	10.7%
Model-based weighting	43.6	7.2	16.4