

# NEWBORN SCREENING AND LIVE BIRTHS RECORDS LINKAGE CONFIGURATION EVALUATION

Mary Kleyn, MSc<sup>1</sup>, Steve Korzeniewski, MSc, MA<sup>1</sup>



<sup>1</sup>Michigan Department of Community Health, Division of Genomics, Perinatal Health and Chronic Disease Epidemiology

## Background

In 2007, the Michigan Department of Community Health's Newborn Screening (NBS) Follow-up Program began routinely linking NBS and live births records.<sup>1</sup> The linkages serve a three-fold purpose: 1) the NBS Program can identify potentially unscreened infants, 2) the Vital Records Program can assess data quality compared to NBS records, and 3) these linkages allowed the NBS Program to determine the feasibility of linking NBS data to other databases. Through linkages with other sources such as Medicaid and Children's Special Health Care Services, the NBS Program can more effectively track the healthcare utilization and morbidity of children diagnosed through NBS. Thus, these linkage methods can be used by maternal and child health programs to provide opportunities for program evaluation through the use of existing databases. However, the best configuration of variables for linking NBS and live births records has yet to be determined.

## Study Question

Which linkage configuration produces the highest number of matches of NBS and live births records?

## Methods

### Study Population

A subset of live births and NBS records of infants born in Michigan from January 1, 2009-March 15, 2009 was used.

### Software

SAS version 9.1 was used to create text files from both the live births records and NBS records.

Link Plus software, a probabilistic record linkage program, was used to match the live births and NBS records. Link Plus was developed initially for cancer registries at the Centers for Disease Control and Prevention's Division of Cancer Prevention and Control.

### Linkage Process

Separate linkages were run for singletons and multiples. The base configuration included child's name and birth date. Birth order was also included in the base configuration for multiples. Identifier variables common to both NBS and live births records were added to the base configuration, and the match rate was examined to determine the best configuration. For all configurations, blocking variables were mother's first and last name. Table 1 reports the blocking variables and associated phonetic systems used, as well as the matching variables and methods included in the linkage process.

Both infant's and mother's names were divided into 2 variables: first name and last name. The variable NBS card number refers to the unique identifier located on each NBS card. Ideally, hospital's will record this number on the birth certificate for each infant.

Table 1. Phonetic system and matching method for blocking and matching variables, NBS and live births records linkage configuration

Blocking Variables		Matching Variables	
Name	Phonetic System	Name	Matching Method
Mom's First Name	Soundex <sup>2</sup>	Mom's First Name	First Name
Mom's Last Name	Soundex <sup>2</sup>	Child's First Name	First Name
		Mom's Last Name	Last Name
		Child's Last Name	Last Name
		Infant's Birth Date	
		Mom's Birth Date	
		Mom's SSN	
		NBS Card Number	Value-specific
		Medical Record Number	
		Mother's Zip Code	Zip-code

Table 2. Linkage configurations and match rates between live births and NBS records among a subset of singletons born January 1, 2009-March 15, 2009, Michigan (N=7,049 total)

Matching Variables	Number Unmatched	Match Rate (%)
Base (Child's Name and Birth Date)	5,722	18.8
Base + Mother's Name	348	95.1
Base + Mom's SSN	2,880	59.1
Base + Mom's Birth Date	2,147	69.5
Base + Mom's Name + Mom's Birth Date	166	97.6
Base + Mom's Name + Mom's Birth Date + Mom's SSN	105	98.5
Base + Mom's Name + Mom's Birth Date + Mom's SSN+ NBS Card Number	91	98.7
Base + Mom's Name + Mom's Birth Date + Mom's SSN+ NBS Card Number+ Medical Record Number	76	98.9
Base + Mom's Name + Mom's Birth Date + Mom's SSN+ NBS Card Number+ Medical Record Number + Mother's Zip Code	58	99.2

Note: All configurations were blocked on mother's name.

Table 3. Linkage configurations and match rates between live births and NBS records among a subset of multiples born January 1, 2009-March 15, 2009, Michigan (N=314 total)

Matching Variables	Number Unmatched	Match Rate (%)
Base (Child's Name + Birth Date + Birth Order)	256	18.5
Base + Mom's Name	37	88.2
Base + Mom's SSN	261	16.9
Base + Mom's Birth Date	86	72.6
Base + Mom's Name + Mom's Birth Date	11	96.5
Base + Mom's Name + Mom's Birth Date + Mom's SSN	7	97.8
Base + Mom's Name + Mom's Birth Date + Mom's SSN+ NBS Card Number	5	98.4
Base + Mom's Name + Mom's Birth Date + Mom's SSN+ NBS Card Number+ Medical Record Number	5	98.4
Base + Mom's Name + Mom's Birth Date + Mom's SSN+ NBS Card Number+ Medical Record Number + Mother's Zip Code	2	99.4

Note: All configurations were blocked on mother's name.

## Results

Overall, we attempted to link 7,049 singleton live births and 314 multiple live births. With the base configuration, 18.8% of the singleton live births records (Table 2) and 18.5% of the multiple live births records (Table 3) were linked to NBS records. Mother's name was added to the base configuration, increasing the match rate to 95.1% for singletons and 88.2% for multiples. Adding Mom's SSN to the base configuration was more effective for singletons than for multiples. For singletons, the base configuration plus Mom's SSN matched 59.1% of the records. For multiples, adding Mom's SSN to the base configuration only matched 16.9% of the records. Adding mom's birth date to the base configuration was roughly equally effective, matching 69.5% of the singletons and 72.6% of the multiples. Adding mother's birth date and mother's name to the base configuration increased the match rates to 98.5% of singletons and 96.5% of multiples. The addition of the following variables improved the match rates to 99.2% for singletons and 99.4% for multiples: mothers social security number (SSN), NBS card number, medical record number, and mother's zip code.

## Conclusions

Due to the large number of NBS records missing child's first name and frequent child's last name changes from the NBS to live births records, child's name and birth date only linked ~18.5% of NBS and live births records. Mother's name was found to be the strongest matching variable between live births and NBS records. The best linkage configuration included the following matching variables: infant's name, infant's birth date, mother's name, mother's birth date, mother's SSN, NBS card number, medical record number, mother's zip code, and birth order for multiples only. If you are unable to access all of those variables, child's name and birth date and mother's name and birth date will most likely match over 95% of live births records to NBS records.

## Public Health Implications

Probabilistic linkages can be used to facilitate program evaluation efforts through the identification of potentially unscreened infants and the enhancement of research opportunities by increased utilization of available resources. We recommend that maternal and child health programs, specifically newborn screening programs, use Link Plus, a free software, to assist with probabilistic linkages between databases.

## Acknowledgments

The authors would like to acknowledge:

-G. Gu (Science Applications International Corporation), K. K. Thoburn (Northrop Grumman), and J. D. Rogers (Centers for Disease Control and Prevention) for their advice and guidance regarding the use of Link Plus software.

-Violanda Grigorescu (MDCH) and Glenn Copeland (MDCH) for their vision, guidance and support in performing this linkage

## References

- Korzeniewski SJ, Grigorescu V, Capeland G, Gu G, Thoburn KK, Rogers JD, Young WI. Methodological innovations in data gathering: newborn screening linkage with live births records, Michigan, 1/2007-3/2008. *Matern Child Health J* 2009 (ahead of print).
- National Archives Trust Fund Board. "The 1910 Federal Population Census: Catalog of National Archives Microfilm". 1996. National Archives Trust Board.