Matern Child Health J
DOI 10.1007/s10995-009-0464-3

Methodological Innovations in Data Gathering: Newborn Screening Linkage with Live Births Records, Michigan, 1/2007–3/2008

S. J. Korzeniewski · V. Grigorescu · G. Copeland · G. Gu · K. K. Thoburn · J. D. Rogers · W. I. Young

Abstract *Objective* To match Michigan birth and newborn screening records to identify and follow-up potentially unscreened infants, assess data quality, and demonstrate the utility of Link Plus linkage software for matching MCH related administrative datasets. *Methods* Birth and newborn screening records maintained by the Michigan Department of Community Health from January 2007 through March 2008 were used in this study. Link Plus, a freely-available probabilistic record linkage software program developed at the Centers for Disease Control and Prevention, was used to match records. Linkage performance was assessed by the linkage success rate (percentage of valid matches). Follow-up of un-matched records was conducted by the Michigan Newborn Screening Follow-up Program. *Results* Nearly all (99.2%) of the 142,178 birth records included in this study were successfully matched to newborn screening records. Following a transition to a web-based electronic birth certificate system and inclusion of a newborn screening card identification number on the birth record in 2008, the linkage success rate increased to 99.6% based on analysis of approximately 18,000 records. Of approximately 600 un-matched records, nearly half had received a newborn screen. Approximately 8% of un-matched records were due to parental refusal of newborn screening. Nine children received an initial screen as a result of this study; one was confirmed as having sickle cell trait. *Conclusions* We have demonstrated that a freely available record linkage software, Link Plus, can be used to successfully match records of MCH databases thereby providing an opportunity for further research and quality assurance investigations.

Keywords Newborn screening · Link Plus · Record linkage · Matching · Epidemiology

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention/the Agency for Chronic Disease.

S. J. Korzeniewski (✉) · V. Grigorescu · W. I. Young
Division of Genomics, Perinatal Health and Chronic Disease Epidemiology Bureau of Epidemiology, Michigan Department of Community Health, 201 Capital View, 4-012, Lansing, MI 48906, USA
e-mail: KorzeniewskiS@Michigan.gov

G. Copeland
Division for Vital Records and Health Statistics, Michigan Department of Community Health, Lansing, USA

G. Gu
CDC/NPCR Contractor, Science Applications International Corporation, San Diego, CA 92121, USA

K. K. Thoburn
CDC/NPCR Contractor, Northrop Grumman, Los Angeles, CA, USA

J. D. Rogers
National Program of Cancer Registries, Centers for Disease Control and Prevention, Atlanta, GA, USA

Newborn screening (NBS) is a highly visible and important State-based public health program that began nearly 40 years ago. Depending on the condition, potential outcomes of disorders in the NBS panel include but are not limited to brain/neurological damage, mental retardation, damage to the liver, eyes, spleen, stroke, or death if not detected early. To prevent such outcomes from occurring, it is imperative that all newborns promptly receive a screening to assure appropriate referral, diagnosis, and treatment prior to the onset of disease symptoms thereby minimizing the risk of long-term sequela. To address the

Published online: 08 April 2009

Table 1 Newborn screening and vital statistics linkage configuration

Blocking variables		Matching variables	
Name	Phonetic system	Name	Matching method
Mother's first name	Soundex [4]	Mother's first name	First name
Mother's last name		Infant's first name	
		Mother's last name	Last name
		Infant's last name	
		Mother's zip code	Zip-code
		NBS card number[a]	Value-specific
		Birth order	
		Mother's birth date	
		Infant's birth date	
		Mother's SSN	
		Mother's medical record number	

[a] NBS card number included only for 2008 electronic birth certificate system records

stringent need to ensure that all newborns in Michigan are screened promptly, the Michigan NBS Follow-up Program re-initiated the linking of live births to NBS records in November of 2007. The first attempts to conduct record linkages lead to revisions of the NBS data to facilitate a linkage process and later attempts were successful. However, these efforts required a significant time commitment and were unable to identify potentially unscreened infants within the first several weeks of life and were therefore stopped.

The current linkage process was initiated as an endeavor mutually beneficial to the NBS and Vital Records programs; specifically, the NBS program gained a means of identifying potentially unscreened infants while the Vital Records program obtained a method for assessing data quality compared to the NBS records. The linkage process detailed herein was also initiated to assess the feasibility of linking NBS records to other administrative databases in hopes of demonstrating opportunities for further quality assurance investigations relative to innovative data gathering. Specifically, the ability to link NBS records to other databases including Children's Special Health Care Services, Birth Defects Registry, Women Infants and Children, and Medicaid data provides a greater opportunity to assess the quality and impact of the NBS program via increased data acquisition relative to service utilization and childhood outcomes. Thus, the aim of this note is to highlight the innovative application of record linkage methods to address maternal and child health program needs by providing new opportunities for program evaluation.

The Michigan NBS Follow-up Program matched NBS and birth records using Link Plus, a probabilistic record linkage program developed for cancer registries at the Centers for Disease Control and Prevention's (CDC) Division of Cancer Prevention and Control in support of CDC's National Program of Cancer Registries (NCPR) [1]. Although originally designed to be used by cancer registries, Link Plus can be used with any type of data in fixed width or delimited format. Link Plus has user-friendly interfaces, and was developed and is provided by the CDC at no cost.

The linkage process in Link Plus involves the computation of a linkage score for each potential match. The probabilistic record linkage scores calculated by Link Plus are based on the theoretical framework developed by Fellegi and Sunter [2]. The linkage score for a comparison pair of records is the sum of the logarithm of odds across all matching variables, based on the probability that a matching variable agrees given that a comparison pair is a match and the probability that a matching variable agrees given that a comparison pair is not a match [3]. Thus, linkage success rates are a function of the cut-off selected for acceptable linkage scores.

Records were matched using demographics and identifier variables common to both NBS and live birth records. Table 1 reports the blocking variables and associated phonetic systems used, as well as the matching variables and methods employed in the linkage process. Blocking variables are used to 'block' (or partition) the two files; matching variables are compared between records matching on the blocking variable. Detailed descriptions of blocking/matching variables, phonetic systems, and matching methods are provided in the Link Plus user's guide provided with the software.

Of note is that infants having a birth record may die prior to being eligible for NBS; thus, infant mortality status (yes/no) was included in the linkage and follow-up process. Specifically, infants ineligible for NBS (those who died in the first 24 h of life) were removed prior to conducting the linkages, and eligible but unscreened infants known to have

died were not reported to follow-up staff for further tracking. Following each linkage, potential comparison pairs of records were manually reviewed, and match status was assigned using the manual review feature available in Link Plus. NBS Follow-up staff met to discuss and resolve questionable matches.

The linkage process evaluation reported herein is a cross sectional study with the primary outcome being the proportion of live birth records successfully matched to NBS records. Births included in the linkage process spanned all of 2007; however, several months of 2008 births were also linked to assess/demonstrate the impact of two changes in the process of registering births in Michigan. First, the transition from a DOS based to a Web based electronic birth certificate (EBC) system began in the fall of 2007. The change to a Web based system accelerated the availability of live birth data from weeks to days. Second, the NBS screening card identification number, a unique identifier assigned to all NBS program cards purchased by hospitals, was added to the electronic record captured by the new Web based system. Thus, 2008 results reported are not directly comparable to those of 2007; however, we feel it is important to note the impact of modifications to our birth and NBS records and therefore chose to include limited 2008 data in this analysis.

Pilot linkages commenced in November of 2007 utilizing live birth certificates and NBS records from January through November of 2007 to construct a final linkage configuration. Multiple combinations of blocking and matching variables along with associated phonetic systems were employed and linkage success, measured as the proportion of live birth records matched validly to NBS records, was used to discern the final linkage configuration. We also conducted linkages stratified by plurality and compared the results to non-stratified linkage; stratification improved linkage results meaning our final configuration included linking singleton and multiparous births separately. This configuration was then used to perform the linkage with NBS records of all resident births occurring in Michigan from January 1, 2007 through March 11, 2008 provided by the Division of Vital Records and Health Statistics. Both, the live births and NBS records were formatted for record linkage using SAS v 9.1 (Cary, NC). Specifically, data were imported into SAS and pertinent fields in both databases were formatted alike for the further conversion into text files. NBS text files served as the base file, and live birth text files were matched to the base file using Link Plus linkage software.

A total of 142,178 birth records were provided to the NBS Follow-up Program for record linkage; because the tally of births had not been finalized for the time period under investigation, we do not report the final number of resident births and instead use the number of unique birth records provided as the denominator of interest. The linkage success rate was 99.2% in total and 99.6% among births received from January 1st through March 11th 2008 ($N = 18{,}197$). Table 2 reports the linkage results by birth year and in total. While the overall non-match rate was 0.8%, it is important to notice the change in the non-match rate from 0.9% in 2007 to 0.4% in 2008 likely due to the transition to the EBC system and the NBS card number that was added to the linkage configuration. However, further study is necessary to justify the assumption that early improvements in linkage success occurring in 2008 are truly derivative of the transition to the EBC system and addition of the NBS card number to the birth certificate.

Table 2 Newborn screening and live births records linkage results, Michigan, 1/2007–3/2008

Birth year	Total	Expired infants	Matched		Un-matched	
	N	*N*	*N*	%	*N*	%
2007	123,981	383	122,515	98.8	1,083	0.9
2008	18,197	51	18,075	99.3	71	0.4
Total	142,178	434	140,590	98.9	1,154	0.8

The MDCH NBS Follow-up Program followed-up the un-matched live birth records by searching the NBS database for these records, case by case, attempting to correct for errors in date ranges, names, or other fields. Among hospital births, if no matches were found then NBS Follow-up staff contacted the birthing hospital to inquire whether the infant was screened; often, this step indicated whether an infant had expired prior to being eligible for screening. Next, remaining live unmatched infants' parents were sent a letter requesting them to either get their child screened or return a completed parental refusal letter that is included along with a stamped envelope with the solicitation for screening. NBS Follow-up staff utilized the Michigan Care Improvement Registry (MCIR) database, a statewide immunization registry with approximately 53 million records, to obtain parents' addresses to send follow-up letters to parents when the NBS database did not have a correct address for the infant's parent. Both the original hard copy NBS cards (scanned and electronically accessible) and MCIR data were used as references to assess the impact of data entry errors (i.e., significant spelling errors and missing data) on record linkage.

While the record linkage statistics were calculated for all resident births from 2007 and 3 months of 2008, the follow-up of the unmatched records was not initiated until December of 2007, when the linkage configuration was finalized. Thus, the total number of children included in the follow-up results does not amount to the total number of

Table 3 Follow-up results of unmatched live birth records received December 2007 through March 2008

Follow-up result	Frequency	%
NBS already completed	286	48.4
Screened after follow-up	9	1.5
Parental refusal of screening at follow-up	48	8.1
N/A (Aged >1 year)	8	1.4
Out of state	14	2.4
Parental refusal at time of screening	32	5.4
Expired	6	1.0
Lost to follow-up	188	31.8
Total[a]	591	100.0

[a] *Note*: The total listed in Table 3 does not match the total number of live birth records unmatched to NBS records because follow-up was not initiated until the linkage configuration was finalized in December of 2007. Thus, follow-up was only conducted on 591 unmatched records from data received December 1st 2007 or thereafter

live births that were unmatched to NBS records in this analysis.

Of the nearly 600 un-matched live birth records that have been sent for follow-up, nearly half had received a newborn screen meaning the records were not matched due either to significant data entry errors or name changes, the latter being quite rare. The most common reason for infants having a newborn screen un-matched is error in the last name of mothers and/or infants. As indicated in Table 3 approximately 14% of un-matched records were caused by parental refusal of NBS either at birth or at follow-up. Of the 32% of records unable to be followed-up, nearly all were due to incorrect addresses in NBS and MCIR databases, or parental refusal to respond to letters sent by NBS Follow-up staff. In aggregate, the number of births lost to follow-up amounts to 0.13% of the 142,178 records eligible for linkage. Of note is that both adoption and name changes play a role in explaining unmatched records, particularly those lost to follow-up. Nine children, 1.5% of those sent for follow-up, were re-screened as a result of the linkage program during the study period, one of which was confirmed as having sickle cell trait.

The results of our linkage efforts indicate that the use of Link Plus is applicable not only for cancer registry data, but also for other types of public health data files, MCH-related in our case. We have developed a program that can successfully link >99% of our live birth records to newborn screens and identify infants missed by NBS. Follow-up of the unmatched records revealed that the majority had been screened or to a lesser degree were derivative of parental refusal of NBS. These results also demonstrate the utility of Link Plus software for matching other MCH administrative databases for further opportunities to conduct quality assurance investigations.

The NBS Follow-up Program is currently working to improve the efficacy of our record linkages by increasing the number of hospitals that include the NBS card number on each birth record. As of March, 2008, there were only about 40% of birth records that included the NBS card number. Considering that our linkage success rate improved following the addition of the NBS card number to the birth record (based on matching of $\sim$18,000 birth records) and that the sole intention of the inclusion of the variable on birth record is to facilitate matching, we feel that increasing the proportion of hospitals entering the NBS card number on birth records may further improve our linkage success rate. Our analysis also indicates that addressing NBS data quality may improve our linkage process. Specifically, verifying the data entry of name fields within the NBS database for accuracy by others than the data entry staff would likely reduce the non-match rate. We are also now making note of records sent for follow-up after manual review of matched records (i.e., false positive matches) as opposed to those unmatched by the application of the Link Plus software to gain a better understanding of the linkage process.

The current linkage process is now an ongoing activity of the NBS Follow-up program. The methods developed for linking live birth and NBS data have been used to match NBS and Medicaid records to assess service utilization among detected cases. In the future we plan to use Link Plus to link NBS data to other administrative datasets within MDCH to provide further opportunities for program evaluation and public health research. In the current economic climate we feel it is imperative to better use *existing* resources, including freely available linkage software, to advance MCH data utilization and quality. Specifically, as we have demonstrated in this study, existing resources can be linked in innovative ways to answer program questions and identify targets for interventions; we are also using our linkages to compare NBS and vital statistics as an indicator of data quality. Thus, the public health implication of this study are that we have demonstrated that freely available software, Link Plus, can be used to successfully match records of MCH databases thereby providing an opportunity for further research and quality assurance investigations. We hope the results of the present study will inspire others to either publish or commence their own linkage research to drive the field of MCH forward.

References

1. Registry Plus, a suite of publicly available software programs for collecting and processing cancer registry data. Atlanta (GA): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease

Prevention and Health Promotion; 2008. Available at www.cdc.gov/cancer/npcr/. Link Plus accessed 12/06/2008.
2. Fellegi, I. P., & Sunter, A. B. (1969). A theory for record linkage. *Journal of the American Statistical Association, 64*, 1183–1210. doi:10.2307/2286061.
3. Fundamentals of Linking Public Health Datasets. Link Plus Probabilistic Record Linkage Software. Retrieved october 12, 2008, from www.nri-inc.org/projects/OSA/LinkPlusOverviewMarch2007.pdf.
4. National Archives Trust Fund Board, "The 1910 Federal Population Census: Catalog of National Archives Microfilm", 1996. National Archives Trust board.