THE NATIONAL IMMUNIZATION SURVEY REGISTRY VALIDATION STUDY: METHODOLOGY FOR MATCHING AND ANALYSES

John Stevenson and Beth Luman, National Immunization Program
Robert A. Wright and Meena Khare, National Center for Health Statistics
Barbara Canavan, Oregon ALERT Immunization Registry Office
David C. Hoaglin, Vicki J. Huggins, Linda Piccinino and Rebecca Strella, Abt Associates Inc.

John Stevenson, National Immunization Program,
Centers for Disease Control and Prevention,
Atlanta, Georgia

Key Words: Registries, Record matching, Population coverage, Measurement error, Best values

INTRODUCTION AND OVERVIEW

The National Immunization Survey (NIS), sponsored by the National Immunization Program (NIP) and the National Center for Heath Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC), is a population-based sample survey of households with children age 19 to 35 months. The survey is conducted by telephone, using random-digit dialing (RDD), in 78 Immunization Action Plan (IAP) areas that cover the 50 states and 28 urban areas. It is designed to support estimates of vaccination coverage levels for each IAP area. In addition to asking household respondents for information on their children's vaccinations, the NIS includes a mail survey that (with consent) asks the children's medical providers to report the vaccinations in the child's medical record. Zell et al. (2000) describe the design and methodology of the NIS in more detail.

In a growing number of states, immunization registries are compiling records of immunizations received by children residing in their catchment areas. Their goal is complete information on all vaccinations that all resident children have received from providers of all types (e.g., pediatricians in private practice, clinics operated by county health departments, and hospital emergency rooms).

For monitoring children's immunizations, surveys and registries take different approaches, even though both are population-based. A survey aims to sample the population and measure various aspects of immunization status in the aggregate. Goals of a registry are to establish a record for each child in its catchment area, to receive a report of each vaccination a child receives, and to allow providers to check a child's immunization status to avoid missed or duplicate shots.

Both surveys and registries may fall short of these objectives. A survey may miss a segment of the population, and it may gather incomplete information on the vaccinations of the children that it does reach. A registry may be less successful in receiving reports from some types of providers (e.g., physicians in private practice), reports may be slow to arrive, or it may contain no information on vaccinations that its children have received outside its catchment area. When a survey and a registry both cover the same geographic area, matching the survey data with the registry's database can help to assess the effectiveness of each. For example, the percentage of NIS children from an IAP area that can be matched to that area's immunization registry is an indication of how successful the registry is in including all children in its catchment area. This assessment is the goal of the National Immunization Survey's Registry Validation Study.

Importance of Complete Registries

The existence of accurate registry information can encourage immunization in at least two important ways. First, by maintaining a complete immunization history for each child, registries offer immunization providers access to comprehensive shot records from all of that child's providers. Omissions of recommended shots could be easily identified, the appropriate shots would be administered, and unnecessary duplication of shots then would be avoided. Registry information might also be used as part of a "tickler" system to remind parents of approaching dates for recommended shots.

In addition to these child-level functions, registries can serve an important surveillance function by monitoring immunization levels for groups of children. Further, registry data for small geographic areas could be used to identify "pockets of need," where immunization levels are critically low.

The Status of Immunization Registries in the U.S.

Many immunization information needs will be addressed by registries at a future time, but in most areas of the country the development of registries is only beginning. As registries mature, as in Oregon, it is important to understand the comprehensiveness of registry data from two perspectives: the extent to which children in a given catchment area are included in the

registry, and the extent to which immunization records for these children are complete.

Data from the 1999 Immunization Registry Annual Report (CDC 2000) suggest that state and local registries have made substantial progress in enrolling children, recruiting providers, and implementing registry functional standards. Forty-three (69%) of the 64 federal government immunization grantees (50 states, the District of Columbia, Chicago, Houston, New York City, Philadelphia, San Antonio, and eight US territories) reported implementing registries. From the grantees implementing registries, 72% reported the capacity to establish a registry record within 2 months of birth for each newborn child residing in the catchment area; 77% reported the ability to automatically produce vaccination coverage reports by providers, age groups and geographic areas. Only 16% reported the ability to consolidate vaccination records from multiple providers, using de-duplication and edit-checking procedures to optimize accuracy and completeness. Data from 42 of the 43 grantees implementing registries showed that the median percentage of providers submitting records to a registry was 96% for public-sector providers and 15% for private providers.

The progress reported in implementing registries underscores the potential of registries as sources of population-based childhood vaccination coverage estimates. An analysis by Bartlett (2000) of the agreement between NIS estimates of immunization coverage and registry-based estimates of coverage was less encouraging. Bartlett stratified registries in three groups based on level of operational status. The registry-based estimates of coverage were significantly lower than the NIS estimates for registries with low to moderate operational status.

STUDY METHODOLOGY
Matching NIS Data to State Immunization Registry Data

In the NIS Registry Validation Study, registry data from participating states will be matched and compared with data from the NIS. The richly detailed demographic and other descriptive data that the NIS collects to describe and characterize immunization levels of young children can be used to help characterize the sources of nonsampling error for those who are more versus less likely to be included in the registry, as well as more versus less likely to have complete immunization data.

Data collected in the NIS from January 1995 to March 2000 will be matched to the state immunization data at aggregate and microdata levels. To preserve the confidentiality of the data, arrangements with the states will ensure that no information identifying NIS children will remain with the registry, and that no information

identifying children in the registry but not the NIS will leave the registry.

Plans call for up to three state registry offices to participate in the NIS Registry Validation Study. The procedures described below acknowledge that multiple states will be involved and attempt to coordinate file preparation and matching activities across all participating states.

A. Build three files from NIS data
1. Child-level file
For children in any of the match states, this file should contain single-valued child attributes such as: Name, parsed into first, middle (or initial), and last; date of birth; gender; and race and ethnicity. The file is intended to capture data on all households/children ever believed to be in any of the match states. This permits children to be applicable to the matching process for more than one state; for example, a child who was born in Oregon, received an at-birth vaccine there, and later moved to another state will have one record in this file.

2. Vaccination-level file
From the record for each child, a file containing one record per vaccination received per child will be created. These records will have the standard vaccine classification L7/CVX (CDC 2000, Fig. 7-3) codes appended to match the vaccination identifications used by the states, as well as a link to the appropriate record in the provider file. The file is developed from those vaccinations received by each child-based on the best knowledge.

3. Provider-level file
For each of the children in the child-level file, this file will contain the name, address, etc., of each of the providers known or believed to have immunization data for the child.

B. Obtain three files from the participating states
1. Child-level file
Though the NIS derived data cover 1995, quarter 1, to 2000, quarter 1, the selection criteria for the states are simply all those children with dates-of-birth in 1992, quarter 1, to 1998, quarter 4. The extra quarters, 1992, quarter 1, to 1994, quarter 4, are needed because a child in the initial NIS quarter, 1995, quarter 1, might be up to 35 months old. This file will be matched to the comparable child-level file created from the NIS data in order to find children in common, those only in the registry, and those only in the NIS.

2. Vaccination-level file
A file of vaccinations for those children in the state registries who actually match a record in the NIS derived child-level file is needed. However, the analysis of the matching process may necessitate vaccination-level data for all or a random subsample of the children in the 1992, quarter 1, to 2000, quarter 1, registry that do not

match to an NIS sample child. Together with the NIS vaccination file, this file will be used to assess vaccination coverage for those children found in both the registry and the NIS.

3. Provider-level file

A provider-level file will be necessary only for those children for whom follow-up reconciliation is attempted.

C. Perform the matching processes

Registry data needed for the full matching operation will be copied from the registry database onto a storage device and given to a programmer representing the contractor, Abt Associates, to perform the matching. The programmer will apply the matching algorithms and compute the matching rates. Once the matched cases have been identified, the registry data for the matched children will be appended to the NIS records for those children. In addition, a random sample of approximately 8,800 children from the registry database that did not match to an NIS record will be added to the research file that contains the matched children. This is about five times the NIS sample for each state and will provide adequate reliability for planned analyses. NIS nonrespondent and partially interviewed cases may be included in the matching process. If so, the vaccination-level data from the registry file are needed for these cases as well. The immunization information for the non-match children will also be attached.

The matching process will consist of three separate steps involving (in order) three pairwise sets of files: Child-level, vaccination-level, and provider-level.

1. Child-level matching

For each state there will be approximately 1,760 child-level NIS records. The child-level file from the state registry will have over 100,000 records. The matching process consists of finding each of the NIS children in the registry file and creating a file of linked record identifiers, or determining that an NIS child is not in the registry.

For the actual matching process, explicit extracts of NIS files containing fields comparable to those of the state registries must be created, as well as extracts from the registry files for computer monitoring purposes and experimentation with the matching program to be used.

One area that will require special investigation and controls during the matching is the identification and tracking of movers. On the NIS side, there will be quarterly samples of children from each test state. Files will contain all of the NIS children in the age cohort who were ever in the NIS and resided in any of the test states. On the registry side, there will be records of children who were not eligible to be in the NIS for a specific quarter because they moved to another state. If this phenomenon is not known or tracked by the registry, and if the registry has no mechanism for including movers

into the state, then the immunization estimates from the registry could differ substantially from NIS immunization estimates, primarily through differences in population definitions. Examining these issues is critical in determining the population coverage of the registries.

The final result from the matching will be a file with linked NIS and registry record identifiers for those NIS children found in the registry files and, by implication, those NIS children not found in the registry (and the residual registry children).

2. Vaccination-level matching

The vaccination records for those children with linked NIS and registry identifiers produced above will be selected from the respective NIS and registry vaccination files, sorted in an appropriately common order, and compared.

The first step in this process is to create a cross-tabulation displaying the number of children with each possible pairwise count of agreement in the two vaccination sets. That is, a table whose rows, for example, consist of 0,1,2,... unique NIS vaccinations and whose columns consist of 0,1,2,... unique registry vaccinations. An intersection, say (1,2), would give the number of children who have one unique NIS-reported vaccination and two unique registry-reported vaccinations. The (i,i) intersection contains the count of children whose vaccination records agreed exactly. This table can then be used to decide which and how many reconciliations to attempt, which cases to reconcile electronically, and which cases to reconcile through telephone contact with providers.

3. Provider-level matching

The respective NIS and registry provider files will potentially be used for two purposes: To analyze potential differences in the number of providers reporting to the NIS and the registries, respectively, for those children whose vaccination histories do not match; and to provide the source of provider information for actual contact. (The NIS file will have more providers named than those who reported.)

The first step in a provider-reporting analysis would be a uniqueness table similar to that described above for the comparison of the vaccination histories. For this process, identifying information for children in the registry who are not matched to NIS children will be needed in order to contact providers. (The non-matched children are now part of the research file for NIS confidentiality reasons. In order to mask which records are in the NIS, we will augment the file with non-match cases from the registry.)

Once the NIS and registry datasets undergo the matching process, data on the immunization histories of all the children originally in either the NIS or state registry will be put in a new data set, the Registry Research File. For each child in the registry who is

matched to an NIS data record, registry information will be attached to the NIS data record. For children in the registry who are *not* matched to an NIS data record, the registry data record for each child (or a subsample of these children) will be appended as a new data record in the matched file. However, the new data record that comes only from the registry file will not contain identifying information about the child. It will be used only as aggregate data for research purposes. A flag will be added to the Registry Research File to indicate the matching status for the case.

Reconciliation of Immunization Information

After completion of the matching operation, cases with immunization data that are not consistent between NIS reports and state registry reports will be flagged. Depending on the number of cases identified and available resources, either all or a sample of the providers for inconsistent cases will be recontacted to verify discrepant data. These cases, or a subset of these cases, will become the sample for reconciliation. Two procedures will be used for the reconciliation. The first is an electronic procedure, and the second involves contact with medical providers to obtain "best values" for reports of immunizations.

Electronic Reconciliation of Minor Inconsistencies

Some minor inconsistencies between registry and NIS information may be resolved by electronic examination of the data. This electronic reconciliation will take place prior to contact with medical providers. This step minimizes the cost of reconciliation, maximizes the number of cases that can be reconciled, and reduces the number of contacts with providers, who may feel burdened by additional requests to supply immunization information.

When comparing matched files from separate sources for the same child, it is anticipated that both minor and major inconsistencies among immunization reports will be identified. Minor inconsistencies would include single immunization dates that vary by less than seven days, single dates that appear to be transpositions of the numbers for the day of the month and the month of the year (e.g., 04-05-1997 and 05-04-1997), and dates with missing information for the day of the month. Additional minor inconsistencies may be identified during examination of the data files. We plan to resolve electronically the minor inconsistencies described above.

The exact criteria for dates to be reconciled can be modified. For example, a criterion of a seven-day difference in dates can be narrowed or widened, depending on the number and variety of inconsistencies actually identified. The relatively low cost and high efficiency of electronic reconciliation argue for the use of this method in any instance where it does not compromise the integrity of the resulting data.

Reconciliation of Matched NIS Cases

Following the electronic reconciliation process described above, paper matching sheets can be printed with all shot dates from all sources, date of birth from each source, child identifying information and provider contact information. Inconsistent dates can be flagged, and the cases sorted for contact with providers.

The first sorting identifies multiple cases requiring contact with the same provider. A second sorting of cases identifies the type of reconciliation necessary:
- missing shot date(s)
- different date(s) for same shot(s)
- shot date(s) prior to date of birth
- different date of birth
- no record for this child

Contact with providers who have multiple cases follows the same steps, but the nature of the interaction will vary with the types of the inconsistencies in the data files. The steps in contacting providers include:
- contact the provider by telephone
- identify the reason for the call
- establish the identity of the child
- verify the date(s) in question
- record the outcome of the verification on the matching sheet
- if the inconsistency is resolved, forward the case to data entry
- if the inconsistency is *not* resolved, verify the source of information in question, verify the source of immunization(s) in question, identify additional possible contacts to aid reconciliation, and forward the case for additional review

Providers with multiple cases requiring reconciliation would be approached with an additional, preliminary contact. An introductory telephone call to the provider would explain the nature of the contact and offer to forward a list of all children whose records need review. This preliminary contact would allow the provider to locate all needed records and schedule a convenient time for resolving all questions. Advance notice should facilitate the retrieval of records and decrease the time that reconciliation staff and providers spend on the telephone.

Cases with missing and/or discrepant dates should be fairly simple to resolve. Reconciliation staff would ask the provider to refer to written immunization records and verify the dates in question. Most of these cases are expected to be recording errors that can be corrected during this telephone contact. Difficulties may arise if the

date in question has been supplied by a previously unidentified provider or from a parent's record and cannot be verified during this contact with the provider. Additional providers will be contacted only if they have been named during the NIS household interview. Any dates from secondary sources will be identified on the matching sheet as "other provider" or "parents' record." These cases will be reviewed before being sent to data entry.

Resolving date-of-birth discrepancies may be simple or may lead to additional problems. Simple recording errors that conform with the pattern of shot dates will be indicated on the matching sheet, and the case will be resolved. In some instances, resolving date of birth inconsistencies may raise issues about the correct identification of the child, or may cause some shot dates to fall prior to the child's date-of-birth. The reconciliation staff will attempt to resolve these cases during the telephone contact. Unresolved cases will be reviewed for possible additional action.

Providers may indicate that no immunization record is available for a child. If a provider reports never having had contact with the child, either the child or the provider may be misidentified. Reconciliation staff will attempt to resolve these problems during the telephone contact. If a provider has forwarded a child's records to a different provider, the new provider will be contacted only if that provider was named during the NIS household interview. If the child's records have been sent to long-term storage, reconciliation staff will gather information about the cost and time required for retrieving the records and refer that information for review.

ANALYSIS

Demographic distributions of the data for the cohort of children 19 to 35 months of age in the years 1995-1998 will be produced in order to assess coverage of the state registry. Distributions of available demographic data will be compared to distributions in the NIS and to Census data. Information from comparisons in the aggregate will help to focus more detailed comparisons of micro-level matched and unmatched data.

From the initial match, the match rate between registry and NIS data is calculated as
Match Rate = (M / N) 100, where M is the number of NIS children matched to registry records for an area, and N is the number of NIS children identified by the survey in the same area. A variety of breakdowns should be possible. For example, it will be informative to compare match rates between children who use public providers and children who use private providers.

For children identified in both datasets, the immunization records available from the registry and from the NIS will be compared to calculate an agreement rate as Agreement Rate = (A / M) 100, where A is the number of cases where the NIS and registry data agree, and M is defined as above. The agreement rate may be considered an indicator of the reliability of both NIS and registry data. More than one definition of "NIS data" can be used. For example, one can use only NIS provider data on vaccinators, or one can use those provider data when available and the household data otherwise. Both of these agreement rates can be stratified by race/ethnicity and by poverty status. The analysis will yield an agreement rate for each vaccine and series that the NIS tracks.

Immunization rates between the two sources, broken out by the matching outcomes and other influential variables, will also be compared. An overall set of immunization rates will be based on five subsets of the data: 1) all NIS children; 2) all children in the registry; 3) matched children, taking immunization status from the NIS; 4) matched children, taking immunization status from the registry; 5) unmatched children in the registry. Each of these can be broken down by calendar year, public vs. private provider, race/ethnicity, and poverty status.

To the extent that NIS and registry data agree for a given child, the two sources may verify each another. Alternatively, analysis of the characteristics of children with non-matching data may point to conditions where one data source is more likely to be correct.

Several secondary analyses are planned, should resources become available. We would like to examine issues related to immunization for the non-telephone population contained in the state registry. This analysis will depend on the extent to which we can determine, from the state registry information on home/business/relative phone numbers, whether a child had a nonzero probability of selection in the NIS.

We plan to examine whether dual-frame or supplemental-frame sampling and estimation approaches using the NIS data and state registry data are a viable option for future improvements to the NIS surveillance program.

Finally, we would like to begin investigating whether and how registry data may be used to replace some or all of the effort in the NIS to obtain provider data on sample children. For example, in the case of Oregon, a process could be set up where registry data on an NIS sample child can be released so that the provider does not have to duplicate efforts for the registry and the NIS provider record check study (PRCS). It is already written into Oregon law that users of the registry are not required to obtain consent for release of information from their registry.

This kind of information ultimately will lead to improvements in the NIS sampling, estimation methods

and questionnaire design.

References:
Bartlett, D. (2000)   The relationship between immunization registry operational status and accuracy of vaccination coverage estimates among children aged 19-35 months.   Presented at the 2000 Immunization Registry Conference, Newport, Rhode Island, March 27-29, 2000.

Centers for Disease Control and Prevention (2000). CDC Vaccine Codes.   National Immunization Program, CDC, Atlanta, GA.
http://www.mcis.duke.edu/standards/HL7/pubs/
PDF, October 2000.

Centers for Disease Control and Prevention (2000). Progress in development of immunization registries - United States, 1999. Morbidity and Mortality Weekly Report, April 7, 2000, 49 (13), 274-278.

Zell, E. R., Ezzati-Rice, T.M., Battaglia, M. P., and Wright, R. A. (2000). National Immunization Survey: The methodology of a vaccination surveillance system. Public Health Reports, 115, 65-77.