AN ASSESSMENT OF MISCLASSIFICATION ERROR IN PROVIDER-REPORTED VACCINATION HISTORIES

Meena Khare (Centers for Disease Control and Prevention), Trena M. Ezzati-Rice (Centers for Disease Control and Prevention), Michael P. Battaglia (Abt Associates Inc.), Elizabeth R. Zell (Centers for Disease Control and Prevention)

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

Key Words: National Immunization Survey, Random-digit-dialing, Best vaccination value

1. Introduction

Immunizations are among the most effective public health interventions to prevent disease and death. In the U.S., most routine childhood vaccinations are recommended for administration before a child's second birthday. The CDC-sponsored National Immunization Survey (NIS) is designed to measure and monitor vaccine-specific coverage rates among children aged 19-35 months in the U.S. Data from the NIS are used to produce timely estimates of vaccination coverage for each of six recommended vaccines for the nation and for each of 78 Immunization Action Plan (IAP) areas, consisting of the 50 states, the District of Columbia, and 27 large urban areas (1). Vaccination histories from the NIS provide a unique opportunity to monitor progress toward Healthy People 2000 and 2010 objectives of 90% of children receiving all recommended vaccines by their second birthday.

Since April 1994, the NIS has been collecting vaccination history data on children aged 19-35 months in the U.S. The NIS is conducted in two phases. 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 IHQ collects data on the vaccination history of the selected child from the medical records maintained by the providers. A summary of the history, sample design, sample weight adjustment methodology, and quality control procedures of the survey are described in Zell *et al.* (2), Ezzati-Rice *et al.* (3), Smith *et al.* (4, 5) and Khare *et al.* (6).

Substantial underreporting has been observed in the coverage estimates from the household-reported vaccination histories. The NIS estimates of coverage rates computed from the provider reports have been shown to be more accurate than the household estimates (2,6,7). However, Battaglia *et al.* (8) reported some misreporting in the provider-reported histories and the up-to-date (UTD) status determined by the number of doses required for a vaccine. As part of the continuous program for monitoring quality of the NIS coverage estimates, recent research is focusing on evaluating potential misclassification errors in the provider-reported vaccination histories. Because provider reported UTD status is used as the 'gold standard' in the NIS, children who are classified as the UTD for a vaccine from the provider reports are not evaluated for misclassification in this paper. In this paper, misclassification error in the provider-reported data is measured by the proportion of children reported as <u>not</u> UTD (NUTD) from the providers and as UTD from the household reports for a vaccine. To assess the impact of the misclassification error on the coverage estimates, the coverage rates computed from the provider reports and the 'best' vaccination values, constructed by combining information from both the household and the provider reports, are compared.

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 system. The household interview information is collected from a parent, guardian or another knowledgeable person who is familiar with the child's vaccination history. The CATI obtains information on demographic characteristics of the mother and child, and vaccine-specific immunization histories from a written vaccination record (i.e., the 'shot card') when available or from 'memory' recall when a 'shot card' is not available.

After obtaining household information and consent to contact the child's immunization provider, the IHQ is used to collect the vaccination history from the child's provider. The NIS staff makes no additional contact with the child's provider or the household respondent to reconcile the observed discrepancies or to collect missing or incomplete vaccination data. Khare *et al.* (6) evaluated the quality of the household-reported histories and showed higher quality of the household data collected from the 'shot card' than 'memory' recall. Also, high ($\geq$90%) agreements were observed between the household 'shot card' and the provider-reported vaccine-specific UTD status(6). Therefore,

only children with household 'shot card' data are included in the analysis to evaluate misclassification errors in the provider reports.

3. Assessment of Misclassification Error

Given the complexity of the recommended childhood immunization schedule (9), the potential for reporting errors in vaccination dates increases due to transcription errors, transposed numbers/digits, poor handwriting, and incomplete vaccination information. Discrepancies in vaccination dates may occur between household- and provider-reported histories as well as between two or more providers reporting vaccination histories for the same child. Also, the potential for incomplete vaccination information is expected to increase when more than one dose is required to be classified as UTD for a vaccine or when multiple providers vaccinate a child. In addition, when two or more providers vaccinate a child or the parents move from the area where the child was born, there is the possibility that only the child's parent may have a 'shot card' with the complete vaccination history. Unique vaccination dates from one or more provider reports are used to determine the vaccine-specific number of doses. Then, vaccine-specific UTD status is determined from the number of doses a child has received for a vaccine. The following 2x2 table provides definitions of selected measures used to evaluate agreement and/or misclassification error in the UTD status.

A 2x2 table showing classification of the UTD from two Sources A and B

UTD status: Source A, 'Gold Standard'	Source B status		
	Yes	No	Total
Yes	a	b	n1
No	c	d	n2
Total	m1	m2	m

Definition of selected measures:

Sensitivity (UTD)= a/ n1;
Specificity (NUTD)= d/n2
Overall Misclassification error=c/m
Overall agreement= (a + d)/m
Overall disagreement= (b + c)/m
Proportion 'UTD' (A)= n1/m
Proportion 'UTD' (B)= m1/m
Net difference= (n1-m1)/m = Difference in proportions

To assess misclassification errors associated with the provider data, vaccine-specific reports of UTD status (e.g., the 4:3:1:3 series of 4+DTP/3+Polio/1+MCV/3+Hib) from the household 'shot cards' are compared with the corresponding provider-reported UTD status. A 2x2 table with 4:3:1:3 UTD status was created among children who had both complete household 'shot card' history and 'adequate' provider data (Table 1). Selected measures [for definitions, see above 2x2 table or pages 319-320 of Friis and Sellers (10)] including, sensitivity, specificity, overall misclassification error, overall agreement, and net difference rates were computed to evaluate agreement and misclassification error in the 4:3:1:3 UTD status from the provider reports.

To adjust for misclassification error in the provider-reported UTD status, a 'best' vaccination history was created. The 'best' vaccination history replaced missing or invalid provider dates with the corresponding vaccine-specific valid vaccination dates from the household 'shot card' using a predetermined rule. For example, all vaccine-specific dates for the children with missing provider data were replaced with the corresponding vaccination dates from the child's 'shot card' history that was 4:3:1:3 UTD from the household. The goal of constructing a 'best' vaccination history was to create a complete history for a child using both household and provider reports.

Vaccine-specific weighted estimates of the coverage rates from the provider-reported vaccination data are compared with those obtained from the household reports and the 'best' vaccination histories among children with household data from 'shot card.'

4. Results from the 1999 NIS

A total of 36,338 households with age-eligible children were identified in the 1999 NIS, of which 33,932 (93.4%) completed the household telephone interview (i.e., the CATI.) The 33,932 households contained 34,442 children aged 19-35 months. In the majority of the households, the respondent was the child's mother or a female guardian. Vaccination histories were reported from a 'shot card' for 16,829 (48.9%) children and for the other 17,613 (51.1%) children from 'memory' recall when a 'shot card' was not available. Consent was obtained to contact the child's immunization provider(s) for 28,936 (84%) children. Provider-reported 'adequate' histories for 22,521 (65.4%) children were used to publish the final estimates of the 1999 NIS coverage rates (1). Approximately 83% of the 'adequate' histories were obtained from single providers and 17% from two or more providers. Of the 22,521 children with 'adequate' provider data, 11,964 (53.1%) children had also vaccine-specific UTD status from the household 'shot card.'

Table 1 shows the distribution of 4:3:1:3 UTD status among the 11,964 children who had vaccination histories from both provider and household 'shot card' reports. The net difference in proportions or 4:3:1:3 UTD shows significant underestimation by 22.8 (=59.7-82.5) percentage points from the household reports when compared to the provider-reported coverage rates.

Table 2 presents changes in the 4:3:1:3 UTD status using the 'best' values among 2092 children in Table 1

who were reported as 4:3:1:3 NUTD from the provider reports. This table shows that of 680 children who were misreported as 4:3:1:3 NUTD by the providers but UTD from the household reports, 642 (94.4%) became UTD after using the 'best' values; the other 38 (5.6%) children remained 4:3:1:3 NUTD due to some inconsistency in the household reports. The proportion of children who were misclassified as 4:3:1:3 NUTD from the provider reports and UTD from the household reports, reduced substantially from 32.5% in Table 1 to 3.1% using the 'best' values in Table 2. Also, an additional 228 children who were reported as 4:3:1:3 NUTD from both household and providers (due to incomplete histories) became UTD using the 'best' values.

Table 3 presents a set of selected measures of agreement and misclassification errors in the 4:3:1:3 UTD status. Using provider data as the 'gold standard' in Table 1, the sensitivity for correctly reporting (i.e., when the UTD status from household agrees with the 'gold standard') a child being UTD was 65.5%, and the specificity for correctly reporting a child as NUTD was 67.5%. The overall agreement on correctly classifying children as UTD or NUTD (i.e. the overall agreement) from both data sources was 65.8%; the other 34.2% (4,089) children had disagreement in the 4:3:1:3 UTD status. Among those 4089 (34.2%) children with disagreement in the UTD status, only 680 (5.7 %) children had misclassification error in their provider-reported histories and the other 3409 (28.5%) children were UTD from the provider reports. The next column shows the corresponding measures using the 'best' values after adjusting for misclassification errors in the provider reports. The sensitivity and specificity increased to 66.1 % and 96.9%, respectively, with the use of 'best' values. The overall misclassification error reduced from 5.7% to 0.3% in the 4:3:1:3 UTD status with the use of 'best' values.

Table 3 also presents the above measures from another CDC-sponsored survey NHIS/NIPRCS (11) where immunization histories were collected from a 'shot card' with a household visit and an in-person interview. A comparison of the results shows similar results in all measures except for the specificity which was higher for the 1997 NHIS/NIPRCS, where interviewers physically see the 'shot card' for the eligible child and enter the immunization history directly into a database.

Table 4 compares the vaccine-specific misclassification errors in the vaccine-specific UTD status for 4+DTP, 3+Pol, 3+Hib, 1+MVC, 3+Hep B, 1+VRC, 4:3:1 and 4:3:1:3 vaccine series. The overall misclassification error ranged from 1.9% for 3+DTP to 6.4% for 3+Hep B vaccine series among children who were NUTD from the provider reports and UTD from the household 'shot card.' For example, the overall disagreement and misclassification error for hepatitis B vaccine (Hep B) are 15.9% and 6.4%, respectively. In a small proportion of children who were UTD for Hep B vaccine, the provider had reported only two doses and checked off the box 'given at birth' on the IHQ form but did not report a vaccination date for the first 'birth' dose resulting in misreporting these children as NUTD (having less than 3 required doses.) The 'best' vaccination value corrected for this misclassification error by using the household-reported vaccination date for the first 'birth' dose and classified them as UTD.

For the DTP vaccine where four doses are required to be UTD according to the vaccine schedule, the overall misclassification error increased from 1.9% for 3+DTP to 4.4% for 4+DTP series. This increase shows that misclassification error due to missing or incomplete information on one additional dose of the DTP vaccine increased the error rate in the provider reports by 2.5 percentage points.

5. Comparison of Coverage Estimates

In the 1999 NIS, among 16,829 children with 'shot card' data, 11,964 children had 'adequate' provider data, and 13,449 children had 'best' vaccination values consisting of either the original provider-reported vaccination data or the 'best' vaccination values from the valid household 'shot card' data. Table 5 presents the weighted estimates of the vaccine-specific coverage rates and the associated 95% confidence intervals among all household 'shot card' children by the source of vaccination information. Figure 1 shows a comparison of the vaccination coverage rates and 95% confidence intervals for the 4:3:1:3 series among 'shot card' children by race/ethnicity and the source of vaccination information.

A comparison of the weighted household coverage estimates with those obtained from the provider data, Table 5 and Figure 1 show significant underestimation in the household estimates. The overall coverage rate for the 4:3:1:3 series from the household reports was 24.4 percentage points lower than the corresponding rates from the provider reports.

A comparison of the weighted estimates from the provider data among 'shot card' children showed that the coverage rate for the 4:3:1:3 vaccine series from the provider reports was lower by 7.7 percentage points from the estimate with the 'best' vaccination values. For individual vaccines, the underestimation in provider-based coverage rates when compared with the 'best' values estimates, ranged from 2.3 percentage points for the 3+Hib to 5.6 percentage points for the 3+Hep B series.

In addition, the misclassification errors in the provider-reported 4:3:1:3 UTD status by race/ethnic groups ranged from 5% to 7% (data not shown.) The underestimation in the provider-based 4:3:1:3 coverage

rates were 9.0 percentage points among Hispanics, 10.7 percentage points among non-Hispanic blacks, and 6.5 percentage points among non-Hispanic whites/other (Figure 1) when compared to the 'best' value estimates.

6. Summary and Discussion

Misclassification errors in vaccination histories occur between household and provider reports as well as between two or more providers who vaccinated the same child. The misclassification error increased when more than one provider vaccinated a child or multiple vaccines are required to report the child as UTD for a vaccine. The majority of the observed reporting errors in the provider data were due to missing or incomplete vaccination histories, transcription errors, poor handwriting, reporting incorrect dates/years, or reporting transposed numbers/digits in the dates (6). This resulted in an overall misclassification error of 5.7% in the 4:3:1:3 UTD status among children who had 'adequate' data from the providers and the household-reported vaccination history from the 'shot card.' The overall misclassification error ranged from 1.9% for 3+DTP to 6.4% for 3+Hep B vaccine series among children who were NUTD from the provider reports and UTD from the household 'shot card.'

A comparison of the coverage rates showed a significant impact of misclassification errors on the estimates from the provider data. The coverage rates increased significantly with the use of 'best' vaccination values, which corrected provider-reported vaccination histories for the misclassification error. The misclassification error in provider-reported histories varied by the vaccine type and underestimated, for example, the 4:3:1:3 coverage rate by 7.7 percentage points due to incomplete or missing vaccination histories. For the individual vaccines, the underestimation in provider-based coverage rates ranged from 2.3 percentage points for the 3+Hib to 5.6 percentage points for the 3+Hep B series.

Khare *et al*. (6) showed that to correct for misclassification errors in the provider-reported histories, the 'best' vaccination value procedure used only vaccination information from the children with household 'shot card' reports and excluded the data collected from 'memory' recall. Therefore, the coverage rates computed from the 'best' vaccination values depend considerably on 'shot card' use (range by IAP area: 18-67%), percent of missing provider data, and other demographic characteristics in an IAP area. Additional methodology research and analysis are planned at the CDC to investigate whether the construction of 'best' vaccination values for the missing provider data would 'truly' improve the vaccine-specific estimates of the NIS coverage rates and further reduce bias due to misclassification errors in the published results by using the 'shot card' histories as well as the 'memory' recall data. Additional research is also planned to use imputation methodology to impute missing provider data for all interviewed children and to evaluate imputed estimates in the NIS.

References

1. 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.
2. Zell, E.R., Ezzati-Rice, T.M., Battaglia, M.P., and Wright, R.A.. "National Immunization Survey: The Methodology of a Vaccination Surveillance System," Public Health Reports, 115:65-77, 2000.
3. 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," Proceedings of the Section on Survey Research Methods, American Statistical Association, VA, 668-672, 1995.
4. Smith, P.J., Battaglia, M. P., Huggins, V., Ezzati-Rice, T., Rodén, A-S., Khare, M., and Wright, R. A. "Overview of the sampling design and the statistical methods used in the National Immunization Survey," American Journal of Preventive Medicine, 20(4S), pp17-24, 2001.
5. Smith, P.J., Rao, J.N.K., Daniels, D., Battaglia, M. P., Ezzati-Rice, T., and Khare, M. "Compensating For Nonresponse Bias In The National Immunization Survey Using Response Propensities," 2000 Proceedings of the Section on Survey Research Methods, American Statistical Association, VA, pp 641-646, 2000.
6. Khare, M., Battaglia, M.P., Huggins, V., Stokley, S., Hoaglin, D.C., Wright, R.A., and Roden, A-S. "Accuracy of Vaccination Dates Reported by Immunization Providers in the National Immunization Survey," 2000 Proceedings of the Section on Survey Research Methods, American Statistical Association, VA, pp 665-670, 2000.
7. Ezzati-Rice, T.M., Zell, E.R., Massey, J.T., and Nixon, M.G.. "Improving the Assessment of Vaccination Coverage Rates with the Use of Both Household and Medical Provider Data," Proceedings of the Section on Survey Research Methods, American Statistical Association, VA, pp 335-340, 1996.
8. Battaglia, M.P., Ezzati-Rice, T.M., and Zell, E.R. "Respondent Characteristics Associated with Misreporting of Vaccinations in a Telephone Survey," 1997 Proceedings of the Section on Survey Research Methods, American Statistical Association, VA, pp 976-980, 1997.
9. CDC, "Recommended Childhood Immunization Schedule, United States, January 2000- December 2000," http://www.cdc.gov/nip/recs/child-

schedule.PDF, 2000.
10. Friis, R.H. and Sellers, T.A. Epidemiology for Public Health Practice, an Aspen Publication, MD, second edition, 1999
11. Peak, R.R., and Cadell, D.M. "Overview of the National Immunization Provider Record Check Study," Proceedings of the Section on Survey Research Methods, American Statistical Association, VA, pp 332-334, 1996.

Table 1: Number and percentage of children with both household 'shot card' and 'adequate' provider reports showing agreement on 4:3:1:3* UTD status, 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	65.5	3,409	34.5	9,872	82.5**
Prov NUTD	**680**	**32.5**	1,412	67.5	2,092	17.5**
Total	7,143	59.7	4,821	40.3	11,964	100

Note: Prov=provider, HH=household, UTD=up-to-date, NUTD= not UTD;
*4+DTP/3+Polio/1+MCV/3+Hib; ** Column percent

Table 2: Changes in 4:3:1:3* UTD status with 'best' vaccination values among children who were reported as 4:3:1:3* NUTD by the providers, 1999 NIS

Best value reported 4:3:1:3 UTD status	Household-reported 4:3:1:3 UTD status					
	HH UTD		HH not UTD		Total	
	n	%	n	%	n	%
Best value UTD	642	73.8	228	26.2	870	41.6**
Best value NUTD	**38**	**3.11**	1,184	96.9	1,222	58.4**
Total (Prov NUTD)	680	32.5	1412	67.5	2092	100

Note: Prov=provider, HH=household, UTD=up-to-date, NUTD= not UTD;
*4+DTP/3+Polio/1+MCV/3+Hib; ** Column percent

Table 3: Selected measures to evaluate misclassification errors in reporting 4:3:1:3* UTD among children with 'shot card' and 'adequate' provider data

	Source of 4:3:1:3 UTD status		
	1999 NIS Shot Card and Provider data	1999 NIS Shot Card and Best values	1997 NHIS/NIPRCS** Shot Card and Provider data
Children with 'adequate' vaccination history	11,964	11,964	427
Sensitivity (UTD)	65.5%	66.1%	64.2%
Specificity (NUTD)	67.5%	96.9%	74.2%
Overall Agreement	65.8%	69.3%	66.5%
Overall Disagreement	34.2%	30.7%	33.5%
Overall Misclassification Error	**5.7%**	**0.3%**	5.5%%
Net Difference rate	-22.8%	-30.1%	-21.8%

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

Table 4: Vaccine-specific misclassification errors in provider reports among children who had both 'shot card' and 'adequate' provider data, 1999 NIS

Vaccine Type	Overall Disagreement in UTD status %	Overall Misclassification Error in provider data %	Net Difference in Coverage rates %
3+ DTP	8.1	1.9	-4.3
4+ DTP	23.6	4.4	-4.7
3+ Polio	14.4	4.7	-5.0
1+ MCV	11.4	4.0	-3.4
3+ HIB	18.5	2.3	-13.8
3+ Hep B	15.9	6.4	-3.1
1+VRC	3.0	5.6	-1.7
4:3:1*	28.1	6.1	-15.9
4:3:1:3**	34.2	5.7	-22.8

Note: Prov=provider, HH=household, UTD=up-to-date, NUTD= not UTD;
*4+DTP/3+Polio/1+MCV; **4+DTP/3+Polio/1+MCV/3+Hib;

Table 5: Weighted estimates of vaccine-specific coverage rates and 95% confidence intervals among children with 'shot cards' by source of vaccination information, 1999 NIS

Vaccine type	Household Estimates (%) (A)	95% Confidence Interval	Provider Data Estimates (%) (B)	95% Confidence Interval	Best Values Estimates (%) (C)	95% Confidence Interval	Difference in coverage rates among children with 'shot card'# (%)	
							(A-B)	(B-C)
Sample size	16,829		11,964		13,449			
4+ DTP	70.4	±1.06	86.5	±1.02	91.8	±0.82	-16.1	-5.3
3+ Polio	85.6	±0.82	91.3	±0.81	95.8	±0.60	-5.7	-4.5
1+ MCV	90.5	±0.66	93.5	±0.75	97.4	±0.48	-3.0	-3.9
3+ HIB	79.4	±0.93	94.4	±0.72	96.7	±0.60	-15.0	-2.3
3+ Hep B	86.4	±0.79	89.9	±0.92	95.5	±0.67	-3.5	-5.6
1+VRC	57.2	±1.11	60.6	±1.37	64.6	±1.25	-3.4	-4.0
4:3:1*	65.3	±1.10	82.8	±1.13	90.2	±0.87	-17.5	-7.4
4:3:1:3**	56.6	±1.13	81.0	±1.18	88.7	±0.95	-24.4	-7.7

*4+DTP/3+Polio/1+MCV; **4+DTP/3+Polio/1+MCV/3+Hib;
All differences are statistically significant at the alpha=0.05 level of significance

Figure 1: Comparison of 4:3:1:3 coverage rates based on data from household, provider, and best values among children with 'shot cards', 1999 NIS