IMPROVING THE ASSESSMENT OF VACCINATION COVERAGE RATES WITH THE USE OF BOTH HOUSEHOLD AND MEDICAL PROVIDER DATA

Trena M. Ezzati-Rice, Elizabeth R. Zell, James T. Massey, Centers for Disease Control and Prevention (CDC)
Mary Gessley Nixon, Westat, Inc.
Trena M. Ezzati-Rice, National Center for Health Statistics, CDC, 6525 Belcrest Road, Hyattsville, MD 20782

Keywords: Response bias, validation study

1. Introduction

Vaccination histories are obtained from household respondents as part of the ongoing National Health Interview Survey (NHIS), a major household survey conducted by the National Center for Health Statistics, Centers for Disease Control and Prevention (CDC). The NHIS is a cross-sectional probability based survey of the U.S. civilian non-institutionalized population. In-person interviews are conducted each week throughout the year. For 1994, data were collected on about 147,000 persons in approximately 49,000 households. One sample child is randomly selected from households with at least one child less than 6 years of age, and data on vaccinations are collected in a special Immunization Supplement. In addition, beginning in 1994, all children 19-35 months of age were selected for the Immunization Supplement. The vaccination data are collected from household respondents and shot records may or may not be consulted. When a shot card is available, the interviewer abstracts the specific dates for five types of vaccines--diphtheria and tetanus toxoids and pertussis (DTP) or diphtheria and tetanus toxoids (DT), poliovirus vaccine (polio), measles-mumps-rubella (MMR) or measles vaccine, *Haemophilus influenzae* type B vaccine (Hib), and hepatitis B vaccine (Hep B). The respondent is also asked to report any additional vaccinations received by the child, but not recorded on the shot card. If a shot card is not available, the interviewer asks the respondent to recall from memory the number of shots, but not the dates for each vaccine. The questionnaire also allows for an unsolicited response of "all" as the number of shots received for a particular vaccine. In the first two calendar quarters of the 1994 NHIS, approximately half of the respondents referred to a shot card to report vaccinations received, while the other half reported from memory (recall). Due to the large proportion of respondents who rely on recall to report their child's vaccination history and concerns about the accuracy of reported vaccination histories even when a shot card is used, the validity of the vaccination coverage level estimates from the NHIS is a major issue. Therefore, beginning in 1994, a medical provider record check study was conducted to improve the accuracy of the national vaccination coverage estimates derived from the NHIS household survey.

This paper describes the process of determining "best" vaccination values for each sample child 19-35 months of age in the NHIS using provider-based and household-based reports of vaccinations. Comparisons of the "best value" and household-based estimates are made, and net and gross difference rates for individual vaccines and combinations are also presented.

2. Response error in a household survey assessing vaccination coverage levels

The Childhood Immunization Initiative, launched by President Clinton in 1993, was established to increase vaccination coverage among 2 year old children in the United States (CDC, 1994). Therefore, producing scientifically valid estimates of vaccination coverage levels is critical to the monitoring of progress towards increasing coverage levels among 2 year old children. It is generally recognized that a household respondent may not always be the best source of medical or health related information. Response errors in surveys can arise for many reasons, including misunderstanding the questions, memory recall errors, prestige bias, and respondent and/or interviewer errors. If the errors are systematic, then the estimates from the survey may be seriously biased.

Previous studies by Goldstein *et al.* (1993) and Valadez and Weld (1992) indicate that parents tend to underestimate the number of doses received for multiple-dose vaccines and to overestimate coverage for single-dose vaccines. Record check studies have been widely used to investigate individual response bias and assess net bias for characteristics from surveys. Therefore, beginning in 1994, a medical provider record check study, the National Immunization Provider Record Check Study (NIPRCS), was conducted in conjunction with the NHIS to evaluate the quality of the household-based reports of vaccinations and to improve the accuracy of the survey estimates. A more detailed description of the NIPRCS is given by Peak and Cadell (1996).

Concerns about data quality in the NHIS Immunization Supplement focus on two different types of response errors -- under-reporting and over-reporting. If a shot card is not used, 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. Further, although one might hypothesize that with the use of a shot card, the accuracy of reporting should increase, especially over that from memory or recall, this will only be true if the respondent's records are up-to-date (i.e., show all the vaccinations that the child has received, legible, and accurate). In addition, interviewer error may result in the mis-recording of vaccines received. For example, a polio vaccination may be listed on the shot record as "pink drops" or "OPV" which may not get recorded by the interviewer as a polio vaccination. Over-reporting can occur with the mis-recording of information by the interviewer or if the wrong child's record was obtained by the household respondent at the time of the interview.

The NIPRCS was designed to reduce the bias associated with national vaccination coverage level estimates derived from the NHIS household-based data. As described by Peak and Cadell (1996), after completion of the NHIS Immunization Supplement questionnaire, the names and addresses of the immunization providers for all 19 to 35 month old children are solicited along with written permission to permit CDC to access the vaccination records of the named providers and to provide enough identifying information on the respondents and their child or children to facilitate access to the records. The data collection for the provider record check was conducted by mail, and providers had the option of responding via mail or facsimile. The provider data were key-entered and edited, and then matched with the household-based data. The basic analytic approach used for the NIPRCS focuses on determining the "true" number of shots that each child has received for each vaccine using the provider-based and household-based reports. The number of shots for each vaccine obtained from the assessment of provider-based and household-based reports is defined as "best values". The process used to compare and reconcile the responses from household respondents and medical providers to determine "best" vaccination values is described in the next section. The data for this analysis are from quarters 1 and 2 of the 1994 NHIS and NIPRCS and consist of 1,230 children 19-35 months of age.

3. Determining "best" vaccination values

The best values for the number of doses of each vaccine that a child had received at the time of the NHIS interview were determined from a combination of provider-based and household-based reports. The household and provider data were matched to determine a best value for the number of doses each child had received for DTP, polio, MMR, Hib, and Hep B. After comparison of the household-based and provider-based reports, best values could be assigned with confidence for the majority of children. In some case, however, the situation was less clear. If after comparison of the household-based and provider-based reports, the reports from the two sources were different and the differences were appreciable, then the households, the providers, or both were recontacted to determine if the differences could be reconciled. This was referred to as the reconciliation process. In some cases, the reconciliation process led to the identification of new providers for the child or to corrections in the original reports of the number of doses of a vaccine. When new providers were identified, they were contacted and asked to provide a vaccination history for the child. All of these data were then used to determine the best values for each vaccine for all sampled children.

The initial process in the matching of the provider and household vaccination values was classification of the cases according to the degree of difference in the reports. The categories were based on whether the reports were identical, the reports had only minor differences (e.g., same number of dose but some different dates), the household reported more doses than the provider, the household reported fewer doses than the provider, or a complex set of differences existed involving discrepancies in both directions. The determination of cases for reconciliation and the best value for each vaccine was implemented in the following manner:

a. If a provider reported the child as 4313 up-to date (4 DTP, 3 polio, 1 MMR, and 3 Hib) and the dates were consistent with the child's date of birth and the household dates (where available), the provider data were taken as the best value.
b. If the provider and household matched all dates and number of shots, even if the child was not up-to-date, the provider information was taken and no further recontacts were made.
c. If the household reported from recall and reported "All" for each vaccine and the provider reported that the child was not 4313 up-to-date, but there was indication that the provider had provided medical care throughout the life of the child, then the provider information was taken and no recontacts were made.
d. If a case was fielded to a provider but no provider vaccination information was available (because of provider nonresponse, provider never caring for the child, or provider having no shot information), AND the child was not 4313 up-to-date from records according to the household report, the household was recontacted for NEW providers.
e. If a household reported more doses than the provider, the household was generally recontacted for additional providers, especially if it appeared that another provider had probably administered vaccinations for the child.
f. If the provider report was inconsistent or discrepant in some way, the provider was recontacted. For example, since DTP is often given in combination with Hib, if the provider reported four DTP shots, but only two Hib

shots as administered on the dates of the first two DTP shots, the provider was recontacted to determine whether Hib 3 and Hib 4 were also given along with DTP 3 and DTP 4.

g. If the household reported "Don't Know" for all immunizations, and the provider reported that the child was not up-to-date, no definite rule was applied. Generally, reconciliation recontact was made with either the household or provider or both.

For a very small number of cases, these general rules could not be applied. In these cases, a decision was made after reviewing the specific pattern of doses reported and comparing them with the recommended pattern of vaccinations for a child's age.

4. Results

Of 1,342 children 19 to 35 months of age in the NHIS for the first two quarters of 1994, an Immunization Supplement was completed for 1,230 children (92%) and provider data were obtained for 852 of these children (69%). Among the 852 children with provider data, 771 were assigned best values based on data from the initial provider reports, reconciliation providers, and followup providers. For these 771 children, referred to as Set A, the assignment of the best value was made with a high degree of certainty. The remaining 81 cases with at least some provider data, but no best values, are referred to as Set B. Rather than treating the Set B cases as "missing best values" since some provider and household data were available, best values were assigned although the assignment was not as certain as it was for the Set A cases. These 81 Set B cases were examined individually to assign the best values. For cases where the household respondent consulted records and reported more doses than the provider, the best value was set to be the household report if it appeared that some of the providers identified by the household had not responded. For cases where the household did not consult records, the provider reported number of doses was generally accepted as the best value. For a few cases, the household-based and provider-based reports were combined to determine the best value. Overall, the best values were the same as the provider reported values for 65 of the 81 Set B cases. Because the best values for the Set B children were assigned with less information and less certainty than the Set A cases, two additional Set B values were assigned to evaluate the robustness of the assignment process -- a conservative value (lower bound value) that was the smaller of the household and provider value and a liberal value (upper bound value) that was the larger of the household and provider value for each vaccine.

Due to the lack of provider data, "best" values were missing for 31% of the children. Hot-deck imputation was used to assign values for the missing best values. The imputation method used in NIPRCS is described in more detail by Nixon *et al.* (1996).

To estimate the response error in the household-based reports of vaccinations, the household based up-to-date vaccination coverage levels are compared with the best value estimates including lower and upper best value estimates treating the best values as "truth". Net and gross difference rates are used to determine the measurement error in the household responses. The definition of these difference rates are illustrated by reference to the table below:

Basic record matrix for comparison of survey interview responses and true (or "best") values

Best value	Household report		Total
	Up-to-date	Not up-to-date	
Up-to-date	a	b	a+b
Not up-to-date	c	d	c+d
Total	a+c	b+d	n = a+b+c+d

The gross difference rate (gdr) is the percentage of erroneous household reports (treating the best value as the truth), that is,

$$gdr = 100 \times b+c/n$$

The erroneous household reports can be of two types, "b" understating, and "c" overstating the number of vaccinations.

The net difference rate (ndr) estimates the net response bias in the household-based estimate of the percentage up-to-date. The household-based estimate of the percentage up-to-date is 100(a+c)/n, and the best value estimate is 100(a+b)/n. The net response bias is the difference between these two estimates. The ndr is thus:

$$ndr = 100 \times c-b/n$$

Table 1 presents the percentage of children up-to-date for single vaccines and combinations. The estimates in column (1) are based on only the NHIS household-based reports. The weighted vaccination coverage levels from the household reports include an adjustment for Immunization Supplement nonresponse. All the Don't Know (DK) or missing household responses were excluded prior to computing the vaccination coverage levels. A key assumption is that the DK responses are missing completely at random. If the vaccination levels for the children with DK responses are different from those with valid responses, then the estimates of the percentage of children that are up-to-date will be biased. The best value estimates are given in column (2). These estimates were calculated using the best values assigned to all children for whom some provider data were available (i.e., the 852 children in Sets A and B) and imputed best values for the remaining 378 children for whom no provider data were obtained. The lower and upper bound best value estimates are given in columns (3) and (4).

For the individual vaccines, the best value estimates when compared to the household-based estimates show increases ranging from 5 percentage points for polio to 16 percentage points for Hib. In the other direction, there were 2 and 8 percentage point deceases for MMR and Hep B, respectively. For the 431 (4 DTP, 3 polio, 1 MMR) combination, the estimated vaccination coverage level increased from 67 percent to 74 percent, while the 4313 (4 DTP, 3 polio, 1 MMR, 3 Hib) combination increased from 59 percent to 72 percent. For the majority of the vaccines and for the combinations, the best value estimate of the percentage up-to-date is, as expected, somewhere between the lower and upper bound estimates and the range between these estimates is not particularly large. For DTP, the range between the lower and upper bound estimates is 5 percentage points and only 2 percentage points for Hep B. These results suggest that the assignment rules used for the Set B cases did not have a large effect on the best value estimates.

To determine the response bias and variance in the household-based estimates, the best value estimates are compared with the household-based data. Specifically, net and gross difference rates, as described earlier, are used to assess the accuracy of the NHIS household-based estimates. Table 2 shows the net and gross difference rates for each individual vaccine and combinations. The net difference rates for some of the vaccines and combinations exceed 25 percent of the coverage level estimate. For example, the net difference rate for Hep B is 7.2 percent or 26.3 percent of the household-based estimate of 27.4 percent. The household-based reports are generally substantial underestimates of the actual level of vaccination for DTP, polio, and the 431 and 4313 combinations. For Hep B, the household-based report is substantially higher than the actual (best) values. The gross difference rates for the percent of children up-to-date for the single vaccines and the combinations were computed excluding all the DK household responses and the observations that were imputed. It was thought that the inclusion of the imputed values would lead to a somewhat overestimated level of response error since in producing the imputed best values, account was taken of the household-based vaccination reports through imputation classes that separated the children into those that were and those that were not up-to-date on the 431 combination. The estimates in Table 2 suggest that about 20 percent of the children are misclassified as not being up-to-date for most of the vaccines -- the exception being MMR, a single dose vaccine.

Although about 20 percent of the 19-35 month old children are misclassified with respect to being up-to-date on their vaccinations for most of the vaccines, the misclassifications are not symmetric. This is why the net response bias estimates shown in Table 2 are relatively large. For some vaccines and combinations, the result is that the household-based estimates are substantially less than the best value estimates. For Hep B, the household-based estimate is biased in the opposite direction. In virtually all cases, except MMR, the biases are large and the bias for Hib is especially large. This may be because Hib is often administered in combination with DTP and household respondents may not be aware that both antigens were administered.

Collecting information from respondents about their children's vaccination providers, contacting and collecting data from providers, comparing the provider-based and household-based reports and recontacting any providers or households with obvious discrepancies, assigning best values for each child with provider data, and then imputing for children with no provider data is an intricate and time-consuming process. To produce more timely provider-adjusted vaccination coverage estimates, a simple ratio adjustment method has been developed that uses only the responses from the initial provider contacts, that is, before any reconciliation. This approach divides the sample into adjustment cells based on two variables: 1) whether or not records were consulted for the household report of vaccinations, and 2) the number of doses reported by the household with the DK and missing responses as a separate cell. The ratio of the weighted number of provider reports of up-to-date children to the weighted number of children with provider data is computed for each cell. This ratio is then multiplied by the total (weighted) number of NHIS Immunization Supplement children in a given cell to compute the estimated number of children who were up-to-date in a given cell. The provider- adjusted estimate of the number of up-to-date children is then the sum of the estimates over all the cells. The estimates computed using this weighting class ratio adjustment method are called the *initial provider-adjusted* estimates.

Table 3 shows the household-based, best value, and the initial provider-adjusted estimates. The initial provider-adjusted estimates are closer to the best value estimates than the household-based estimates. The differences between the best value and initial provider- adjusted estimates are all small, most being less than 2 percentage points.

5. Summary

The magnitude of the net and gross difference rates indicate that the NHIS household-based reports of vaccinations are subject to a variety of response errors. The response errors are not one-directional; household respondents under-report some vaccinations and over-report others. However, the errors generally result in the vaccination coverage level for specific vaccines and for combinations being biased downward. The underlying reasons for the errors are likely complex. In another paper by Zell *et al.* (1996), different subgroups are examined to determine if certain subgroups of the population are bigger contributors to the errors than others.

The initial provider-adjusted method of estimating the percentage of children who are up-to-date can be applied many months before the reconciliation can be completed and best value estimates computed. The biases of the initial provider-adjusted estimates are relatively small compared to the best value estimates, therefore this approach appears to be a reasonable one for producing advance estimates. This finding is important since initial provider-adjusted estimates are also used in a large scale random-digit-dialing survey, the National Immunization Survey, to improve the accuracy of the estimates produced from that survey. The initial provider-adjusted estimates and the best value estimates are both clearly more accurate than the same estimates produced from just the household-based reports. The results from this study show that household-based reported vaccination information without provider verification has a high degree of response error. This error may be a much larger contributor to the error associated with reporting of true vaccination levels than even the sampling error. Thus, provider verification should become a routine data collection methodology in all household-based surveys assessing vaccination coverage levels.

References

1. CDC. (1994). "Reported vaccine-preventable diseases--United States, 1993, and the Childhood Immunization Initiative." *Morbidity and Mortality Weekly Report* 43:57-60.
2. Goldstein, K.P., Kviz, F.J., Daum, R.S. (1993). "Accuracy of immunization histories provided by adults accompanying preschool children to a pediatric emergency department." *Journal of the American Medical Association* 270:2190-2194.
3. Valadez, J.J., Weld, L.H. (1992). "Maternal recall error of child vaccination status in a developing nation." *American Journal of Public Health* 82:120-122.
4. Peak, R.R. and Cadell, D.M. (1996) "Overview of the National Immunization Provider Record Check Study." *1996 Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
5. Nixon, M.G., Kalton G., Brick, J.M., Ezzati-Rice, T.M., Zell, E.R. (1996) "Variance estimation with missing best values in the NIPRCS." *1996 Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
6. Zell, E.R., Ezzati-Rice, T.M., Massey, J.T., Brick, J.M. (1996) "Response errors associated with household reports of immunization - analysis of subgroup differences." *1996 Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Table 1. Vaccination coverage levels among children aged 19-35 months by alternative methods of estimation: National Health Interview Survey, Quarter 1 and Quarter 2, 1994

Vaccine (Doses)	Household	Best value	Conservative value	Liberal value
Single antigens				
DTP (4+)	68.8	75.4	73.3	78.5
Polio (3+)	77.9	83.0	81.0	85.7
MMR (1+)	90.6	88.7	87.1	90.3
Hib (3+)	73.0	89.0	86.5	90.0
Hep B (3+)	27.4	19.0	19.0	21.1
Combined Series				
4 DTP/3 Polio/ 1 MMR	66.8	73.8	71.6	76.8
4 DTP/ 3 Polio/ 1 MMR/ 3 Hib	58.5	72.2	70.2	74.5

Table 2. Net bias, net difference rates, and gross difference rates for single antigens and combinations-comparison of household and best value estimates: National Health Interview Survey, Quarter 1 and Quarter 2, 1994

Vaccine	Household estimate	Net bias	Net difference rate	Gross difference rate[1]
Single antigens				
DTP	68.8	-6.6	-7.6	22.3
Polio	77.9	-5.1	-6.5	18.6
MMR	90.6	+1.9	+1.0	9.5
Hib	73.0	-16.0	-16.8	19.8
Hep B	27.4	+8.4	+7.2	17.8
Combinations				
431	66.8	-7.0	-8.3	22.6
4313	58.5	-13.7	-16.1	27.7

[1]Percentages are based on weighted data. Observations with household response of don't know and observations with imputed values were eliminated from the computations.

Table 3. Percentage up-to-date for household, best value, and initial provider adjusted estimates for single antigens and combinations: National Health Interview Survey, Quarter 1 and Quarter 2, 1994

Vaccine	Household	Best value	Initial provider adjusted
Single antigens			
DTP	68.8	75.4	74.3
Polio	77.9	83.0	81.9
MMR	90.6	88.7	86.7
Hib	73.0	89.0	88.2
Hep B	27.4	19.0	17.8
Combinations			
431	66.8	73.8	72.1
4313	58.5	72.2	70.9