# ADJUSTING FOR RESPONDENT BIAS ON VACCINATION STATUS IN A TELEPHONE SURVEY

**Elizabeth R Zell, Centers for Disease Control and Prevention (CDC); Trena M. Ezzati-Rice, CDC**
**David C. Hoaglin, Abt Associates, Inc.; James T. Massey, CDC**
**Elizabeth R Zell, National Immunization Program, MS-EOS,**
**Centers for Disease Control and Prevention, Atlanta, Georgia 30333**

**Key Words: Response Error, Validation Study**

## 1. Introduction

One of the Healthy People 2000 objectives is to have at least 90% of 2-year-old children fully vaccinated with the recommended schedule of vaccines. Timely estimates of vaccination levels for children 19 to 35 months of age are needed to monitor progress in these levels. The National Immunization Survey (NIS) is being conducted in 78 Imrmmization Action Plan (IAP) areas, consisting of the 50 states, the District of Columbia, and 27 metropolitan areas, to obtain timely quarterly estimates of vaccination coverage. Using the same methodology in each IAP area, this survey aims to produce rates of vaccination coverage that are comparable among IAP areas.

Beginning with the second quarter of 1994, the NIS data collection effort involves independent quarterly surveys in each of the 78 IAP areas. This design, described in more detail by Ezzati-Rice et al. (1995), makes it possible to combine four consecutive quarters of survey data to provide annualized estimates of the coverage rates for nine antigens (diphtheria and tetanus toxoids and pertussis vaccine DTP], poliovirus vaccine [polio], measles mumps and rubella vaccine [MMR], *Haemophilus influenzae type* b vaccine [Hib], and hepatitis b vaccine [Hep B]) within each of the 78 IAP areas with an acceptable degree of precision.

To locate households with one or more children 19 to 35 months of age, a quarterly random-digit-dialed (RDD) sample of telephone numbers is selected from each IAP area. If a sample household is eligible, the interviewer collects information on the vaccinations received by all age-eligible children. Although the interviewers urge the respondent to refer to the child's immunization record or "shot card" if one is available (and are prepared to schedule a call-back to facilitate use of a shot card), only about 45% of respondents use a shot card. Even when available, the shot card may not show all the vaccinations that the child has received. Without a shot card the complexity of the recommended vaccination schedule in the first two years of life makes it difficult for a respondent to recall the child's vaccination history accurately. Thus the households' reports of vaccination status are subject to potentially large response bias.

To assess such biases, the National Immunization Provider Record Check Study (NIPRCS), part of the 1994 National Health Interview Survey (NHIS), collected vaccination information for children 19 to 35 months of age from their providers. Provider reports, in most cases, are considered to be an accurate measure of the vaccinations actually received by children. Preliminary NIPRCS results suggest that household reports often contain errors (CDC, unpublished data 1995). Accordingly, the NIS includes a second-phase sample of providers to improve the accuracy of the vaccination coverage estimates. Interviewed households are asked to give the name and address of their child's health care provider(s), and to give verbal consent for contact with the provider(s), although some households are unable or unwilling to furnish the requested information. A mail survey of the identified providers collects vaccination information on the children. The provider survey thus yields vaccination information for a subsample of children identified in the first-phase (RDD) sample.

## 2. Design and Data Collection

The provider record check study for the NIS was developed and implemented in three phases during data collection for the fourth quarter of 1994 (Q4/94). In the first phase, a small pilot was conducted to assess the feasibility of obtaining provider data from previously interviewed respondents in a recontact interview. The feasibility study was conducted during December 1994, using 700 households that had completed the immunization questions during the RDD interview. This study demonstrated that provider information and verbal consent could be collected in a telephone interview, and, therefore, a full-scale data collection effort began immediately.

In the second phase of household data collection for the provider study, respondents who had completed an RDD immunization interview prior to the implementation of the provider study during Q4/94 were recontacted retrospectively to obtain information on the providers of their children's vaccinations and consent to contact the named providers. Because of time constraints, children with a household report from

a shot card that documented receipt of 4 DTP, 3 polio, 1 MMR, and 3 Hib were excluded from this phase of the study. After the implementation of the provider component into the NIS, the third phase collected provider data prospectively (as part of the RDD interview) during Q4/94 from eligible households completing an RDD immunization interview. In both the retrospective and prospective phases, respondents were asked to permit CDC to access the vaccination records of providers of vaccinations for all eligible children and to provide enough identifying information on the respondents and their children to facilitate access to the records.

The information from the households was subsequently used to contact the providers of vaccinations for the children in the RDD sample. Written requests for vaccination information were then submitted to the named providers in order to obtain reports of vaccinations from medical records. The data collection for the provider record check study was conducted by mail, and providers had the option of responding via mail or facsimile. The provider data were key-entered and edited, and the resulting data file was compared with the household data.

Sample Size Considerations

Because the schedule for producing the 1994 estimates left only limited time for the provider study, an initial analysis considered the minimum number of RDD cases that needed to be recontacted to obtain provider data. The provider record check study falls under the general area of models for the use of verification information. Anderson *et al.* (1979, Chapter 10) discuss the use of verification information for all participants in a survey and describe a two-phase sampling approach that collects the verification information from only a subsample of the survey participants. Cochran (1977, Chapter 12) details the two-phase sampling approach from the viewpoint of using the first-phase sample to develop the stratification variable for the second phase of sampling. For example, in the RDD survey the immunization status of the child and the household's use of a shot card can serve to stratify the second-phase subsample of children for the provider record checks.

To determine the minimum acceptable sample size for the provider study, estimates were developed for the standard error and coefficient of variation (CV) at the IAP level, assuming a sample size of 50 completed provider vaccination forms per IAP area, using the formula for the variance of the adjusted proportion from a two-phase sample. The initial second-phase sample size of households drawn from the RDD interview file had to be larger than the desired number of completed provider forms to account for: (1) inability to obtain provider identification information and/or verbal permission to contact the provider, (2) provider identification information that is insufficient or inaccurate for contacting the provider, and (3) refusals by providers to fill out the second-phase data collection instrument.

As the outcome variable, this analysis used the proportion of children who were up-to-date on the 4:3: 1 series of vaccinations (at least 4 doses of DTP, at least 3 doses of polio, and a least 1 dose of MMR), guided by preliminary results from Quarter 1 of the NIPRCS. The households' responses (from the first phase of the survey – in the NIPRCS this was the 1994 NHIS) led to the creation of four strata:

- shot card, up-to-date;
- shot card, not up-to-date;
- no shot card, up-to-date (includes "ALL");
- no shot card, not up-to-date.

On the basis of the preliminary NIPRCS data, a decision was made to set the estimate of the proportion (of children whose providers reported them as up-to-date) for the fmt stratum equal to 1.0 (that is, if the household reported, from a shot card, that a child was 4:3:1' up-to-date, it was assumed that the child's provider(s) would report the child as 4:3: 1 up-to-date). This assumption made it unnecessary to allocate sample to the first stratum. For the other strata, the entire 1994 RDD sample was subject to sampling for the provider study.

The standard error of the proportion from the second phase of a two-phase sample will generally be larger than the standard error of the estimate from the first-phase sample because the second-phase sample is smaller than the first-phase sample (e.g., 50 versus 400 cases per IAP area per year). This calculation, however, does not take into account the potential bias in tie unadjusted estimates from reporting error in the survey. The mean squared error of the adjusted estimate may very well be considerably smaller than the mean squared error of the unadjusted estimate, because the adjusted estimate will be unbiased with respect to reporting error (assuming the provider data are accurate). For a sample size of 50 completed provider record check forms per IAP area, the coefficient of variation was estimated to be in the range of 6% to 10%. This is considerably larger than the original design specification for the RDD survey of 2.5% (5% CV at the 95% confidence level), but it ignores a possibly larger bias from reporting error that is very likely to be associated with the estimates

without provider data. Viewed in this light, the provider-adjusted estimates will have a coefficient of variation that is considerably smaller than the estimates without provider data, even though the latter will be based on about 400 sample children, versus 50 second-phase observations.

In order to minimize the coefficient of variation of the provider-adjusted 1994 estimates, given the considerations outlined above, it was determined that the sample size for the provider study should be as large as cost and time would permit. In addition, the up-to-date restriction (i.e., the subset of children, whose report was based on a shot card, for whom it would be assumed that the provider(s) would report them as up-to-date) was changed to exclude from sampling only those cases who were 4:3:1:3 up-to-date (4:3:1 up-to-date and at least 3 Hib), rather than 4:3:1 up-to-date. Thus the final design attempted to obtain provider data for all RDD cases from Q2/94,Q3/94, and Q4/94 for whom complete immunization data were obtained in the initial RDD interview and who were not 4:3: 1:3 up-to-date according to shot records during the interview.

## 3. Estimation of Response Bias

To assess the respondent error, provider data are assumed to be the gold standard (i.e., correct in determining the true vaccination status). Respondent error is defined as the difference in the proportion of children who are up-to-date in their vaccinations as defined by the provider data versus the household respondent report. This information can be summarized by the gross difference rate, which is the percentage of inaccurate household reports when treating the provider reports as truth. The overall impact of the respondent error is best summarized by the net difference rate, which is the percentage point change in the vaccination coverage estimate.

Respondent reports of vaccination information can be obtained from in-home shot cards or from recall. In-home shot cards are believed to provide the best estimate of vaccination coverage for an eligible child. Ideally, the shot card includes the date of each vaccination, but it may not be correct if it is not taken to the vaccination provider at each visit Problems may also arise if the respondent provides the dates from the card of a non-eligible child in the household. Prior to the second birthday a child should receive 11 to 15 doses of vaccine. Household respondents tend to underestimate the number of doses for multiple-dose vaccines and to overestimate coverage for single-dose vaccines (Goldstein *et al.* 1993, Valadez and Weld 1992). To understand respondent bias, it is best to evaluate vaccination coverage based upon the source of reported information (e.g., shot card, recall). The gross and net difference rates serve as an excellent summary measure of the degree of response bias for the two sources of household reported vaccination information.

## 4. Provider Study: Nonrespondents

For the data collection period April through December 1994, 25,247 age-eligible children participated in the survey. Of these, 6,768 (27%) reported receiving 4 DTP, 3 polio, 1 MMR, and 3 Hib from an in-home shot card. The remaining 18,479 (73%) were eligible for the provider study. Consent was obtained from 11,204 (61%) of the eligible respondents with an average of 1.2 providers per child. Vaccination information was received for 7,594 (68%) of these respondents (range: 49 to 139 per IAP area). Of those eligible for the provider study, 10,885 (59%) did not have any provider data available for use in the estimation procedures among respondents eligible for the provider study.

It is important to determine the comparability of those with and without provider data. Basic demographic characteristics reported during the telephone interview were compared at the national and IAP level. Nationally, whites and those with a higher education were slightly overrepresented among those with provider data. However, comparing the telephone report of vaccination status, those with and without provider data had almost identical vaccination coverage levels at both the national and IAP level. These similarities suggest that applying the results of the provider reports to all children will not introduce a bias.

## 5. Adjustment for Response Bias

The procedure for combining household and provider data to produce provider-adjusted estimates of vaccination coverage involves three steps:

1. Categories of household respoves are formed based on the availability of a shot card (yes, no) and the response to vaccination status on the 4:3:1 combination (e.g., up-to-date on 4:3:1, don't know). The resulting response categories (e.g., up-to-date on 4:3: 1, reported as numbers of shots from a shot card; missing on 4:3:1) each contain adequate provider data.

2. Adjustment factors are calculated using the provider data as the gold standard. Within each response category, the adjustment factor for each vaccination (or combination) is the proportion of

children in the provider sample who, according to their providers, are up-to-date on that vaccination.

3. The adjustment factors are applied to the entire NIS sample. The estimation process multiplies the adjustment factor by the number of children in that response category (in the NIS as a whole) to produce an estimate of the number of children in that category who are up-to-date. Summing these numbers over the set of 4:3: 1 response categories and dividing by the total number of NIS children yield an overall estimate of the proportion of children who are up-to-date on the particular vaccination.

Table 1 illustrates the calculation of the rate of vaccination coverage for the 4:3:1 series, using data from the two-phase sample for the United States as a whole (the details of the calculation for each of the 78 IAP areas are similar). As the table indicates, the actual calculations are somewhat more complicated than the steps outlined above. Mainly, the numbers and proportions of children are weighted, because each child in the NIS sample receives a weight, as discussed by Battaglia et al. (1995).

Also, on the first line for each vaccination or combination ("Shot Card, up-to-date on 4:3: 1" in Table 1) the factor is calculated, not from the data for each individual IAP area, but from the combined data from all IAP areas (on the other lines the factor for an JAP area uses only data from that IAP area). The reason for this special calculation lies in the design of the provider study. Initially it was assumed that essentially all children whose telephone reports, when based on a shot card classified them as 4:3: 1:3 up-to-date (at least 4 DTP, 3 polio, 1 MMR, and 3 Hib) would also be reported by their providers as up-to-date. Thus, reports generally were not sought from the providers of these children. However, some limited provider data for such children indicated that a small percentage of them were actually not up-to-date. These data, together with provider data for children whose telephone reports (based on a shot card) classified them as 4:3: 1 upto-date but not 4:3: 1:3 up-to-date, form the basis for a national adjustment factor. The sample sizes in the individual IAP areas are too small to support IAP-level adjustment factors on the first line. Beginning with the first quarter of 1995, the provider study is collecting data on such children. The use of a national adjustment factor for this group of children is appropriate because the proportion reported as up-to-date by providers should vary little from IAP area to IAP area.

In Table 1 the telephone responses for 4:3:1 comprise five categories. For each of these categories the succeeding columns give the actual (unweighted) count of cases in the provider study, the total weight associated with those cases, the total weight associated with cases that are up-to-date (4:3:1 series), the (weighted) proportion up-to-date, the (unweighted) count of NIS cases (i.e., children in the RDD survey), the total weight associated with those children, and, finally, the total weight for NIS cases estimated as being up-to-date for the 4:3: 1 series. The "Total" line contains column totals, and the line for the "Adjusted Estimate" shows the ratio of the entries on the "Total" line in the two rightmost columns.

More specifically, the 401 children from the provider study on the "Shot Card, up-to-date on 4:3 : 1" line have a total weight that indicates they represent 76,278 children, of whom 69,253 are considered up-to-date by provider reports. The ratio of these total weights, 69,253/76,278, yields 0.896. The 7,460 children who were "Shot Card, up-to-date on 4:3: 1" in the RDD survey represent 1,596,254 children. Multiplying this by 0.896 yields an estimated 1,429,817 children who are up-to-date on the 4:3: 1 series.

On the second line of Table 1, 1,616 children in the provider study were not up-to-date on 4:3:1 (as reported from a shot card in the RDD survey). Their weighted proportion uptodate is 0.583. Applying this factor to the total weight of the corresponding 3,685 children in the RDD survey yields an estimate of 503,371 children who are considered up-to-date.

Similar calculations on the remaining four lines lead to the totals of 5,523,451 and 4,128,620 in the rightmost two columns and thus to the estimate that, for the United States as a whole, 74.7% of children 19 to 35 months of age are upto-date on the 4:3: 1 series.

To assess the validity of the reported vaccination levels at the national level, findings were compared with data from the NHIS (CDC). For January-June 1994, NHJS data were supplemented with provider information in the same manner as the MS. Differences in antigen-specific and the 4:3:1 series coverage levels between the two surveys ranged from 0 to 3 percentage points (Table 2).

## 6. Conclusions

Monitoring vaccination levels at the national, state, and local level is an important public health function. Accuracy of reported vaccination levels is essential for the usefulness of the results. Respondent error is the largest source of error in a survey evaluating vaccination coverage levels for children 19 to 35 months of age across the United States. These errors arise from in-home shot cards that often omit

vaccinations actually received and from poor recall of the correct number of doses a child has received. GveralL a gross difference rate of 35.4% was observe 26.4% for those reporting from an in-home shot card and 45.0% for those relying on recall. Use of provider data increased the national 4:3: 1 coverage level by 21 percentage points. With respondent error far outweighing sampling error, routine provider verification of self-reported vaccination data in surveys of childhood vaccination should be an integral component to accurately assess vaccination levels.

The NIS has shown that a telephone survey can provide valid estimates of vaccination levels when provider data are combined with household data. The NIS yields current, population-based, state-specific estimates of vaccination coverage from a standard methodology. As an ongoing survey the NIS will provide timely and routine state-specific and national vaccination levels for continued monitoring and improvements in an important national health program.

References


Anderson, R, Kasper, J., Frankel, M.R, and Associates (1979). *Total Survey Error. San* Francisco: Jossey-Bass.

Battaglia, M.P., Malec, D.J., Spencer, B.D., Hoaglin, D.C. and Sedransk, J. (1995). "Adjusting for Noncoverage of Nontelephone Households in the National Immunization Survey." *Proceedings of the Section on Survey Research Methods* Alexandria, VA: American Statistical Association.

CDC (1995). "State and National Vaccination Coverage Levels Among Children Aged 19-35 Months -- United States, April-December 1994." *Morbidity and Mortality Weekly Report* 44,6 13-623.

Cochran, W.G. (1977). Sampling Techniques, third ed. New York: John Wiley & Sons.

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

Goldstein, K.P., Kviz, F.J., and Damn, RS. (1993). "Accuracy of immunization histories provided by adults accompanying preschool children to a pediatric emergency department." *Journal of the American Medical Association,* 270,2 190-2 194.

Valadez, J.J. and Weld, L.H. (1992). "Maternal recall error of child vaccination status in a developing nation." *American Journal of Public Health,* 82, 120-122.

Table 1. Number and Percent of NIS Vaccination Reports That Are Provider Verified as Up-to-date for the 4:3:1 Series, with Estimated Total NIS Cases, 1994-Q2/Q3/Q4 - U.S. Total

| | Provider Records | | | | Total Sample | | |
|---|---|---|---|---|---|---|---|
| 4:3: 1 Response Category | Unweighted Total Count | Weighted Total Count | Weighted Verified Up-to-date | Proportion Verified Up-to-date | Unweighted NIS CaseS | Weighted NIS CaseS | Population Total Up-to-date on 4:3:1* |
| Shot Card up-to-date | 401 | 76,278 | 69,253 | 0.896 | 7,460 | 1,596,254 | 1,429,817 |
| Shot Card not up-to-date | 1,616 | 375,605 | 219,117 | 0.583 | 3,685 | 862,868 | 503,371 |
| No Shot Card up-to-date | 2,229 | 450,531 | 352,326 | 0.782 | 5,538 | 1,125,096 | 879,852 |
| No Shot Card not up-to-date | 2,030 | 444,104 | 297,311 | 0.669 | 5,167 | 1.185.030 | 793,334 |
| Missing | 1,261 | 268,138 | 185,672 | 0.692 | 3,397 | 754,203 | 522,247 |
| Total | 7,537 | 1,614,657 | 1,123,679 | | 25.247 | 5.523.451 | 4.128.620 |
| Adjusted Estimate | | | | | | | 74.7% |

*Adjustment based on proportion verified up-to-date from provider records.

Table 2. Vaccination Coverage Levels among Children 19-35 Months, by Selected Vaccines, United states, 1994

| | NHIS* | | NIS+ | |
|---|---|---|---|---|
| Vaccine/Doses | Percent | 95% CI' | Percent | 95% CI |
| DTP/DT | | | | |
| >= 3 Doses | 93 | (90.8,95.2) | 93 | (92.3,93.7) |
| >= 4 Doses | 76 | (72.6,79.4) | 77 | (75.9.78.1) |
| Poliovirus | | | | |
| >= 3 Doses | 83 | (80.0,86.0) | 83 | (82.0,84.0) |
| *Haemophilus influenzae type* b | | | | |
| >= **3** Doses | 89 | (86.4,91.6) | 86 | (85.1,86.9) |
| Measles-Mumps-Rubella | 88 | (84.2,91.8) | 89 | (88.1,89.9) |
| Combined Series | | | | |
| 4 DTP/3 Polio/l MMR | 72 | (68.6,75.4) | 75 | (73.8,76.2) |

* 1994 National Health Interview Survey, January-June.
+ 1994 National Immunization Survey, April-December.
fi Confidence interval.