

# DATA INTEGRATION AND WAREHOUSING: COORDINATION BETWEEN NEWBORN SCREENING AND RELATED PUBLIC HEALTH PROGRAMS

Bradford L Therrell Jr

US National Newborn Screening and Genetics Resource Center, Austin, Texas 78757, USA, and  
Department of Pediatrics, University of Texas Health Science Center at San Antonio,  
San Antonio, Texas 78229, USA

**Abstract.** At birth, patient demographic and health information begin to accumulate in varied databases. There are often multiple sources of the same or similar data. New public health programs are often created without considering data linkages. Recently, newborn hearing screening (NHS) programs and immunization programs have virtually ignored the existence of newborn dried blood spot (DBS) newborn screening databases containing similar demographic data, creating data duplication in their 'new' systems. Some progressive public health departments are developing data warehouses of basic, recurrent patient information, and linking these databases to other health program databases where programs and services can benefit from such linkages. Demographic data warehousing saves time (and money) by eliminating duplicative data entry and reducing the chances of data errors. While newborn screening data are usually the first data available, they should not be the only data source considered for early data linkage or for populating a data warehouse. Birth certificate information should also be considered along with other data sources for infants that may not have received newborn screening or who may have been born outside of the jurisdiction and not have birth certificate information locally available. This newborn screening serial number provides a convenient identification number for use in the DBS program and for linking with other systems. As a minimum, data linkages should exist between newborn dried blood spot screening, newborn hearing screening, immunizations, birth certificates and birth defect registries.

## INTRODUCTION

At (and sometimes before) birth, patient demographic and health information begin to accumulate in varied databases. Even in public health programs that reside within the same agency, there are often multiple sources of the same or similar data. It is not unusual for these databases to exist separated from one another, often without the responsible parties having knowledge that other databases containing similar elements exist. In cases where such knowledge is present, there are often political or administrative reasons that influence the creation of separate data "silos." Additionally, new public health programs are often created without considering possible linkages to data with common elements that might exist in another program. Sometimes, persons developing new public health programs may not even think to consider the existence of databases that could be useful in supporting the new program.

For example, in recent years newborn hearing screening (NHS) programs and immunization programs seeking to develop databases for patient tracking have virtually ignored the long time existence of newborn

dried blood spot (DBS) newborn screening databases containing relevant demographic information eventually duplicated in their 'new' data systems. At a minimum, all of these programs (NHS, DBS, and immunizations) have some basic newborn patient demographic information in common since all three deal with the newborn soon after birth. In addition, common demographic elements exist between these three programs and birth certification programs, among others. It is a waste of time and effort when all of these systems duplicate data fields that could be easily shared if an overall data integration plan were thoughtfully developed and implemented.

In order to decrease duplicative data entry, database maintenance, and human errors in data handling, some progressive public health departments are developing data warehouses of basic, recurrent patient information, and linking these databases to other health program databases where programs and services can benefit from such linkages. Thus, in data warehouses, patient demographic data (and possibly other common data elements) are uploaded from any one of a number of possible sources to reside in a centrally accessible database (ie the data warehouse)

potentially accessible to multiple users. These individual users (ie other health programs) may have separate specialized databases and/or software for unique program-specific purposes separate and apart from the data warehouse, but access to basic patient information available in the data warehouse makes duplicate entry of this information unnecessary. Demographic data warehousing not only saves time (and money) by eliminating duplicative data entry for programs utilizing it, but it also reduces the chances of errors within the data since human interactions with data entry (the largest source of potential error) are minimized or eliminated.

In most US newborn screening programs, the DBS specimen is collected within the first 1-3 days after birth. Thus, newborn screening information is usually the first patient demographic data available, even earlier than official birth certificate information in most systems, including electronic birth certificate systems. In a data warehousing system, therefore, it is logical to construct the system to take advantage of these early data. While newborn screening data are usually the first data available, they should not be the only data source considered for early data linkage or for populating a data warehouse. Birth certificate information should also be considered along with other data sources for infants that may not have received newborn screening, who may have been born outside of the jurisdiction or whose birth certificate information is not locally available. In any case, it is prudent to consider the possible linkages of newborn screening information with other newborn or child health programs and birth certificates in order to improve overall public health program efficiency and cost effectiveness.

## NEWBORN SCREENING SYSTEMS

### **Newborn dried blood spot screening**

In processing DBS newborn screening specimens, the first step is to obtain blood from the newborn for the required testing and submit the specimen with identifying information to the newborn screening laboratory. In the US, there is a published national standard that describes this process in detail including: quality assurance steps, minimal data to be obtained, sampling procedures, and related processes (Hannon *et al*, 1997). Newborn demographic information is usually printed by hand onto the collection device. In some electronically advanced newborn screening/hospital systems, the patient demographic information may be internally downloaded from admissions records to the facility laboratory or nursery, the serial number of the DBS collection card and the date of collection added to the data, and a label printed and attached to the collection card prior to mailing. Thus,

the manual step of completing the demographic information portion of the dried blood collection device is eliminated. Often, for quality control purposes, a handwritten or electronically maintained logbook is also kept in the newborn nursery or birthing facility laboratory, into which information tracking the newborn screen is maintained including date of sample collection and result of the testing when it has been completed. Soon after birth (usually 24-72 hrs), a DBS sample is collected, dried, packaged for shipment, and sent to the screening laboratory by mail or courier.

Upon receipt at the testing laboratory, the patient identifying information is usually keyed into a laboratory database [or uploaded from electronic transmittal systems, in more technologically advanced systems usually linking on collection device serial number], the laboratory tests are performed, the results are recorded, and the test results are reported back to the submitting facility and/or to the physician of record, depending on program rules and regulations. If abnormal or unsatisfactory results are obtained, then notification is also given to a follow-up coordinator responsible for ensuring that confirmatory diagnostic or repeat testing occurs, depending on the urgency of the testing results. This coordinator may contact subspecialists for assistance, or in emergency situations, may directly contact the patient's family.

### **Newborn hearing screening**

In a typical NHS system, the newborn demographic information in the hospital admissions database is accessed for input into the hearing testing equipment in the newborn nursery [alternatively, this information may be obtained directly from the parent(s)]. Testing is performed and the testing results are recorded in the patient's chart (and may also be stored in a database at hospital or associated with the testing equipment). If a second test is required by the testing protocol, then these test results are also entered into the patient's chart (and database). If further testing is needed, then this information is usually shared with the parents and audiological follow-up is recommended (perhaps even scheduled). In many hospital-based NHS systems, the hospital is responsible for following up on results of the confirmatory testing and ensuring that the patient is aware of any needed intervention services. The process of obtaining intervention (hearing aids, etc.) may also be a hospital responsibility, but often this step in the system is left to others. If centralized data reporting is required by the government agency overseeing NHS, then the hospital is responsible for maintaining and transferring testing results from the hospital database to a central data repository.

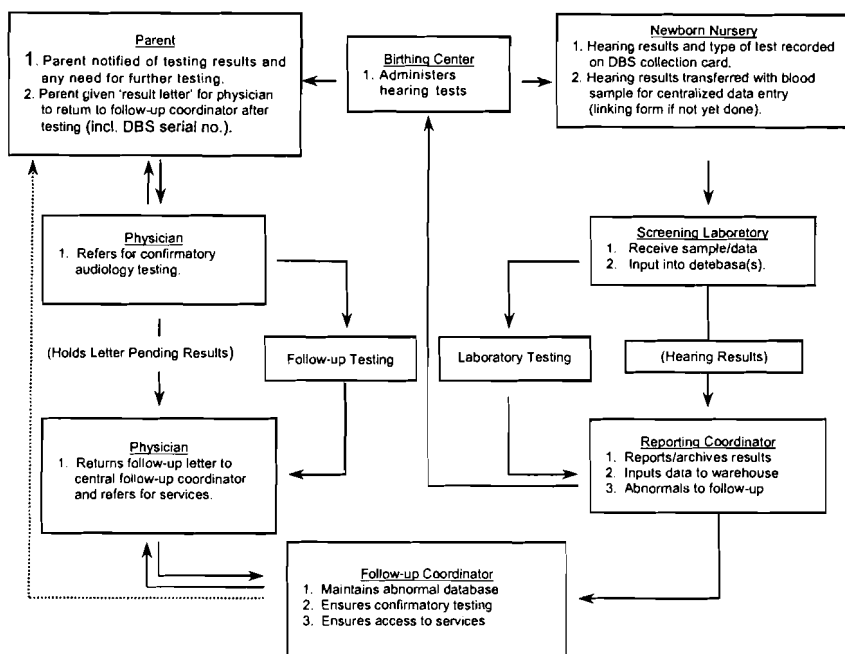


Fig 1. Diagram showing one mechanism of integrating newborn hearing screening and data into an ongoing dried blood spot (DBS) screening program.

Alternatively, as an aid to data management and centralization of the data, some NHS programs have added a limited number of extra data fields to the DBS collection form so that the hearing screening data can be recorded and transferred with the DBS data (see Fig 1). In such systems, NHS results are recorded on the DBS collection card, along with the patient's demographic information, and submitted to the central newborn screening testing laboratory.

Here they are recorded along with the other patient information and then transferred to a hearing follow-up data system. If the hearing results are not available in the nursery at the time the DBS sample is submitted, a tear out sheet is usually included as part of the DBS collection device so that this sheet can be removed and submitted later after hearing testing has been completed. The tear out sheet contains the NBS serial number for linking to patient's demographic information and can be color-coded to aid in recognition. In programs where NHS and DBS follow-up are combined, the data may reside in the same system, but this is not necessary if follow-up and program evaluation are facilitated through other databases. The DBS card merely provides a mechanism for transmitting data in a fast and efficient way to a central database. The

addition of NHS data fields on the DBS card also serves as an educational reminder to hospitals not yet performing newborn hearing screening. Data submitted in this way, cover essentially all newborns in the jurisdiction and provide not only a mechanism to enhance follow-up, but also allow for a count of patients not yet receiving NHS services.

## DISCUSSION

A unique linking number is often cited as the critical missing element in linking or integrating data systems. Patient names or other data elements or combination of data elements have been used for linking, but they are usually complex and subject to a number of different caveats in order to make them useable. For example, if a name is used, then the spelling must be exact. Some programs have tried to develop unique linking numbers from selected information fields in the data, and some have even tried to create pseudo-social security numbers, based on the mother's social security number, but problems have arisen in the case of multiple births and/or early infant deaths. The simpler and more effective approach to data linking has been to provide and use the unique serial numbers preassigned and preprinted on the DBS newborn screening collection devices. In addition to providing a progressive

sequential numbering system, 'smart' DBS serial numbers can include data elements for other program components including the year in which the cards were ordered and the birth location (ie the particular state in which the birth occurred). A check sum character can also be included at the end of the serial number to ensure the accuracy of the number when it is entered into a computer system. Newborn screening collection devices can be manufactured with multiple serial number stickers printed and attached to the collection devices. These stickers can be formatted with serial numbers and/or barcodes, and can be easily peeled away from the DBS collection device and attached to other paperwork. In this way, the identification number can be easily appended to other linkage documents. For example, the linking serial number can be affixed to the birth certificate, to the patient's medical record, and/or to the tabular listing of patients maintained in the nursery for newborn screening submission and result record keeping. In cases where newborns receive more than one newborn screen, the initial screening collection device provides the permanent tracking number to which all subsequent specimens from that patient are linked.

In the US, essentially all programs now include a serial number on the DBS collection device, and many have it bar coded for quick entry into automated data systems. While many programs use the DBS serial number as a means of tracking the patient's collection device from the birthing facility to the testing laboratory and beyond, the DBS serial number was originally created for use in inventory management. In inventory management systems, the serial numbers of collection devices shipped to various birthing facilities are recorded so that the facility can be easily identified should its identification fail to be recorded on a submitted specimen (eg if a facility loaned some collection devices to another facility). In cases where such an inventory control system is linked to the laboratory data management system, submitter data (uploaded from the inventory system) can be automatically supplied to the data management system at the time the serial number is entered. In this way, the keystrokes necessary to type the submitter's identity and address can be conserved during the data entry process. Editing overrides can be provided in data entry software so that address changes can be made if the submitter information on the form differs from that contained in the inventory control system. Additionally, bar coded serial numbers that include check sum characters automatically validate the number at data entry so that bar codes not only speed the entry process but also improve the accuracy of data entry.

As noted earlier in this report, the current NHS environment includes a large number (approximately 50%) of newborns with positive hearing screening tests who do

not receive confirmatory testing and/or intervention services. By integrating newborn hearing results with DBS newborn screening programs already in place, and by using existing DBS follow-up systems as a model, NHS follow-up should be able to improve in effectiveness and efficiency (American College of Medical Genetics, 2000). In NHS programs the overall data flow is similar to that of DBS screening. While many NHS programs are organized so that screening and follow-up responsibilities reside with the birthing facility, especially in small states and in pilot programs, the need for improved follow-up, monitoring of service access and delivery, and centralized program evaluation data creates a data flow that is operationally similar to that of DBS programs. Thus, it seems reasonable to consider using the DBS collection device as a mechanism for transmitting NHS data to a central database utilizing the data capturing processes already in place in an existing DBS program. Indeed, some state NHS programs have already taken advantage of this mechanism for data transmittal. Currently, data fields for NHS information are included on the DBS collection devices in: Delaware, Maryland, Michigan, Minnesota, Mississippi, Missouri, Nevada, Oklahoma, Oregon, Tennessee, Washington, and Wisconsin. Basic information such as type of test and test results can be easily added to most collection forms using 1-digit coded responses, and entry of these data have little impact on the data entry workload. For example, minimal data may include recording a coded response for whether a test was performed, the hearing results for each ear, and/or the equipment used for testing. While modification of an existing DBS data system to accommodate added data fields is often required, the modifications are usually straightforward, easily accommodated, and advantageous to the overall screening system. At least one state, Utah, has initiated data linking/integration using the birth certificate as the primary data source to which all programs link, but utilizing the DBS serial number as the unique linking number. While the experiences thus far in Utah have been positive, in some states where a provision for recording the DBS serial number on the birth certificate exists, completion of this data field has not been required, and therefore its potential for linking has been limited because the field is often left blank.

The patient demographic information required for newborn hearing follow-up is similar to that required for routine DBS follow-up. The minimal data elements suggested for DBS newborn screening are specified in a national standard (now in its fourth revision) and are limited to the essential data elements needed for identifying patients considered at risk as a result of screening (Hannon *et al*, 1997). Already captured in most DBS databases are: infant's name, address, phone number, physician of record,

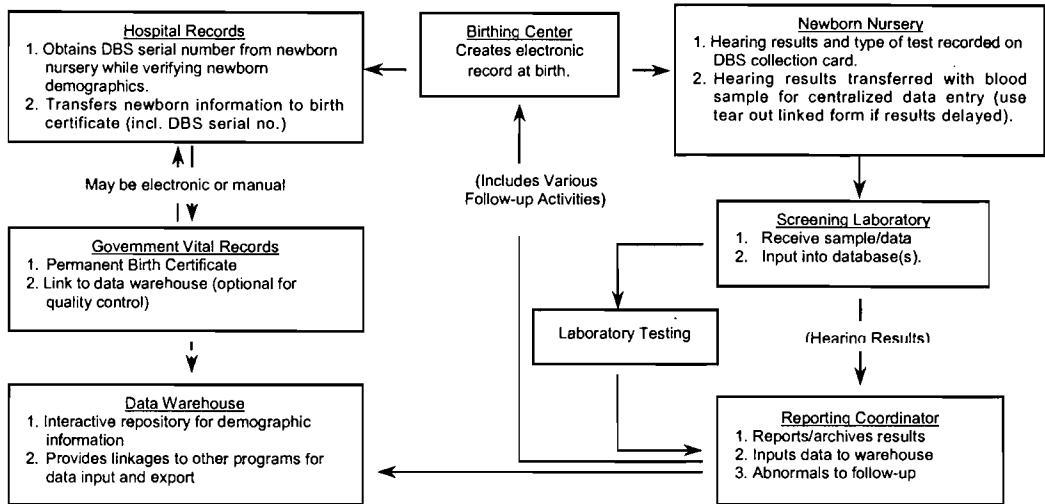


Fig 2. Diagram of newborn screening data flow using the warehousing concept and linkages with vital records as a means of ensuring that all newborns receive both a newborn screen and a birth certificate.

physician's phone number, etc. By limiting the information captured on all patients to essential elements, the mass of population data entered into the data system is minimized. Case specific information on the small number of patients with abnormal test results can be obtained as part of the follow-up process. Limiting the case specific data in this way adds to the overall efficiency of the data collection/data entry part of screening by leaving non-critical information to be obtained only on the less than 1% of patients for whom it is needed. Thus, for example, additional data elements

specific to hearing loss can be recorded in a case management database (or other appropriate file), in a process similar to that generally employed in the DBS newborn screening follow-up. In cases where a screening program may wish to monitor data on risk factors, additional data elements could be added to the DBS form. However, care must be taken to ensure that the information anticipated from added data is valid and useful, since data entry expenses will be increased by any additional information that must be input. Overall, the logistics of recording NHS results on the DBS form at

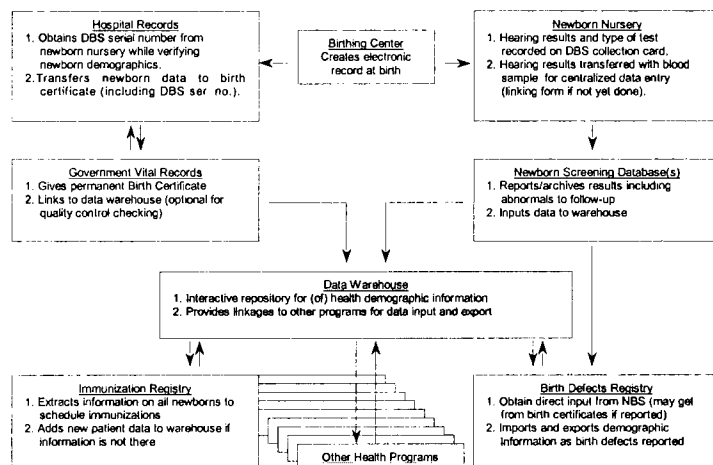


Fig 3. Diagram showing data flow into and out of a data warehouse, with particular attention to interactions with newborn screening, birth certificates, immunization registries, and birth defects registries.

the newborn nursery may be more complex than a stand-alone system for NHS, but data transfer between the screening systems can be managed by most hospitals with little additional effort, and the system savings in reducing duplication of effort is cost beneficial.

Timely information available from DBS and NHS early screening can also provide demographic information that could be useful for databases associated with birth certificates, childhood immunization programs and birth defects registries, among others. If newborn screening data are not used to populate the birth certificate database, they can still be used alternatively as a quality control check to ensure that birth certificates exist for each newborn receiving a newborn screening test (see Fig 2). Reverse validation may also be beneficial in assuring that each recorded birth has received an appropriate newborn screening test (although programs should be sensitive to the fact that birth certificate information is not collected to be used punitively).

A truly comprehensive data warehousing system would theoretically include mechanisms for integrating initial patient information from any program that may have the data available, whether or not it originated in the newborn screening program (see Fig 3). Thus, for example, if a child was to be given an immunization, an inquiry of the warehouse should indicate whether or not there was basic demographic information available, and additionally whether or not there was an immunization history. If demographic data were missing, then they would be input at that time and would be available for future inquiries, whether or not the inquiry originated with the immunization program.

In the typical birthing facility billing (and tracking) system, demographic information on the newborn is available almost immediately following the birth (technically a new hospital admission), and this information is available to hospital personnel in both the nursery and the hospital laboratory through internal communication pathways. This information has the potential to form the nucleus of a patient information data record from which basic patient information can be accessed and extracted as health-related encounters occur within the hospital. Therefore, these data offer the potential for uploading pertinent patient information to the DBS request card, the NHS data form and the birth certificate. Even though most hospitals now have electronic databases of newborn demographic information, and could share this information electronically within the hospital, it is still the practice in most birthing facilities to manually record patient

information on the newborn DBS screening collection device or other patient records. Additionally, many facilities also keep manual tracking records of DBS specimen submissions in order to ensure that results are received on all newborns transiting the newborn nursery.

Birthing facilities maintain a supply of DBS collection cards, and in almost every state, the DBS collection card contains a preprinted unique serial number defined by the state newborn screening program, primarily for inventory purposes and patient tracking within the system. This serial number provides a convenient identification number for use not only in the DBS program, but also for patient linking to other systems. It is usually the case, that other health program data systems also have identification numbers that could be used in such a linking system or data warehousing system, but since the DBS serial number is the first one to be used chronologically, and since it can exist as multiple preprinted labels that could be peeled off the DBS card and affixed to other patient records, it seems to be a simple solution to the problem of an identifier that can link systems to the patient.

While to some it may be appealing to develop new programs from the ground up without interference from outside influences, it is usually the case that new programs can benefit from the experiences of older programs. This is particularly true of data systems in which there has historically been a rapid change in data technology resulting in systems modifications that are expensive and complex. The experiences of 'old' data systems such as those associated with DBS newborn screening provide a wealth of development experience, and the opportunities for linkage and 'exploitation' by other health systems with similar data needs should not be overlooked. It is incumbent on all public health programs to continue to explore ways in which data duplication can be minimized with consequent savings in funds previously used for data entry so that health monies can be better utilized for service delivery.

## REFERENCES

- American College of Medical Genetics. Policy statement of the American College of Medical Genetics on universal newborn hearing screening. *Genet Med* 2000;2:149-50.
- Hannon WH, Boyle J, Davin B, *et al.* National Committee for Clinical Laboratory Standards. Blood Collection on filter paper for neonatal screening programs - 3rd ed. Approved Standard. NCCLS document LA4-A3. Wayne, PA: NCCLS, 1997.