ACCURACY OF VACCINATION DATES REPORTED BY IMMUNIZATION PROVIDERS IN THE NATIONAL IMMUNIZATION SURVEY

Meena Khare (National Center for Health Statistics, CDC), Michael P. Battaglia (Abt Associates), Vicki J. Huggins (Abt Associates), Shannon Stokley (National Immunization Program, CDC), David C. Hoaglin (Abt Associates), Robert A. Wright (National Center for Health Statistics, CDC), Ann-Sofi Roden (Abt Associates)

Meena Khare, National Center for Health Statistics, Centers for Disease Control and Prevention (CDC), 6525 Belcrest Road, Room 915, Hyattsville, MD 20782

Key Words: RDD, Immunization history, Vaccination coverage, Best vaccination value

1. Introduction

Immunizations are among the most effective public health interventions to prevent disease and death. Currently, vaccine-preventable diseases are at their lowest level in U.S. history (1, 2). The availability of a large, high-quality, and reliable population-based immunization database provides a unique opportunity to monitor vaccination coverage. The CDC-sponsored National Immunization Survey (NIS) is designed to measure and monitor vaccine-specific coverage estimates for the nation and within each of the 78 Immunization Action Plan (IAP) areas, consisting of the 50 states, the District of Columbia, and 27 large urban areas. To monitor progress toward Healthy People 2000 and 2010 objectives, NIS data are used to produce timely estimates of vaccination coverage within each IAP area for each vaccine type and series of vaccines among children aged 19-35 months (3, 4).

Since April 1994, the NIS has been collecting vaccination history data on children aged 19-35 months in the U.S. The NIS is based on a two-phase survey design. The first phase uses a list-assisted random-digit-dialing (RDD) sample design to screen and select a sample of telephone households with age-eligible children. A computer-assisted telephone interview (CATI) is then administered to obtain demographic information, vaccination history, and name(s) of the immunization provider(s) for each eligible child. In the second phase, after obtaining consent from the parent or guardian, an Immunization History Questionnaire (IHQ) is mailed to the identified immunization provider(s). The mailed questionnaires collect data on the immunization history of the selected child from the medical records maintained by the providers. The ongoing quality control (QC) procedures ensure the validity and accuracy of the vaccination coverage estimates. Details of the history, sample design and quality control procedures of the survey are published in Zell *et al.* (3).

For the first time, CDC is planning to disseminate public-use data files (PUF) to allow other public health researchers to analyze the NIS data. While planning for the PUF, the quality of the data was evaluated to identify inconsistencies between the household-reported and the provider-reported dates. Two critical dates that potentially impact the eligibility of the children, the age at immunization, and the coverage estimates are the date of birth (DOB) and the date of immunization. The provider-reported dates were compared with the corresponding household-reported dates to determine the accuracy and completeness of the immunization histories. This paper provides a summary of the current QC procedures, the observed discrepancies in the reported DOB and vaccination dates, and the proposed enhancement procedures to reduce or eliminate these discrepancies.

2. Collection of Immunization Histories

CDC and its contractor (Abt Associates Inc.) have implemented continuous quality improvement procedures for the state-of-the-art automated CATI data collection and monitoring systems, data preparation, processing and editing systems, and a post-edit quality review system. The CATI system is a carefully designed data collection system with hundreds of built-in edits and validation procedures. In addition to recording the date of interview and the child's DOB, name, gender, and other socio-demographic information, it also obtains vaccination dates from the written vaccination record (i.e., 'shot card') or vaccination status from 'memory' recall for each antigen by a parent or guardian. The trained telephone interviewers enter the household information, immunization history, and contact information for the immunization provider(s) in a database.

The IHQ forms that are mailed to the immunization providers also collect the child's DOB and vaccination dates for each antigen. An office assistant or nurse practitioner generally completes the IHQ and returns it to the contractor by mail or facsimile. The returned medical records or IHQ forms from the individual provider first go to editors for manual scan-edit, visual verification of completeness of the reported DOB, name, and gender, and evaluation of legibility of the vaccination dates. Then these forms are forwarded for data entry. After data entry these records go to data processing for further recoding, back-coding, editing, imputation, and consistency checks. Vaccination dates from the IHQ forms are used to determine the number

of doses received for each recommended vaccine type (7) and to compute age at immunization for each shot. Currently, provider-reported immunization histories are considered the 'gold standard' and are used in calculating coverage estimates.

Given the large size of the NIS and the need to have timely analytic data available to produce estimates, the NIS relies primarily on automated procedures to edit and modify the provider data. The automated Provider Record-Check (PRC) edit software compares the data from the IHQ forms with the household-reported vaccination history. Often, children have more than one immunization provider, and some of them do not respond. This results in an incomplete provider-reported history. A set of disposition codes and more than 200 error flags summarize the amount of provider data obtained for each child, the household respondent's use of a 'shot card' or reliance on recall during the telephone interview, the agreement between the provider- and household-reported data, and the number of fields in error. Children with unreliable provider data are excluded, and only children with reliable or 'usable' provider data are included in analyses.

To assess quality of the immunization histories, vaccine-specific reports of up-to-date (UTD) status (e.g., 4:3:1:3 series of 4DTP/3Polio/1MCV/3Hib; for details see references 3, 4, or 7) from household 'shot cards' and 'memory' recalls are compared with provider reports. Selected measures including odds ratios (OR), sensitivity, specificity, gross difference rates, and net difference rates are computed to evaluate quality of the household data (3,5). Agreement on UTD status between household-reported 'shot card' data and the 'usable' provider data are also evaluated.

3. Observed Discrepancies in the IHQ

Because of the complexity of the recommended childhood immunization schedule (7), the potential for errors in the reported vaccination dates is very high; they may arise from transcription errors, transposed numbers/digits, or poor handwriting. Since the IHQ is self-administered, discrepancies in reported DOB and vaccination dates often occur. Occasionally, a report received from a provider has missing data for vaccinations administered to the child by other providers. Differences between the provider- and household-reported DOB may occur if: i) the household respondent gave an incorrect DOB, ii) the provider entered an incorrect DOB, or iii) the provider filled out the IHQ for a 'wrong' child. Therefore, the first step involves a visual scan of the returned IHQ forms by the editors to identify whether 1) a 'wrong' child has been reported, 2) the reported child is age-ineligible based on the provider-reported DOB, 3) the provider-reported DOB differs from the household-reported DOB by more than 14 days, or 4) one or more reported shot dates are before the child's DOB. Such inconsistencies are used to mark IHQ forms for follow-up evaluation.

Next, the automated PRC edit program combines the household and the IHQ data from one or more providers into a single child-level record and assigns a 'best' DOB. Currently, the provider-reported DOB is most often used as the 'best' DOB for the child.

Discrepancies in vaccination dates occur between household- and provider-reported histories as well as between two or more providers. Inconsistencies also occur in vaccine-specific dates within a single report (e.g., vaccination dates incorrectly recorded before the DOB or incorrect recording of calendar year makes two consecutive shots within 2 days.) The PRC edit program compares the vaccination dates and checks for inconsistencies in vaccination dates within and between vaccines, within and between providers, between household and providers, and also compares the age at immunization for each dose. It also checks for the number of days between two consecutive vaccine-specific shots and flags the dates that are less than 30 days apart.

Furthermore, some providers fail to report a date for the first Hepatitis B dose but check off the 'given at birth' box on the IHQ form. Usually, this shot is given at the hospital where the child was born; the hospital staff may have given the 'shot card' to a parent and the child's provider has no information. The sample distribution of age at the 'birth' dose revealed that it ranged from 0-7 days, and 91% of the shots are given within first two days of the DOB. Cases with missing provider-reported dates for the 'birth' dose of Hepatitis B (i.e., with a missing date of first dose within 7 days of the DOB) are flagged and compared with the household 'shot card' reported dates for completeness.

For all cases identified for a follow-up, the PRC edit program prints a specially designed *matching* sheet with data from all relevant sources for the child for a manual review. The contents of a *matching* sheet include the child's demographic information, PRC error flags, disposition code, and immunization histories as reported by the provider(s) as well as from the household respondent. Manual *matching* sheet reviews are extensively used in the NIS to compare and edit individual items in the immunization histories.

In addition, to determine sources of date-related data-entry errors, the NIS data editors manually reviewed a sample of cases with discrepancies and compared each data item from the original IHQ form with the items entered in the database. To continue this effort, the editors also computed the rate of data-entry error by blindly double-keying a sample of IHQ forms and compared the first data entry with the second for each field entered in the database. Results of these reviews are provided in the next section.

4. Results

A total of 36,338 households were identified with eligible children in the 1999 NIS, and 33,932 (93%) of them completed the household interview (i.e., the CATI). These 33,932 households contained 34,442 children aged 19-35 months. A majority of the household respondents were the child's mother or a female guardian. Consent was obtained to contact the child's immunization provider(s) for 28,936 (84%) children. Provider-reported 'usable' histories for 22,521 (65%) children were used in analyses. Approximately 83% of the 'usable' histories were obtained from single providers and the other 17% from two or more providers. Table 1 shows the distribution of children with completed interviews by 'usability' of the provider data and sources of the household reports. Approximately 49% of the household-reported immunization histories were obtained from a written 'shot card,' and the other 51% from 'memory' recall. A total of 11,964 children had 'usable' provider data and immunization histories from the household 'shot card.'

Table 2 shows the agreement on 4:3:1:3 UTD status between histories reported from the household 'shot card' and by the providers. Table 3 presents a set of selected measures to evaluate quality of the household-reported histories. Tables 2 and 3 show that for households that reported a child as 4:3:1:3 UTD from a 'shot card,' provider-reported history agreed more than 90% of the time (sensitivity >90%.) Conversely, when a household reported a child as not UTD for 4:3:1:3 from the 'shot card', the provider(s) disagreed more than 70% of the time (specificity <30%.) Similar agreements were observed in UTD status for individual vaccines. A comparison of odds ratios shows that when a household reported a child as 4:3:1:3 UTD, the UTD status from a provider-reported history was twice as likely to agree with the household 'shot card' (OR=3.9) than when the household reported from recall (OR=1.6). Table 3 also shows that the rates of disagreement, gross difference and absolute net difference were higher for the 'memory' recalls (75%, 61%, and 53%, respectively) than the 'shot card' reports (71%, 34%, and 23%, respectively). A comparison of the household 'shot card' and 'memory' recall reports revealed poor quality of the recall data and somewhat higher reliability of the 'shot card' data. The net difference in 4:3:1:3 coverage estimates shows a substantial underreporting (-23%) from the household 'shot card' reports. The 'shot card' utilization rates vary substantially by IAP area (ranging from 18-67%). Similar results were reported elsewhere by Zell *et al.* (3,5) and Ezzati-Rice *et al.* (6).

In the NIS, the provider-reported DOB is primarily used as the 'best' DOB. However, in the 1999 NIS, after the *matching* sheet review of inconsistent provider-reported DOBs, the household-reported DOB was used as the 'best' DOB for 65 (0.3%) children who had 'usable' provider data; household-reported DOB was also used for an additional 779 (3.4%) children who had missing provider-reported DOB and 'usable' provider data.

Among the children with 'usable' provider data, the PRC edit program also identified a total of 565 (2.5%) children with one or more discrepancies in dates. The *matching* sheet review resulted in deleting some or all of the provider-reported data for 50 children because the IHQ was completed for a 'wrong' child, and it removed the entire record for 23 children who were determined to be age-ineligible. Vaccination dates were edited for 127 children who had reported dates before their 'best' DOB, and for 256 children who had vaccine-specific reported dates 'close' to each other and dose-intervals within 30 days for at least one vaccine.

In addition, in the 1999 NIS a total of 545 (2.4%) children with 'usable' provider data had a missing date for the 'birth' dose of Hepatitis B, but the box 'given at birth' on the IHQ form was checked off by the provider. Among those 545 children, vaccination dates for the 'birth' dose were available for 231 children from the household 'shot card.' The 1999 NIS coverage estimates count these 545 children with missing 'birth' dose, as being not UTD for Hepatitis B (having fewer than three shots.)

To identify the sources of response or data-entry errors associated with provider reports, a sample (from July 1998 - June 1999 NIS) of 172 IHQ forms where some of the shot dates were entered prior to the DOB, was reviewed. Of the 218 individual errors found on the 172 forms, approximately 81% (176) of the errors occurred in the dates recorded on the original provider forms (e.g., transposed dates, missing month/day, or incorrect calendar year), 12% (26) of the errors were made by the editors in transcription or correction of faded/illegible dates on the forms, and the other 7% (16) of the errors were made in data-entry. To continue efforts to identify sources of data-entry errors, a few batches of the completed IHQ forms were re-sent to the data-entry contractor for a blind data validation. A total of 195 forms were sent in four batches for double data keying. Only 91 (0.36%) fields out of the potential 25,116 fields entered in the database were found to be in error. Based on these evaluations, to reduce data-entry errors, a corrective action was implemented to immediately call the provider's office to verify and correct the plausible errors in the IHQ forms prior to data entry.

5. Enhancements to the Current Quality Control Procedures

The overall goal of the QC procedures is to ensure the highest-quality data possible by reducing or

eliminating data discrepancies. The first steps in reducing the data discrepancies in the NIS included some fairly simple changes made to the clerical edit procedures at various stages of the data collection and data processing. The use of *matching* sheets in the NIS has significantly improved the quality of the immunization histories. Currently, *matching* sheets are being extensively used in the NIS to review, evaluate, and edit discrepancies in the immunization histories.

The National Immunization Provider Record Check Study (NHIS/NIPRCS) uses provider-reported immunization histories to improve the coverage estimates obtained from the Immunization Supplement of the National Health Interview Survey (8, 9). In addition to combining the household- and provider-reported immunization histories, the NHIS/NIPRCS also uses reconciliation and nonresponse follow-up procedures to improve coverage estimates. Given its large size, the NIS does not plan to use nonresponse follow-up or extensive callback reconciliation procedures with providers or households. The reconciliation procedure in the NIS is currently limited to those cases where a provider is suspected to have reported data for a 'wrong' child or some of the dates are illegible or inaccurate. Some of the procedures proposed in this section are direct enhancements to the current NIS procedures, and others are derived from the NHIS/NIPRCS. The NIS is planning to develop and implement editing, evaluation, and estimation procedures similar to the methods used in the NHIS/NIPRCS. A comparison of the methods used in the NHIS/NIPRCS and the NIS is described in Stokley *et al.* (10).

Since January 2000, all returned IHQ forms that are flagged by the editors for reconciliation are also compared with the household records before forwarding for data-entry. Also, IHQ forms observed with discrepancies in DOB or vaccination dates (as described in Section 3) or suspected to have reported data for a 'wrong' child are pulled, and the reporting provider is immediately called back to verify the information or to resubmit the IHQ form. The missing date of the 'birth' dose of Hepatitis B will be imputed with a valid household-reported vaccination date from the 'shot card,' or with a date within seven days of the 'best' DOB using the current sample distribution of the 'birth' dose, if and only if the provider had checked the 'given at birth' box on the IHQ. The remaining discrepant cases will be resolved with a manual *matching* sheet review, using a set of rules recommended by CDC experts who are familiar with the childhood vaccination schedule (7). The contents of the *matching* sheets have been significantly enhanced to compare and improve completeness and accuracy of the provider-reported vaccination history, specifically related to the shot dates that were reported before the child's DOB and/or vaccination dates that are too 'close' to each other (within 30 days).

The results of analyses show that despite all efforts to improve the data quality, immunization histories were incomplete for a small number of cases using data from any single source. Because agreement on immunization histories was higher than 90% among UTD children from the provider and household 'shot card' reports, it has been proposed to supplement the missing or incorrect 'usable' provider data with the household 'shot card' data to create a 'best' vaccination history. The NHIS/NIPRCS has previously used 'best' values only for the number of doses to create UTD status and to produce coverage estimates. The NIS is planning to modify the 'best' value methodology used in the NHIS/NIPRCS (10).

Three major categories were proposed in the NIS to construct the values for the 'best' vaccination dates: 1) cases where children are UTD from a household 'shot card' and not UTD from provider reports, 2) cases where provider reports are missing the 'birth' Hepatitis B dose, and 3) cases where children are UTD for a vaccine from the household 'shot card' report and all provider-reported dates are missing. The first category was further expanded to investigate whether a child would become UTD by combining incomplete immunization histories from the two sources. An example of this would be a child who had two valid DTP shots reported from the household, and the provider reported two more shot dates such that combining the four shot dates makes the child UTD for DTP. Finally, the 'best' DOB and the 'best' vaccination dates would be used to compute the 'best' age at immunization, the 'best' number of doses for each vaccine, the 'best' UTD status, and the 'best' age-appropriate immunization status for a child.

The changes in up-to-date status and the impact of the 'best' vaccination values on the coverage estimates will be carefully evaluated because 'shot card' use (18-67%) and missing provider data (52-77%) vary considerably among the IAP areas. Also, the use of 'best' vaccination values could substantially change the patterns of missing immunization histories. For example, in an IAP area where 'shot card' use is low, the impact could be very small. However, in areas where 'shot card' use is high and a large proportion of the provider data is missing, the impact could be significant. Preliminary results of the feasibility study using the 'best' vaccination values show that the national coverage of the 4:3:1:3 vaccine series increased by approximately 3.5 percentage points and by 0.3-6.8 percentage points across the 78 IAP areas (10) in the 1999 NIS.

6. Summary

Overall, a very small proportion of the children

with 'usable' provider data in the 1999 NIS were observed with date discrepancies. This confirms the high quality and reliability of the provider-reported immunization histories in the NIS. The underreporting in the household coverage estimates strongly supports the use of a provider record-check study to improve coverage estimates. Currently, provider-reported histories are used as the 'gold standard,' and therefore, accuracy of the coverage estimates depends on the immunization histories reported by providers. Most of the observed reporting errors in the provider data were due to transcription errors, poor handwriting, writing incorrect dates/years, or recording transposed numbers/digits in dates. Starting in January 2000, to improve completeness of the immunization history of Hepatitis B and to reduce bias in the coverage estimates, the missing date for the 'birth' dose in the provider report will be imputed with a valid vaccination date from the household 'shot card' or a date within 0-7 days of the child's DOB (using the distribution of the first dose), if and only if a provider had checked the 'given at birth' box on the IHQ.

The number of discrepancies increased when more than one provider immunized a child. An average of 1.35 providers immunized a child in the 1999 NIS, and 61% of the 565 children with discrepancies had two or more providers. Observed discrepancies between household and provider reports, and between multiple providers indicate that no single source maintained a 100% accurate or complete immunization history.

The results in this paper indicate that reliability of the household-reported histories was higher if the household respondent used a 'shot card' to report; and for more than 90% of the UTD cases, the status agreed with the provider reports. Provider reports were twice as likely to agree with the household reports when the child was reported UTD from a 'shot card' than when reported UTD from 'memory' recall. Usually, for a new patient, providers use the household 'shot card' as a reference to complete their medical record. These results and high agreement in the UTD status indicate that household 'shot card' information could supplement the missing or discrepant provider-reported immunization history in order to construct the 'best' vaccination values. The goal of constructing a 'best' vaccination value is to create a more complete immunization history for a child and therefore, to improve the accuracy of the vaccine-specific coverage estimates.

The NIS is planning to release the first public-use data file in late 2001. Prior to disseminating the public-use data file, CDC is planning to conduct a feasibility study to construct the 'best' vaccination values for all vaccine types. Coverage estimates will be compared with and without the 'best' vaccination value to assess the impact on the estimates. With 'best' vaccination values, the coverage estimates are expected to increase at the national and IAP levels. After completion of the feasibility study, a decision will be made whether to release the 'best' vaccination values in the data file.

Finally, CDC is also planning to conduct a feasibility pilot study in selected areas to match the NIS database with a state-based registry database (11). This study will help in evaluating the quality of the coverage estimates and overlap of the NIS children in the state-based registry database and vice versa.

References

1. CDC, "Achievements in Public Health, 1900-1999, Impact of Vaccines Universally Recommended for Children, United States, 1990-1998". Morbidity and Mortality Weekly Report, 48(12):243-248, 1999.
2. CDC, "Ten great public health achievements--- United States, 1900-1999". Morbidity and Mortality Weekly Report, 48(12):241-243, 1999.
3. Zell ER, Ezzati-Rice TM, Battaglia MP, and Wright RA. "National Immunization Survey: The Methodology of a Vaccination Surveillance system". Public Health Reports, 115:65-77, 2000.
4. CDC, "National, State, Urban Area Vaccination Coverage Levels among Children Aged 19-35 Months--- United States, 1999, Morbidity and Mortality Weekly Report, 49(26):585-589, 2000.
5. Zell, ER, Ezzati-Rice TM, Massey JT, and Brick MJ. "Response Errors Associated with Household Reports of Immunizations: Analysis of Subgroup Differences", 1996 Proceedings of the Section on Survey Research Methods, ASA, Alexandria, VA, pp 341-346.
6. Ezzati-Rice TM, Zell ER, Massey JT, and Nixon MG. "Improving the Assessment of Vaccination Coverage Rates with the Use of Both Household and Medical Provider Data", 1996 Proceedings of the Section on Survey Research Methods, ASA, Alexandria, VA, pp 335-340.
7. CDC, "Recommended Childhood Immunization Schedule, United States, January 2000- December 2000", http://www.cdc.gov/nip/recs/child-schedule.PDF, 2000
8. Peak RR, Cadell DM. "Overview of the National Immunization Provider Record Check Study", 1996 Proceedings of the Section on Survey Research Methods, ASA, Alexandria, VA, pp 332-334.
9. Massey JT, Moore TF, Parsons VL, *et al.* Design and estimation for the National Health Interview Survey, 1985-1994. Hyattsville, Maryland: US Department of Health and Human Services, Public Health Service, CDC, Vital and Health Statistics; series 2 (110), 1989.
10. Stokley S, Battaglia MP, Khare M, Daniels D, and

Ezzati-Rice TM. "Comparison of the NIS and NIPRCS Estimation Methods", to appear in the 2000 Proceedings of the Section on Survey Research Methods, ASA, Alexandria, VA, 2001.
11. Stevenson J, Luman B, Wright RA, Khare M, Canavan B, Hoaglin DC, Huggins VJ, and Strella R "The National Immunization Survey Registry Validation Study: Methodology for Matching and Analyses", to appear in the 2000 Proceedings of the Section on Survey Research Methods. ASA, Alexandria, VA, 2001.

Table 1: Usability and sources of immunization histories among children with completed household interviews, 1999 NIS*

Usability of provider-reported immunization history	Source of household-reported immunization history					
	Shot Card		'Memory' Recall		Total	
	n*	%	n*	%	n*	%
'Usable'	**11,964**	71.1	10,557	59.9	22,521	65.4
Not usable or not obtained	4,865	28.9	7,056	40.1	11,921	34.6
Total	16,829		17,613		34,442	100

*Number of children with completed household interview

Table 2: Agreement on 4:3:1:3* UTD status between 'usable' provider- and household-reported 'shot card' data, 1999 NIS

Provider-reported 4:3:1:3 UTD status	Household-reported 4:3:1:3 UTD status					
	HH UTD		HH not UTD		Total	
	n	%	n	%	n	%
Prov UTD	6,463	90.5	3,409#	70.7	9,872	82.5**
Prov not UTD	680#	9.5	1,412	29.3	2,092	17.5
Total	7,143**		4,821		**11,964**	100

Note: Prov=provider, HH=household, UTD=up-to-date; *4DTP/3Polio/1MCV/3Hib
**P(HH UTD)= 59.7%, P(Prov UTD)= 82.5%, and net difference rate=-22.8%; # Gross difference rate= 34.2% (see Table 3)

Table 3: Selected measures of agreement on the 4:3:1:3* UTD status between household- and provider-reported immunization histories by source of household reports, 1999 NIS

	Source of household-reported 4:3:1:3 UTD status	
	Shot Card	'Memory' Recall
Children with 'usable' provider-reported history (n=22,521)	11,964	10,557
Odds Ratio (Prov UTD): Odds of (Prov UTD| HH UTD)/ Odds of (Prov UTD/ HH not UTD)	3.9	1.6
Agreement (sensitivity): P (Prov UTD| HH UTD)	90.5%	82.7%
DISagreement (1-specificity): P (Prov UTD| HH not UTD)	70.7%	75.0%
Gross difference rate: P(Prov UTD and HH not UTD) + P (Prov not UTD and HH UTD)	34.2%	61.1%
Net difference rate = P(HH UTD) – P(Prov UTD)	-22.8%	-52.8%

Note: P=proportion, Prov=provider, HH=household, UTD=up-to-date; *4DTP/3Polio/1MCV/3Hib