DETECTING DATA VARIATION IN DISPARATE PERINATAL CLINICAL SYSTEMS USING A TRIANGULATED APPROACH OF DATA CONCEPT ANALYSIS, CLINICIAN PERCEPTION STUDY, AND PATIENT RECORD REVIEW

by

Val Norman Hicken

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Medical Informatics

The University of Utah

May 2004

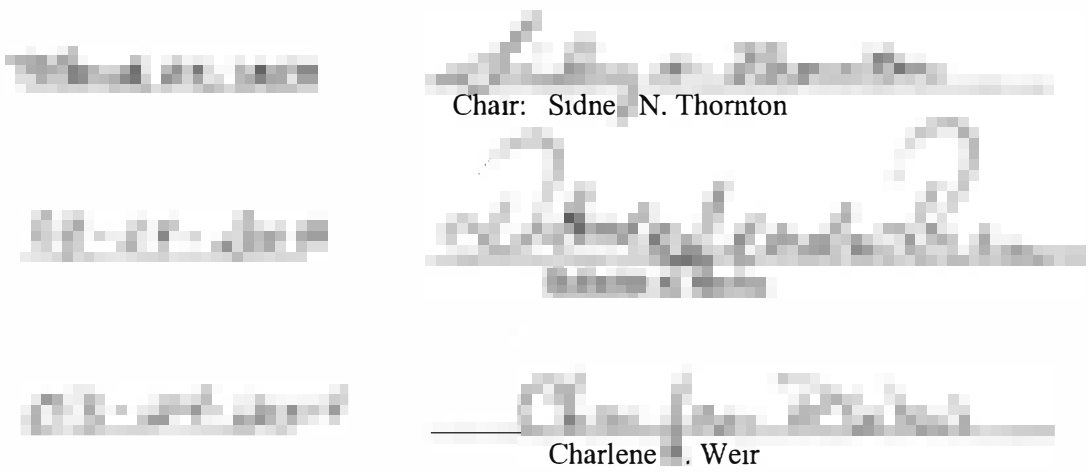
THE UNIVERSITY OF UTAH GRADUATE SCHOOL

SUPERVISORY COMMITTEE APPROVAL

of a thesis submitted by

Val Norman Hicken

This thesis has been read by each member of the following supervisory committee and by majority vote has been found to be satisfactory.



Chair: Sidne N. Thornton

Charlene . Weir

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

FINAL READING APPROVAL

To the Graduate Council of the University of Utah:

I have read the thesis of Val Norman Hicken in its final form and have found that (1) its format, citations, and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the supervisory committee and is ready for submission to The Graduate School.

March 24, 200

Date

Sidney N. Thornton
Chair: Supervisory Committee

Approved for the Major Department

Reed M. Gardner
Chair/Dean

Approved for the Graduate Council

David S. Chapman
Dean of The Graduate School

ABSTRACT

Delivery of high quality health care requires access to complete and accurate patient information. Variation in data context and content across disparate clinical systems adversely affects the integration of information needed for effective patient care and outcomes research. This study detects the extent and nature of data variation across three disparate clinical systems used along different points of the perinatal care continuum at Intermountain Health Care (IHC).

Three analytical methods were used to examine data variation: data structure analysis; clinician perception of missing data elements; and patient record review of key data values. Knowledge acquisition techniques and consensus among clinical domain experts were used to select sample data elements for the data structure analysis. Findings revealed only 17% of the sample data elements had compatible structure and meaning across the prenatal, labor and delivery (L&D), and newborn intensive care (NICU) clinical data systems. Impact on clinician efficiency from missing and contradicting information in nonintegrated perinatal systems was captured and analyzed using a Critical Incident Technique-based clinician survey. In a 1-month period, 75% of responding clinicians reported missing data and 34% reported contradicting data. The time taken to resolve problems from 1 month's missing data was estimated to be 231 hours for 23 clinicians. Data values from patient records for eight laboratory results were compared across the three perinatal systems. The best match across any two systems was

88% (blood type) and the worst was 0% (antibody screen, chlamydia). The highest incidence of contradicting data was 2.5% for blood type.

Comparing agreement of the three methods, "triangulation," gave additional insight into IHC's data variation problem. The data model study and the patient record review study showed missing data element problems beyond what clinicians perceived. In all, the consistency of data capture in the three perinatal systems at IHC is worse than expected. The data necessary to computationally execute the logic of the perinatal care process models is intermittent and unreliable. Rework of the perinatal applications based on a uniform data model and standard terminologies will provide an infrastructure to achieve IHC's vision of interdisciplinary care.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Page

ACKNOWLEDGMENTS

I acknowledge the thorough instruction and substantial time outside of class by the professors in the Medical Informatics Department at the University of Utah, and the expertise and insight of the IHC data managers, analysts and clinicians on the Women and Newborns Clinical Program, and to Intermountain Health Care for their funding and use of facilities for this project. I also want to thank my wife, Marjann, and my children for their love and support.

INTRODUCTION

Data variation in the health care industry

Clinicians have always gathered and integrated patient information so they could take appropriate action for medical care. Data were gathered from various sources, and if the desired information was not available, or was questionable in one source, another source was sought. Organized methods for data capture, storage and retrieval have evolved into what is now referred to as systems. This study will use the word "system" to refer to a group of devices and methods forming a network dedicated to gathering and distributing information, and can be paper-based or computer-based. As systems grow in scope, they often have concepts that overlap with other systems. The overlap across systems introduces the potential for data variation through duplication of data entry and differing concept definition or context of use. Precise definitions of commonly used terms in this study are found in Appendix A. The term "perinatal continuum" will be used in this study to describe the continuum of care that ranges from prenatal through well baby.

Magnitude and prevalence

In 1991 the Institute of Medicine (IOM) issued a report stating the need for patient data to be integrated into a computerized patient record (CPR) so that key clinical information could be readily shared across settings and among clinicians.[1] Subsequent IOM reports have reemphasized the need for systems to share accurate and complete

clinical data.[2-4] The CPR has evolved in purpose to become an electronic medical record (EMR). The terms EMR and the electronic health record (EHR) will be considered synonymous within this study. Functional requirements of the EHR have been recently published by the IOM.[5, 6]

Since 1991 there has been a substantial growth of the number of integrated delivery systems (IDS). An IDS is composed of health care providers, service providers, and facilities organized to provide a continuum of health care services to a defined population. The 1997 IOM report explained, that to manage the delivery of care in an IDS, a health system must have efficient and accurate ways of capturing, managing, and analyzing clinical data collected at all the different sites where care is provided. With each new IDS formed, new disparate systems are brought together that have potential data overlap.

Ambulatory care data systems frequently record data that are inadequately organized, lack documentation of key aspects of care, and show inaccurate diagnostic coding. Outpatient records may exhibit greater variance in quality than inpatient records. Neither established standards nor review organizations exist for outpatient records like those established for inpatient records.[2] The 2001 IOM report focused more on errors occurring in hospitals, while the 2003 IOM report on safety points out that serious safety issues existing in outpatient settings could dwarf the number errors in inpatient settings.

Failure to share patient information across data systems can lead to inefficiency and reduce quality of care.[5] A study by Allen et al. categorizing clinicians' information needs, supports the assertion that errors may occur if key information needs go unanswered when clinicians are making clinical decisions.[7] When the context of a

concept differs across systems, mapping is hindered and aggregate knowledge about that concept is reduced. Studies show that redundant records lead to errors, extra effort, misdirected data, over reliance on the spoken word, inaccuracies, information loss, limited standardization, miscommunications, decision changes, and limited outcomes evaluations.[8] One of the leading causes of medical errors occurs when clinician's unmet information needs result in delayed and uninformed decisions.[9] The 1991 IOM report, "The Computer-based Patient Record: An essential technology for health care," brought attention to the need for the implementation of computer-based patient records.[1]

The long-term solution for accessible data requires electronic clinical data systems that span health care settings. New health care information infrastructures must provide methods to capture and share the data currently in paper medical records in order to contribute to quality reporting and improvement.[4] Access to all relevant components of the patient medical record provides clinicians and multidisciplinary teams with the information necessary to facilitate coordination of care and timely response to changing patient conditions.[2] Bates et al. identified that data format and interface issues are key barriers to implementing decision support.[10] According to the 1997 IOM report, several studies found that paper records were frequently lost or inaccessible.

Watt and others from the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) recently finished a study that showed a need among many hospitals for better accuracy and completeness of core measures data collection.[11] The same conclusion can be made for completeness and accuracy of prenatal risk factors and laboratory results.

Some medical errors are caused by system failures, and the most frequent errors share a theme of impaired access to information.[12] Building on this, Stetson et al. developed an ontology of concepts related to medical errors. The ontology provides a common ground on which informatics project members can discuss and measure the impact of variation in clinical communication.[13]

Costs of data variation

Data variation can be represented by missing data, inconsistent format of data, or incorrect data. The results from studies highlighted in the "To Error Is Human" IOM 2000 report, imply that at least 44,000 and as high as 98,000 Americans die in hospitals each year as a result of medical errors.[9] In 1996 the estimated total national costs (lost income, disability, health care, lost household production costs) for preventable adverse events were between $17 billion and $29 billion. The IOM committee felt that many of the preventable adverse events could have been avoided had better systems of care been in place.

Errors that result from inadequate information have not been widely studied except for the area of adverse drug events (ADE). There is a scarcity of studies on variation of data from disparate sources. An explanation for the lack of published studies could be from the fact that integration of disparate systems is still a relatively new field. An institution must have established systems for several years or decades to be in a position to need to evaluate the costs/benefit of interfacing old system modules, interfacing a new commercially available module, or internally developing a new module. Few enterprises are at that juncture. The complexity, breadth of medical

information, and dependency on that information for effective care, however, will require all health care institutions to interface disparate systems at some point.

Data variation is often a factor in adverse events, and the costs from adverse events are substantial. A study by Rothschild et al. looked at costs of defending medication-related malpractice claims.[14] They found that the ADEs represented 6.3% of claims and that system deficiencies (causing data variation) were some of the most frequent causes of preventable ADEs. Mean costs of defending outpatient ADEs were $64,700-$74,200, while inpatient ADEs had a much higher mean cost of $376,500. Bates et al. conducted a study on the additional resource utilization associated with an ADE. Annual costs attributable to preventable ADEs for a 700-bed teaching hospital were $2.8 million.[15]

Some costs caused by inadequate data and data systems are difficult to measure. An important factor with patients and their families is that effective systems should be in place for transferring patient-related information so that the information is accurate and available when needed. When disparate systems do not share patient information, it is not only inconvenient for both the patient and the clinician, but creates concerns and patient doubt in the ability of the provider to give quality care.[3]

Inconsistent data affect the ability to do outcomes research. Kogan et al. looked at the changing trends in prenatal care use and its varied short-term and long-term outcomes. One of the trends indicates that there has been an increase in the number of visits that women receive that exceed recommendations by the American College of Obstetricians and Gynecologists (ACOG), yet the increase in prenatal care to low-risk women did not reduce the rates of low birth weight and preterm birth in the United

States. Their study stressed the need for more investigations on the cost benefit of prenatal care utilization.[16] The lack of standard codes and historical data stifles decision making at the point of care and outcomes research.

Reasons for data variation

Lack of standards. Data for a patient are collected by a variety of source systems including admissions, pharmacy, laboratory, and radiology. Specialty clinical data systems such as perinatal, cardiology, and respiratory also collect patient data. Each system at each health care organization could potentially store the data in a unique structure with various levels of granularity. The possible variation in data structure is infinite. Data standards are necessary for organizing, representing, and encoding clinical information so it can be accepted and understood by another system. A number of standards exist, yet the lack of adoption of standards has prevented information sharing between laboratories and health care facilities; between pharmacies and health care providers regarding prescriptions; and between health care organizations and payers for reimbursement. A lack of understanding of the importance of these standards by both the general public and policy makers has impeded their adoption. Standards for health care data involve: definition of data elements, data interchange formats, terminologies, and knowledge representation.

Data models may express concept representation in various ways. Clinical researchers and health care managers tend to conceptualize data where all the observations for a given encounter are represented as fields within a single flat record.[17] In contrast, clinical computer system developers usually conceptualize information as stacked records. The difference in conceptualizing the data shows the various ways a

user can organize the data for particular purposes. One method for representing and storing data elements is by establishing name-value pairs. Huff et al. explained the interdependency of a semantic data model and a vocabulary that became more apparent as they created the Logical Observation Identifiers, Names and Codes (LOINC) vocabulary. They list two representation styles for data models called value-style names versus variable-style names.[18] Two examples of data models are:

Model 1 (History of diabetes as a value)

Context (Name)	Value
Finding	History of diabetes

Here the *assertion* is made that the finding exists in the patient. The value name style treats the field as more open-ended and the value of the field may be the specific diagnosis. The model is often used when a broad variety of assessments are being asked and only the conditions found will be reported.

Model 2 (History of diabetes as a name)

Context (Name)	Value
History of diabetes	Yes

Here the *question* is asked, "Does this finding exist in the patient?" The variable name style treats the field as a more limited set, which could include the value domain of Yes, No, Unknown. These are derived conclusions rather than the raw information. When one system asks for information using Model 1 and another system uses Model 2, the different models prohibit the computer from understanding the data elements, even though the human mind can quickly convert the meaning. If one data model is not consistently used within the application, the complexity is compounded when attempting to map the data elements and values with other systems. Using common data models

supports data exchange and knowledge exchange between heterogeneous clinical information systems, and avoids arduous conversion of decision logic. Although some standard data models are gaining attention, adoption is very slow and a multitude of models are actually being used. The proliferation of models produces redundancy and the task of sorting out which models are right for an institution may lead to confusion.

Terminologies are another component of meaning, and custom terminologies abound. Users of a terminology find it easier to create their own terms with tailored definitions, rather than to adapt to a standard. Local modifiers and extensions make mapping to an outside terminology difficult. When an external standard terminology is updated, the institutions using it must dedicate time and resources to revise their own dictionary to stay synchronized. Standards for meta models such as ontologies, archetypes and the semantic web are not well established or supported.

Complexity of data. The shear number of clinical concepts in health care demonstrates part of the complexity of interoperability and the potential for variation in data. The Systematized Nomenclature of Human and Veterinary Medicine, Clinical Terms (SNOMED CT) is a comprehensive medical vocabulary system with over 300,000 fully specified concepts and 450,000 supporting descriptions. Despite its comprehensiveness, SNOMED CT requires additional terminologies that provide further levels of granularity. The LOINC vocabulary, which has over 30,000 codes, provides the laboratory subset for SNOMED CT.

Kleinke walks through the queries involved in the process of clearing a typical insurance claim to show increased complexity when compared with that of a credit card transaction.[19] He points to the complexity of health care data, not a lack of technological

infrastructure as the reason major health care connectivity companies have not achieved the acceptance of their products.

Clinician attitudes on using information technology. Several reasons physicians are reluctant to adopt information technology are the lack of trust of information systems, the extra time required in their workflow, and the perceived risk of losing patient data entrusted to vendors who may not be in business for the long term.[20-23] Kleinke expresses the view that when the survival of a human being is the product of your work, an information system that is not 100 % reliable is zero percent useful.[19]

The 2001 IOM study, "Crossing the Quality Chasm" emphasizes the need for the appropriate use of information technology to support clinical and administrative processes.[3] Cooperation among clinicians is one of the recommendations from the 1998 Advisory Commission on Consumer Protection and Quality in the Health Care Industry. The report states, "clinicians and institutions should actively collaborate and communicate to ensure an appropriate exchange of information and coordination of care." Despite a recognized need to capture and effectively use medical record data, physicians will tend to use computer systems less often if there is limited structuring and integration causing the user to repeatedly enter the same information.[24] Benson reports on computer usage by British physicians in general practice versus in hospital settings. His findings support the assertion that because of the lack of standards and lack of interoperability, what works for a small practice does not work for a big hospital or across the primary-secondary care divide.[25]

Lack of standard datasets for clinical domains. Defining the appropriate information for clinicians at different stages of care is a challenge for clinical information

systems developers. Various organizations exist to recommend what data should be gathered. Health Level Seven (HL7) is an ANSI-accredited Standards Developing Organization (SDO) operating in the health care arena that establishes standards of how clinical information can be shared.[26] It is worth noting that the scope of the recent initiatives in which IOM and HL7 are focused, deals with functions (add, update, logical delete, archive, time-stamp, audit, etc.), but these initiatives are not focused on specifying content for any particular area.[5] Professional societies such as ACOG, AAP (American Academy of Pediatrics), AWHONN (Association of Women's Health, Obstetric and Neonatal Nurses) have published their own versions of forms and lists of guidelines.[27, 28] Forms from these organizations are often used to develop the data element content of the computerized record data sets for specific areas of obstetrical and perinatal care. In the search for a national standard for content for a perinatal longitudinal patient record, it was concluded that no single standard exists. Discussions with officers from two pertinent organizations confirmed this fact, (Williams, vice president Education, ACOG, 7/2/2003, and Kogan, director of Office of Data and Information Management, Maternal and Child Health Bureau of Health and Human Services (HHS), 7/11/2003).

Government regulations requiring specific kinds of data in specific formats are constantly changing. Starting in January of 2004, the JCAHO will increase the number of core measure sets required of hospitals from two sets to three, including the set related to pregnancy outcomes.

Disparate clinical data systems. Aggregating information from health care providers in the community is problematic because of the lack of standard report formats, the costs involved for the provider to collect and send it, and the political/legal issues of

who owns the patient medical record information. For example, no defined set of data elements for representing patient safety information, nor the specification of allowable values for those data elements exists. Each institution determines its own report format and may use its own terminology to represent the information. Each State has developed its own classification system to track adverse events. The report from a provider in some States may go to four or more agencies as well as to specific institution internal systems for collection. Analysis of what factors cause or prevent certain events would be extremely difficult as many sources store the information, the formats of the reports vary greatly, and the types of reported events vary.

Another example of variation is the process by which hospitals receive or fail to receive prenatal record information from outpatient clinics. Each state has regulations that measure whether the prenatal record was delivered to the hospital before the birth. Grants to health care institutions are dependent on the results of these and other performance metrics. Despite the importance placed on sharing the prenatal information, it is frequently inaccessible, incomplete, or unusable.

Data variation at IHC

The existence of many facilities and many patient data systems has created challenges of data variation for Intermountain Health Care (IHC). As a not-for-profit integrated delivery network of 22 hospitals and 72 clinics in Utah and southern Idaho, IHC is upgrading and developing an enhanced information systems infrastructure on an enterprise-wide scale. The vision of interdisciplinary care at IHC is that all IHC direct care providers use the IHC established care processes to plan, document, and deliver

patient care. The patient is at the center of care; problems and goals are clearly identified, and all clinicians work collaboratively as a team.

IHC is redefining its organization along clinical service lines such as Women and Newborns, Cardiovascular, Neuromuscular, Primary Care, Pediatrics, etc. The organizational structure is referred to as the Clinical Programs. Considerable time has been invested by Clinical Programs to standardize data structure and representation of concepts. An extensive Health Data Dictionary (HDD) has been built to establish relationships and synonyms of terms and concepts. Integration of many of the systems developed have earned IHC national recognition in 2000 and 2002 as the top integrated health system.[29] As separate applications were developed over the last 20 years, some were designed with integration in mind, while others were developed independently, often without the vision that a particular application would grow to be used across facilities or across platforms. As independent development occurred, terms and data models specific to the needs of a department or specialty were used that were never checked against or mapped to the standard medical vocabulary.

Magnitude and prevalence at IHC

Figure 1 illustrates the numerous systems that exist in some form for perinatal care at IHC. Integration exists in various degrees of maturity, yet a longitudinal view with data from each system does not exist. Much work is yet to be done to tie clinical data with financial and administrative data.

Labor and delivery (L&D) system. L&D at IHC is representative of the independent systems that were developed for specific functions and used their own data

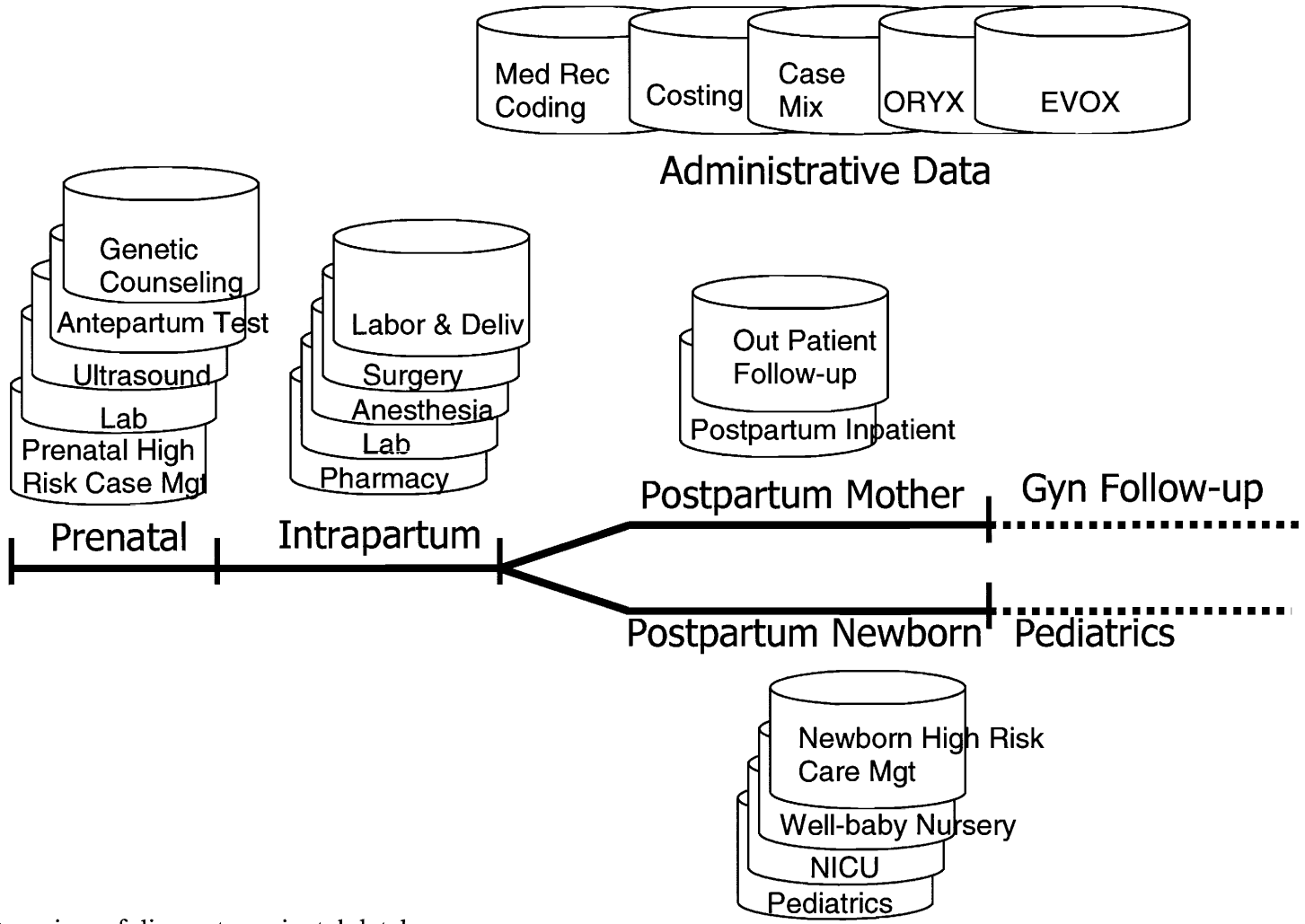


Figure 1. Overview of disparate perinatal databases.
Sources of maternal and infant data that have not been fully integrated. Many of these systems do not share a standard data model or standard data dictionary.

dictionaries. Nurses in the labor and delivery units at most IHC facilities use the PC-based L&D system. It was originally developed for intrapartum monitoring and nursing documentation at the LDS Hospital in 1984.[30] Although LDS Hospital had a strong history of developing an integrated clinical data system (HELP), the L&D system was developed independently. Original functional requirements for the system did not consider interoperability, and consistent information models were not used. Design constraints led to the use of an independent operating system. The system was developed using Microsoft® FoxPro®, a database development application which no longer offers technical support. In 1989 the L&D system was used for all routine clinical care in the Labor and Delivery Department at LDS Hospital. In 1996, a more user-friendly front end was created in Visual Basic® and the application was moved to a more reliable Windows NT® operating system.[31]

The initial objective associated with the L&D system was to provide a stand alone, dependable system for nursing documentation using coded terms for tracking clinical care and associated administrative data. Data element representations that were meaningful to the users (nurses) were displayed and embedded in the information system itself, as opposed to residing in an enterprise data dictionary. Being a screen-based approach, similar concepts from slightly different clinical contexts were added to increase usability in relation to data capture. Examples include five different care situations where rupture of membranes can be documented. Although very effective at documenting intrapartum care, the L&D system was not designed to be integrated with the HELP information system, the HDD, or other enterprise applications and services. Patient demographics, encounter and patient identification keys were created and stored in the

system in ways specific to one hospital. As nurses receive the paper-based prenatal record from the provider clinic, they manually reenter the information into the L&D system. If an infant's situation requires NICU services, the information from the L&D system is reentered manually into the NICU EMR system. The development approach with the L&D system was designed to be very accommodating to users allowing user-defined values to become coded data elements for ease of use.

Use of the L&D system at IHC has been extended to thirteen hospitals. Since 1989, information from clinical care documentation has been captured and stored on over 340,000 maternity visits, representing 187,000 births to date, resulting in a rich body of clinical knowledge. Over the past 19 years, specific clinical information systems for prenatal, L&D, and NICU have evolved separately from the HELP system at IHC.

Pediatric cardiology. Another database developed independent of the enterprise data dictionary and standard data models is a repository for the Pediatric Cardiology Department at the Primary Children's Medical Center (PCMC). The database stores codes for events, diagnoses, and procedures for pediatric cardiac catheter and surgery. It has been used by cardiologists for about 15 years and presently tracks about 9,000 cases per year. The application is used mainly to generate monthly summary reports, quality assurance reports, and outcomes research. Since the system is not linked to any patient tables or enterprise patient indexes, the users must key in patient demographic information for each case. Regular maintenance to correct variations in patient information and the resulting duplicate records takes about 3 hours per month and is handled by one of the cardiologists.

Cardiology users have expressed a desire for a new system with more functionality, but the purchase of a new system has been postponed because of concerns that they would lose access to the historical data in the present database. Tracking patient reports that are frequently sent from clinic to clinic is mostly a manual task. The physicians do not have the ability to access a single longitudinal view of a patient, as the patient data are in many locations. Each division is duplicating the storage and maintenance of much of the same information. The Pediatric Cardiology Department is uniquely positioned to implement a longitudinal record since 80-90% of the pediatric cardiology work in Utah is done at PCMC. Much of the non-PCMC work is done in very close proximity at the University Hospital clinics adjacent to PCMC. Discussions with corporate information systems personnel concerning the use of an enterprise common data dictionary or standard information models have been secondary to debates over shared patient data stewardship.

Costs of data variation at IHC

Terminology and interface maintenance. Developing and maintaining integrated systems comes at a cost, and becomes a major factor in the direction of information resources in many institutions. Building and maintaining a health data dictionary, a common data repository, and interfacing the various systems throughout the institution requires extensive research and knowledgeable resources. As of early 2004, IHC employed 19 full-time people who maintain and build interfaces, and eight people working full-time on creating and maintaining vocabulary codes that are stored in the dictionary. It is estimated that over 36 person years have been invested in defining the terms, synonyms and hierarchy and for maintenance of the data dictionary. Of the 12.5%

of IHC's overall information systems budget, 4% is spent on interfaces and vocabulary.[32] When an institution has the support of senior management to invest the costly initial outlay of resources to create this kind of infrastructure, the database content of a well-structured approach using controlled medical vocabularies remains useful over a broader user base.

IHC has adopted an open architecture approach and is actively working to interface multiple best-of-breed individual systems.[32] The cost, however, is substantial to create interfaces with older systems that were not developed using a common repository and data dictionary. Presently much of the information captured in the prenatal record is not interfaced, and has to be reentered manually into other data systems. Effects from the prevalence of missing or conflicting data across systems have not been measured. The existence of disparate, complex systems, each containing many parts of the infant patient puzzle, provide a unique environment and opportunity to detect and report the extent of data overlap and gaps in information across applications.

Process costs. Business and legal issues of patient medical record ownership create some important factors that have an effect on the variation of data. IHC affiliated/non-IHC employed clinicians agree contractually to provide a prenatal record for patients who are IHC Health Plan members. The prenatal information identifies early in the pregnancy those patients who have risk factors for possible complications and would likely benefit from special prenatal education and monitoring. Reducing variation in data collection methods has met some obstacles regarding legal requirements for independent clinicians to share their patient information with IHC. If various prenatal forms are used, the reported data varies. If the clinician does not submit the prenatal

form, the accessibility of data varies. The paper version of the IHC prenatal record is a compilation and combination of forms from the State of Utah, ACOG, and Hollister, a medical forms distributor. Prenatal forms are distributed to clinics to be used for IHC Health Plans patients. The form was shortened from previous versions based on feedback from physicians that a shorter form would lessen the effect on their workflow, and increase the likelihood of getting more records. Justification for the shorter form was verified by experiences from studies at Brigham and Womens' Hospital, where speed of use and low impact on workflow were shown to be key factors in getting acceptance of clinical decision support systems.[21] The form was created and modifications coordinated by the medical director of Women and Newborns Clinical Program as well as with the chair of NICU development team at IHC. Clinical experts at IHC estimate that 80 % of the information gathered on the prenatal record is relevant to clinicians during the labor and delivery stage of care.

In 1999, the data elements and associated values from the paper version of the prenatal record were used to create an electronic version. The prenatal EMR uses a MicroSoft® Windows®-based infrastructure called Clinical Workstation® developed by IHC and 3M Health Information Systems to access and store data to the Oracle-based Clinical Data Repository (CDR). The prenatal module was originally designed by a group of doctors, nurse midwives, medical informaticists and analysts. The system was pilot tested 3 years ago by a group of eight nurse midwives. Many modifications were made to suit the nurse midwives' workflow. Following the completion of the pilot, the application has continued to be used by the majority of the midwives. Acceptability was curtailed by some functional and nonfunctional limitations, such as slow response time,

difficulty in navigation, inability to automatically calculate gestational age, and the absence of interactive charting. Due to technical and security limitations of the Clinical Workstation, the prenatal record application has been limited in its rollout to users within the IHC network of physicians and midwives. The user base of IHC employed family practitioners has slowly increased, while feedback on system functionality from a large base of potential non-IHC obstetrician users has not been received. The creation and maintenance of standardized prenatal forms, and computerized applications to capture prenatal information are a few of the bridges being built to manage data variation from disparate data sources.

Solutions to data variation in the health care industry

Missing data and having differing data across systems are problems faced by most health care institutions. In search of possible solutions, many in the health care industry have placed high hopes on the computer-based record to facilitate integration of patient information. The 1991 IOM report lists 12 desirable attributes for computer-based patient records.[1] Three of those attributes are particularly pertinent to data variation problems discussed in this study. A computer-based record should:

- link to other clinical patient records,
- be accessible in a timely manner by all who have authorized access
- use a defined vocabulary and support structured data.

Essential information for a patient (the mother) may need to be stored on another patient's record (the infant). The ability to link clinical and outcomes information from one patient to another is critical in providing effective health care.

Although clinical applications may store the text from dictation or summary narratives in an electronic block, unless the information is structured at the proper level, rapid retrieval and decision support will most likely be limited, and the time to find specific data may not be much faster than looking in physical patient records.

Elson and Connelly's [33] research strongly suggests that when physicians have the operational support of a properly configured computerized patient record system, it has a major impact on physician compliance with standard practice guidelines. Increased compliance, in turn, leads to more consistent patient care particularly when physicians need to cover for each other. Research by Shortell et al. showed how coordination and communication among clinicians and across settings resulted in greater efficiency and better clinical outcomes.[34]

The need for sharing data both internally and externally is an area of emphasis targeted by the 2001 IOM study, "Envisioning the National Health Care Quality Report." Improving health care quality requires the ability to access data by utilizing electronic clinical data systems that span health care settings. Clinical decisions at the point of care depend on clinicians having immediate access to complete and accurate information about the patient from all relevant sources. With knowledge of the patient's physiological state from laboratory results and other provider notes, the treatment that has been given, as well as what the reaction to that treatment has been, a physician can make decisions more quickly and more accurately. Integrated, longitudinal electronic records are a key part of enabling point of care decision support, measuring effectiveness of care, timeliness of care, and examining specific safety problems. Electronic data sharing of the huge amount of information documented in the paper medical records should be part of

each provider's information infrastructure. As part of their vision statement, the American Health Information Management Association (AHIMA) stated the future of health information is electronic, patient-centered, longitudinal, accessible, and credible.[35]

The Hospital Corporation of America (HCA), Nashville, Tenn., has benefited from its standardized nomenclature. HCA's well-defined clinical data repository and unified nomenclature has allowed them to collect, access, and share core measure reports with JCAHO with less effort.[36] Sharing information about prenatal risk assessment allows for anticipatory planning, individualized education, and appropriate referral. Information used for outcomes of risk assessment facilitates the refinement of guidelines by which the effectiveness of the care can be evaluated.

Maloni et al. point out that comprehensive care involving multidisciplinary caregivers can lead to an improvement in maternal and infant care outcomes.[37] Patient care must be coordinated from the prenatal period through the first year of infant life when the needs of the mother, infant and family adjustment are the greatest. The care must be combined with information systems that allow the effective sharing of patient needs and treatment across the perinatal continuum, further justifying an integrated longitudinal record.

In May 2003, the Department of Health and Human Services (DHHS) asked the Institute of Medicine (IOM) to establish guidelines on the key care delivery-related capabilities of an EHR system.[5] Among many points discussed, the IOM stated an EHR system must be a longitudinal collection of electronic health information for the individual. A single longitudinal view of the patient data will provide more complete and accurate information to the user.

Work by Cimino and others to create and implement the Medical Entities Dictionary (MED) showed how a central terminology facilitates interoperability among disparate data systems.[38] It demonstrated how strings and codes from legacy information systems can be translated into a format that other departmental systems could process. The MED represents an ontology, a predefined set of concepts, relationships among concepts and constraints on those concepts. Ontologies provide reference models that specify all the concepts that comprise the target domain for specific information systems. Ontologies form the basis for human-computer interaction and for information exchange throughout the health care environment. Ontologies offer a human-readable and machine-processable description of the data modeling assumptions the computer systems need to achieve interoperability.[39]

Archetypes are another proposal to close the semantic gap between generic information models such as the HL7 reference information model (RIM) and the vocabulary.[40, 41] Archetypes allow modeling of domain concepts external to the system information model and preserve meaning that would be lost otherwise if the semantic representation were embedded into the information system. Archetypes for selected concepts that are linked to an ontology can also facilitate coordination of the EHR structure with computer interpreted clinical guidelines. Barretto presents a model and uses archetypes to link the needed EHR content for decision making to standard care guidelines.[42] He states that care guidelines allow specification of what needs to be recorded (in the EHR), when to record, how to evaluate/make decisions, and what needs to be done. Generic data models have been developed that show usefulness in mapping between structured clinical vocabularies. Three elements that must be common across

independent systems so information can be transferred without complex mapping are: 1) the representational model with its associated structure, 2) a structured vocabulary that has been organized by domain, 3) links that bind the domains of the vocabulary to attributes of the model.[43] A variety of generic information models have been developed and adoption of a standard is not likely to happen soon. There are those who feel that the combination of strong terminology models and a flexible information model will address their needs for maintaining semantic interoperability. Under the proposal that the HL7 Templates special interest group is building, archetypes would be a valid formalism in HL7 for constraining RIM artifacts.

Integration options within the institution

A fundamental step to reduce variation in disparate systems is to have an integrated view of the data. Clayton et al. describe three options for institutions trying to achieve integrated systems: build, buy, and interface.[32]

Buy integrated systems. The first option is to buy all the applications from a single vendor to ensure ease of use and smooth integration. The "buy" approach is common for smaller health care institutions that may lack knowledgeable integration specialists. With over 233 vendors advertising EHR products in the 2004 Resource Guide in the Health Informatics trade magazine, the "buy" approach can be challenging. A major concern with this approach is that a small percentage of these companies will be in business to support a complex system that may take 5 to 10 years to fully implement. As the scope of the system and the number of facilities increases, the conversion to a new system can pose significant risks. Anderson and Stafford describe the operational concerns faced when the University of Missouri Health Care (MUHC) replaced an

outdated system with a new integrated information system from Cerner.[44] MUHC's various hospitals and clinics, previously independent of each other, evolved with disparate computer systems. A key reason for choosing a single vendor system and a "big bang" implementation approach was the increased difficulty for staff to deal with the several disparate computer systems simultaneously given their tightly integrated patient care processes. In 2003 when the University of Utah Health Science Center began the replacement of their legacy systems, they adopted several lessons learned from the University of Missouri experience.

Information in health care is constantly changing and the need for integration with other systems will always be present. The "buy" approach does not fully insulate the purchaser from the need to integrate. The user is dependent on the ability, responsiveness, and priorities of the vendor to meet the user's needs.

Build monolithic systems. The second option is to build all the necessary applications and integrated database from scratch. Monolithic systems are typically developed in a common environment by the same tools, databases, vocabularies, structures, user interfaces, etc. Systems that have been successfully built this way have taken decades. Examples of internally built integrated hospital information systems include the early HELP[45] system at LDS Hospital, systems developed by the Regenstrief Institute of Health Care,[46] clinical data systems at Beth Israel Hospital,[47] and the VISTA system used by the Veterans Administration.[48] Creating such large comprehensive systems from scratch is not always practical today, given the economic emphasis on rapid return on investment at most institutions.

The monolithic approach can work well along a continuum of care. Nielsen et al. reported on an obstetric electronic record charting system that integrated antepartum, intrapartum, and postpartum care records. Completeness and accuracy of data as well as rapid access to obstetric outcomes were improved.[49]

Escobar et al. describe the reasons Kaiser Permanente in Northern California chose to build a NICU database instead of purchasing an off-the-shelf product.[50] First, the database products available at the time of their decision did not meet their specific needs with respect to type of information captured. Second, commercial products often have high hidden costs because they required data entry by a physician or by some other highly trained personnel.

Interface "best of breed." A third approach to integration is to interface disparate components allowing each subcomponent system to feed data into or retrieve from a central longitudinal repository. The interfaced approach attempts to take advantage of data from legacy systems regardless of whether they were built or purchased, by use of standard data exchange protocols such as HL7. The approach requires substantial expertise in integration and tools to maintain the interfaces.

Integration tools

"Integrated data" are data that are stored in a uniform encoded form and structure. Commercial tools and systems exist that facilitate sharing data and reducing data variation in clinical information systems.

Longitudinal Data Repository (LDR). A longitudinal patient record consists of a lifelong chain of patient medical events stored in a central data repository that is accessible across settings and care providers. The record links clinical, demographic, and

encounter data with a single patient identification number and is stored in database tables. It also includes the services needed for storing and accessing that data. The EHR requires some form of LDR from which to pull medical information about a patient. Records from many sources are formatted and translated according to the rules of the repository through the use of interface applications and health data dictionaries. The LDR, in concert with unique patient identifier, ties all medical record and encounter numbers used at local points of service. The enterprise master patient index is used as the key when storing the patient record in a clinical data repository. The LDR may contain problem lists, medication lists, allergy lists, patient history, physical evaluations, physician progress notes, nursing assessments, imaging reports, radiographic images, therapy reports, and laboratory results from both inpatient and outpatient clinical care facilities. LDRs are commonly referred to as Clinical Data Repository (CDR).

Health Data Dictionary (HDD). A data dictionary is a structured collection of words or terms with information about each of them. The HDD is a set of tables, which serves as a reference source for interface programs that encode and interpret information in the LDR. These tables link and store data with consistent meanings and form relationships between data items. The interface programs, along with data storage and retrieval services, use the information stored in the HDD to translate data as it comes into and out of the LDR. The HDD is a vocabulary server that ensures that the data are stored in the LDR in a common encoded language. Data dictionaries can be used as a stand alone database tool, or used as a reference source embedded within other health care applications.

Commercial interface engines. Data that come from departmental computer systems, such as admissions, laboratory, and pharmacy, are typically in a format unique to each system. The translation of data from the original system format into a standard encoded repository format occurs through interface software. To translate the values of individual data elements, the interface software may access tables to convert each element into the values understood by the repository. Accurate mapping, or linking of the elements from various sources with the correct identifier in the translation tables, is an essential aspect of data integration.

Industry Data Standards

No current single terminology has the breadth and depth needed for all health care data. The Department of Health and Human Services (DHHS) will be receiving advice from several entities with the goal of making decisions on national standards for medical terminology for the EHR.[6] The National Library of Medicine (NLM) houses the largest database of standardized terminologies known as the Unified Medical Language System (UMLS). Under direction of the DHHS, the National Committee on Vital and Health Statistics (NCVHS) has evaluated and selected a group of terminologies from the UMLS that will serve as the national standard for medical terminology for the EHR. The Consolidated Health Informatics (CHI) initiative is also evaluating these terminologies. These two organizations will work together to determine which terminologies best represent the clinical domain areas, with their joint recommendations to be accepted for federal government-wide implementation. The NCVHS has heard presentations from several terminology developers with the intent to establish "a core set of medical terminologies that together are sufficiently comprehensive, technically sound, mutually

consistent, and readily available to deliver most of the envisioned functionality of a national standard medical terminology for the EHR."[6] In November of 2003, NCVHS officially recommended that the DHHS adopt five core medical terminologies for use by federal health care service programs. The terminologies include SNOMED CT, the laboratory subset of LOINC, RxNORM, NDF RT and the FDA terminology sets for drug ingredient name, dosage form and package form. ICD-10 was recently recommended for adoption as the new coding system under the Health Insurance Portability and Accountability Act of 1996 (HIPAA). Even when standards are agreed upon by federal government agencies, the adoption and implementation of standards throughout the private sector takes years.

External sharing consortium

Solutions to address health care system data sharing challenges are being sought by a number of organizations. Not only must institutions be able to share information effectively across the enterprise, but also as patients have become more mobile and see providers at multiple institutions, the need increases for cross-institutional sharing of information.

Internationally, efforts are underway to develop and test standards for information sharing. The Netherlands have established national domain models based on HL7 RIM version 3, and have created a perinatal domain message information model as a national pilot.[51] National standards for content in specific clinical domains need to be established that can be used with currently accepted messaging standards such as HL7.

The National Health Information Institute (NHII) of the Department of Health and Human Services (DHHS) has been working to establish guidelines, networks,

standards, technologies, and promote applications of a comprehensive knowledge-based network of interoperable health information systems.[52] The key to a national information infrastructure lies in successes of local community applications of health information sharing. Dr. Yasnoff, senior advisor to the NHII, spoke of two successful prototype implementations of common infrastructure in separate communities (Indianapolis and Santa Barbara) to share health data from disparate systems.[53] He commented that as important as the technology standards are, the most difficult work is to establish cooperation of political, and organizational factors within the community. A common motivation must be to make these information infrastructures work for the good of the community.

The Indianapolis Network for Patient Care (INPC) has put together a single network of databases to share health information from multiple institutions. The consortium of major hospitals, homeless care organizations, County and State public health departments, primary care providers, subspecialists, and public school clinics, spans much of the State of Indiana. INPC participants cover over 95% of the acute and inpatient and nonoffice based outpatient clinical care within INPC. Participants contribute laboratory reports, discharge and admission summaries, radiology reports, inpatient medications, immunization registry data, and so forth. The network spent a considerable amount of time on data standardization. All data are sent as, or converted to HL7 compliant messages, and then standardized through LOINC, CPT, ICD, coding structures. The data standardization is key to aggregation across institutions facilitating ease of clinical review and public health reporting. Some of the other success themes for the network include: it was driven by physicians toward a clinical information focus;

physicians provided the consensus of what information was needed; the business community provided strong support because it was for the good of the community; and the partners worked closely with senior level corporate management to develop trust. Nancy Lorenzi, in a report on the Strategies for Creating Successful Local Health Information Infrastructure Initiatives, noted that the INPC network grew primarily as the result of a single clinical champion who developed informatics tools and strategies as well as instilled cooperation and community belief in a regional system.[54]

The Santa Barbara County Care Data Exchange (CDE) was organized as a countywide regional health information exchange. It allows access by authorized patients, and health care providers for regional patient information using a single web-based interface. The California Health care Foundation awarded a 10 million-dollar grant to develop CDE over 5 years starting in 1998. The CDE is currently administered by a private corporation and has been organized as a public utility since December 2001. Different financial models have been considered, as the project needs to be financially sustainable when the grant funding ends.

The CDE participants are not as numerous as the program in Indianapolis, but plans exist to expand the region. Clinical information is exchanged for laboratory reports, radiology reports and images, clinical notes, pharmacy data, eligibility and administrative data. Lorenzi also compiled several success themes for the project: seed money to begin a complicated and time consuming initiative of the scope; the need for a strong entrepreneurial instigator to get the grant within a short time period; preliminary infrastructure to address the legal issues of data exchange; and the value of having a neutral third party manage the project.

The creation and adoption of the HIPAA has had an effect on reducing variation of patient health care transaction information exchanged across insurance payors.[55] Developed by the HHS, these new standards provide patients with access to their medical records and more control over how their personal health information is used and disclosed. The standards represent a uniform, federal floor of privacy protections for consumers across the country. State laws providing additional protections to consumers are not affected by this new rule.

Comprehensive database applications for a particular domain can reduce variation within that domain. Although not used for point of care decision making, there are systems that bridge what had been several disparate systems. Dr. Kogan from the Department of Maternal and Child Health Bureau (MCHB), noted that the most comprehensive and proven perinatal dataset of which he was aware, was one created and implemented by Thomas Hulsey MSPH, ScD, Department of Pediatrics/Epidemiology, at the Medical University of South Carolina, Charleston, SC.[56] Hulsey has established a statewide regional perinatal information system with the breadth of the infant care continuum needed for an infant longitudinal patient record.

Sistema Informatico Perinatal (SIP) is a perinatal clinical database in use as part of the technical cooperation of Pan American Health Organization's (PAHO/WHO). SIP has been gathering individual clinical records of pregnant women in 20 countries in Latin America and the Caribbean for over 18 years. The dataset of 170 data elements has become recognized as an international standard.[57]

Solutions to data variation at IHC

IHC has implemented many of the approaches available to the health care industry for reducing data variation. A combination of internally developed integration tools along with commercial tools and systems have been used to share information across settings and among clinicians.

Solutions through interfacing at IHC

Common data model. Huff and others have built upon the clinical event model to integrate data associated with events linked to domains in a controlled medical vocabulary.[43, 58] The clinical event model permits changes to the information model without having to recompile the clinical applications.[58] Steiner et al. report on the development of a logical data model called "ClinicalElement" that is able to encompass structures beyond clinical events to be fully recursive (allowing events within events).[59] The flexible abstraction model has been instantiated using simple XML constructs to take advantage of the growing set of tools that facilitate the growing standard. By using the model and other abstraction mechanisms, an application platform is created that is not affected by data evolution and migration. Maintenance costs and obsolescence of data should be reduced for applications using this platform. When asked if archetypes should be used in conjunction with IHC's Clinical Element Model (CEM), developers at IHC replied that archetypes require more modeling overhead for minimal clinical gains; therefore, they have chosen not to use archetypes.

Build integrated systems. Starting in the early 1970s, the HELP hospital information system at LDS Hospital was a pioneer clinical information system in the use of decision support logic. A numerical coding scheme for medical terms facilitated

computerized interpretation, patient-specific warnings, diagnostic suggestions, and limited management advice. A key component of the system was a data dictionary that integrated medical terms with other applications across LDS hospital's departments.[45] The HELP system was originally designed for one facility to provide patient monitoring, order communications, information management, physician guidance, and clinical research support. HELP has been effective for over 20 years in improving patient safety by flagging contraindicated medications and generating alerts when patient conditions meet predefined criteria. HELP's strong point is the sharing of inpatient information across departments at LDS Hospital. Ancillary service computer systems that transfer data to the HELP system through interfaces include the clinical laboratory system, the Marquette MUSE ECG system, and the PC used in the cardiac catheterization laboratory.

Interface of custom and commercial. IHC has set up a system architecture that allows collecting and sharing of data from the many disparate systems that have been developed at IHC, or purchased commercially. The architecture includes several components that will be described in subsequent paragraphs. IHC's plan is to have a vocabulary and data model that allow them to define content in the EHR in a way that is independent of a single application.

In 1995 IHC began the focused development of an HDD in collaboration with 3M Health Information Systems.[60] The purpose for building the HDD is to identify and maintain core clinical data concepts across the health care institution, as opposed to the hospital centric data dictionary used in HELP. Many of the concepts from the HELP data dictionary have been transferred to the HDD. Clinical system applications that have been

developed since 1995 have used concepts and data elements defined in the HDD. New concepts are added as needed.

IHC has implemented a CDR with an enterprise master member index (EMMI) identification. Hospital and clinic patient registration systems use the same unique identifier. The CDR uses a data model based on clinical events that are structured using an ASN.1 syntax.[43] Newer data models based on clinical elements are being developed that will allow the data to be stored in XML format in the CDR.

Over 850 point-to-point interfaces have been established to exchange data between the CDR and other systems used at IHC hospitals and ambulatory clinics. HELP, Labs, Pharmacy, Radiology, Billing and Financial, and Insurance Plans are a few of the sixty different applications interfaced with each other. Approximately 20 unique applications are integrated with the CDR. Two-thirds of the interfaces are for clinical applications, and the remainder connect to financial systems.

Results Review (RR) is a Web-based application that provides users access with existing systems to display data from the CDR, and to enter patient care documentation, or receive guidance on alerts and decisions. Both inpatient and outpatient patient data in a comprehensive, longitudinal view can be accessed via the Internet.

Clinical integration outcomes tracking and special project analysis are achieved by accessing data marts in the Enterprise Data Warehouse (EDW). The data warehouse pulls its data from the CDR and IHC's financial and claims systems. Clinical outcomes reporting and analytics from the EDW relieve the CDR of large population queries and preserve CPU usage for responsive clinical transaction needs. Aggregate data from all sources can be optimized in the EDW for a larger population view to improve health care

methods, to optimize utilization of resources and to provide valuable analysis for public health issues.

Clinical Workstation (CW) allows IHC employed physicians as well as some IHC affiliated/non-IHC employed clinicians access, to the CDR. Physicians from the ambulatory clinics use CW to enter problems, visit notes, allergies, prescriptions, phone log messages, and template-based forms for select visits and procedures.

IHC has established a Web-based user interface for clinical applications called the Clinical Desktop. It uses standard utilities and messaging to allow the user to have one login and access multiple sources of patient information through different clinical applications.

A project was undertaken and completed in 1998 that sent L&D data in HL7 formatted messages to the CDR.[61] The CDR was able to accept the records, yet because the concepts did not match with ones in the HDD, the meaning of the exchanged data was lost. Although syntactic interoperability was achieved, the more complete semantic interoperability was not. The L&D HL7 exchange project was a good illustration of the need for a common understanding of meaning between "interoperable" systems when sharing concept-based information.

Thornton, Yu, and Gardner were able to reduce errors and eliminate variability in charge capture by embedding time or data-driven logic within the L&D system to capture charges systematically.[62] Their results showed how clinical documentation entered at the point of care can also trigger logic to determine consistent patient charges.

Gibson and Haug built an interface that allowed patient findings from the HELP clinical information system to be processed by a commercial computerized severity index

(CSI) system to produce a severity index score.[63] When comparing the computerized severity index score with that reached by manual chart review, scores differed approximately one third of the time. Results showed that neither the paper-based records, nor the computer-based system were correct all of the time. Reasons for variation were found and interventions were put in place to reduce the disparity. An important finding was that significant errors were found in both the automated and the manual CSI methods, to the extent that neither system could be considered a gold standard. The manual system that was expected to be accurate was determined to be incomplete or incorrect in at least 119 of 278 cases.

Longitudinal EHR as a solution to data variability at IHC

Although inpatient reports, laboratory results, and some vital signs are integrated with the ambulatory data, an integrated longitudinal patient view of prenatal and maternal information does not presently exist. Wide-scale automated sharing of data captured in the various perinatal systems has not been accomplished. Clinicians obtain clinical information about the patient from several disparate specialty systems, as well as from physical charts, notes and verbal communication with other care providers, patients and patient relatives. The process clinicians use to access and compare patient data from each source could be greatly simplified with a single longitudinal view of the patient. IHC is reviewing the requirements of building longitudinal Electronic Medical Record (EMR) system that bridges the relevant information across the maternal and infant care continuum. Data would be entered directly into and pulled directly from the CDR.

External sharing at IHC

IHC has been proactive in sharing patient data concerning immunizations with the local health care community. Immunization history is a prime area where data about patients may reside in several disparate sources. The Utah State Immunization Information System (USIIS) is a central repository that contains immunization records for all children in Utah that can be accessed only by authorized users.[64] It assists health care providers in tracking, recalling, and reporting of immunization information for patient care. Parents benefit by having access to a complete and updated record of their child's immunization history. The USIIS repository has been planned and developed by partnerships in the community, IHC being one of the contributors.

WebKIDS is an application used by subscribers to look up, print out, and download information from the USIIS database via the Internet. In 2002 IHC implemented the WebKIDS application at one of the pediatric clinics. A study was conducted at the clinic to measure the time taken by the staff's activities with immunization encounters. After the implementation of WebKIDS the time study was repeated. The findings proved WebKIDS has been an effective tool in a clinical environment. A few of the study's findings show: time to determine the patient's need for immunization was reduced 90% to 10 seconds, time to create a new individual immunization record was reduced 85% to 1:29 minutes.[65]

Several IHC employees participate in key roles in the ongoing development and management of this community program. IHC Operations confirmed that the year IHC began using the data of WebKIDS, IHC's Health Plan Employer Data and Information Set (HEDIS) measure of fully immunized children at age 24 months went from the mid 60s to the mid 70s percentages.

Relevance and timeliness of this study

While a few studies have looked at accuracy of a specialty system, no formal studies have been completed on the extent of variation of data that exists across disparate clinical systems at IHC. This study compares the data variation across three core IHC perinatal clinical data systems using perspectives from a) data model compatibility, b) clinician perception, c) prevalence of missing and contradicting data values in patient records across systems.

Timeliness of this study is supported by the heightened attention on data quality and the need for increased efficiency of health care providers. Effective capture of data and the ability to easily access complete infant patient data in a longitudinal electronic health record are important steps to achieving that efficiency. Infant-centric views are a prototype for the development of discipline-specific patent views in other domains. The long-range goal is to use the results of this study as a baseline, and for developing and selecting new informational support tools at the point of care that facilitate positive clinical outcomes.

Triangulation approach

Triangulation is a method to verify study findings. Gaining knowledge of where, how, and why data variation occurs in health care systems is a broad and complex task. Knowledge comes through application and verification of what is learned from multiple views and experiences. The term triangulation comes from the field of navigation and is a technique to determine the precise position of a ship or aircraft by using several reference points. Triangulation in evaluation generally refers to the use of multiple sources of data, observers, methods, or theories in studies of the same phenomenon.

Both triangulation and meta-analysis use multiple studies of the same phenomenon to draw conclusions more powerful than those obtainable from any single study. Meta-analysis traditionally refers to an objectivist strategy using more formal statistical techniques for combining quantitative study results across a set of comparable completed studies.[66] Through triangulation the subjectivist researcher considers different kinds of information from multiple sources or methods to assess if a consistent picture emerges from the results.

Triangulation in the health care industry

Ammenwerth et al. explain that clinical information and communication applications are inherently complex and challenging to evaluate because of the changing nature of clinical, human, and technical environments.[67] Their study presents a background on the theory of triangulation and expounds its well-established use in sciences other than medical informatics. Their work shows how the use of both quantitative and qualitative methods can achieve a more balanced approach to evaluate the technical and social aspects that exist in the field of medical informatics. The use of triangulation as a comprehensive evaluation methodology will lead to the development of better information systems and better support of patient care.

A comprehensive plan is required to adequately evaluate variation across medical information systems. The use of multiple methods focusing on a variety of technical, economic, and organizational issues is recommended by Kaplan for two reasons.[68] The first is the diverse and broad nature of information system's effects. The second reason is to provide a more complete understanding of causal links by collecting a variety of data showing different perspectives. Triangulation involves combining data from various

methods and from various sources to strengthen the robustness of research results. The methodology also allows the flexibility to include issues and concerns that may not have been evident in early stages of the design.

Triangulation at IHC

This study is unique in its application of triangulation to assess characteristics of clinical information systems at IHC. The justification for using multimethod triangulation is to detect the extent of data variation, the reasons it exists, and its effect on clinicians. Reasons for data variation in clinical systems are complex and multifaceted. Development of a complete longitudinal patient EHR is a vast and complex undertaking which requires a broad and comprehensive method of evaluation. It is expected that such a wide-ranging assessment will help create more useful systems that will lead to a greater adoption by clinicians. The three methods for this triangulation study are: a data model comparison to show variation in data structure and meaning across perinatal data systems; a clinician perception survey to examine the user's perception of problems with missing and contradictory data in perinatal data systems; and a patient record comparison to assess variation in key data values across perinatal data systems.

Data model comparison (structure and meaning)

The first of three methods considered for triangulation will be the data model comparison between disparate data systems to show the extent of variation in structure and meaning of data elements. Variation of data structure and meaning determine the level at which different computer systems can understand data that is exchanged. The ability to preserve the meaning of data in health care when exchanging information across information systems is known as semantic interoperability. Heard states,

"meaning is the product of context and content."[40] The combination of data models, terminologies and agreed conventions of users determine the effectiveness of transferring meaning. Comparing the context and content of concepts across systems will illuminate modeling limitations of sharing meaningful data. Within this study the term "data model comparison" will be used to describe that process.

Data model comparisons in the health care industry

The methods Logan et al. describe for measuring completeness and correctness of clinical data recorded, rely heavily on subjective judgments about the clinician's level of expertise.[69] Their study used clinical experts who viewed the encounter to help determine the gold standard of data elements to be compared. Classifications of data element comparisons included, element was equivalent, element was incorrect, element was missing, and element was extra. One method of measuring completeness and correctness was described, but it was acknowledged that there are many perspectives on the best ways to evaluate medical record quality.

Chute et al. used concept comparison to determine the extent of completeness of various classification systems. Each classification was given a score based on whether there was a concept match or partial match to each concept in a data set of 3061 distinct concepts.[70] McClay and Campbell compared the text entry of reasons for visit to the emergency room against concepts in ICD-9-CM, SNOMED-RT, and SNOMED-CT for accuracy.[71] Text entries were judged to be a match, a broader or narrower concept, or no match. The method demonstrated that the SNOMED coding of the data element was more accurate than ICD-9-CM coding.

Data model comparisons at IHC

In the interest of improving common interface installation processes such as vocabulary matching, Rocha and Huff compared SNOMED terms to components of LOINC names.[72] The LOINC model for identifying test result names was used as a common interface to compare decomposed name segments with SNOMED terms. Their study showed how using a detailed data model to support vocabulary mapping provides the ability to better handle "many to one," "one to many," and "many to many" mappings. The extended model also allows the data dictionary to identify the discrete concepts that when combined, correspond to the complete meaning of all distinct test result names.

Clinician surveys

The second method considered for a triangulation study involves using clinician surveys to provide new insights in several areas of data variation across systems. Surveys can be used to identify the extent of data variation in the different systems from the user's perspective, and surveys can provide a detail description of the clinician behavior when faced with missing or contradicting data. Clinician surveys are effective data collection tools in the clinical setting, to the extent they are valid (the degree it actually measures the characteristic it is supposed to measure), and reproducible (the degree to which a variable has nearly the same value when measured several times).[73] Open-ended questions are useful in developing an understanding of the perceptions of the respondent in their own words. Disadvantages of qualitative methods are that they take more time to code and analyze, and the results are more subjective. Closed-ended questions are easier to answer, tabulate, and analyze. Lists of possible answers can often

assist the user in understanding the question. Disadvantages include the tendency to lead the respondent in a certain direction, and the possible answers may not be exhaustive.

Clinician surveys in the health care industry

Contrasting views of physicians and nurses about the impact of computerized information systems were measured by Weiner and others using questionnaires.[74, 75] The questionnaires used 7-point Likert scales as well as open-ended questions to describe advantages, disadvantages, and desired improvements. The study revealed barriers to implementation of computer systems and showed physicians were more adverse to using the system than were nurses.

McKnight and others studied the perceived information needs by physicians and nurses at a New York hospital using surveys and follow-up focus groups.[76] A common theme expressed by physicians focused on information gaps in the system pertaining to medications, laboratory test results, patient problem lists. Although specific needs between nurses and physicians differed, both groups expressed difficulty in identifying and contacting other health care providers to get clarification about patient care given.

The Critical Incident Technique (CIT) methodology was originated by Dr. John Flanagan and associates in the Aviation Psychology Program of the Army Air Corps during World War II.[77] CIT is effective in collecting and categorizing incidents of behavior, and of events in a specific environment. The technique involves the collection of detailed reports of "incidents" in which an individual did something that was especially effective or especially ineffective in achieving the purpose of an activity. CIT has been used in hundreds of studies for improving performance in several diverse industries. The National Board of Medical Examiners used it to collect examples of

effective and ineffective clinical practice by interns and residents. Another study by the National Library of Medicine (NLM) used CIT to evaluate the impact of MEDLINE on medical decision making and medical outcomes.[77]

Clinician survey approach at IHC

Over the last fifteen years at IHC, hundreds of surveys have been conducted with a variety of audiences including patients, clinicians, employees, and community members. At least 12 periodic surveys specific to clinicians have been conducted at IHC. All of them contain at least one question that asks for the clinician's general perception about their experience with IHC. Surveys in 2001 and 2003 to IHC-employed physicians (n=251 and 291) included two questions related to a particular clinical computer application known as Clinical Workstation (CW). The first question asked the clinician to rate the extent that "the Clinical Workstation improves my ability to practice medicine," using a 6-point disagree/agree scale. The mean scores of 4.66 and 4.85 respectively were not significantly different. The second question asked the physicians to best describe how the physician personally uses the Clinical Workstation computer. Five statements were provided to gauge level of use ranging from "Full use at the point of care" to "Do not use CW." The responses showed that access to Clinical Workstation improved significantly over the 2-year period. From 2001 to 2003 physicians increased their use at the point of care (18% to 35%) and for data entry (12% to 26%). Speaking with the IHC Manager of Research and Planning, it was found that there has not been a survey conducted within the last 15 years that focused specifically on one or more IHC clinical information systems. User feedback on functionality of clinical information systems has been sought via other methods such as focus groups.

Patient record review to compare data values

The third method considered for triangulation is the review of patient records to compare data values across several data systems. Data value comparison provides a quantitative method to measure the degree of disparity across systems. Comparing data values found in patient records traditionally involves the use of a paper-based patient record system as the gold standard. Few studies compare the same data element across more than two systems at a time.

Patient record review in the health care industry

Nielsen et al. compared data values in patient records specific to several quality assurance measures. They assessed the variation between reports generated from an automated obstetric information system, and from a manual paper-based obstetric system used at a U.S. Military medical center.[49] Comparisons showed a 99% correlation between the data in the two reports.

Costakos et al. compared data values of 32 variables from the patient paper medical record with the same variables found in a computerized perinatal database at a Mayo health system hospital in Wisconsin.[78] The study examined the correctness of the clinical data and results show varying levels of agreement for variables compared. They concluded that users of clinical database information should be critically aware of the source and quality of the information they consume, and that every institution should undergo similar processes to audit data quality.

A study conducted by Goodwin et al. looked at issues obstructing data mining for improved outcomes. They found the main factors to be data quality problems, missing data, and data inconsistency.[79] It was noted that data coming from various sources often

leads to inconsistent data being stored in a common repository. Data element problems that prohibited patient records from being useable for factor analysis included incomplete dates, missing values, and free-text.

McGlynn and others abstracted data from patient records from several hospitals to compare whether patients actually received the care that is recommended. The overall results showed that patients received about half of the recommended processes for basic care in twelve metropolitan areas across the United States.[80]

Patient record review at IHC

Patient data values contributing to severity index scores found through manual chart review was compared to an automated CSI score to find the extent and causes of variation.[63] The most common reason for disparity was the lack of adequate codes representing CSI concepts in the HELP system. The other main cause was from nurses not using established system codes to identify the attributes used by the CSI system.

Wallace and others reported on the creation and use of an electronic logbook to replace the paper-based logbook for intensive care units at three IHC hospitals.[81] Verification methods included comparisons of select data elements between the paper-based documentation and the computerized patient record. Results showed specific differences between the two systems and that the electronic version improved data access, data quality, and ability to conduct quality improvement as well as enabled clinical research activities.

An IHC internal study in 2003 by Wallace, Stanfield, and Clayton, looked at electronic documentation quality in acute care.[82] They used patient records to compare data elements to the HELP system. The findings revealed instances of redundancy and

outdated (erroneous) standard treatment protocols being recommended. One of the most important conclusions was that automated methods to review electronic records are desperately needed for the evaluation of documentation quality, completeness and meaning.

Literature supports that data variation is a growing concern in the health care industry as well as within IHC. Challenges of integration and inaccessible data that accompany disparate data systems have been reviewed. Analyses of inconsistent data models and clinician survey studies help clinical systems developers better understand the technical and social implications of effective interfaces. Using patient record review to analyze missing and contradicting data will complete the triangulation objective of understanding the extent of data variation from multiple perspectives, and what must be done to create more effective clinical information systems.

METHODS

This study is divided into three aims, each showing from a different perspective the effects of using disparate perinatal clinical data systems in a health care institution.

Data Model Study (Aim 1): Assess the level of consistency of the data models used for a set of 30 data elements in three core perinatal clinical data systems at IHC.

Clinician Perception Study (Aim 2): Assess the types of missing and contradicting data problems as well as their impact on clinicians from a clinician perception, in the perinatal clinical data systems at IHC.

Patient Record Study (Aim 3): Assess the level of consistency of key data element values entered in patient records across the three core perinatal clinical data systems at IHC.

After analyzing each aim separately, the three aims will be triangulated to contrast, compare, and validate results. These analyses will provide a multidimensional understanding of the extent and effects of data variation in the three perinatal systems.

Methods: Data Model Study (Aim 1)

Data Model Study (Aim 1): Assess the level of consistency of the data models used for a set of 30 data elements in three core perinatal clinical data systems at IHC.

The Data Model Study is a descriptive study that evaluates the extent of the disparity between three core data systems: prenatal, L&D, and NICU EMR. Although there are many systems that are not integrated in the Women and Newborns Clinical Program, these three systems contain the core of the data that will make up the longitudinal record for the perinatal care continuum.

Criteria for sample selection

The selection process of the data elements in Data Model Study (Aim 1) used a combination of knowledge engineering methods, a model for data relevancy, as well as consensus by domain experts. The desired knowledge was a dataset with each data element being representative of relevant data elements across the perinatal care continuum. Buchanan, Barstow, Bechtal, Bennett and others present a framework for knowledge acquisition that identifies the major stages as: 1) identification, 2) conceptualization, 3) formalization, 4) implementation, and 5) testing.[83]

During the identification stage, preliminary interviews were held with the data manager, clinical project development lead, clinical nursing programs director, and various clinicians in the IHC Women and Newborns Clinical Program domain. These discussions established a foundation for understanding the characteristics, participants, resources, and goals of the domain. Data system components and key problem aspects were discussed as well as domain experts identified. Access was obtained to the L&D computer system, the NICU EMR computer system, and samples of the prenatal paper-based record were gathered. Preliminary high-level comparisons of data element categories across the three core systems were made to assess what information would be common. Individuals who were instrumental in the formulation and ongoing maintenance of the core perinatal systems were identified as possible domain experts. Clinical data systems used at each major facility and users categorized by clinician type were identified. Data elements for each system were identified from prenatal records, data element files, and SQL database queries of appropriate elements as illustrated in the Ven diagram in Figure 2. The prenatal record contains approximately 400 unique data

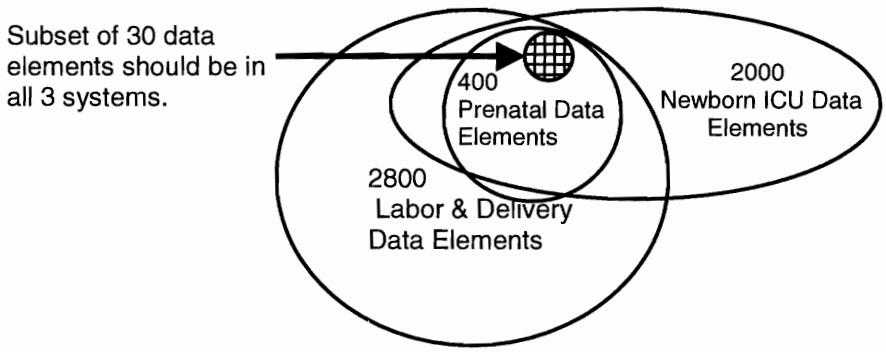


Figure 2. Overview of all data elements involved.
The majority of prenatal data elements are pertinent to both Labor & Delivery, and NICU according to recommendations by ACOG and IHC clinician review. Approximately one third of the elements in Labor and Delivery are also relevant in NICU.

elements, over 2800 elements are found in the L&D system, and approximately 2000 elements are in the NICU EMR system.

During conceptualization, domain experts identified some of the primary concepts, data integration problems, and key relationships among the data systems in the perinatal domain. Mapping of the concepts across systems gave a general picture of the assumed organization of domain knowledge.

In the formalization stage, sections from the prenatal record such as risk assessment factors were mapped to corresponding template screens in the NICU EMR, as well as to menu groupings in the L&D system. The collection of conceptual categories became the formal representation mechanism.

The implementation stage involved the use of the formalized knowledge to formulate rules. Two rules were used as sample data element selection criteria. The dataset elements must be:

1) commonly used data elements from the current prenatal record, that are relevant across the prenatal, delivery, discharge care continuum,

2) data elements, which if missing or are contradicting other sources of clinical documentation, would have a high potential to lead to a medical error.

Using these rules an initial set of 20 data elements was arbitrarily selected from the prenatal record. Iterative rounds of discussion and consensus by a panel of clinical domain experts validated that the elements met the criteria. In the process of discussion, several elements were eliminated while others were added. The consensus process increased the number of data elements to thirty. See Appendix B for the complete dataset. Comparing the dataset to national standard references further validated each data element. Standard references consisted of data sets from national associations specializing in women and newborns care information, specifically, "Perinatal Nursing" from AWHONN,[27] the "Guidelines for Perinatal Care" which is a joint publication from AAP and ACOG,[28] and a 2001 IOM study on quality of care.[3] See Appendix C for the comparison. The panel of clinical experts at IHC that refined and validated the sample data set was comprised of two physicians (medical directors of the Women and Newborns Clinical Program at IHC) and three nurses (directors and managers of Women and Newborns Nursing, Data Management, and Health Data Dictionary teams).

The testing stage involved a comparison method to assess both the degree of variation and form of variation in data structure (data model) across three core perinatal data systems. Thirty data elements were evaluated across three systems for a total of ninety comparisons. For example one comparison would evaluate the structure and meaning of the data element "personal medical history of hypertension" to see if it was the same in the prenatal record and in L&D. When comparing the prenatal to L&D, the prenatal would be considered the source system and the L&D would be the destination system.

A categorization method created similar to that used by Humphreys et al. to compare concepts in different terminologies.[84] The result of each data element comparison between perinatal systems was assigned a numeric score from 1 to 7 representing the seven following categories: 1) matching, 2) related but less granular in the destination, 3) related but more granular in the destination, 4) missing in the destination, 5) missing in the source, 6) missing in 2 of the 3 systems, and 7) entered as free-text. The scores are nominal and are not to be interpreted that a score of 2 is "more matching" than a score of 6. Descriptive statistics demonstrate the proportions of matches, mismatches, and missing data elements. The term "more granular" would be used when the options to define the data element values are either more numerous or contain more detail about the data it represents. For example, *"Chronic Hypertension"* is more granular than *"Hypertension."* Dates that describe when a laboratory test was ordered would make that laboratory test data element more granular. The documentation of several specific drugs used would be more granular than the term "Drug Use."

If the systems were perfectly integrated and if they shared the same data models, the result would be perfect matches in all 90 comparisons. A perinatal nurse administrator and a neonatal nurse practitioner, both very familiar with the use of these clinical systems, validated the element match results.

Methods: Clinician Perception Study (Aim 2)

Clinician Perception Study (Aim 2): Assess the types of missing and contradicting data problems as well as their impact on clinicians from a clinician perception, in the perinatal clinical data systems at IHC.

A descriptive study was conducted using the responsive-illuminative approach[66, 85] to gather viewpoints of the users of the data systems. Clinicians were asked about the

prevalence of missing and contradicting prenatal data in data systems. Their perceived prevalence of data problems was documented. Clinicians were asked about their behavior when dealing with data systems that had missing and contradicting data. Effects on clinician efficiency as a result of data variation across perinatal data systems were estimated.

In August of 2003, a clinician questionnaire was designed based on the Critical Incident Technique (CIT) to capture and analyze experiences of actual users of the perinatal clinical systems at IHC. CIT was the data collection method used within the responsive-illuminative approach. Traditional CIT methods were altered for this study in that a semistructured questionnaire was created and administered instead of interviews.

Criteria for sample selection

The target population for this study involved clinicians from various roles and locations who were currently using at least one of the three perinatal data systems involved. Selected IHC affiliated/non-IHC employed clinicians meet monthly in development teams and practice councils to collaborate on process improvement and quality of care issues. Arrangements were made to present and solicit participation at two physician development teams, and at three nursing practice councils. Attendees included clinicians who are focused both on nurse and physician administration, as well as direct users of clinical information systems.

Representatives from all regions (rural to large urban facilities) of IHC attend and bring a wide breadth of experience with different levels of information systems, both manual and automated. These users have the clinical and operational background to intimately know when lack of data or contradicting data occurs, and the subsequent

clinical and patient safety implications. Representation was sought from clinicians with different roles and from different facilities to take advantage of lessons learned from past experience with clinical system development at IHC. The HELP clinical practice management system, developed exclusively at LDS Hospital, was later implemented at multiple facilities in IHC's network. Discussions with IHC Operations confirmed that sufficient user acceptance of electronic clinical information systems requires frequent and active involvement from users at all levels and facilities involved.

Development teams are separate workgroups that focus on the development of clinical processes and best practice guidelines for specific areas. Practice councils are workgroups that focus on operationalization of protocols and best practices created by the development teams. Other tasks include care documentation forms creation and electronic data capture methods to track and follow-up on care process models for specific areas. The five teams/councils that received the questionnaire are as follows:

1) The NICU development team focuses on clinical processes for NICU and Nursery care process models. It consists of neonatologists from each region, pediatricians, a neonatal nurse practitioner, and nurse administrators.

2) The obstetrician development team focuses on clinical processes for obstetrical and gynecological care process models. It consists of maternal fetal medicine physicians, obstetricians from all regions, a nurse midwife, and quality representatives.

3) The NICU nursery practice council focuses on implementation of nursery care process models and involves NICU nurse leads from five hospitals.

4) The NICU mom-baby practice council focuses on implementation of antepartum, postpartum, and gynecological care process models.

5) The labor and delivery practice council focuses on implementation of care process models for patients in labor and delivery.

The coordinator for each team or council was contacted and arrangements were made to present the purpose and instructions of the questionnaire. The study was

introduced as an effort to identify gaps and inconsistencies in the present systems so that new systems could be designed and implemented in a way that clinicians would have relevant information when they needed it at the point of care. An attached cover letter explained the purpose, importance, who to contact if there were questions, the Institutional Review Board (IRB) statement about rights as a research subject, confidentiality, and estimated time to complete the survey.

The questionnaire was administered on September 17 and 18, 2003 to the appropriate clinicians. Each questionnaire included a self-addressed, postage-paid envelope with the investigator's address at IHC. Although the survey was confidential, a role code on each survey identified the team or council of the respondent so results by group and role could be determined. Only active users of the perinatal systems were requested to respond to the survey. Attendees who were not current users were excused.

Approval for this study was received by the IHC Urban Central Region Institutional Review Board (IRB) on August 19, 2003. The University of Utah IRB approved the study on October 15, 2003.

Data collection

The questionnaire was specifically designed to use both open-ended questions to flexibly capture the respondents' experiences, and also structured questions to measure scale and frequency of the incidents of missing data and contradicting data. Two usability analysts were consulted to validate the questions from a human factors engineering standpoint. Several interviews were conducted with the IHC nursing director and data manager for the Women and Newborns Clinical Program, both registered nurses, to ensure the questions assessed the correct characteristics of the study (face validity).

Various problems associated with missing or contradicting values of prenatal information were identified. Questions concerning the effects of the problems were modified and documented. The system medical director, IHC Women and Newborns Clinical Program also reviewed the revised questionnaire.

The clinician questionnaire is found in Appendix D. A two-column format was used to help the respondent separate the "missing" type incidents from the "contradicting" types. The first question on the questionnaire regarding missing data was,

"In the last month, did you have a problem with **not being able to find adequate *prenatal*** information about the patient?"

The first question for the contradicting data was,

"In the last month, did you have a problem involving **contradicting *prenatal*** information about the patient?"

Following the initial question, if answered yes, a series of additional questions was asked to collect information in a consistent and uniform way about what specific information was sought, what was the source, number of patients seen in a month, the number of patients with this kind of problem (missing data), potential for danger to patient, reason for danger, time required to resolve problem, and steps taken to resolve problem. The same sequence of questions for contradicting data problems was in the adjacent column.

Respondents were instructed both during the presentation and on the questionnaire itself that the scope of the survey focuses on the current prenatal, L&D, NICU patient care documentation environment. They were instructed to answer the questions as if the computer systems were working as intended and not down because of a power outage or unexpected virus. The intent of the instructions was to minimize the

responses that related to confounders of the study that were outside the focus of using nonintegrated systems.

A spreadsheet was created to record and tabulate all responses. Summary descriptive statistics will show the response rate by clinician group, percentage of respondents who encountered data problems, and data problems by the system used. Data was sorted by problems with highest frequency, problems by clinician group, specific context of the data problems with steps taken to resolve. Estimated count of problems and the time to resolve was aggregated. Processes are identified that clinicians use to obtain patient information when it is not available through the primary data system. Clinician perceptions of patient safety issues related to data problems are gathered.

Methods: Patient Record Study (Aim 3)

Patient Record Study (Aim 3): Assess the level of consistency of key data element values entered in patient records across the three core perinatal clinical data systems at IHC.

Patients with records in the three core perinatal systems were identified and the target patient records were accessed. Comparison rules for consistency were created and results of the comparisons were entered into a spreadsheet. Eight laboratory test result values from patient records were compared between two systems in different system combinations. Where records were available for all three systems, a three-system comparison was conducted. Descriptive statistics were calculated to show the percentage of matching comparisons between the following system combinations:

1. All three systems (prenatal, L&D, NICU EMR),
2. Prenatal record and L&D,
3. Prenatal record and the NICU EMR,

4. L&D and the NICU EMR.

Descriptive data were aggregated to show the prevalence of contradicting data values for all three core systems. Additional views of the data by the type of laboratory test were examined to determine if trends or patterns were evident.

Criteria for sample selection

Patient record. Medical records from LDS Hospital were used based on considerations that, a) LDS Hospital deals with the most births of any hospital in the IHC network, b) LDS Hospital uses both the L&D and the NICU EMR system. Table 1 shows the number of births and beginning dates of NICU EMR system use at several IHC facilities.

Initial sample. A search was run in the L&D system using an attribute called "Newborn to NBICU with Team." The search returned a list of the maternal record IDs and the infant record IDs for all babies that were transferred to the NICU. Records from the list were validated with the tables in the Oracle database to ensure the baby was the correct child of the specific mother using the maternal enterprise master member index (EMMI) with a linked infant EMMI. The medical record number for mothers at LDS Hospital were identified that related to delivery and maternity domains.

Table 1. Births by Hospital Location.

Numbers of births from January 1, 2003 to November 3, 2003 are shown at various hospital locations, and when each site started using the NICU EMR.

Facility	No. of births	Started using NICU EMR
LDS Hospital	3907	February 2003
Utah Valley	3799	Not used
McKay-Dee	3074	June 2002
Cottonwood	2964	Not used
Alta View	1627	Not used
Primary Children's	No Labor & Delivery	September 2002

Supplemental sample. Due to lower than expected number of patient records that were accessible from the NICU EMR, additional records were sought from the time period between August 1 2003 and September 30 2003. A new search in the L&D system was conducted using the new date range with the previous criteria and resulted in finding 87 new records. From that 87, 14 records were duplicates, 11 records were missing the mother data which was needed for this study's comparisons. There were 44 records in the L&D that did not exist in the NICU EMR. Eighteen records were in both systems and 9 were retrieved by the LDS Hospital Medical Records Department. These nine records combined to provide 38 patients that had records in all the three systems of prenatal, L&D, and the NICU EMR. None of the 9 records were duplicates of the earlier sample records selected. Ven diagrams in Figure 3 and Figure 4 show the differences between the best case available record count if systems were fully integrated, and the actual available record count in the three perinatal systems.

Data collection

The aim was to assess whether the data entered by clinicians at a variety of clinics using several different prenatal record formats, were consistent with data reentered in the L&D system, and again reentered in the NICU EMR system. To compare the data, a standard data model was needed. Escobar used dichotomous fields when possible to simplify the minimum dataset for the Kaiser Permanente Neonatal Minimum Data Set.[50] Following that method, a data sheet was created to convert values from the various

Availability of records across the 3 systems *should* be 100%.

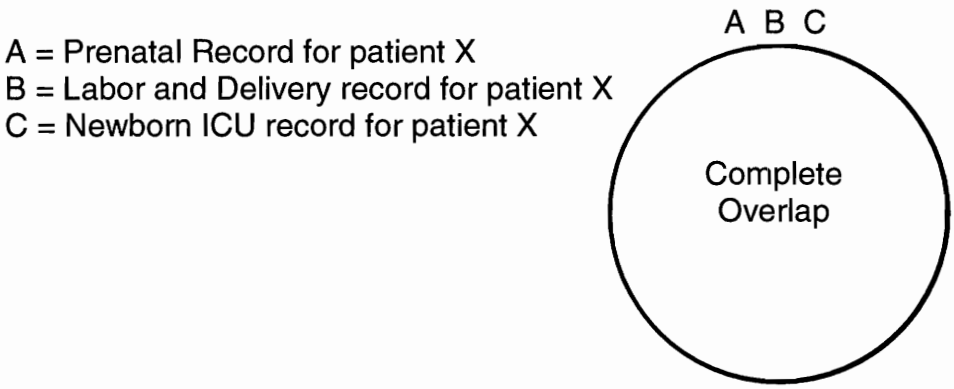


Figure 3. Best case record availability.
Records from all three systems should be available if data systems were fully integrated.

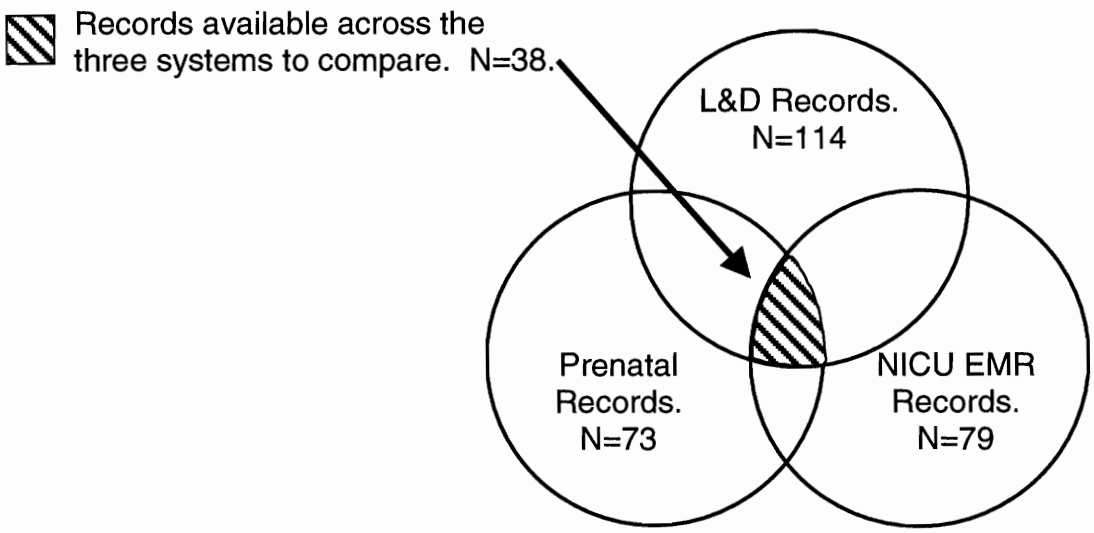


Figure 4. Actual available records.
Only a fraction of the patients in the sample had records in the prenatal record, the L&D, and the NICU EMR.

prenatal record formats to a consistent data model, which used dichotomous, numeric and categorical values.

There were at least eight different record formats encountered, each with its own data element position on the form and terms representing the concepts. Finding a common denominator for the target data elements was done by comparing the data elements across the prenatal record formats and splitting composite data elements into single data elements. Composite data elements in this study that appeared on the most recent version of the prenatal record were:

- Drug , medication, herbal therapy or radiation exposure in 1st trimester
- Drug / latex allergy
- HIV, herpes
- Previous cesarean section, scar type
- Smoking, packs per day
- All of the labs had result value and dates given

Appendix E Paper Prenatal Collection Sheet details the separation of each data element found on various prenatal record formats.

The data sheet went through several iterations of validation to ensure it would accurately and consistently capture the data values. Detailed rules for interpretation and abstraction were documented and reviewed and approved. Interpretation rules for each of the target data elements are found in Appendix F. Strategies to reduce Systematic Error included: 1) standardizing the measurement methods, and 2) refining the data capture instruments. Content validity was achieved by consulting with clinical experts and data structure experts. They assisted in defining the rules and use of data sheets to record the existence of data values on the prenatal record. Questions arose concerning variation in

allowable values for certain laboratory results. Experts at the LDS Hospital laboratory were consulted on how to interpret variations in values for chlamydia and rubella status screenings.

Data comparison

Categorization schemes were defined to identify matching and nonmatching data value comparisons. A categorization method similar to that used in the Data Model Study was used, but was adapted to compare values of data elements between the three core perinatal systems. Codes representing the result of the comparison were defined and are listed as follows:

1) values had an exact match,
2) values had a partial match but had narrower granularity in the destination,
3) values had a partial match but had broader granularity in the destination,
4) values were missing in destination ,
5) values were missing in the source,
6) values were missing in both source and destination,
7) values were contradicting.

Methods of triangulation

The objectives for using multiple methods of data variation detection were to elucidate the extent of data variation from several perspectives, and to use findings from one perspective to validate those from the other perspectives. Dependencies and interdependencies of data used by an EHR can also be more completely understood. Questions that triangulation can help answer include:

1) How does prevalence of mismatched data from a data model viewpoint compare to prevalence of missing data from patient record review?

2) What is the effect on clinician efficiency when incompatible data models and disparate data dictionaries are used?

3) Are the data elements that are not shared because of data model limitations the same elements that are perceived as missing by clinicians?

4) Is the clinician perceived prevalence of missing and contradicting data validated by patient record review?

5) What data elements should be compared across patient records based on the clinicians' perception of data most often missing?

6) How is the prevalence of missing data shown by other approaches perceived by clinicians in different clinician roles, i.e., nurses versus physicians?

7) What methods of adaptation are used by clinicians when presented with inadequately integrated information systems?

RESULTS

Results: Data Model Study (Aim 1)

Data Model Study (Aim 1): Assess the level of consistency of the data models used for a set of 30 data elements in three core perinatal clinical data systems at IHC.

The data model study is the first of three comparison methods used to detect the extent of data consistency and variation in data models across the prenatal record, L&D, and NICU EMR data systems. Categorization schemes were created to identify which data elements matched across the systems, which data elements did not match, and what is the form of data model variation for those elements that did not match. The following paragraphs describe the assignment of each of the 90 element comparisons (30 data elements x 3 data systems) into one of seven categories defining data variation. Data element comparisons and summary statistics are also given to display the proportion of differences between the prenatal record and L&D, prenatal record and NICU EMR, as well as L&D and NICU EMR systems.

The first category of data variation looks at element matches between two systems. For the 30 data elements compared, only 5 (17%) data elements had the same data structure and meaning across the three systems. Table 2 shows shaded areas for each comparison where a match occurred. The five data elements with shaded match indicators in all three columns were Gravida, Term (full-term birth), Premature (premature birth), Living, and Patient blood type. Those five elements were responsible for the 15 (5 X 3 systems) matches of the total 90 comparisons made.

Table 2 - Data Model Comparison of Data Elements Using a Categorization Scheme. The seven categorizations explain different forms of data variation when data elements were compared between systems.

	Data Element Conceptual Comparison	Prenatal to L&D	Prenatal to NICU	L&D to NICU
Gravidity				
	Gravida	1	1	1
	Term	1	1	1
	Premature	1	1	1
	Miscarriages/Abortions	3	1	2
	Living	1	1	1
	Multiple Births	4	4	6
	EDD by LMP	2	3	3
	EDC by US	2	3	3
Family History				
	Drug medication, herbal therapy or radiation exposure in 1st trimester	4	7	7
	Family history of birth defects	1	7	7
	Age > 35 at time of delivery	2	2	3
Obstetric History				
	Previous stillborn / neonatal death	1	7	7
	Previous infant admitted to NICU	4	7	7
	Previous Cesarean section, scar type	3	7	7
Personal Medical History				
	Diabetes mellitus	3	7	7
	Hypertension	3	7	7
	Kidney disease	3	7	7
	HIV, Herpes	3	7	7
	Drug / Latex Allergy	3	7	7
Risk Factors for Preterm Birth				
	Prior preterm birth (<37 weeks)	2	7	7
	Drug Use (including alcohol)	3	3	2
	Smoking pk /day	3	2	2
Lab Results				
	Patient Blood Type & Rh factor pos/neg	1	1	1
	Rhogam given date	3	2	2
	Antibody screen pos/neg/ titer / date	4	2	5
	Hepatitis B AG pos/neg / date	2	2	3
	GBS Culture pos/neg / date	3	2	2
	Rubella immune/not immune / date	2	2	1
	Chlamydia Screen immune/not immune / date	2	2	3
	Serology pos/neg /date	2	2	1

CATEGORIZATION KEY				Total
1= elements are complete match	7	6	7	20
2= elements are related, but more granular in source	8	9	5	22
3= elements are related, but more granular in destination	11	3	5	19
4= element is missing in destination system	4	1	0	5
5= element is missing in source system	0	0	1	1
6= element is missing in two of three systems	0	0	1	1
7= element in destination may be entered as free text	0	11	11	22
Total Comparisons:	30	30	30	90

It is interesting that four of the five matches were concentrated in area of gravidity. Findings show that an additional 5 data elements out of 30 (17%) had matches between just *two* systems. Combining the three-system matches with these additional two-system matches totaled 20 of 90 (22%) comparisons that resulted in a complete match. Data elements that matched between two systems are Miscarriages/Abortion, Family history of birth defects, Previous stillborn/neonatal death, rubella (laboratory test result), and serology (laboratory test result). The matches were seen in equal proportions among all three pairs of data systems. No predictive pattern for matching was observed.

Results for the second and third categories of data variation deal with more or less granularity in the destination system. Of the 90 data elements, 22 (24%) were less granular in the destination system while data elements that were more granular in the destination system totaled 19 of 90 (21%) as shown in the Key of Table 2. There were 20 of 30 (67%) data elements with some type of granularity difference when compared across the three systems. Data elements that had granular differences are shown on Table 2 as those comparisons with values of 2 or 3. The highest proportion of granular differences occurred between the prenatal record and L&D.

Results for the fourth, fifth and sixth categories of data variation deal with missing elements in one or more systems. Findings show missing data elements in 7 of 90 (8%) of the comparisons. The 4 of 30 (13%) data elements that were missing in at least one system were multiple births, drug medication, previous infant admitted to NICU, and antibody screen (laboratory test result). Data elements with categorizations for missing are shown in Table 2 as those comparisons with values of 4, 5, or 6. The

majority of these missing elements were found when comparing the prenatal record to L&D.

The seventh category of data variation identifies data elements that are entered as free-text in at least one of the systems. Twenty-two of 90 (24%) comparisons resulted in a free-text entry of the data element as shown in Table 2. The 11 of 30 (37%) data elements that were entered as free-text were all found in the NICU EMR. Free-text data elements are shown in Table 2 as those comparisons with values of 7. These data elements came from the Family History, Obstetric History, Personal Medical History, and Risk Factors for Preterm Birth contexts.

Graphic representation for the matching data elements between systems is shown in Figures 5, 6, and 7 between prenatal record and L&D (23%), between prenatal record and NICU EMR (20%), and between L&D and NICU EMR (23%), respectively. The largest difference between the three comparisons between systems occurred because of free-text entry in the NICU EMR.

Results: Clinician Perception Study (Aim 2)

Clinician Perception Study (Aim 2): Assess the types of missing and contradicting data problems as well as their impact on clinicians from a clinician perception, in the perinatal clinical data systems at IHC.

The second method to show data variation across perinatal data systems used a clinician questionnaire to look at the user's perspective of missing and contradicting prenatal record data over a 30-day period (September 2003). The following paragraphs give descriptive statistics for response rate by clinician group, percentage of respondents who encountered data problems, and data problems by system. Tables display data problems by highest frequency, problems by clinician group, context of the data

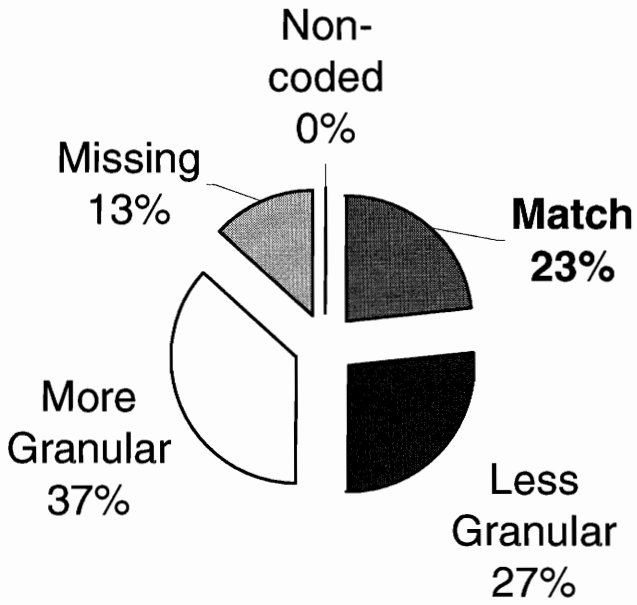


Figure 5. Prenatal to L&D data model comparison
Data model comparison illustrating the proportions of match, mismatch, and missing data elements between prenatal and L&D data systems.

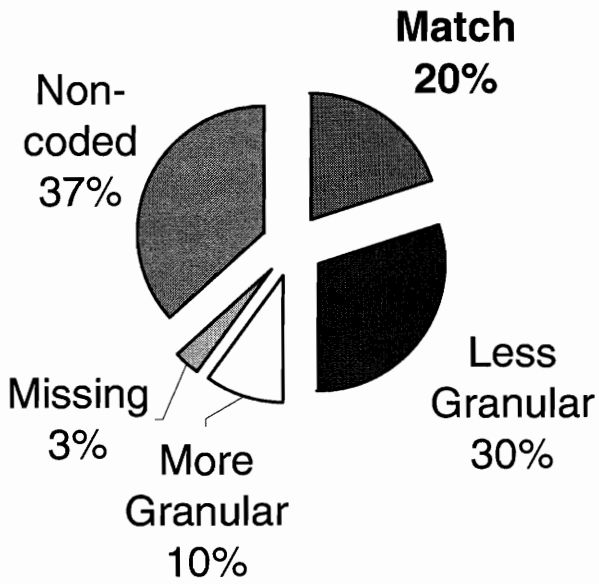


Figure 6. Prenatal to NICU EMR data model comparison
Data model comparison illustrating the proportions of match, mismatch, and missing data elements between prenatal and NICU EMR data systems.

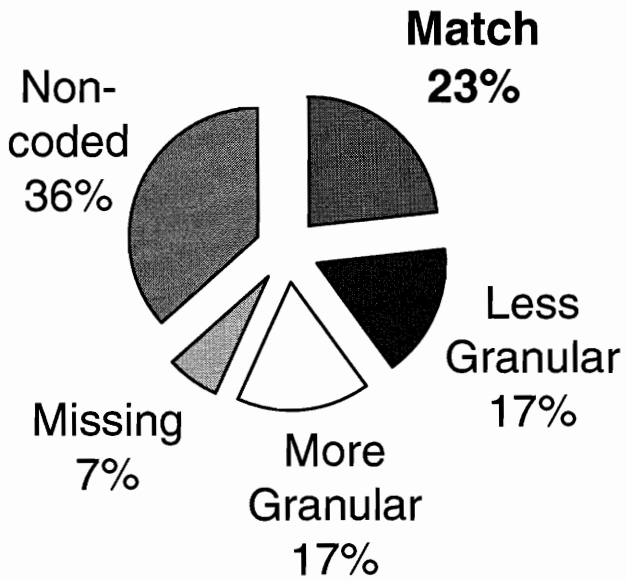


Figure 7. L&D to NICU EMR data model comparison
Data model comparison illustrating the proportions of match, mismatch, and missing data elements between L&D and NICU EMR data systems.

problem as well as steps taken to resolve missing and contradicting data. Processes are identified that are used by clinicians to get patient information when it is not available through the primary data system. Estimated counts of missing and contradicting data problems as well as estimated time to resolve those problems are given. Clinician perceptions of patient safety issues related to data problems are addressed.

The clinician questionnaire involved clinicians from five different care focus groups as detailed in Table 3. Eligible clinicians represent those clinicians who are actively using at least one of the three perinatal systems presently (prenatal record, L&D, NICU EMR). Within the five clinician groups, 47 clinicians met the criteria for eligible current users. Responses from 32 (68%) were received. All five groups responded with at least 50% participation from each group. Some responses were received from attendees of the councils who admitted they were not involved in the actual use of the systems and felt they were not qualified to complete the questionnaire, but were interested in the results from a quality assurance perspective. Those responses were not included in the analysis. Of the 32 clinicians that responded, one response was determined to contain outlier values and was withdrawn from the sample.

Table 3. Response Rate by Clinician Group.
Sixty-eight % (32/47) of the clinicians from five development groups participated in the questionnaire describing incidents of missing and contradicting prenatal data in different data systems.

Clinician Group	**Responded**	**Eligible**	**Percent**
NICU development	5	9	56%
Obstetrical development	8	11	73%
Nursing practice	2	4	50%
Mom-baby practice	10	12	83%
Labor & Delivery practice	7	11	64%
Total	**32**	47	68%

Results show that 23 of 31 (74%) clinicians reported they encountered problems with missing data, and 10 of 31 (32%) clinicians reported they encountered problems with contradicting data within a given 30-day period. Because the questionnaire was anonymous, any ambiguous responses could not be returned for clarification. The responses, however, were largely complete and most ambiguities were verified with the nursing director, Women and Newborns Clinical program. The only incomplete fields in the responses involved the number of patients seen during the month. If the field was not entered, that response for that respondent was not used in the total percent of patients affected by this problem.

Data for both missing and contradicting problems were charted by systems used. The prenatal record had the highest frequency of any sources for both missing and contradicting prenatal data. L&D, Tandem/HELP, Results Review and Clinical Workstation followed as shown in Figure 8 and Figure 9. Two systems (prenatal and L&D) accounted for 70% of the missing and contradicting data problems.

Results show the missing data problems with the highest perceived frequency were laboratory test results, specifically those for GBS. The highest frequency of contradicting data problems also involved the values for GBS test results. Table 4 contains the users descriptions and frequencies of the most common types of missing and contradicting data value problems. Specific laboratory test results were the main concern of clinicians, although there were some responses that indicated patient history and medication information was missing.

Clinician groups differed greatly in the number of data problems recalled. The total number of missing data problems reported (228) was much higher than those

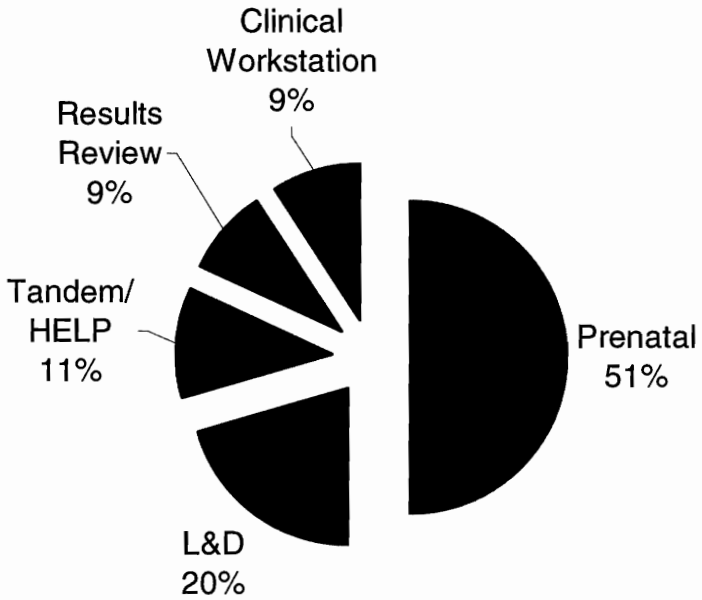


Figure 8. Missing data by system.
Clinicians indicated which systems they encountered incidents of Missing data within the last 30 days (n=24).

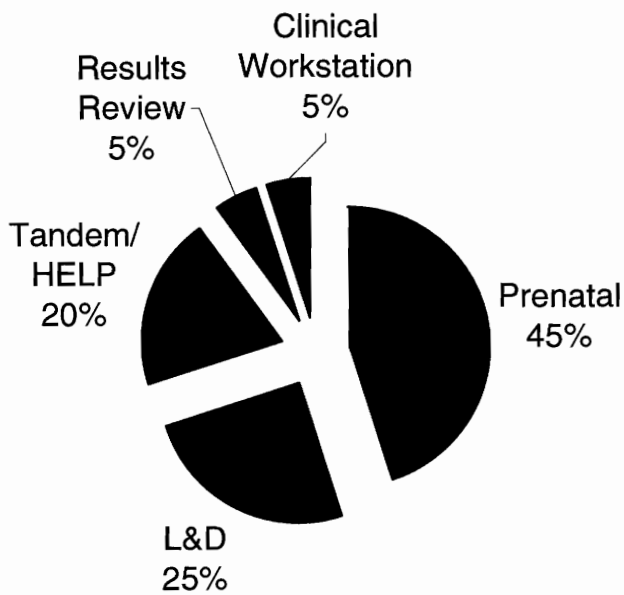


Figure 9. Contradicting data by system.
Clinicians indicated which systems they encountered incidents of contradicting data within the last 30 days (n=11).

Table 4. Problems Reported in Order of Frequency.
Clinicians listed the most common problems they encountered relating to missing and contradicting data.

Frequency	Order of Missing Problems Reported	Frequency	Order of Contradicting Problems Reported
9	Labs	4	Group B Streptoccus(GBS)
9	GBS	3	Hepatitis B Virus (HBV)
4	Blood / Rh	2	Blood / Rh
4	Rubella	1	HIV, Sexual Abuse, Prenatal Diagnosis, Ultrasound, Medical History
2	HBV		
2	History		
1	Medications Taken		
2	Prenatal (not specific)		

reported (18) for contradictory data problems. The values in Table 5 represent the clinicians' perception of the total number of patients that had (missing or contradicting) data problems in the 30-day period. Median scores for clinicians that recalled patients with data problems were 7.5 patients with missing data and 3 patients with contradicting data.

Clinicians described a variety of processes used to obtain missing data. Table 6 lists the context of the missing data encountered and the corresponding steps used to resolve it. The comment, "had to call Dr. office four times" was indicative of many of the resolutions stated. Table 7 shows the different laboratory tests reported as having contradicting data problems, as well as the corresponding steps taken by clinicians to resolve each problem.

The survey revealed a variety of informal communication channels and sources through which clinicians obtain patient information when present information systems do not provide needed data.

Table 5. Problems Reported by Clinician Group.
Total missing data problems and total contradicting data problems are shown for each clinician group.

Clinician Group		Missing Data Problems	Contradicting Data Problems
NICU Development	(n=5)	28	4
Obstetrical Development	(n=8)	10	0
Nursing Practice	(n=2)	9	1
Mom-Baby Practice	(n=10)	40	7
Labor & Delivery Practice	(n=7)	141	6
Total		**228**	**18**

Those channels are summarized in Table 8. Sixteen channels or methods to obtain missing or validating information were identified.

Time spent to resolve the problems for missing data ranged from five minutes to three hours per patient with each problem. Figure 10 shows total potential time spent resolving problems for missing data. Time was calculated for each respondent using the reported number of patient records missing data per month multiplied by the number of hours per patient problem. The number of contradictory data problems and time to resolve is shown in Figure 11.

Respondent entries were validated by a perinatal RN to ensure time spent would be per-patient times and not total time for all patient problems during the month. Twenty-three clinicians reported 227 cumulative missing data problems, which they estimated took 231 hours to resolve. Ten respondents reported 18 cumulative contradicting data problems, which they estimated took 77 hours to resolve.

Several survey questions gauged the clinician's perceived potential impact of data problems on patient safety. Clinicians were asked to use a 7-point Likert scale(1=not at all dangerous; 7=very dangerous) to rate the degree of danger to the patient if

Table 6. Missing Data Problems and Steps Taken to Resolve

	Missing data context	Missing: steps used to resolve problem
1	antenatal screening tests	Phone calls
2	blood type, rubella	call Dr's office, look in Tandem for lab results
3	different blood types reported	checked with lab
4	Down syndrome screening	phone call physician's office
5	GBS	look up lab results on Tandem, call Dr's office.
6	GBS status	look up data on RR, call office. If at night this would take longer, Patient would usually be treated.
7	GBS status	if information was found, then treated in accordance. If not, treated the risk.
8	GBS status, blood type, rubella status	call office to get reply
9	GBS status, prenatal lab results	called L&D, physician office, and lab
10	GBS status, rubella	checked Storkbytes for info, look in chart for prenatal record, check L&D for prenatal record, ask patient, notify nursery for unknown GBS status
11	GBS status, rubella, lab data	called office, pulled office and hospital labs, spoke with physicians, spoke with patient
12	GBS test done in physician's office	many of our BFS cultures are run at LDSH lab. We had positive results appearing on our lab results review, but not negative results. It was just blank. Our secretary called LDS lab to get resolved. Also, most prenatal records arrive in our department at 35-36 weeks gestation for each patient. the GBS culture isn't done by then. the office Drs. call for results, but there is a margin of error.
13	HBV and other ID information	I usually write an order to have the office record (often unavailable) checked within 24 hours
14	labs	call lab
15	labs	call Dr's office, but if weekend or nights - not possible
16	labs, GBS status	checked the prenatal record, checked the RR, checked with patient, called MD's office
17	labs, HBV	phone calls to L&D unit at referring hospitals
18	labs, history	called clinic, called lab, calling lab at other facilities for items sent out, redraw labs
19	labs, medical prenatal history	had to call Dr. office 4 times
20	No Response	Support staff helps with this
21	office prenatal information	call office, discuss with MD or personnel. Arrange for fax. Wait for fax. Check Clinical Workstation
22	prenatal meds taken	found copied chart and searched through record when had time to leave bedside
23	prenatal record info	notify clerk
24	Rh factor	called 2 different labs, called OB's office and patient needed to have her blood redrawn and tested.

Table 7. Contradicting Data Problems and Steps to Resolve

	Contradicting data context	Contradicting: steps used to resolve problem
1	antenatal screening tests	phone calls
2	HBV	some times just give up
3	had one record that said there was sexual abuse and this was found to be incorrect	finally asked social worker if the information was correct, and corrected our patient history form.
4	HBV	call Dr's office, look in Tandem for lab results
5	GBS status	called office and hospital labs, spoke with patient
6	Rh test results	called 2 different labs, called OB's office and patient needed to have her blood redrawn and tested.
7	HBV status	Discuss with MD who went to office for current update.
8	GBS status, Blood Rf different in two sources	if information was found, then treated in accordance. If not, treated the risk.
9	GBS status; prenatal diagnosis / medical history; Breast/Bottle feeding; ultrasound findings	Rummaged through chart to sort out information, called L & D and in some cases, asked the patient.
10	GBS was positive on computer, and negative on Prenatal	we trusted lab report
11	HIV testing on mother	phone call to the attending obstetrician

Table 8. Clinician Channels and Sources.
Responses listed by clinicians as the methods used to get information when present clinical data systems do not provide it.

Phone call to doctor's office
Phone call directly to physician
Phone call L&D, at referring hospital
Phone call L&D, internal
Phone call to Lab
Fax clinic to send records
Look up lab on Tandem/HELP
Look up lab on Results Review
Look up on Labor & Delivery system
Look up lab on Clinical Workstation
Pulled office charts
Left patient to search through physical files
Redraw labs
Asked patient for results
Discuss with MD or other personnel
Get support staff to locate and retrieve the record/information

Figure 10. Missing data estimated hours to resolve.
Clinicians encountered an estimated there were 227 cumulative missing data problems over a 30-day period, and that it took 231 hours to resolve them (n=23).

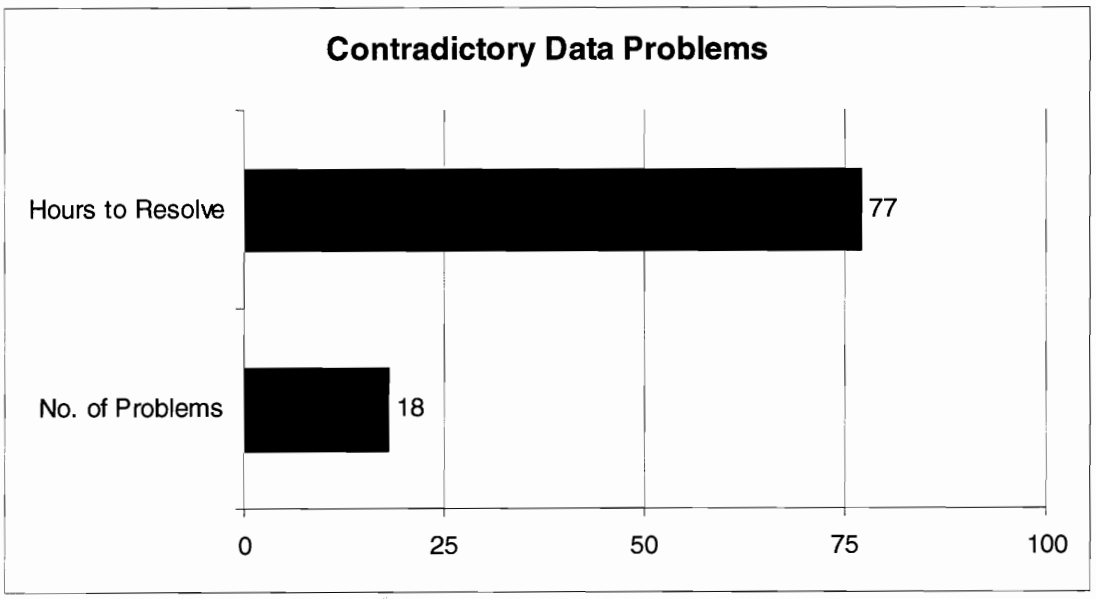


Figure 11. Contradicting data estimated hours to resolve.
Clinicians encountered an estimated there were 18 cumulative contradicting data problems over a 30-day period, and that it took 77 hours to resolve them (n=10).

this problem (one they had identified) had not been resolved. The mean score for "missing" data was 6.0 and the mean for "contradicting" data was 4.2. Several clinicians rated the danger level of missing and contradicting data as lower than the mean. They stated their reasons were because of the routine practice of treating for the assumed positive test, and erring on the safe side when information is missing. An example of a response about missing GBS status follows: "not dangerous because we treated 2- 3 people unnecessarily, just to be safe."

Results: Patient Record Study (Aim 3)

Patient Record Study (Aim 3): Assess the level of consistency of key data element values entered in patient records across the three core perinatal clinical data systems at IHC.

The third method used to measure data variation compared specific data values from patient records across the three perinatal systems. The following paragraphs describe the summary descriptive statistics reflecting the proportion of matches and occurrences of contradicting data element values for eight laboratory results. Proportions of matches are shown for: comparisons across all three systems, comparisons between prenatal and L&D, comparisons between prenatal and NICU, and comparisons between L&D and NICU. The number of occurrences of contradicting values when compared across all three systems is given. It is important to note that these results do not portray completeness or accuracy of data in any one system, only the degree to which the systems contain the same information. Matching comparisons for laboratory result findings were grouped from highest range of match to lowest.

A three-system match occurs when the data value of blood type in the prenatal record matches the blood type in both the L&D record and in the NICU EMR record. Results of 38 records show that blood type had the highest match percentage (71%). A

three-system matching graph for eight laboratory tests is shown in Figure 12. Matching percentages for rubella, HBV, and serology were grouped at around 50%. GBS was at 11%, while RhoGAM, and chlamydia had 0% match.

Data value comparisons between two systems (prenatal and L&D) for 73 records showed these two systems had the highest match range of the comparisons between any two systems in this study. Figure 13 shows blood type had the highest match percentage with 88%. Rubella, HBV, and serology were closely grouped around 68%. GBS was at 27%, RhoGAM was at 5%, with chlamydia and antibody screen at 1% and 0%.

Matching comparisons between prenatal and NICU EMR for 38 records show similar results as the three-system match range with the exception of antibody screen, which was at 34% matching compared to 0% in all other system comparisons. Figure 14 shows blood type matched 74%, while rubella, HBV, and serology was around 68%. GBS was at 16%, RhoGAM was at 5% and chlamydia had 0% match.

Comparisons between L&D and NICU EMR were based on 79 records and are shown in Figure 15. Blood type had the highest match percentage with 74%. Rubella, HBV, and serology were again grouped around 53%. GBS was at 14%, with chlamydia and antibody screen at 1% and 0% respectively. RhoGAM was significantly higher in this comparison at 52%.

Occurrences of contradicting data values and the associated systems are found in Table 9. A contradicting categorization occurs when a specific value (not including "unknown" or "undocumented") is different than the value in the other system. An example would be RhoGAM Given = "yes" in system A and RhoGAM Given = "no" in system B. While blood type had the highest range of match it also had the highest

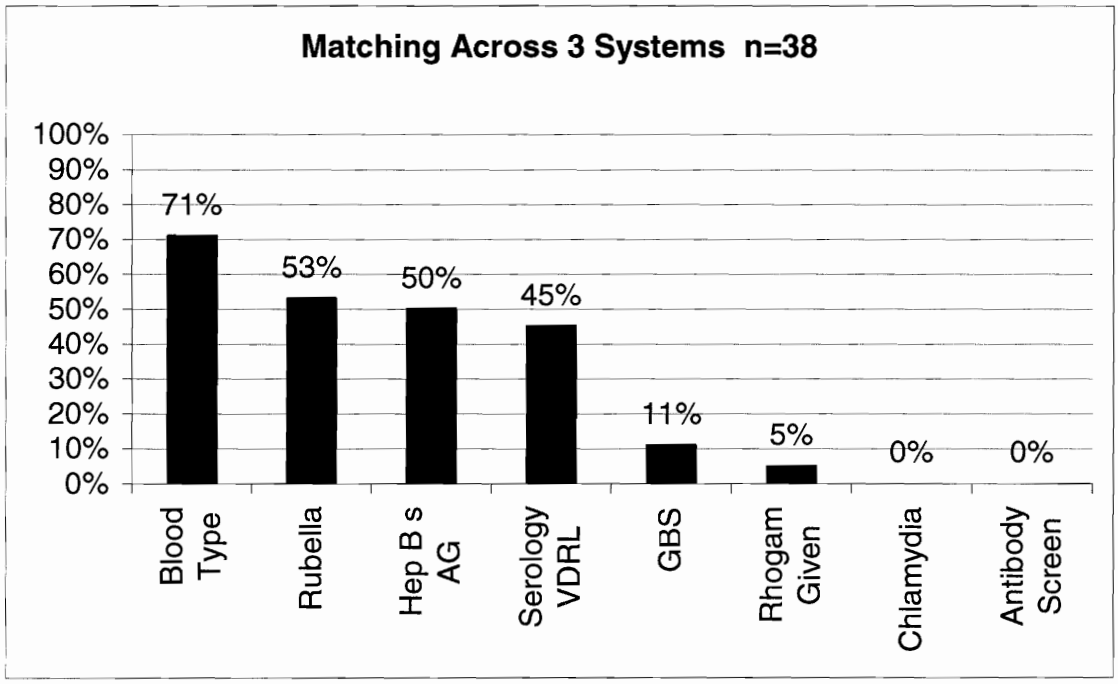


Figure 12. Three-system matching graph for eight laboratory tests.
Percentage of the time the data element value was a match in each record in all three systems.

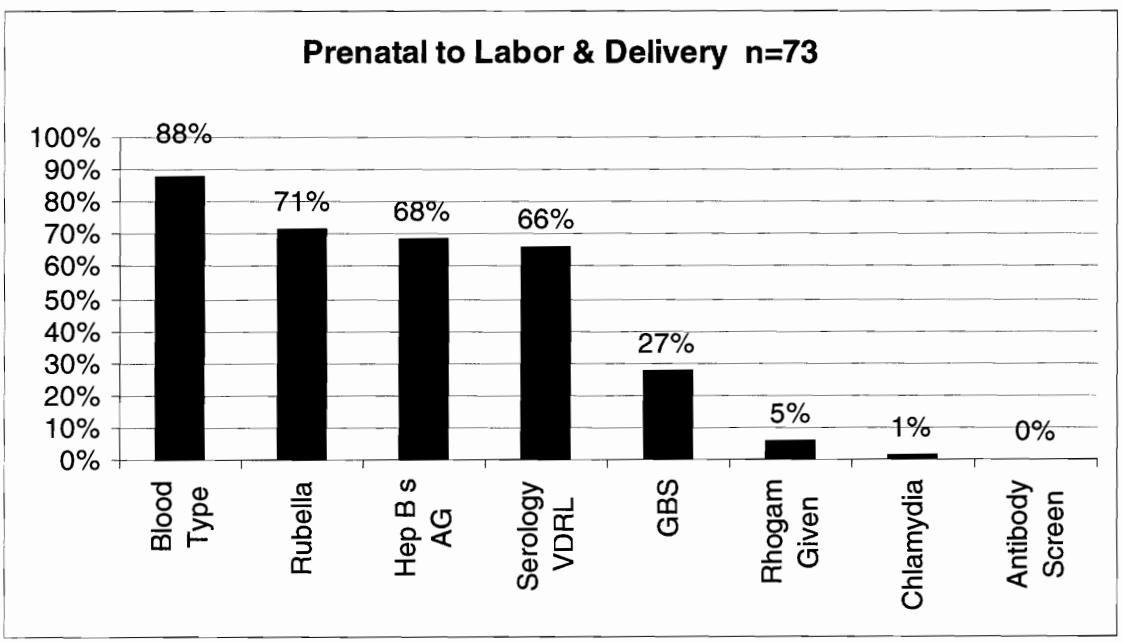


Figure 13. Prenatal to L&D graph for eight laboratory tests.
Percentage of the time the data element value was a match in each record in the prenatal record and in the L&D system.

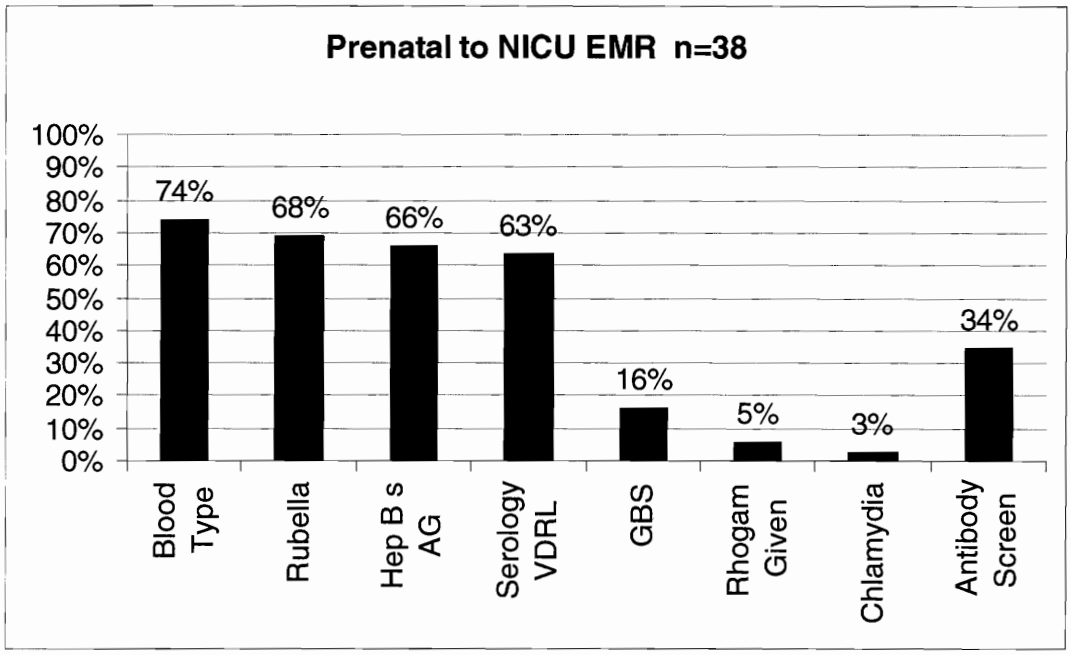


Figure 14. Prenatal to NICU graph for eight laboratory tests.
Percentage of the time the data element value was a match in each record in the prenatal record and in the NICU EMR system.

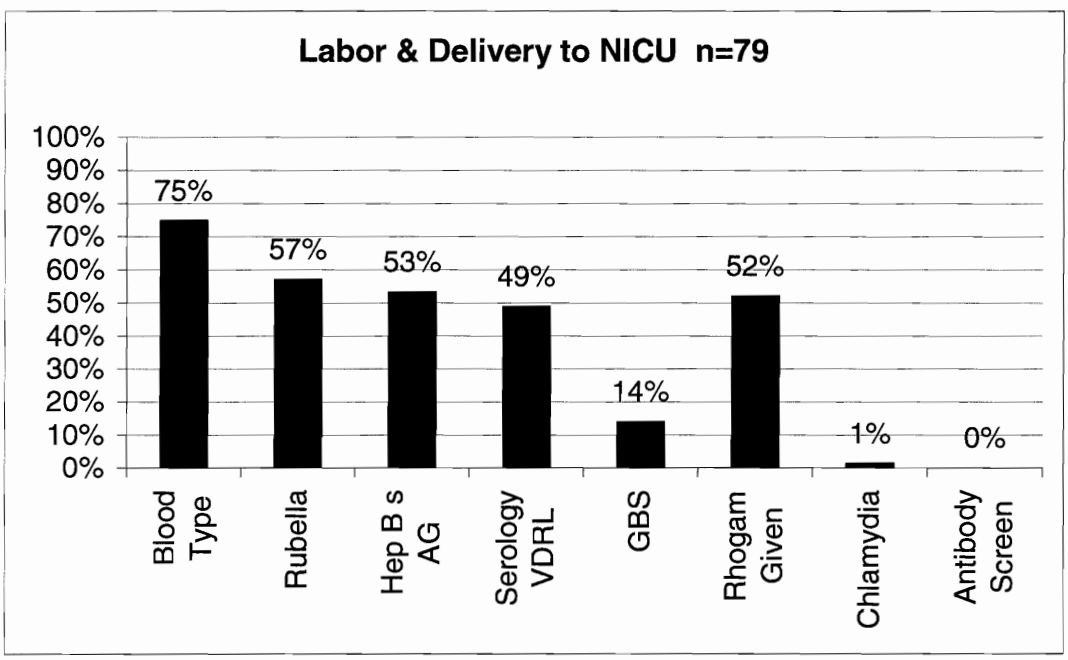


Figure 15. L&D to NICU graph for eight laboratory tests.
Percentage of the time the data element value was a match in each record in the L&D system and in the NICU EMR system

occurrence of contradicting values at 2.5%(2/79) of the tests compared as shown in the L&D – NICU column of Table 9. Six of the eight laboratory tests had at least one contradictory value detected in the cross-system comparisons.

The lowest range of match occurred when all three systems were compared, and the highest percentage of match occurred between the prenatal record and L&D as shown in Figure 13. The range of match is defined from the highest matching percentage value to the lowest matching percentage. Prenatal to NICU, and L&D to NICU had middle ranges as seen in Figure 14 and Figure 15. The trends stay consistent with the comparisons between prenatal to L&D being the highest levels of match for the top four laboratory tests. The last four ranking laboratory tests do not show any consistent pattern.

Table 9. Systems Associated with Contradicting Values.
Occurrences of contradicting data values for eight laboratory test results when compared between three core clinical systems.

	Prenatal - L & D	Prenatal - NICU	L&D – NICU
Blood Type	0	0	2
Rubella	1	0	0
Hepatitis B s Ag	0	0	0
Serology VDRL	1	0	1
GBS	1	0	0
RhoGAM Given	0	0	1
Chlamydia	1	0	0
Antibody Screen	0	0	0
Total	4 n=584	0 n=304	4 n=631

Tests for GBS(Group B Strep), RhoGAM given, chlamydia, and antibody screen will require individual analyses as to their low matching range.

Analysis comparing three data systems for data variation identified particular systems that appear to be more prone to contradicting data. In the sampled data, four occurrences of contradicting values were documented between prenatal and L&D systems, and four additional occurrences of contradicting values were detected between L&D and NICU systems. No value contradictions were found when comparing prenatal to NICU values. These findings infer that the L&D system is the highest source of disagreement of the three systems evaluated.

Results of triangulation

The three approaches used in this study show the existence of missing data and data variation across perinatal systems. Both qualitative and quantitative methods were used to validate each approach. Relevant data were shown to be missing from a data model perspective, from a clinician user perspective, and from a patient record data value perspective. The existence of data variation was congruent in each of the three approaches: first, as shown by the inconsistent way data are organized and stored (data model); second, as shown by the clinician perception and recollection of incidents of contradicting data in the last month; and third, as evidenced by the existence of differing data values in patient records across systems.

The data model study showed many of the data elements associated with medications, personal medical history, obstetric history, risk factors for preterm birth, and laboratory test results were missing or had variations in data models. In contrast, the

clinician survey showed that clinicians were mainly concerned with laboratory test results, and in only a few responses mentioned patient history.

Most of the clinicians acknowledged a problem of missing data across the perinatal systems. Of the eight survey responses received from the OB development team, which are mostly physicians, only two respondents recalled missing data incidents and there were no contradicting data incidents recalled by that group. Clinicians perceived that laboratory values in general and specifically GBS, blood type, rubella, and HBV results were missing and had variations. The patient record comparison study confirmed that the data from these test results were often missing or contradicting; thus, validating one study with another. Table 10 contains comparative metrics showing that missing data and variation in data occur in all three studies from different perspectives.

Table 10 Missing and Variable Data from Three Perspectives

Comparison Across Systems	**Missing Data**	**Contradicting Data**
Data model study (Aim 1)	13% (4/30) elements were missing in at least one system	70% (63/90) mismatched data model.
Clinician perception study (Aim 2)	74 % (23/31) of the clinicians recalled missing data	32 % (10/31) of the clinicians recalled contradicting data
Patient record study (Aim 3)	100% (8/8) Lab test results had missing data values. Best case comparisons (blood type) 29% missing. Worse case comparisons (chlamydia) 100% missing across systems	75% (6/8) Lab test results had contradicting data values. Worse case comparisons were 2.5% (2/79) contradicting (blood type) across systems

DISCUSSION

Discussion: Data Model Study (Aim 1)

Data Model Study (Aim 1): Assess the level of consistency of the data models used for a set of 30 data elements in three core perinatal clinical data systems at IHC.

This study shows the extent of several forms of data variation that exist across three disparate clinical data systems. The following paragraphs explore the veracity, relevance, and impact of the findings for the Data Model Study pertaining to matching data elements, granularity differences, missing data elements, and data entered as free-text. Limitations of the Data Model Study will be examined before discussion of the next aim. Recommendations will be covered for all three aims at the end of the Discussion section.

The sample size (n=38) is small and this research becomes a pilot study for future data variation studies. No similar studies were found from which to derive sample size and power. Clinical experts validated that the sample data set was a representative list to test the effectiveness of the existing systems in passing key and relevant information.

Matching data elements

Only 17% of the sample data elements matched across all three data systems. Only 24% of the data element comparisons resulted in a complete match between two systems. Four of the five elements that matched across all three systems are simple numeric data types related to gravidity. All of those data elements are passed electronically from the L&D system to the NICU EMR system. IHC clinical application

development teams have spent eight years refining and building consensus on standardized content and tying concepts to a common Health Data Dictionary. Much progress has been made; yet interfacing legacy clinical information systems has proven to be a larger challenge than was earlier anticipated. Efforts to resolve variations in data element types, contexts, and granularity across disparate systems have been very time-consuming. Difficult decisions about correct interpretations of meaning and clinical contexts have involved numerous discussions with clinicians and developers so that subjective assumptions are minimized. Although efforts are ongoing, the electronic medical record still does not contain all the information that is available and envisioned.

During IHC's project to interface data from the L&D system to the CDR, a clinical expert panel defined an initial data set of 50 data elements out of a potential 2,800 in the L&D that should be interfaced. To date 30 have been mapped and are routinely stored in the CDR. The difficulty in mapping supports this study's findings of data variation levels across the perinatal systems. Very similar issues are being addressed by the HL7 Vocabulary Technical Committee and terminology developers as they seek to realize semantic interoperability, and facilitate information exchange and use among disparate systems in all clinical domains.[86] Variation in data models in each of three core perinatal clinical data systems has created a substantial barrier in allowing systems to communicate key patient data to each other. These barriers result in much greater demand on the time and resources needed for integration.

Granularity differences

Granular variations in data elements are the effects of using inconsistent data models and data dictionaries. Of the 30 sample data elements, 67% had granularity

variations when compared across the three perinatal systems. In a study based on a perinatal data repository at Duke University's Medical Center, Goodwin found that 8.74% of the data was unusable for factor analysis despite the fact that data comes from multiple sources. One of the reasons for a successful integration was that the database had a clearly defined data structure that has remained stable for 20 years.[79]

The highest percentage of total granularity differences (60%) occurs between the prenatal record and the L&D systems. Data elements that should share meaning and structure across the two systems are not equivalent. Although humans can interpret the meaning of two similar but unequal elements, computer logic does not have that ability. With the creation of a longitudinal EMR, elements must be defined and share common references in a data dictionary.

All health care institutions are under pressure to become more efficient in using and sharing patient data for patient care; as well as exchanging data externally, as required by administrative, regulatory, and market forces. When concepts and data elements are inconsistent across systems, decision support tools cannot be consistently triggered, and outcomes research is compromised. If the L&D system had been built using a consistent concept-based data model and a vocabulary organized by domain, the mapping with other systems would be much more manageable. Hindsight from the development of different perinatal systems will benefit IHC and others who are engaged in the implementation of a longitudinal patient record in other domains.

The levels of granular variation are high in this study as a result of users being given flexibility to define data values and alternative data models in a system to such an extent that it became a collection of free text values. The result is that the system has

limited use for meaningful decision support with other systems, and limited value for longitudinal outcomes research. Inconsistent data models resulted in integration efforts that have been prohibitively time and resource consuming. Allowing such flexibility, creates a system incompatible with later integration needs and illustrates what Cimino described as "patchwork of terms with inconsistent granularity and organization."[87]

Many of the monolithic clinical information systems have been created from acquisitions of separately developed modules. This study's findings have important implications for users and integration specialists for those systems including: incompatible information models and mismatching terminologies between old and new systems need to be addressed; and unless a disciplined approach using a common data dictionary is used, concerns are warranted that those systems may face similar daunting conversions. Adoption and use of common data standards would reduce the risk when building or buying an EMR system because the transfer of data from a legacy system to a newer system would not be as great of a disruption in workflow.[22]

Missing data elements

The existence of missing data elements in a system shows that the singular focus of development for that system was on specialized functions and not on a longitudinal patient view. Of the 30 sample elements compared, four (13%) were missing in the L&D system. Developers of the L&D system did not see the relevance of these elements at the time of creation. The impact of not having all the relevant necessary data elements in an integrated system is that it places an additional burden on the clinician user to know which of multiple systems must be accessed to find particular data. As was shown in the Clinician Perception Study, without a single longitudinal view, users have to spend time

and effort navigating several systems to quickly obtain the information they need for decision making and patient care.

Studies by Gibson et al. have also found that determining the relevant concepts and codes needed in system for various purposes remains a constant challenge.[63] IHC domain experts verified that the sample data elements gathered from the prenatal record, such as maternal and family medical history, laboratory test results, and risk factors, were relevant and should be accessible to appropriate clinicians across the perinatal continuum. Questions on relevant data elements bring to the forefront content management issues that affect data quality, patient safety, reduced costs of integration, and efficient use of clinician time. The existence of missing relevant data elements also illustrates that because the field of medicine and health care will continually change, new concepts and representations will need to be added. Flexible data models and terminology models are necessary to accommodate those additions.

Free-text elements

Currently the NICU EMR system requires users to review paper printouts and reenter summarized prenatal patient information into one of several free-text boxes. The findings show 37% (11/30) of the sample data elements would be manually reentered into four possible text boxes in the NICU EMR. Taking coded data from one system and transforming it into text in another system often results in loss of meaning, and decreases the clinician's face time with the patient, while introducing potential for errors because of human interpretation during conversion as well as differences in spelling. As was stated previously, in any system, it is inevitable that new terms will need to be added.

Flexibility to add terms, however, needs to be balanced with the discipline to use a formal methodology adhering to structured data models and existing coded terms.

Data Model Study limitations

The three perinatal data systems used in this study included a paper-based system (prenatal record), a 20-year old specialty legacy system (L&D), and a more recent commercial clinical application (NICU EMR). These systems may not be representative of other separately developed systems in other institutions. Sample data elements were arbitrarily selected and validated by clinical experts. Clinical experts each have their own biases on views of what determines relevancy for data elements across a care continuum.

Discussion: Clinician Perception Study (Aim 2)

Clinician Perception Study (Aim 2): Assess the types of missing and contradicting data problems as well as their impact on clinicians from a clinician perception, in the perinatal clinical data systems at IHC.

The clinician survey study allows another perspective of the prevalence of missing and contradicting data in the different perinatal clinical data systems at IHC. Discussion points for this aim will look at the veracity, relevance, and impact of: survey response rate by clinician group; overall percentage of respondents who recalled data problems in the 30-day period; data problems by specific data systems; data problems with the highest frequency; count of data problems by clinician group; processes clinicians use to get information when it is not available through the primary data system; estimated time to resolve data problems; and clinician perception of potential safety issues of data problems. Limitations of the Clinician Perception Study will be examined before discussion of the third aim.

Response by group

Response to the survey by clinician groups ranged from a low of 50% (2/4) for nursing practice to a high of 83% (10/12) for mom-baby practice, with an overall rate of 68% (32/47). The overall response rate of 68% provides a valid representation of the different spectrum of clinicians who use the various perinatal data systems. The responses from the three practice councils represent nurse's view while the NICU development and OB development teams represent the views of physicians. The response by physicians (NICU, OB teams) was 65% (13/20), and the response rate for nurses was 78% (21/27); thus, showing a better involvement in the survey by nurses. To minimize confounders because of recall, a recent time period of the last 30 days was used. Although the sample population size for the groups was small, the recipients were all qualified users of the perinatal data systems and each was asked to be a member of the team because of the clinician's proven ability to represent the larger population.

Overall response

Distinction by problem type was divided into either missing data problem or contradicting data problem. Of the 30-one respondents, 74% indicated that they had encountered problems with missing data, while 32% recalled incidents of contradicting data. Other studies confirm finding problems of patient information availability,[12] and that both nurses and physicians had difficulty in obtaining certain types of data.[76]

Problems by system

Missing and contradicting data problems were identified by the data system in which they occurred. Findings showed that the prenatal record was the data system with the highest percentage of perceived missing and contradicting data. The L&D system

was second in frequency of problems encountered. The NICU EMR system has substantially fewer users, which should correlate to fewer problems reported. The many different formats in which prenatal information is received from outpatient clinics, illustrate needs for standards of data capture and the challenges of interaction with other services. Even when clinics use prenatal forms designed by IHC, differences in the data models used on those forms have introduced variation in what data can be effectively understood by other data systems. These findings are supported by Leape and others who showed interservice communication was one of the top seven reasons for system failures that can cause adverse errors.[12]

Highest frequency problems

Laboratory test results and patient history were the general contexts of data problems identified by clinicians as having the highest frequency. Studies by McKnight and others found the same foci in their studies when looking at perceived information needs of clinicians.[76] Several survey responses included multiple problems and were not always clearly associated with the system involved; hence, a breakdown of specific kind of problem such as GBS test result by specific system was not possible.

Specific laboratory tests data problems reported most often were for GBS, HBV, blood type, and rubella. The relevance of not knowing the patient status for GBS, HBV, and blood type will be covered in more subsequent detail. Respondents stated that when clinical information system records do not inform the clinicians of key patient information, additional laboratory tests are required and medical risk can be increased. Clinicians depend on integrated and timely information to provide correct treatment often under critical time constraints. The data problem with the highest frequency reported for

both "missing" and "contradicting" was GBS status. Standard care protocols require substantially more clinician attention in monitoring the infant when the mother has tested positive for GBS and has received intrapartum antibiotics to prevent early-onset GBS disease. Ensuring patient safety requires that there be 48-hour labor-intensive baby monitoring, bed checks every 4 hours, and frequent vital sign checking. Clinicians state that prophylactic treatments, which may be unnecessary, are routinely done to ensure patient safety. Several responses indicated that a general attitude exists that it is better to run tests again to be conservative.

Several implications should be noted concerning taking action without complete information. Drawing blood for tests on a 500-gram premature infant, for example, can have complications and actually cause harm. Some very small infants do not have sufficient muscle mass to accommodate injections without resulting in muscle tissue damage. Knowing the mother's HBV status in such a case may save the infant from that unnecessary harm. In contrast, if contradicting data were used and the administration of the vaccine was omitted based on an erroneous patient record, far more harm could come to the patient.

Another example of critical time constraints for clinicians involves preventative treatment of infants for HBV. When a mother has tested positive for HBV the current protocol requires that the infant receive the HBV vaccine in the first 12 hours after birth to prevent infection to the infant. Depending on the exposure circumstance, it is recommended that the infant also receive a treatment with hepatitis B immune globulin (HBIG) in the first 12 hours after birth. Women admitted for delivery whose hepatitis B surface antigen (HBsAg) status is unknown should have blood drawn for testing. While

waiting for the test results, the infant should receive the first dose of HBV vaccine without HBIG within 12 hours of birth. If the mother is found to be HBsAg positive, the infant should receive HBIG as soon as possible but not later than 7 days of age. Premature infants (<2000 grams birth weight) whose mother's HBsAg status is unknown should receive hepatitis B vaccine and HBIG within 12 hours of birth, unless results of HBsAg testing can be available in less than 12 hours.

The lack of patient information increases the demands on time and cost for rushed laboratory results. In addition to laboratory activities, clinician time and attention are required to research whether the previous provider has the results, submit a new order for a test, analyze the new results, consult with patient, and administer the vaccination/treatment. Missing data from such a scenario could cost the health care institution $100-200 considering the material and clinician labor. With inadequate medical history of the mother, physicians may not be sure what is wrong with the infant; thus, requiring closer monitoring such as with the GBS protocol that requires vital sign checking and capillary refill every 4 hours for 48 hours.[88]

Bates looked at the theory of error and pointed out that, "Investigation often indicates that an operator 'erred' because the system was poorly designed."[10] These situations raise the question of how often unnecessary treatment or tests occur and what is the impact. Several discussions were held with data analysts knowledgeable with IHC Health Plans actuary databases as well as the EDW. After some initial research in the EDW, it was found that measurements of prevalence or magnitude of unnecessary laboratory tests run, have not been done, and would be possible only with considerable manual effort using variable methods. Future studies are needed in that area.

Problems by group

Knowing problems by clinician group can help identify user patterns and attitudes by role. The questionnaire results showed that only two of the eight clinicians from the OB Development team encountered missing data problems and none reported any contradicting data. A closer look at the responses was taken, as other clinician groups did not share the same differentiation. It is possible that some clinicians don't see the problems with missing data because they rely on others to access and research the information. One of the OB team members responded that "few" missing problems were encountered, and then entered the following important comment in the Steps to resolve section: "Support staff helps with this."

There are personal costs of time and effort to learn to navigate the many data systems for necessary information. Some clinicians may choose not to take the time for learning or searching the various data systems and ask others to do it instead. There is an ongoing need for research in human-computer interaction studies and workflow analysis of clinical processes to find the best ways to make systems helpful for the clinician, and to discover why some systems are not used.[89, 90]

This study looked at a snapshot of how clinicians use clinical data systems to find prenatal information, and it would be incorrect to generalize that all doctors rely on staff for all computer access to data. Obstetricians do not perceive the same gaps and inconsistencies of data in the systems, as does the nursing staff. Responses obtained from the questionnaire support the existence of an elaborate hierarchy of verbal and duplicative systems for obtaining information when the primary clinical data systems are not adequate in providing the necessary data.

Considering the role of personal communication between clinicians, the distinction needs to be made between intradiscipline consultation with specialists regarding a complex case, as opposed to the task of actively polling several sources to get laboratory results or patient history that should have been routinely passed on with an integrated patient record. It is not to be expected that personal communication between clinicians be eliminated. Instead, inefficient communication should be replaced with a higher level of interchange. When complete laboratory test results and patient medical history information is automatically passed among clinicians, they are able to progress through the decision making process faster and are freed to seek and discuss more complex patient care issues if necessary.

Informal, interpersonal methods of getting information to make clinical decisions are in themselves inconsistent practices, and have a high potential for variation. Variation creates a greater potential for medical error, which adversely affects patient safety. Costs to provide care are increased by the need to support sufficient staff to make the calls, send the faxes, search through various records, etc.

If physicians do not directly benefit from using an automated data system, they are not likely to change their habits and workflow from relying on a paper system, especially if they perceive it takes them less time. Until entering and accessing patient information electronically is as fast and more dependable than using dictation, or looking at a clipboard with paper charts, some will not make the change. Studies by Bleich et al. found that integrated systems at Beth Israel Hospital were used heavily by clinicians because the systems were helpful.[47]

Use of a system where data are missing or response time is slow can also have an effect on clinician satisfaction. Sittig et al. showed that if users do not like the system they will not continue to use it.[89] To the extent that staff can provide the data as well or better than disparate computer systems, it is reasonable that staff will be the network of choice to get the needed patient information.

Processes to resolve

The specific steps or processes to resolve missing and contradicting data problems for each data context show the many choices clinicians must make in getting necessary information. Table 8 summarizes the different channels and again shows the potential for variation when clinicians seek patient data. Coiera also described situations when clinicians prefer conversational methods to obtain clinical information over computer systems.[90] Person-to-person discussion is preferable when the two agents communicating have established common ground. Common ground refers to the knowledge shared by the two agents about what data each other needs. It is very comfortable for a clinician to use the telephone to contact a laboratory support person.

When computational systems do not meet the clinicians' information needs, the default verbal informal channels and methods are used. The responses from the questionnaire reflect that clinicians are dedicated to the extent that they will go to many sources if necessary to get the key information needed for proper patient care. One of the definitions of the term mode is "a customary, or preferred way of doing something." When a data system is lacking the necessary data, clinicians react by actively seeking the information from a number of sources. When using disparate perinatal data systems, clinicians must poll multiple unstructured data sources for the missing data elements that

are needed. Each individual has his own mode of polling the various data sources as well as varying personal contacts.

Multiple channels exist in the process including, direct access to laboratory systems, radiology systems, fax machines, telephone, email, thumbing through paper charts, and delegation to staff to take on the search effort. Additional variations exist regarding access rights depending on whether the clinician is an IHC employee or not. Access rights stem from technology limitations related to the use of operating system related firewall and security issues. Temporal factors such as the time of day, whether the inquiry occurs after hours, or during the weekend, affect the polling process and can change the information obtained. Sequence variation of who gets polled in what order depends on each clinician's experience. Clinicians need to decide which systems or settings are most likely to have the information, be it for laboratory results, radiology, mother-related, infant-related, etc.

Another variable is the level of trust in the different data sources, which influences the clinician to go where he or she has had the best success in the past. McKnight's study also showed variation in information transfer between systems, and the occurrence of medical errors and near misses as a result.[76] The alternative informal mode of data access is noticeably more complex because it is infinitely more variable than a structured, well-defined procedure. Each individual has her own process for gathering certain kinds of information. Those who are comfortable with the arrangement, perpetuate the informal networks. Each clinician, however, still has to convert dissimilar data types, compare dissimilar values from multiple sources, and deliver it to a final decision making destination. Coiera, rightly supported the need for informal and

interactive conversations between clinicians when answers to questions are poorly structured and become clear only through the act of conversation.[90] Laboratory result values are fairly predictable and are readily understandable to clinicians. Verbal communication should not be required simply to obtain laboratory test results. The problems with clinician polling are that the methods are undocumented, ever changing and most importantly, the data are not stored, which limits the effective long-term sharing by other stakeholders in the patient care or outcomes research.

Some physicians argue that they prefer being called by laboratory personnel when results are ready as opposed to having to actively log into the system to check availability.[91] The impact of this arrangement is additional cost and time to notify the clinician. The elaborate system of polling modes comes at a cost in clinician and staff time and energy. It costs the institution to adequately staff such a communication channel. The variability in the channel is extensive since no two clinicians have exactly the same methods and personal communication habits, capabilities, personal connections, etc. It stands to reason that if variation is introduced through the use of personal networks; the variation in personal communication can lead to variation in clinician efficiency, variation in quality of care, and variation in patient safety.

An inconsistent electronic clinical information system perpetuates the use of the informal network. It is reasonable to conclude that even if problems are fixed, the user is hesitant to go back to the formal data system if the informal network is perceived to be more reliable or easier to use. Many studies have shown that EMRs are used more when they are complete, or have a critical mass of the patient data.[47, 92]

When the same data are captured multiple times in different systems, there is greater potential for contradicting data to occur. Earlier it was noted that Rosenal's study showed how redundant records could lead to errors and extra effort, misdirected data, overreliance on the spoken word, and inaccuracies. Some critics of rigid computerized clinical systems have maintained that redundancy has its place in clinical data systems. There are times when redundancy of systems can be positive and improve the likelihood for safer patient outcomes. Richard J. Croteau, MD, the Joint Commission's executive director of strategic initiatives, explained the role of redundancy by design in the Failure Mode and Effects method of analysis as:

> a systematic approach for identifying ways that a process can fail, why it might fail, and how it can be made safer. We need to protect the patient from results of errors, or failure of a safe design. We not only need a 'fail-safe' design, but **redundancy** in the system, because a backup system protects the process. This system must possess two independent, redundant steps, not interdependent steps.[93]

This view supports a strategy of careful workflow process analysis to identify which data elements must be verified by planned follow-up processes. These data elements would be determined by their potential to be used for key decisions or treatments such as blood transfusions or RhoGAM administration. Selective redundancy of capture and verification on certain data should be designed into a clinical data system.

The verification step is very different than the redundancy of capturing critical information because it was not integrated, or not initially entered in disparate systems. The first is reactive redundancy because of system or process inefficiency while the second is proactive redundancy as an added safety mechanism.

Time to resolve

The findings in this study show a sizable number of hours spent resolving missing and contradictory data problems. These are cumulative estimates based on the respondent's recall over the last thirty days. There are no audit logs or other reports established with which to validate the numbers. Distinction of hours spent by role (clinical staff, nurses, physicians) was not part of this study and it is acknowledged that the total time is only an estimate. An IHC perinatal RN verified the results for estimated hours as being reasonable. Despite imprecise numbers, the clinician's perceptions establish that a substantial amount of time is spent by clinicians resolving both missing data and contradicting data in the perinatal data systems.

Time is spent searching for data that should be readily accessible, and time is spent verifying data that is questionable because it contradicts data from other sources. Whether physicians rely on their staff to search for patient information, or they poll the different sources personally, several issues warrant future studies regarding:

- Amount of time the physician is kept waiting while data is searched
- Amount of time the patient is kept waiting while the staff polls various sources
- Amount of time, the staff spends on establishing and maintaining a personal network of trusted contacts for verbal communication to get information that should be easily accessible to the clinician at the point of care
- Frequency of delays when the verbal network, is "down" because inquiries occur after regular hours or on weekends when the source of the information is not available

The time clinicians spend searching for data is time that cannot be used for directly administering care to the patient. These findings have impact on clinician efficiency,

resource utilization, and institution efficiency. Cost issues are an important factor in an environment of reduced reimbursement and rising costs to do business.

Patient safety concerns

Clinicians' concerns about patient safety were gathered through question #7 on the survey. Findings of a high value of 6 on the Likert scale of 1 to 7, show the clinician's concern about the potential of adverse events if treatments or vaccinations were not given at the right time because of using incomplete or inaccurate information. Concerns were shared that inaccurate information from data systems might be unknowingly used that would result in a near miss or medical error.

Root cause analysis has shown that one of the potential factors in adverse events is lack of integration.[94] Richard I. Cook, MD, director of the Cognitive Technologies Laboratory at the University of Chicago, explained the significance of small failures when he stated,

> minute failures within a system align to produce a greater fissure: a problem, an incident, or a bad outcome. Accidents and incidents do not differ in their process, only in their outcome. An accident is an incident that has a bad outcome. Near-miss incidents and those with harmful results happen because of latent small system failures that combine to produce a large system failure.[93]

Clinicians must continue to go with their best judgment when complete information is not available. This study's findings show a need to reduce the variation in data systems so that the clinicians' judgment can be based more on complete information combined with evidence based best practices. Clinicians have the responsibility to decide if they will reject a validated protocol instruction or not, yet they need adequate and accurate patient-specific information if they are to select the best explicit protocol or guideline.[95]

The findings of the Clinician Perception Study show that when clinical data systems lack the completeness or availability of patient information, one of the common alternative methods used is to ask the patient. When the patient's memory replaces the provider's clinical data repository for treatment received and personal medical history, obvious limitations exist that can affect patient satisfaction and safety. Repeated questioning of the patient for information that has already been given, or for information that should be on record leaves the patient with the feeling that the provider is inept at capturing and sharing vital information needed for effective performance as a care provider. The 2001 IOM report states it is important to patients and their families that effective systems for transferring patient-related information be in place so that the information is accurate and available when needed.[3]

The clinician survey findings show the adverse impact on both patients and providers of having to deal with information systems that do not adequately capture and allow optimal use of patient information. The following response from the questionnaire expresses a clinician's frustration:

> Information determines potential treatment for mom and baby and can have a huge impact on care. It is also very dissatisfying for patient and family. It leads to families being asked the same questions over and over again, leaving them wondering why the information is not communicated correctly. Can be time consuming to sort out the information.

There are ethical and patient safety issues of relying on patient's memory. Consider instances when the patient may not be fully capable of giving reliable clinical information because of mental or emotionally instability. Patient safety is a factor when the reliability of clinical information is questionable. Information reliability may be suspect, for example, when coming from a mother about her medical history; or about the infant's condition just as the infant is taken from her arms right after birth to be

transported to a different facility for NICU care. Transport nurses often do not have time to collect the prenatal record to be included in the patient chart. Some patients have memory recall disabilities and should not be the default source of clinical information when data systems fall short. A foreign language speaking mother, or someone with a language barrier who does not understand the meaning or implications of jaundice, will not know how to get medical attention when it is appropriate.

Clinician Perception Study limitations

Physicians may not be representative of the larger population. IHC affiliated/non-IHC employed physicians, make up a substantial percentage of the providers who care for maternal and infant patients who are IHC Health Plans members. These physicians may feel pressured and inconvenienced to use IHC's systems in addition to using their own systems and methods for documenting patient care.

Questionnaires lack the ability to get complete responses on all questions. Some responses were left blank, several questions were misunderstood, and some answers were vague to the point of being unusable. An interview setting could have increased the validity and the completeness of the responses, yet time to administer would have increased and resulted in fewer clinician responses.

Discussion: Patient Record Study (Aim 3)

Patient Record Study (Aim 3): Assess the level of consistency of key data element values entered in patient records across the three core perinatal clinical data systems at IHC.

Comparisons of patient record data values are used for the last of the three studies to show data variation across the core perinatal clinical data systems at IHC. Sample laboratory test results are the data values compared across the three systems. The next

few paragraphs describe the veracity, relevance, and impact of the proportions of matching comparisons for the following: values across all three systems, values between prenatal and L&D, values between prenatal and NICU, and values across L&D and NICU. Implications regarding occurrences of contradicting values will be given. Challenges for data capture, patient record access, and data value comparisons will be examined. External validity and limitations of the Patient Record Study will be discussed.

This study is focused on the consistency of the data values between systems, not on the quality or completeness of data within a particular type of record or system. When looking at the data element example of RhoGAM given, the study is not concerned whether RhoGAM treatment ***should have*** been given, only whether it was documented as given, documented as not given, or not documented. The issue of whether a test should have been given and documented is outside of this study's scope and will not be addressed.

Comparing values across all three systems

Findings from the three-system comparison of the Patient Record Study (Aim 3) involve the comparison of specific values of data elements from the prenatal record with the same data elements in the L&D system, and again with what should be the same data in the NICU EMR system. Results showed substantial variation in data values across the three systems and are consistent with findings by Rosenal[8] and others who found that redundant records could lead to misdirected data and inaccuracies.

Several of the laboratory test value comparisons showed very low match percentages because of variations in procedures to enter data on the prenatal paper record, variations in the multiple versions of the prenatal record, and variations in data models

used between data systems. RhoGAM given, for example, was never marked as "No" on the prenatal record. The only option for the user was to enter a date indicating when RhoGAM was given. If there was no date entered, there was no way to definitely know if RhoGAM had not been given, or if the clinician simply did not know whether it had been given. The other two core systems asked the question in a different way (was RhoGAM Received?) that called for a different data element type and allowable values of "Yes," "No," and "Unknown." The high number of "Unknown" values in the prenatal record accounts for part of the reason why the match range was so low (between 5% and 52%). Even with data elements that shared common data models across all three systems such as blood type, the fact that 29% of the time blood type values were missing in at least one system indicates that data capture in disparate systems is not adequate.

Substantial challenges were encountered during the process of obtaining the patient record data for this comparability study. Data capture in the various systems had several process weaknesses such as time needed to enter data, institution policy on which patients records should be entered, training of users on systems, multiple versions of data entry forms, redundant data entry, and more. Once the data were captured, new challenges of access to the different systems had to be resolved and involved complexities of system integration and mother-infant patient identifier linkages. Nearly 300 maternal patients and corresponding infants were identified as potential sample subjects, yet records for only 38 patients had records in all three systems available for use in this comparison. When patient information has been captured, yet for various reasons is not accessible, the end result is that it is missing. Challenges and issues concerning capture and access of data will be further examined later in this section.

After patient records were obtained, the variation in data format and structure required adjustments so comparisons could be made. Findings from the Patient Record Study showed not only variation in the data values, but also substantial variation in the processes by which those values were obtained. The effect of inadequate processes for data capture and data accessibility was a reduced number of records that could be used for this study. Poor availability of patient data has a strong impact for other studies that deal with outcomes research, and process improvement for quality of care.

The nature of the RhoGAM data element requires a more in-depth look. RhoGAM is only given to mothers who have an Rh negative blood type. Of the 73 prenatal records that were reviewed, only 11 (12%) patients had an indication of Rh negative blood type. It would be expected that the three different system records for all eleven patients would be consistent. There were only 4 of 11 (36%) that had matching data values across all three core perinatal systems. The rest of the comparisons had missing data results. There were three additional patients with Rh negative blood where comparisons were made between the L&D system and NICU EMR. Of those three comparisons, only 1 (33%) was a match. From these findings, it is reasonable to infer that the clinicians needed to check several systems or sources to verify that the actual medical need was present for a RhoGAM treatment to be given, and to know whether the appropriate care was given. Secondly, without the use of consistent data models and consistent methods of data entry, data quality across several systems is jeopardized. When data accessibility and data quality are reduced, decision making by both clinician and by expert systems is negatively affected.[21]

The antibody screen test as a data element was found on most of the prenatal record formats, and is on the NICU EMR record. Values of positive, negative, and unknown are the allowed values that were compared. The L&D system is the only one of the three core systems that does not have a data element for antibody screen. Comparisons with the L&D system will always have a "missing" categorization. The low percentage of match (34%) for antibody screen between prenatal record and NICU EMR may be from clinicians feeling that the data is already in the prenatal record and is not critical to be reentered into the NICU.

Comparing prenatal and L&D values

Many of the inferences stated in the three-systems study also apply here. Records from 73 patients were available for comparison and provided a better base than the three-system study. Findings show blood type match (88%) and GBS match proportions (27%) were much higher between these two systems than in other system comparisons. Possible explanations may be that the prenatal record is more likely to be available to labor and delivery staff just before birth, than it would be to the NICU staff when the infant comes into that setting. Even though the overall match proportions are the highest between these two systems, the data shows missing data is still very substantial. Impact from these findings would lead us to infer that clinicians are not relying on data from either the prenatal record or the L&D system for laboratory results. Findings from the Clinician Perception Study confirmed that many clinicians are going directly to the laboratory results in Results Review or calling the laboratory directly for this information.

Comparing prenatal and NICU values

Many of the inferences stated in the three-systems study also apply here. As was in the case with the three-system study, only 38 patients had accessible records from the sample set to compare. The only substantial difference with this comparison versus the others was that the antibody screen match proportion was 34%. As was mentioned the L&D system did not have a data element for antibody screen so the comparison between these two systems will be the only place it would be greater than zero. Reasons for why it would not be higher could be that the test for GBS is not done until later in the pregnancy (week 35-37) so some prenatal records would have been sent to the hospital L&D Department before the GBS test was done. This brings up another implication that records from several systems are less likely to be updated with new information if the data is entered in the later system. Policies for updating all systems are inconsistent and even when policies are set, compliance to those policies is inconsistent.

Comparing L&D and NICU values

Many of the inferences stated in previous system comparisons also apply here. This comparison had the highest number of patients (79) with available records for comparison. The RhoGAM given match proportion (52%) was far higher in this comparison than in any others. Reasons may be that this data value is simply not updated on the prenatal record because it is documented in two other systems (L&D and NICU) that are further down the road in the care continuum and clinicians don't feel the need to update it on the paper record also. The fact that GBS had the lowest match proportion between these two systems was surprising. Since GBS status requires time-critical care for both mother and infant around the time of birth, it would be expected that the

mother's GBS status would be known and documented more often in the L&D and NICU EMR systems than in the prenatal record, yet findings show match proportions twice as high between prenatal and L&D.

Comparing contradicting values

Implications of potential mismatch scenarios based on the total number of births at IHC facilities are shown in Table 11. Extrapolated from 28,000 births per year, one would expect the blood type value to be missing in the prenatal record, yet be present in the L&D system 9.6% of the time (2,685 births). The value would be expected to be in prenatal but not in L&D 2.7% of the time (767 births). Depending on which of the two systems was referenced, the value for blood type could be missing as often as 12.3% of the time (3,452 births) over a 1-year period. Values for HBsAg status could be missing in the prenatal, or in the L&D, or in both systems as much as 31.5% of the time (8,822 births). GBS had the highest percentage of variation of values between the two systems. Values could be missing in one of the systems for over 71% of time (19,945 births) for GBS. Rubella, GBS, and serology tests also each had contradicting values between the two systems 1.4% of the time (384 births for each test).

Table 11. Potential Mismatch of Data Values for Births.
More than 28,000 births occur in IHC facilities per year. Numbers in parentheses represent potential mismatch scenarios between the Prenatal – L&D.

	Missing	Contradicting
Blood Type	12% (3,452)	
RhoGAM Given	64% (17,920)	
HBV Status	32% (8,822)	
GBS Status	71% (19,945)	1.4% (384)
Rubella	27% (7,671)	1.4% (384)
Serology VDRL	33% (9,205)	1.4% (384)

Implications from the variation in data values between the L&D system and the NICU EMR system also showed a large number of potential mismatch scenarios as shown in Table 12. Treatment of an infant for GBS should occur within the first week after birth, yet of the 2,800 infants cared for annually in NICU units at IHC, the GBS test result value would be expected to be missing in either the L&D system, the NICU EMR system, or both 86% of the time (2,408 infants cared for). Blood type values were missing in at least one of the two systems 23% of the time (638 infants cared for), and RhoGAM Given values were missing in at least one of the two systems 47% of the time (1,311 infants cared for).

Contradicting data values were also present for serology, RhoGAM given, and blood type. A clinician who refers to a printout from a system where the value is missing will routinely search other sources for the laboratory result, or would likely order a new test run. The contradicting value results, however, are particularly alarming in that a clinician accessing the system with the incorrect value, would not know it is incorrect, and would have much less reason to verify the value. Speaking with clinicians about the situation, it was conveyed that many of them routinely access laboratory results directly

Table 12. Potential Mismatch Data Values for NICU Infants.
More than 2,800 infants are admitted to the NICU per year. Numbers in parentheses represent potential mismatch scenarios between L&D – NICU.

	Missing	Contradicting
HBV Status	47% (1,311)	
GBS Status	86% (2,408)	
Blood Type	23% (638)	2.5% (70)
RhoGAM Given	47% (1,311)	1.3% (36)
Serology VDRL	49% (1,400)	1.3% (36)

from the Results Review application much of the time, and do not rely solely on the prenatal or L&D systems for laboratory results.

When reports are disregarded and additional sources are accessed, extra time is expended and variable methods are used to gather patient information. With the variation in clinicians' data verification methods, comes an element of risk that some clinicians may rely on a system with an incorrect result value. The risk of using inadequate information systems is one of the factors discussed in the recent Institute of Medicine report on patient safety. The report stresses that adequate, accurate information is critical to the provision of safe health care—care that is free of errors of both commission and omission. Efforts to improve and ensure patient safety must go beyond measuring adverse events by creating methods to measure and reduce near misses which the report defines as "an act of commission or omission that could have harmed the patient, but did not do so as a result of chance, prevention, or mitigation."[6] The follow-up tasks of verifying missing or questionable clinical system data qualify as "mitigation." Remedies must be actively identified for the chain of small system failures that often lead to a serious adverse event.

Data capture challenges

Throughout the Patient Record Study detecting data variation across clinical data systems was made more complex because of variability of systems and methods within systems to capture patient data. The findings show the effect on variation of the many paper-based prenatal records. Even when data capture is moved to an electronic channel, new challenges of time to enter data, user attitudes towards responsibilities must be managed.

Data capture challenges: Paper-based record. Paper-based records by their nature are not integrated with the other electronic perinatal systems, and introduce additional processes. Various methods for capturing patient data on the prenatal form were documented. Some clinics have the patient fill out the prenatal form themselves, which can lead to incomplete and erroneous information depending on the patient's understanding and recall about obstetrical terminology. Some clinics have their staff complete the forms. Here again, the user's clinical knowledge determines the data quality. Inconsistent data quality results from variations in the personal styles of users doing the data input on the prenatal record forms.

Variation of data representation was greatly increased by the use of multiple versions of the prenatal form. Eight formats were identified in the sample of prenatal records accessed for this study. Descriptions and differences of the various forms are found in Appendix F, Rules for interpretation of the prenatal record. Some non-IHC forms used formats and data elements from outdated IHC forms. Using out-dated form versions perpetuated differences in the data elements used. Three pages of the Hollister forms were commonly used with the first page of the IHC prenatal record. On one form, the value for RhoGAM Given was found in a comment section instead of in a separate data element, which was the standard method on most of the other record formats.

Names for tests and categorization and organization of data differed substantially on the different versions. The complexity of data interpretation was compounded by the variety of prenatal forms. In an effort to simplify and standardize the data capture, IHC updates and manages the version control of the prenatal record forms, as well as making them available for purchase through a third-party distributor. Many of the clinics

photocopy the forms when supplies are low to reduce costs, as was evident by the poor image quality of many of the records encountered in this study. Ideally the clinics would use the most recent form, but this is often not the case.

Simplifying the form completion for the clinicians will likely increase the use of the correct version. Some facilities have taken the IHC prenatal form, customized it with their logo, and printed enough to meet their inventory needs. One clinic has created an electronic form based on an outdated version. Clinics will deplete an existing inventory of costly paper forms even though versions are outdated. Of the prenatal forms encountered in patient records for this study, the majority were not the most current version, but were dated 05/01(May 2001).

Clinics see patients from various health insurance programs, each requiring a special form to be sent to the health plan organization, and also faxed to the hospital where the patient plans to give birth. Receiving the correct prenatal record from all the clinics is problematic in that IHC receives old versions of forms, non-IHC forms, or no form at all. Figure 16 shows the inefficiencies that attend the process of sharing prenatal record information with all the appropriate parties.

Prenatal records are regularly sent by the clinics to health insurance institutions for the purpose of early identification of high-risk cases, and to promote patient education on prenatal care issues. This study illustrates how automation and integration of the information found in the prenatal record might affect the health plans and the clinical stakeholders.

Certain data contained in the prenatal record is useful to IHC Health Plans objectives and has to be manually reentered into the Health Plans database. Health Plans

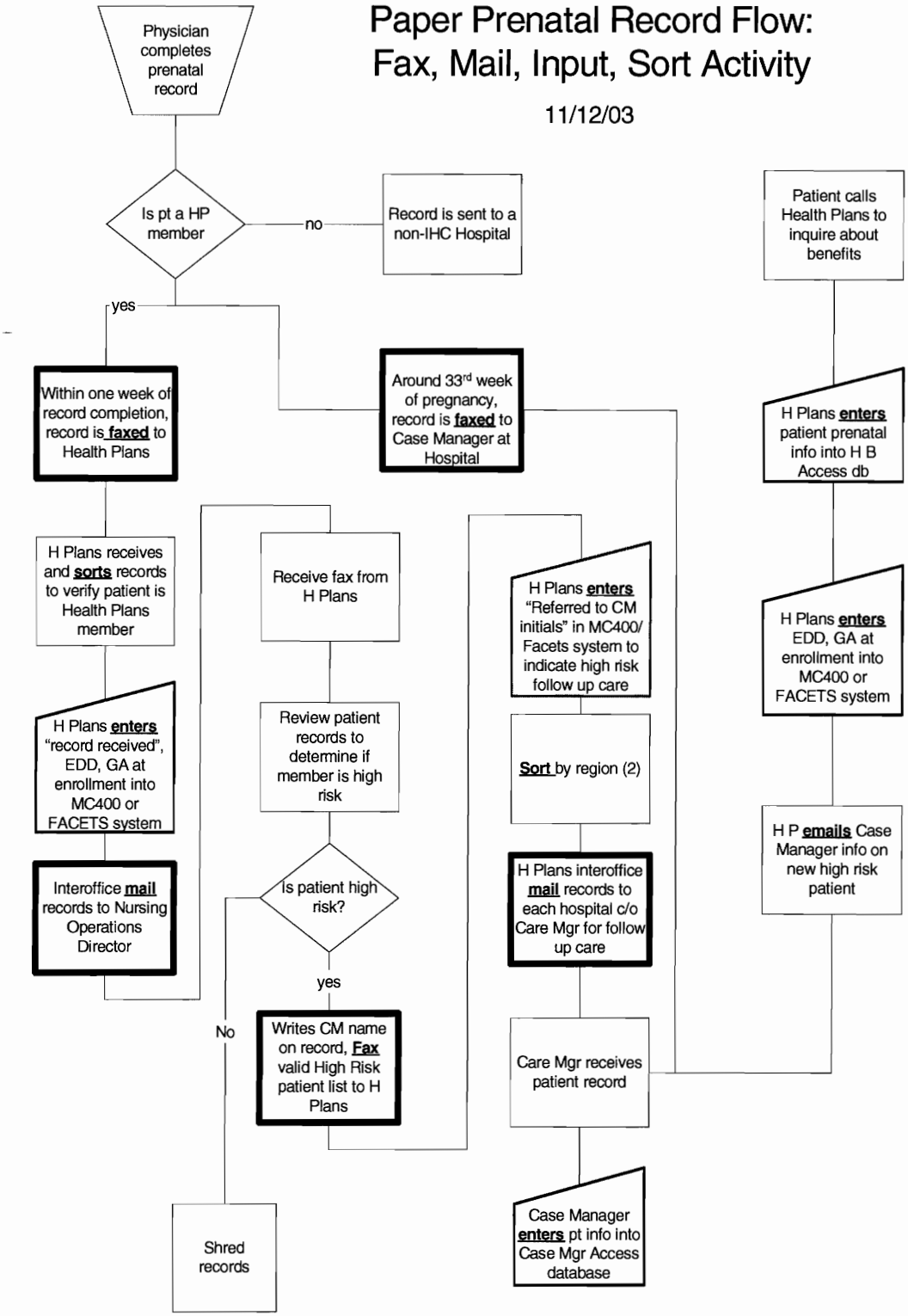


Figure 16. Paper prenatal record flow.
This diagram shows the multiple instances when the prenatal record is faxed, mailed, input into a database, or mailed.

intake coordinators use risk factors to identify patients who should receive prenatal educational literature about adequate prenatal care. Clinical perinatal care managers use prenatal risk factor data to identify and educate high-risk patients about early precautionary treatment. Prenatal records are supposed to be sent to the hospital L&D unit where the mother expects to give birth. A prenatal record takes a circuitous route once it is filled out at an outpatient clinic. Some of the maternal information entered on the prenatal record at the clinics, is reentered into two other systems during the prenatal stage (Health Plans Database and Care Manager Database). Much of the clinical information is reentered at the antepartum stage into the L&D system and into the NICU EMR at the postpartum stage. With four episodes of data entry into different systems, the likelihood of missing or contradicting information increases.

Efficiency and resource utilization are decreased when there is duplication of effort. The bold boxes on Figure 16 show the five different times the prenatal record is faxed. Fax technology is inconsistent and requires much more time to send and receive than would an electronic record that is integrated to receive appropriate information. Three separate sorting processes are used to determine where the prenatal record should be routed next. These processes could be automated if the data were in electronic form and integrated to appropriate systems.

When clinicians must use multiple systems and one of those systems is paper-based, it is common to printout results from the electronic systems and combine those printouts with the paper-based form, so that all records are in one place. Although the clinician may have all the information in one stack of printouts, a more effective method would be to capture the data electronically and generate a custom summarized view from

the aggregate information. Much time is used inefficiently by assembling paper printouts from multiple sources.

With all these variations in the data capture methods, it is difficult to measure the additional time spent, and frustration experienced by a clinician who has to search through up to four pages of at least eight prenatal record versions to find relevant data needed for making clinical decisions. Reaching consensus on standards for content and data models among a majority of the stakeholders who use the prenatal record would be a huge step in simplifying the data collection.

Data capture challenges: Computer-based record. Findings from this study elucidated the various technical, political/social, and clinical aspects that are a part of new clinical data system implementation. Initial access to the NICU EMR system revealed a number of missing sample records. Further research showed that many of the clinicians had not been trained to use the NICU EMR system and were not using the system. While talking with clinicians about the system, it was made known that IHC had set policies that required neonatal nurse practitioners and pediatricians to enter only qualifying infants into the NICU EMR. Infants qualified to be entered into the NICU EMR if they weighed less than 2500 grams, or the infant was on assisted ventilation. Data related to these situations were captured to meet the data requirements of the Vermont Oxford report.[96]

Many of the physicians using the system have their own methods in place to document the care for their patients, and the additional time to use IHC's system is regarded as an added burden on their time with no reward. Time to enter new patient data into the NICU EMR for pediatricians has been reported to take 15 to 30 minutes if

all the information is in hand. Gathering the necessary data for admission can take 1 or 2 additional hours.

Specific areas of limited functionality, slow response time of the application, and system down time have also been factors, which reduced the willingness of the pediatricians to enter NICU EMR records. As a result of these conditions, only 30% of the sample population had records in NICU.

Findings indicate that clinicians enter only those data elements they feel are important. Of note were the differences in the relatively consistent high rate of matches with blood type as compared with the low rate of matches with chlamydia. The findings showed that clinicians selectively choose what data elements are documented and which ones are left blank. Selective documentation will be the case whenever a response is not required regardless of whether a paper-based system or an electronic system is used.

Data quality suffers when variation exists in data capture methods, and when there is little control over the enforcement to enter the data. Data entry of patient care should not be regarded as optional since someone downstream in the data flow may rely on a particular data element for a crucial decision. To borrow a phrase, "what is one man's trash is another man's treasure," could be applied to data and information. The more consistent an institution is in capturing, storing and optimally using relevant patient clinical information, the better it will be in solving problems and making correct medical decisions.

Patient record access challenges

Several reasons for the inaccessibility of patient records in the three perinatal systems are examined. Records from the L&D system were accessed without much

problem, but determining which mothers were linked to babies that had visited the NICU, turned out to be much more problematic. Programming errors in the EDW were discovered that had caused the failure to store mom's identification in the baby record. Additional errors related to the mom/baby link in the NICU EMR system were found that caused a number of incomplete records that were missing portions of the mother's clinical data. In both cases, records have been updated with correct information, and the level of data quality and accessibility will have been improved as a result of this research. The fact that several tables associated with the EDW did not have the desired data illustrates points made by Moyer as he describes Data Warehouse pitfalls.[97]

Implications of a clinician not having access to a prenatal record are that the doctor might know important background such as a patient's history of congenital heart problems. Important symptoms may be missed and the patient may be sent home too soon. The mother may not be adequately educated on certain topics and may not know to watch more closely for certain symptoms. If either the mother or the infant has kidney problems and that information is not documented, it creates important gaps for the pediatrician who will care for the infant later.

Gathering patient records from multiple systems was more difficult than anticipated. Once the patient record for the mother and corresponding record for the infant were accessed, it was assumed that the mother would have a prenatal record in the physical patient chart. That assumption was met with disappointment as a number of patients in the sample did not have prenatal records on file. Reasons include they either did not have any prenatal care visits, or the documented care was not sent to the LDS Hospital. These were some of the roadblocks to access of both paper-based and

computer-based records. Similar problems of record availability have been documented by the 1997 IOM report.[2] The challenges the author faced with data accessibility was consistent with observations of Wallace et al.[81] who reported, "the acquisition of high quality, accessible clinical data is one of the significant barriers encountered in the quest for clinical practice improvement."

Data value comparison challenges

The original intent of the Patient Record Study (Aim 3) was to compare only the data values of data elements that had consistent data types across all three systems. The findings in the Data Model Study (Aim 1), however, showed only five data elements that met the above criteria, and four of those five were parity-related data elements. It was decided that the study could be more valuable if data elements with slight granular variations could also be compared. The findings from the Clinician Perception Study (Aim 2) highlighted specific data elements that clinicians perceived to be frequently missing or contradicting. With the results of the first two aims in mind, the laboratory results were selected to be the focus of the actual data value comparison.

The structure of certain data elements was modified to make the comparison of those data elements more meaningful. Instead of comparing the date when RhoGAM was given, the RhoGAM given values of "yes" or "no" were compared. The date modifier was dropped from laboratory test result values because the L&D system and the NICU EMR system did not have the option of recording dates when laboratory tests were given. A limited value set of positive, negative, or unknown for laboratory test values was compared. These adjustments and comparisons were made to more realistically reflect the environment in which clinicians actually function. It illustrates the need for

individual subjective decision making that clinicians must do when information is incomplete or different data models are used.

Comparison of data values for the purpose of assessing data quality raises several questions: can the data element value be derived from other data on the prenatal record, and what are the rules the clinicians must follow in documenting the data element? Appendix F Rules for interpretation covers criteria for deriving data values. Although some elements can be grouped with related items or similar data structures, each data element requires individual attention as to whether its value can be derived from existing data.

As data value comparisons were conducted, it became apparent that many of the data elements had complexities and ambiguities to the extent that a valid comparison of values would not be feasible. Many of the original sample set of 30 data elements were excluded from value comparisons because of these reasons. Measuring data quality relies on a consistent comparison process. Complexity of the data element definition has a substantial effect on these comparisons. Factors such as variations in the names of medications, and determining whether a specific medication is the same as one documented in another system, are issues that require clinical expertise and judgment. To compare terms related to diseases, an in-depth clinical knowledge of the various terms and synonyms of all diseases is necessary; and even then, interpretation by clinician can vary. Comparison of data values for medications and disease history across systems was also excluded from this study for those reasons stated.

Several data elements were examined where two or more concepts were represented by one data element, i.e., HIV/herpes. The clinician completing the prenatal

form is given one check box to indicate that either condition exists. If the user checks the box, it is ambiguous whether HIV, Herpes, or both conditions were selected. Later follow-up is required by the next provider in line to interpret and provide appropriate care. Using combined concepts on documentation can cause loss of meaning and loss of clinician time. Combining concepts was done by the request of clinicians in an effort to reduce the number of sheets needed on the form. The "shortening" has introduced ambiguity down stream in the interpretation of the documentation. Use of atomic (postcoordinated) data elements makes the input form longer. If molecular (precoordinated) data elements are used they must be mutually defined and understood by both the sending and the receiving systems. Findings by Costakos et al. emphasized the importance of designing data elements and databases that have clear definitions to minimize variation.[78]

Another example of inconsistent data models that causes data variation involves ambiguity of personal or family history. Frequent confusion occurs when a heading lists "Personal or Family Medical History." As patients check a box, it is unclear if they are referring to themselves, or to a family member. Ambiguity is increased when clinicians make up their own codes. Several different methods were used by clinicians to add codes or comments to data elements on the various prenatal records. Although some methods had the potential to add helpful information to the record, it depended on the clinician's ability to read and interpret the handwriting that was written in a very small area on the form. Examples from the Personal Medical History section found during this study include: a check mark would be found next to Cardiac Disease and to the right of the term, the letters "MOM" or "MGM" or "FAM" were written. A section on the form for

Family History exists, but the specific disease was not included in that section. The specific disease, however, was included in the Personal Medical History Section. Only when other clinicians accurately interpret these codes as the patient's mother or the patient's maternal grandmother or the patient's immediate family that had cardiac disease, is the information relevant.

In the Obstetrical History section, patients may have family history of twins and will check this box. The box should be marked only if the patient personally has had twins. Shortcuts to achieve faster completion of the record, or to use less space can cause loss of meaning and ambiguity. The ability to read and accurately interpret these notations is not universal, and it is information that cannot be used in electronic decision support or outcomes research if there are not consistent data elements defined in a controlled medical vocabulary.

External validity

The findings of this study emphasize the need for designers of clinical applications to consider how integration could be facilitated early in the development cycle. Challenges from disparate systems stress the importance of a concerted migration toward the use of common terminologies and common data models when updating existing systems as well as when developing new systems. IHC is in an interim state where the new platform has not been completely defined, yet specialty systems continue to be developed or updated. Database development at local department levels have taken place over the last several years in areas such as the Pediatric Cardiac Catheter Department, fetal ultrasound tracking, and case managers of high-risk pregnancies. These

systems have been developed without the coordination of corporate resources knowledgeable with standard medical vocabularies or interface guidelines.

Patient Record Study limitations

Sample patient records were used from only one hospital and may not be generalizable to other facilities. Variation in polices at different hospitals was evident and introduced confounders in the consistency and data quality of the patient records in the systems evaluated. In some facilities, neonatologists were required to document patient care in the NICU EMR at certain facilities but not at others. Medical Records personnel at one hospital comply with a local policy to routinely make copies of the prenatal record and see that those copies are delivered to the L&D unit at that hospital. Other facilities do not share the same procedure. Variation in policies across facilities was discussed and verified by clinician workflow experts familiar with each of the Women and Newborns information systems covering multiple IHC facilities.

The NICU EMR system was new to the LDS Hospital and the training of all clinicians had not been completed before the NICU EMR records for the sample were selected. The result was a smaller sample size of available NICU EMR records to be used for comparison.

The study used a relatively low number of patient records from each of the three core perinatal systems. Limited availability of records illustrates the practical challenges in using patient information for studies.

There may have been patient sample bias as all of the patients in the sample were associated with an infant with problems serious enough to warrant a visit to the NICU. It is possible that clinicians were more attentive of these patients along the care continuum.

The additional attention may have a confounding effect in that the clinicians may have been more likely to enter the documentation about prenatal information than for other patients who did not have the same severity level.

Discussion of triangulation

Triangulation of methods and data sources can provide a more complete picture of how clinical data is captured and organized within an institution. It is helpful to briefly look at the evolution of methods and motivations for the data capture process. The focus and objectives of computerized clinical information systems have evolved since their inception. Much of the initial focus was on coding data for billing and charge capture. Gradually that focus included more information about clinical care. With that shift in focus the nature of the data changed to accommodate the clinician's workflow and vocabulary. Clinicians began defining the data to be captured and the terms used were those particularly meaningful to the clinician user base. The data acquisition process sometimes introduced redundancy and ambiguity of data elements stored because consistent data models and common data dictionary were not always used.

It is difficult for clinicians to immediately see the problems caused by inconsistent data model usage. The informaticians involved in actively mapping data elements see those effects more vividly. There is an indirect effect on clinicians as they would not have the ability to use the data for outcomes research, decision support systems, and viewing a complete patient EMR. Hripcsak et al. pointed out that decision support tools such as clinical event monitors are only useful to the degree that the clinical encounter is captured, both as clinical events and as data in the patient database.[58]

The decision of what data to store and how to store it requires a collaborative effort, and clearly points to the importance of involving experts from both the clinical domain and from the medical informatics domain when designing and implementing clinical information systems. Without collaboration when data are captured for a specific domain, their use is often limited to that domain only.

Systems with disparate and incompatible data models are created. At a later point efforts to integrate those disparate systems are found to be problematic. When integration is prolonged or delayed, duplicate data entry is perpetuated and redundant data in separate repositories exist. The users' diligence in consistently entering data gradually decreases because they know it is available if needed in another system. Inconsistent data entry causes gaps in patient records.

Clinicians try to fill in the pieces by developing informal polling methods and networks to obtain segments of patient data from various sources. These informal networks take more time and require more resources to maintain. Data access methods begin to evolve and differ by clinician. Some clinicians need only a narrow view of the longitudinal patient record and do not want to take time to enter data that is not directly useful to them even though it may be useful to another clinician later in the care continuum. When user interfaces are slow or cumbersome, they hinder the willingness of clinicians from entering more complete information about the patient. Variation in access methods leads to variation in decisions about care and care practice. Risk of medical error is increased when there is variation in practice. Best practices are difficult to identify without accurate and complete longitudinal aggregate data. Costs are increased when inefficient practices or unnecessary practices are used.

Recommendations for reducing data variation

Promote standards early in development

The importance of using common data elements with consistent levels of granularity across applications should be addressed early in the development process. Regardless of the initial scope of a new clinical application, eventual sharing of data needs to be facilitated by using enterprise-common data elements and data models. This is true of applications developed in any discipline or setting. New application development should benefit from lessons learned to avoid the interface challenges faced by those presently converting the L&D system. Cooperative use of shared data dictionaries must take place before longitudinal EMRs can be effectively implemented.

Establish definition guidelines

The author recommends that guidelines must be established to assist the effective definition and maintenance of core concepts and minimize unnecessary variation in data elements. The guidelines presented in Table 13 are compatible with the formal ontology spoken of by Stead and others[39] who describe a method used by system developers to make it easier for the users to interact with the computer while preventing the meaning of concepts from drifting over time.

Focus on relevant information capture

There is a tendency for information systems designers to require clinicians to ask many questions of unknown informational value. These add to the time constraints of clinicians and cause them to view this data gathering as an unnecessary task that should not be part of the functional requirements of a data system for clinical care. Careful study is needed to determine the predictive value of the many history and physical data

Table 13. Concept Guidelines for an Integrated Longitudinal Patient EMR

Create a conceptual metadata structure which identifies:
1. Necessary clinical core data elements
2. Appropriate personnel to define the element values (domain of concepts for each data element)
3. Appropriate personnel/instruments to enter the element values (instances of concepts reflecting the conditions observed in the patient)
4. Optimal time to capture the element values (stage of the clinical process where the individual concepts are first observed or measured by a clinician)
5. Potential users of the data element by role
6. Potential uses of the element both at point of care and later for outcomes research
7. Effective utilization of the element with other elements to derive new elements and/or concepts

elements that are collected. Relevant data elements based on evidence will reduce the unnecessary data capture, while the use of common clinical data models and standard vocabularies will enable universal information sharing. Such collaboration will help diminish the challenges of management and governance regarding trade-offs between integration and functional requirements.

The definition of relevant data that should be captured at the point of care needs to be a collaborative effort by clinicians and medical informatics specialists. Clinicians should select data elements whenever possible from standard practice guidelines to most effectively function with EHRs within and across institutions. The domain coverage of any single terminology may be inadequate and additional terms may be needed. Institutions should look to the clinical measures of evidence based clinical guidelines and the narrative text documentation of physical examinations as sources of data for the development of new terms to represent clinical information. Care process models (CPM) are IHC-specific models of best care that are used to integrate evidence-based processes into the EHR. Disease management systems are being defined in virtually every

discipline with the goal of using CPMs to make delivery of care easier and strengthen the analysis of practice patterns. With these tools, clinicians can be prompted to document care based on content in a care process model.[98]

Comparing outcomes across locations, platforms, clinical domains, and longitudinal care continuums require the terms to be linked to a common reference vocabulary. Coordination from informaticists is needed to verify formal methodologies are used in selecting and defining data elements that are meaningful to the clinician and still represent concepts from a standard terminology referenced in the data dictionary. The collaboration process will ensure that information and knowledge can be derived in a systematic way and optimally used by clinicians, researchers and administration.

Provide application development guides

The creation of development guides for new clinical applications along with active education of developers will facilitate the adoption of common vocabularies and data models throughout an enterprise. Early coordinated efforts help ensure effective data modeling for data capture and data sharing with interdisciplinary teams, and at levels beyond the institution. Part of that development guide should be guidelines for using clinical concepts drawn from standard medical terminologies, i.e., SNOMED, UMLS that are mapped with terms in the HDD so that applications can be easily integrated. Integration and collaboration will go forward if user-friendly access to the core concepts and data element registers are facilitated.

Relatively few health care institutions have management and staff who understand the importance of data modeling and terminology. Information specialists who have the expertise to create and maintain a central data repository, health data dictionary, and can

oversee integration are even more rare. Even at institutions where an interface approach has been taken, applications continue to be developed without knowledge of the common terminologies and data models in place.

Utilities need to be continually developed and refined to help clinical application developers know what terms and concepts have been defined. Tools such as the ConcepT Scanner have been developed at IHC to facilitate new application development.[99] The tool reports how often and where concepts have been used in the CDR at the name value pair level. Application developers without knowledge of SQL or CDR structure as well as experienced developers will save time using it to track concept usage as compared to querying the CDR directly.

Recognize the changing nature of data sets

The variation presented in the nine different prenatal record formats, as well as the differences in data elements with the L&D and NICU EMR systems, show that data sets are an evolving tool as attested by Escobar's work with Kaiser Permanente (KP). The KP neonatal data set has undergone several more revisions, most recently with new elements from the Vermont –Oxford Neonatal network, and one maintained by the Canadian NICU network that focuses more on processes. Defining the needs of each institution and keeping current with regulatory bodies is an ongoing task. Kaiser Permanente's approach was to select from what others had proven as sound, modify the dataset with the needs of their institution, and move forward with the use of the best possible dataset knowing that it will evolve.

Conduct ongoing workflow analyses of perinatal processes

Providing systems that help clinicians do their work faster is the answer to getting them to fully participate. Effective design and implementation of the necessary systems requires a solid understanding which clinicians by specialty are not using the present systems, and why. Continued usability studies must focus on specific areas such as what data are the pediatricians not getting that should be captured by the obstetricians or neonatologists. Conduct periodic audits of data quality and system usefulness from the clinician's perspective using combinations of questionnaires, focus group discussions, and direct observation. Identify more completely what systems different clinicians use and for what purposes. Starting with surveys and moving toward direct observation facilitates resolution to system data variation problems with an increasingly narrower focus.

Implement multichannel notification systems

Some clinicians prefer personal communication for notification that a patient's test results are ready instead of logging into and accessing a clinical data system. Instead of the laboratory test personnel calling a clinician of completed results, a better option would be to create a system that automatically pages clinicians to notify them of high-priority lab results, and also populates the online clinical patient record with those laboratory results. In this way information can be shared with multiple appropriate parties without the time needed to reach the physician personally via traditional phone call.

Implement merits and incentives for clinicians to enter data

Before data can be shared it must be entered or captured by a system. Capturing more complete and higher quality patient data is key to increasing the clinicians' desire to use and improve the data systems. The motivation for clinicians to spend the extra time to enter information into the data systems has been a challenging issue. Additional remuneration to general practice physicians in the United Kingdom has had a positive impact on computer use.[100] Other incentives might include adjusting downward the number of patients physicians are expected to see each day, or receiving higher reimbursement rates for clinicians who comply. Effective user interfaces as well as effective motivation methods will facilitate improvement of both care processes and the capture of vital data.

Implement internal and external strategies for interoperability

Several of IHC's hospitals serve as tertiary care facilities and as such receive a large number of high-risk patient transfers. In many of these cases, the original clinic or hospital is out of state, and the prenatal record is not sent. Many "send-out" tests like rubella, GBS, and HBV are not easily or quickly done, so the tests have to be rerun. If there is not time, the baby is given the immunization or treatment to be safe. No method exists to track how often an unnecessary test is done. Without integrated records and systems, the process of identifying and alerting clinicians when a test for a patient is about to be duplicated, is a manual, time consuming task. Because most of the IHC inpatient admits are paid on a diagnosis related group (DRG) basis, some of the detail linked to the claim is not captured. Aggregate analysis of duplicate tests is not possible without a time-consuming search of laboratory data marts to identify whether specific

tests were rendered. Possible solutions involve early identification of tests with a more complete designation of reasons for the test, so that the test can be tied to a specific diagnosis from the ordering physician's notes or to a specialty such as maternity. Knowing how to identify the tests requires further analysis about what are the necessary comparisons (reasons for use) for laboratory test data. Most of the maternity admits are at IHC facilities; however, the availability of laboratory result data is inconsistent depending on where the tests were rendered as some tests are done at non-IHC facilities. Integration with laboratory test results from IHC sites is not nearly as challenging as getting results into a system from non-IHC facilities. Active education and facilitation with external laboratories to use HL7 and LOINC formats and codes is a starting place that will promote better integration to address the problem. Metrics to measure duplication of procedures or treatments should be included in performance standards for safety and efficiency.

Organizations are forming partnerships to develop a standard specification for the continuity of care record (CCR). The intent of the initiative is to improve continuity of care when patients are transferred, discharged or referred from one provider to another.[101] A specialized maternity-focused CCR, containing prenatal information could be defined with the most relevant and timely facts about the patient. If the patient had possession of the record, or the record could be quickly accessed from a secure data repository, then key patient information could be used to minimize unnecessary treatment and care could be administered more quickly and efficiently.

Secure data repository models are presently taking place in the form of Local Health Information Infrastructure (LHII).[53] The NHII is active in educating and

promoting these as a building block of a national program. IHC is in a unique position to contribute to the establishment and implementation of programs such as a CCR and LHII. IHC's expertise in the women and newborn care continuum would be a good place to start such an initiative. Obstetrics has been suggested before as natural place to facilitate electronic health record use and acceptance. A patient who goes to a different hospital than what was originally specified, would avoid the inconvenience and time to recount her medical history upon admission. It would help the facility by getting minimum information and avoid time and resources being used to rerun tests.

IHC could provide a leadership role in the industry joining with other national organizations already involved in the initiative. It is of prime interest to regional public health organizations because it allows for better health care for all patients and helps share information potentially with all care providers.

Develop a core concept data set for the perinatal continuum

A longstanding debate among clinicians centers on just what clinical information is relevant and should be shared across the perinatal care continuum. One of the by-products of this study has been the collection of standard references, which provide clinical perinatal information data sets. As was stated earlier, work done by Simini et al. and Hulsey provides additional models from which to learn. These various data sets can and should be used for consensus building not only among clinicians within an institution but across institutions to move to an industry standard for data sharing at a content level for the perinatal arena. Tables that utilize these and other external standard references could be developed to identify the most relevant data elements, and form a standard perinatal dataset. Years of experience regarding proven data sets with the accompanying

granularity and domain specifications defined could save many from reinventing the wheel for many data elements. The knowledge of experts on what content is needed from both internal and external sources, combined with key standards of Health Level 7 messaging and SNOMED common clinical vocabulary terms will move an organization closer to an integrated longitudinal patient record.

IHC is one of many institutions seeking to define the necessary data elements for the women and newborn care continuum. The methods of this study can serve as a prototype to select, compare and reach consensus on which elements should be used in the longitudinal patient record. The number of total data elements used in the three perinatal systems involved in this study was over 5,000. Defining a reduced set of necessary data elements will also reduce problems of ambiguity. The simplification by reduction of variables results in reduced maintenance costs. Time spent by clinicians retrieving data from other systems can be redirected to direct care of the patient or to necessary consultation with clinicians on complex issues.

Clayton et al. described IHC's strategy of using a common data dictionary and data model that allows them to define the content of the patient record so that it is independent of any single application.[32] The flexibility and extensibility of the approach enables IHC to interface with most applications. Time and cost limitations, however, exist in an interfaced approach. When a system or application has been developed in a way that the concepts and data model are not consistent throughout, it is counterproductive to try to interface all concepts to another concept-based system. Mapping interpretations that apply to inconsistent data models result in haphazard results, the time to map will be extensive, and the existence of data model exceptions will result

in limited mapping accuracy The L&D legacy system is such a scenario as was shown from findings in the Data Model Study. Data elements from the L&D system that have been prohibitively complex to map do not exist in a structure that allows an accurate conversion. Examples include medications that use the same code and cannot be differentiated between whether they were used for tocolysis, anesthesia, or induction. Instead of prolonging the work to convert these remaining elements, it is recommended that further effort to map L&D system concepts to the CDR be curtailed. Through a short analysis of the remaining data elements, the list should be prioritized. The top five crucial elements should be passed to the CDR to the extent possible. After that point all efforts should be focused on determining the more broad data needs of the longitudinal perinatal EHR.

CONCLUSIONS

Consistency of data capture in the three perinatal systems at IHC is worse than previously expected. Current perinatal data continuity at IHC is intermittent, unreliable, and unsuitable for use with the care process models necessary to improve interdisciplinary care. Without consistent patient data, clinical systems cannot fully measure what is the best process from an aggregate data perspective, nor can the clinician know what is best for a specific patient. Clinicians do not fully realize the gaps of data continuity. The need for clinicians to access multiple systems and verify data accuracy fosters clinician inefficiency and potential for medical error.

Disparate clinical data systems allow and perpetuate variation in data models, variation in data quality, and result in lower data accessibility. When data quality and accessibility are reduced, the following are adversely affected: computer decision support, integration, outcomes research, clinician decision support, operational management, clinician support of systems, and patient satisfaction.

Computer decision support is most accurate and helpful when based on high quality data from all possible sources. Interpretation and processing of complex data is limited when there is variation in data models between disparate systems. The result is lost meaning between data elements, computer decision support tools cannot be consistently triggered and expert systems tools cannot be used without costly interfaces.

Extensive mapping of dissimilar concepts and restructuring of conflicting data types between the L&D system have consumed substantial resources without solving the

fundamental problem; integrating multiple incompatible data models. Modification of structures in both the CDR and in the L&D system to meet somewhere in the middle puts at risk the compatibility and interoperability with other disciplines. Tying prenatal systems to the CDR is not as difficult, but passing key, relevant data from prenatal into L&D is hampered because of differing data types and vocabularies. Clinicians often refer to the printout of one system and if the data was not manually keyed into that system. As a result, clinical decisions are sometimes made without access to pertinent patient information.

Outcomes research suffers from missing data, inaccessible records, and incompatible data models that degrade data quality and hinder data sharing. Without the sharing of accurate information, conclusions drawn from aggregate data are deficient. Population studies using common data repositories are less reliable because they lack the information that is locked in the islands of disparate systems whose data is prohibitively expensive to convert or simply not available. Outcomes research efforts to improve care processes are compromised because of the need to settle for partial views of the total population.

Clinician decision support at the point of care is best achieved by ready access to information about the patient and all interdisciplinary care administered. When data such as patient history and clinical laboratory results are inconsistent and fragmented, it forces clinicians to make their best judgment based on a partial view of the whole patient picture. Even when using best practices, if they are based on a partial or possibly inaccurate view; then, decisions are questionable. With disparate information sources and systems, clinicians need to go to many sources to get a more complete view, which

means using an ever-evolving verbal network to find the missing patient information. Reliance on a partial view of the patient status becomes routine and accepted. Variation in information gathering, logically leads to variation in care practices. Because time is a finite resource in many medical decisions, clinicians frequently must go with the data they have and proceed to care for the next patient.

From an operational management perspective, the administration of unnecessary treatments increases costs and the potential for complications, yet has become a common procedure. Clinical staff engaged in the verbal polling network to track down information, distort the resource allocation picture. Human effort is constantly exerted to rediscover and maintain these verbal networks. Maintenance of incompatible data systems and user support demands, drain time and energy from development resources that could be better utilized if focused on one coordinated solution. Without system integration, profiling and audit systems that benchmark performance at a clinician level and facility level are incomplete.

Clinician support and satisfaction will increase as clinical documentation demands are simplified, redundant data entry is minimized, and the assurance that what is entered will be shared, used, and not lost in a technological silo of data. Trade-offs between integration needs and functional requirements will be diminished. Clinicians will respond to the improved data quality and accessibility with increased use of the information systems. Time will be transferred to providers to care for the patient instead of clerical data searching.

The increasing reliance on the patient or the patient's relatives as sources of medical care history, has become an unreliable bridge for the system information gaps.

Questioning the patient excessively for clinical information causes dissatisfaction, loss of confidence and raises questions of impropriety and patient safety. The patient or patient's spouse acting as advocate often becomes the care coordinator because the documentation of care is not adequately being shared across the care continuum.

Ultimately, providing the best quality of care at a reasonable cost is the goal. Using standard data models, standard vocabularies, and standard policies are the keys to assembling the best possible clinical data and information content. Having an integrated longitudinal view of the patient gives the clinician all the relevant information in one source to determine what the appropriate care process should be. Once gathered, the data can be easily shared from a common repository to discipline-specific patient views for enterprise sharing of information. Using a common repository as a longitudinal EHR, an environment is in place to have the patient be at the center of care; problems and goals be clearly identified and documented; and all clinicians work collaboratively as a team.

APPENDIX A

DEFINITION OF TERMS

Definition of Terms

Defining the terms used throughout this study is necessary to understand the nature of the challenges inherent when integrating disparate data systems.

Concept: A unit of thought constituted through abstraction on the basis of properties common to a set of objects.[102]

Data Element: Data elements represent concepts and are synonymous to attributes that describe a context.[103] Examples of context/data element relationships are: Risk Factors/Alcohol used, Laboratory Test/GBS Culture, Gravidity/Number of premature births. All data elements have one or more data values.

Data Element Type: Composite data element types are combinations of two or more simple data elements. The more granular the data element, the more specific or narrow is the definition. Derived data element types examples are calculated age, or time or date.

Data Element Value: Data element values are the allowable values for a data element. Examples of data element/data element value relationships are: Alcohol used/ (yes), GBS Culture/unknown), Number of premature births/2.

Data Model: Data modeling produces a formal description of the data of a given enterprise – a "conceptual schema." It represents concepts of a given domain (people, places, equipment, events, etc.), and how these entities are conceptually related to one another. There should be agreement among the participants in the enterprise about the facts and constraints in a domain. Current understanding of the process of health care is incomplete, and consensus lacking in certain areas, which makes agreement on data model formalization difficult.[104]

Dictionary: Structured collection of lexical units, with linguistic information about each of them.

Interoperability: the ability of two or more systems or components to exchange information and to use the information that has been exchanged.[105]

Object: Any part of the perceivable or conceivable world.

Term: Designation of a defined concept in a special language by a linguistic expression.

Terminology: Set of terms that are elaborated according to preestablished naming rules.

Vocabulary: Dictionary containing the terminology of a subject field

Term.

APPENDIX B

DATA ELEMENT LIST

The data elements evaluated in this study are as follows:

Gravidity
Gravida
Term
Preterm
Miscarriages/Abortions
Living
Multiple Births
EDD by LMP

Family History
Drug medication, herbal therapy or radiation exposure in 1st trimester
Family history of birth defect
Age > 35 at time of delivery

Obstetric History
Previous stillborn / neonatal death
Previous infant admitted to NICU
Previous Cesarean section, scar type

Personal Medical History
Diabetes mellitus
Hypertension
Drug / Latex Allergy

Risk Factors for Preterm Birth
Prior Preterm Birth (< 37 weeks)
Drug Use (including alcohol)
Smoking pk /day

Laboratory Results
Patient Blood Type
Patient Blood Rh factor
RhoGAM given date
Antibody screen, titer
Hepatitis B s AG
GBS Culture

APPENDIX C

STANDARD REFERENCE COMPARISON

Data Set Compared to Standard References

Data Element		IOM 2001 Rpt	AWHONN	ACOG
Gravidity				
	Gravida	4	4	1
	Term	4	4	1
	Premature	4	1	1
	Miscarriages/abortions	4	3	3
	Living	4	4	1
	Multiple births	4	1	1
	EDD by LMP	4	2	1
	EDC by US	4	2	1
Family history				
	Drug medication, herbal therapy or radiation exposure in 1st trimester	4	4	1
	Family history of birth defects	1	4	4
	Age > 35 at time of delivery	4	4	1
Obstetric history				
	Previous stillborn / neonatal death	4	1	4
	Previous infant admitted to NICU	2	4	4
	Previous cesarean section, scar type	4	4	2
Personal medical history				
	Diabetes mellitus	4	4	4
	Hypertension	4	3	1
	Kidney disease	4	1	1
	HIV, herpes	4	4	3
	Drug / latex allergy	4	2	1
Risk factors for preterm birth				
	Prior preterm birth (<37 weeks)	4	1	4
	Drug use (including alcohol)	1	3	3
	Smoking pk /day	1	3	3
Laboratory results				
	Patient blood type & Rh factor	1	1	1
	RhoGAM given date	1	1	1
	Antibody screen / titer	1	1	1
	Hepatitis B s AG	1	1	1
	GBS culture	1	1	1
	Rubella	1	1	1
	Chlamydia screen	4	1	1
	Serology	4	1	1

KEY

1= value level match

2= element not in this reference

3= more granular in Standard Reference

4= less granular in Standard Reference

APPENDIX D

CLINICIAN QUESTIONNAIRE

Questionnaire Cover Letter

Purpose

The purpose of this research is to determine the value of an infant-focused view of a longitudinal patient Electronic Medical Record (EMR). I will evaluate the prevalence of missing or contradicting data occurrences in several separate nonintegrated perinatal systems. The results will be used to improve patient safety and efficiency of clinician access to patient clinical information.

Importance

This research is important in that it provides justification and baseline measurements for a fully integrated EMR to provide a single view for clinicians that ties maternal information to fetal information with the goal of improving decision making and the quality of care.

Contact person for questions

Val Hicken, BA, a Medical Informatics graduate student at the University of Utah and Research Assistant at the Medical Informatics department at IHC is conducting this research. He can be contacted at 801-450-7202 or 801-442-5956. Email: lpvhicke@ihc.com, mail to 4646 West Lake Park Blvd. S4 East 103, Salt Lake City, Utah 84120. This research is sanctioned as part of IHC's clinical application development and operational process improvement initiatives. Paul Clayton, PhD, Chief Information Officer, IHC. Sidney Thornton, PhD, Clinical Systems Operation Manager, IHC.

IRB Statement

If you would like to speak to someone not associated with the study about your rights as a participant, or any other matter related to the study, contact Kendell Nelson, Assistant to the Hospital Administrator at (801) 408-1968.

Confidentiality

All surveys will remain anonymous and confidential. De-identified findings from the survey will be summarized by specialty groups, and reported to the IHC Women and Newborn Clinical Program, PCMS Design Team.

The return of this survey serves as your consent to participate.

Time it will take to complete survey

This survey should take approximately 6 minutes to complete.

If you feel a question is too sensitive and prefer not to answer, you may omit that question.

We sincerely appreciate your time and thoughtful participation in this research and will work to use the acquired information to improve the quality, safety and efficiency of IHC's clinical information systems.

Val Hicken
IHC Medical Informatics
Wk 801-442-5959
Cell 801-450-7202

The scope of this survey focuses on the current prenatal, labor & delivery, and NICU patient care documentation environments. Please answer the questions as if the computer systems are working as intended and not down because of a power outage or unexpected viruses.
An example of Contradicting Data would be when a record in one system indicated Antibody screen = positive, and in another system the record indicated negative.

	Missing Data	***Contradicting Data***
1	a) In the last month, did you have a problem with **not being able to find adequate *prenatal*** information about the patient? (Ex. labs, medical history, risk factors, parity, etc.) Circle: Yes or No	b) In the last month, did you have a problem involving **contradicting *prenatal*** information about the patient? (Ex. labs, medical history, risk factors, parity, etc.) Circle: Yes or No
2	a) If yes, for what specific prenatal information were you looking?	b) If yes, for what specific prenatal information were you looking?
3	a) What was the source of data used? (ex. Prenatal, Storkbytes, CW, RR, HELP, Other.)	b) What was the source of data used? (ex. Prenatal, Storkbytes, CW, RR, HELP, Other.)
4	a) About how many patients do you see in a typical month?	
5	a) In the last month, estimate how many patients had this particular problem of ***missing*** data?	b) In the last month, estimate how many patients had this particular problem of ***contradicting*** data?
6	a) What was the degree of danger to the patient if this problem had not been resolved? (1=not at all dangerous; 7=very dangerous).	b) What was the degree of danger to the patient if this problem had not been resolved? (1=not at all dangerous; 7=very dangerous).
7	a) Why was it dangerous, or not dangerous?	b) Why was it dangerous, or not dangerous?
8	a) Approximately how long did you spend trying to resolve this situation per patient? ____ days, or ____ hours, or ____ minutes	b) Approximately how long did you spend trying to resolve this situation per patient? ____ days, or ____ hours, or ____ minutes
9	a) What steps did you use to resolve it?	b) What steps did you use to resolve it?

Please complete and return this in the self-addressed stamped envelope no later than **Oct. 15.**

APPENDIX E

PAPER PRENATAL COLLECTION SHEET

	Record _____ Form Version _____ 1st Trim Y / N			
	Data element	**Possible values**	Std	Derived
Gravidity				
	Gravida	Numeric ND		
	Term	Numeric ND		
	Premature	Numeric ND		
	Miscarriages/abortions	Numeric ND		
	Living	Numeric ND		
	Multiple births	Numeric ND		
	EDD by LMP	___/_____/___ ND		
	EDC by US	___/_____/___ ND		
Family history				
1	**Drug medication in 1st trimester**	1=yes / 0=no		
1	**Herbal therapy in 1st trimester**	1=yes / 0=no		
1	**Radiation exposure in 1st trimester**	1=yes / 0=no		
2	**Family history of birth defects**	1=yes / 0=no		
4	**Age > 35 at time of delivery**	1=yes / 0=no		
Obstetric history				
16	**Previous stillborn / neonatal death**	1=yes / 0=no		
17	**Previous infant admitted to NICU**	1=yes / 0=no		
21	**Previous cesarean section**	1=yes / 0=no		
21	**C-sec scar type (Ver, Tra, Unkn, other)**			
Personal medical history				
33	**Diabetes mellitus**	1=yes / 0=no		
35	**Hypertension**	1=yes / 0=no		
41	**Kidney disease**	1=yes / 0=no		
47	**HIV**	1=yes / 0=no		
47	**Herpes**	1=yes / 0=no		
39	**Drug allergy**	1=yes / 0=no		
39	**Latex allergy**	1=yes / 0=no		
Risk factors for preterm birth				
55	**Prior preterm birth (<37 weeks)**	1=yes / 0=no		
59	**Drug use**	1=yes / 0=no		
59	**Alcohol use**	1=yes / 0=no		
60	**Smoking**	1=yes / 0=no		
60	**Smoking pk /day**	Numeric		
	comments			
Lab results				
	Patient blood type and Rh factor	_____ Pos Neg ND		
	RhoGAM given date mo/day/yr	___/_____/___ ND		
	Antibody screen	1=pos/ 0=neg ND		
	Antibody screen titer	Numeric ND		
	HBsAg	1=pos/ 0=neg ND		
	GBS culture	1=pos/ 0=neg (NR) ND		
	Rubella	1=immune (pos) / 0=not immune ND		
	Chlamydia screen	1=Pos / 0=Neg ND		
	Serology	1=pos/ 0=neg ND		

APPENDIX F

RULES FOR INTERPRETATION OF PRENATAL RECORD

All prenatal records were accessed at the Medical Records Department at the LDS Hospital. At least 8 different versions of prenatal records were observed in being used for patients in the sample population. Differences in data element content and format necessitated a set of rules so that the study could analyze the same concepts across perinatal clinical systems. Appendix F describes the rule set which was validated by clinical experts to provide consistent interpretation and extraction methods. Data from the prenatal records were compared with data from the L&D system and from the NICU EMR system.

Data elements needing abstraction rules

Prenatal record versions

Codes are entered on the Data Sheet to track frequency of use of the various Prenatal record formats.

"I 02" is the code that represents the most current version IHC HC 4005 modified on 12/02. The version on the IHC prenatal forms is found on the lower left corner. If the date is missing, the version can still be identified by several changes unique to that version. One of those is that item #47 in Personal Medical History is "HIV, Herpes."

"I 01" is the code that represents the version dated 05/01 and is the next most recent version of the IHC prenatal form.

"WGOC" is the code for forms from Western Gynecological and Obstetrical Clinic Inc. The form has the IHC form version code, "IHC HC 4006/6 99" at the bottom, indicating that these forms have the same content and format with IHC forms as of June of 1999 and are still being used now in November of 2003.

"ACOG" is the code for forms from the American College of Obstetricians and Gynecologists. The ACOG Antepartum Form consists of four pages labeled Form A through Form D.

"Holl" is the code used for Hollister Maternal/Newborn Records Inc. form #5872-100. Three pages of the Hollister forms were commonly used with the first page of the IHC prenatal record. Laboratory results, charting and patient visit information were entered on the Hollister forms.

"Gran" is the code used for Granger Medical Clinic records. Finding the value for RhoGAM Given was found in a comment section instead of in a separate data element, which was the standard method on most of the other record formats.

"LSWC" is the code for Life Spring Women's Clinic records. Blood type, RhoGAM given, antibody screen and HBV data elements could not be found on this form.

"ICHC" is the code for Ingham Community Health Center, Lansing Michigan. The ICHC form was quite similar to the ACOG version but listed RPR screening test instead of the more common VDRL, which has been treated as synonymous with serology.

First trimester. At the top of the data sheet, enter "Y" or "N" if the record date shows that it was filled out in the first trimester of pregnancy (weeks of gestation is 13 or less). The information was gathered to determine whether the medications were taken in the first trimester.

No prenatal care given. Several times the prenatal record was not in the physical Patient Record. Validation was achieved by checking in the L&D record under

Patient Information, to see if it contained the statement, "Prenatal Record not Available." When that statement was present; then, "No prenatal record" would be recorded on the Data Sheet. If there was a prenatal record, the L&D system recorded, "Prenatal Record Reviewed (IHC Form- Yes)."

Data value derived. In certain instances, physicians can derive values for some data elements from various places on the form. The data sheet provides two columns to record and evaluate where the data was recorded on the prenatal record:

1) **standard** and expected place on the form
2) value may be **derived** from other values on the prenatal record, given the user has a clinical background and understanding of the interrelationships of the other data elements.

When reviewing **each data element** on the patient records, a value of 1 will be entered in the appropriate column indicating where the value was found. It is hoped that the information is helpful in evaluating the actual use patterns and improving the ease of use with which clinicians both document and obtain information from the prenatal record.

Not documented. If a data element value cannot be found or derived, enter the letters "ND."

Gravidity. Enter the numeric value if documented. If the notes state that this is the first or prima gravida and there are no other numbers, documented for Term, Premature, Abortions, Living, or Multiple Births; then, those values can be derived to be zero. If a value is not recorded in the appropriate place, but can be derived, enter the derived number, and enter a "1" in the Derived column. If Multiple Births cannot be derived from the other parity values, enter "ND" for Not Documented.

EDC, EDD by LMP. These elements are found on the second page in the Gestational Age Determinant section. It will be shown right after LMP.

If LMP and the GA were documented, the EDC by LMP can be calculated. Record the LMP and GA, but do not calculate the date; just put the 1 in the Derived column. GA will be calculated and recorded at a later time. Either enter the date or if it cannot be derived, circle ND (Not Documented). If there is an EDC date on a non-IHC form, but it is not known if it was determined by US or by LMP because a record of US being done is not seen; then, assume that it was done by LMP.

EDD by US. This is found on the second page in the Gestational Age Determinant section. It will be shown right after LMP. Either enter the date or circle ND

Family history, obstetric history, personal medical history, risk factors

For these four sections, the box will be checked only if the condition/diagnosis is positive or abnormal. If the box is not checked, that means that the condition does not exist, or is normal. From the existing forms it was not possible to determine if these data element values are "unknown" or "not documented." Unless the system has a specific option to enter unknown or not documented, consider it as No or not present. If the user writes a circle with a slash through it in each section, that is interpreted as the section was reviewed and there were no concepts with abnormal conditions and all in that section should be marked No. If the user checks the box marked "Risk Assessment Reviewed, No Problems Identified"; then, mark 0 (no) for all the values in these four sections. Some data elements on the prenatal record are combinations of two or more concepts. Those concepts will have individual boxes on the data sheet for recording.

Family history and genetic screening

#1Drug, medication, herbal therapy or radiation exposure in 1st trimester. The Data Sheet separates the combination concept into the following three concepts:

- Drug/medication taken during 1st trimester
- Herbal Therapy taken during 1st trimester
- Radiation Exposure during 1st trimester.

Enter a "1" on the Data Sheet when the box #1 on the prenatal record is checked, and there are further clarifications in the comments section relating to at least one of the concepts, such as notes defining what the medications or herbal therapy or radiation exposure were. In the comment section on the data sheet, write down the name(s) of the medication to the side of the appropriate concept, such as: "1- Drug 1st trimester: Phenergan." Enter a zero when the box #1 is not checked and there are no other indications that any of the values exist. If one of the three concepts is further clarified, enter a "1" for that concept on the data sheet, and Enter a "1" in the Derived box. If a concept is not further clarified, enter a zero for that concept. If the box #1 is checked but there are no further clarifications to know which concept is being noted, enter the ***numeric value of .33 in each of the 3 boxes.*** Verify that the drugs were in taken in the first trimester by seeing that the estimated weeks of gestation is 13 or less. If the weeks of gestation are greater than 13, it is not conclusive that the drugs were taken in the first trimester, and .33 would be entered for this concept. This method will be used to show there is some information value, but not enough to be complete on its own.

#2 Family history of birth defects. Enter a "1" when the box is checked. Enter a zero if it is not checked.

#4 Age > 35 at time of delivery. Enter a "1" when the box is checked. If the box is not checked, but the form has the age of the mother or the date of birth, calculate the age at time of delivery. If the patient is 35 years old or older at age of delivery, and this information was derived, enter a "1" and also enter a "1" in the column titled Derived

from other concept values. Enter a zero if it is not checked, and the age cannot be derived.

Obstetric history

 # 16 Previous stillbirth/neonatal death. Enter a "1" when the box is checked. Enter a zero if it is not checked.

#17 Previous infant admitted to NICU. Enter a "1" when the box is checked. Enter a zero if it is not checked.

#21 Previous cesarean section, scar type. This is a two-part concept. Enter a "1" in the Previous C-section box, when the box is checked. Enter a zero if it is not checked. If this box is not checked but the prenatal record indicates on another section that the patient had a previous c-section, enter a "1" here and also enter a "1" in the column titled "Derived from other values." If there is a check in box # 21 but there is not documentation of the Scar type, enter "1" in the previous cesarean section box and enter the category of "UNKN" (unknown) in the scar type box. If the scar is specified, enter the scar type on the data sheet. Options from which to choose include: vertical, transverse, unknown, and other.

Personal medical history

#33 Diabetes mellitus. Enter a "1" when the box is checked. Enter a zero if it is not checked.

#35 Hypertension. Enter a "1" when the box is checked. Enter a zero if it is not checked.

#39 Drug / latex allergy. Enter a "1" in the Drug allergy box if the Drug Allergy box #39 is checked, and there further indications specific to Drug allergies. Look at the

general allergy section at the top of the form and in the comments section below to see if there are drug or latex allergies listed there. Enter a zero when both the Drug Allergy box #39 is not checked and the Allergy section at the top are blank, and there are no indications on the form of drug allergies. If there are specific notes to indicate there are latex allergies, enter a "1" in the latex allergy box. Enter a zero in the Latex Allergy box when both the Drug/Latex Allergy box #39 is not checked, when the Allergy section at the top are blank, and when there are no indications on the form of Latex allergies. Enter a zero in the Latex Allergy box when the form does not list Latex Allergy, and when there are no indications on the form of Latex allergies. If the box #39 is checked but there are no further clarifications to know which concept is being noted, enter the *numeric value of .50 in each of the 2 boxes.*

#41 Kidney disease. Enter a "1" when the box is checked. Enter a zero if it is not checked.

#47 HIV, herpes. Enter a "1" when the box #47 is checked and the comments indicate specifically that HIV history exists. Enter a "1" when the box #47 is checked and the comments indicate specifically that Herpes history exists. Enter a "1" when there is an indication of HIV even if the box #47 is not checked, also Enter a "1" in the Derived box.

Enter a "1" when there is an indication of Herpes even if the box #47 is not checked, also Enter a "1" in the Derived box. Enter a zero in the HIV box when there are no indications on the form of HIV. Enter a zero in the Herpes box when there are no indications on the form of Herpes. Enter a zero in the Herpes box if there is no specific indication of Herpes, and the form does not list Herpes. If the box #47 is checked, but

there are no further clarifications to know which concept is being noted, enter the ***numeric value of .50 in each of the 2 boxes.***

Risk factors

#55 Prior preterm birth (<35 weeks). Enter a "1" when the box is checked. Enter a zero if it is not checked.

#64 History of GBS. Enter a "1" when the box is checked. Enter a zero if it is not checked.

#59 Drug use. Enter a "1" when the box is checked and there are specific notes indicating recreational drug use has occurred. Enter a zero if it is not checked, and there are no other indications of recreational drug use. Enter a "1" when the Alcohol Use box is checked and there are specific notes indicating alcohol use has occurred. Enter a zero in the Alcohol Use box if there is no indication specifically of alcohol use. If the box #59 is checked but there are no further clarifications on Drug or Alcohol use to know which concept is being noted, enter the *numeric value of .50 in each of the 2 boxes.*

Laboratory data

For many of the laboratory data, there are times when the laboratory test was not done because it may not have been necessary or appropriate. The appropriateness of conducting the test is not being determined. The focus is on whether it was recorded, and if the recorded value matched the same data element in another system. If it was not run, the value "ND" for not documented, will be recorded.

Patient blood type and Rh factor. These concepts will be treated as one data element. Enter the letter(s) and the Rh factor. If there is no entry on the prenatal record, enter "ND" on the data sheet.

RhoGAM given date. Enter the date using two digits for the month, two for the day and two for the year. If there is no entry on the form, enter "ND" on the data sheet.

Antibody screen. This value must be either positive or negative. Enter "1" for positive and "0" for negative. If there is no entry, enter "ND" on the data sheet.

Antibody screen titre. Enter the numeric value of the titre. If there is no entry, or if titre was not on the prenatal form, enter "ND" on the data sheet.

HBsAg for HBV. Clinicians using the Hollister form have used NR (nonreactive) as a value, which is equivalent to negative. Enter "1" for positive, and "0" for negative or NR. If there is no entry, enter "ND" on the data sheet.

GBS culture. Clinicians using the Hollister form have used NR (nonreactive) as a value that is equivalent to negative. Enter "1" for positive, and "0" for negative or NR. If there is no entry, enter "ND" on the data sheet.

Rubella. The IHC Electronic form has options of positive and negative, while most forms, and the correct value is immune and nonimmune. Nonimmune = negative. Enter a "1" for immune or positive. Enter a zero for nonimmune or negative. Enter "ND" if no indication was made. If the word "transitional" is used enter "0" for nonimmune.

Chlamydia screen. Enter a "1" for positive. Enter a zero for negative. Enter ND if no indication was made.

Serology (VDRL). Clinicians using the Hollister form have used NR (nonreactive) as a value, which is equivalent to negative. Enter "1" for positive, and "0" for negative or NR. If there is no entry, enter "ND" on the data sheet.

REFERENCES

1. Institute of Medicine, Committee on Improving the Patient Record, Dick RS, Steen EB, editors. The computer-based patient record: an essential technology for health care. Washington, D.C.: National Academy Press; 1991.

2. Institute of Medicine, Committee on Improving the Patient Record, Dick RS, Steen EB, Detmer DE, editors. The computer-based patient record: an essential technology for health care. Rev. ed. Washington, D.C.: National Academy Press; 1997.

3. Institute of Medicine, Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington D.C.: National Academy Press; 2001.

4. Institute of Medicine, Committee on the National Quality Report on Health Care Delivery, Hurtado MP, Swift EK, Corrigan J, editors. Envisioning the national health care quality report. Washington, D.C.: National Academy Press; 2001.

5. Institute of Medicine, Committee on Data Standards for Patient Safety. Key capabilities of an electronic health record system: letter report. Washington, D.C.: National Academies Press; 2003.

6. Institute of Medicine, Committee on Data Standards for Patient Safety, Aspden P, Corrigan J, Wolcott J, Erickson SM, editors. Patient safety: achieving a new standard for care. Washington, D.C.: National Academies Press; 2003.

7. Allen M, Currie LM, Graham M, Bakken S, Patel VL, Cimino JJ. The classification of clinicians' information needs while using a clinical information system. Proc AMIA Symp 2003:26-30.

8. Rosenal TW, Forsythe DE, Musen MA, Seiver A. Support for information management in critical care: a new approach to identify needs. Proc Annu Symp Comput Appl Med Care 1995:2-6.

9. Institute of Medicine, Committee on Quality of Health Care in America, Kohn LT, Corrigan J, Donaldson MS, editors. To err is human: building a safer health system. Washington, D.C.: National Academy Press; 2000.

10. Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T. Reducing the frequency of errors in medicine using information technology. J Am Med Inform Assoc 2001;8:299-308.

11. Watt A, Scott W, Lee K, Robertson J, Koss R, Loeb J. Keen eye on core measures. Journal of AHIMA 2003;74(10):20-25.

12. Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T, et al. Systems analysis of adverse drug events. ADE prevention study group. Jama 1995;274(1):35-43.

13. Stetson PD, McKnight LK, Bakken S, Curran C, Kubose TT, Cimino JJ. Development of an ontology to model medical errors, information needs, and the clinical communication space. Proc AMIA Symp 2001:672-6.

14. Rothschild JM, Federico FA, Gandhi TK, Kaushal R, Williams DH, Bates DW. Analysis of medication-related malpractice claims: causes, preventability, and costs. Arch Intern Med 2002;162:2414-20.

15. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse drug events prevention study group. Jama 1997;277:307-11.

16. Kogan MD, Martin JA, Alexander GR, Kotelchuck M, Ventura SJ, Frigoletto FD. The changing pattern of prenatal care utilization in the United States, 1981-1995, using different prenatal care indices. Jama 1998;279:1623-8.

17. McDonald CJ, Schadow G, Suico J, Overhage JM. Data standards in health care. Ann Emerg Med 2001;38:303-11.

18. Huff SM, Rocha RA, McDonald CJ, De Moor GJ, Fiers T, Bidgood WD, Jr., et al. Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary. J Am Med Inform Assoc 1998;5:276-92.

19. Kleinke JD. Vaporware.com: the failed promise of the health care Internet. Health Aff (Millwood) 2000;19(6):57-71.

20. Bates DW, Gawande AA. Improving safety with information technology. N Engl J Med 2003;348:2526-34.

21. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10:523-30.

22. Bates DW, Ebell M, Gotlieb E, Zapp J, Mullins HC. A proposal for electronic medical records in U.S. primary care. J Am Med Inform Assoc 2003;10:1-10.

23. Kassirer JP. Patients, physicians, and the Internet. Health Aff (Millwood) 2000;19(6):115-23.

24. Laerum H, Karlsen TH, Faxvaag A. Effects of scanning and eliminating paper-based medical records on hospital physicians' clinical work practice. J Am Med Inform Assoc 2003;10:588-95.

25. Benson T. Why general practitioners use computers and hospital doctors do not--Part 2: scalability. Bmj 2002;325:1090-3.

26. Health Level 7 SIG. Electronic health record functional descriptors, draft document. Web documents. Ann Arbor: Health Level 7; 2003 1/7/2004.

27. Association of Women's Health Obstetric and Neonatal Nurses (AWHONN), Simpson KR, Creehan PA, editors. Perinatal nursing. 2nd ed. Philadelphia: Lippincott Williams and Wilkins; 2001.

28. American Academy of Pediatrics (AAP) ACoOaGA. Guidelines for perinatal care. 5th ed. Elk Grove Village, IL 60009-0927, Washington, DC: American Academy of Pediatrics (AAP), American College of Obstetricians and Gynecologists (ACOG); 2002.

29. Moon S. Improving care through integration. Mod Healthc 2002;32(1):32-3.

30. Twede M, Gardner R, Hebertson R. A PC-based system for intrapartum monitoring. Contemporary OB/GYN 1984(Special Issue -- Technology 1985):24:13-17.

31. Microsoft Corporation.Microsoft Product Information.2004. Available from: http://www.microsoft.com/. Accessed Jan 13, 2004.

32. Clayton PD, Narus SP, Huff SM, Pryor TA, Haug PJ, Larkin T, et al. Building a comprehensive clinical information system from components. The approach at Intermountain Health Care. Methods Inf Med 2003;42(1):1-7.

33. Elson RB, Connelly DP. Computerized patient records in primary care. Their role in mediating guideline-driven physician behavior change. Arch Fam Med 1995;4:698-705.

34. Shortell SM, Zimmerman JE, Rousseau DM, Gillies RR, Wagner DP, Draper EA, et al. The performance of intensive care units: does good management make a difference? Med Care 1994;32:508-25.

35. AHIMA e-HIM Task Force. A vision of the e-HIM future: a report from the AHIMA e-HIM task force. Chicago: American Health Information Management Association, AHIMA; 2003. Available from: http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_020477.pdf Accessed Jan 7, 2004.

36. Ball MJ, Farish-Hunt H. Standards at center stage safe, efficient care delivery depends on standardization. Healthcare Informatics 2003;20 No. 11(Nov 2003):52-54.

37. Maloni JA, Cheng CY, Liebl CP, Maier JS. Transforming prenatal care: reflections on the past and present with implications for the future. J Obstet Gynecol Neonatal Nurs 1996;25(1):17-23.

38. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Inform Assoc 1994;1:35-50.

39. Stead WW, Miller RA, Musen MA, Hersh WR. Integration and beyond: linking information from disparate sources and into workflow. J Am Med Inform Assoc 2000;7:135-45.

40. Heard S, Beale T, Freriks G, Mori AR, Pishev O. Templates and archetypes: how do we know what we are talking about?: HL7; 2003 2/12/2003. Available from: http://www.hl7.org/library/committees/ehr/documents/templates%20and%20archetypes1.pdf. Accessed Jan 7, 2004.

41. Beale T. Archetypes constraint-based domain models for future-proof information systems.: 2001 08/21/2001. Report No.: 2.2.1. Available from: http://www.deepthought.com.au/it/archetypes/archetypes.pdf. Accessed Feb 7, 2004.

42. Barretto SA, Warren J, Goodchild A, Bird L, Heard S, Stumptner M. Linking guidelines to electronic health record design for improved chronic disease management. Proc AMIA Symp 2003:66-70.

43. Huff SM, Rocha RA, Bray BE, Warner HR, Haug PJ. An event model of medical information representation. J Am Med Inform Assoc 1995;2:116-34.

44. Anderson LK, Stafford CJ. The "big bang" implementation: not for the faint of heart. Comput Nurs 2002;20(1):14-20; quiz 20-2.

45. Kuperman GJ, Gardner RM, Pryor TA. Help: a dynamic hospital information system. New York: Springer-Verlag; 1991.

46. McDonald CJ, Overhage JM, Tierney WM, Dexter PR, Martin DK, Suico JG, et al. The Regenstrief medical record system: a quarter century experience. Int J Med Inf 1999;54(3):225-53.

47. Bleich HL, Beckley RF, Horowitz GL, Jackson JD, Moody ES, Franklin C, et al. Clinical computing in a teaching hospital. N Engl J Med 1985;312:756-64.

48. Safran C, Perreault LE. Management of information in integrated delivery networks. In: Shortliffe EH, editor. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001. p. 359-384.

49. Nielsen PE, Thomson BA, Jackson RB, Kosman K, Kiley KC. Standard obstetric record charting system: evaluation of a new electronic medical record. Obstet Gynecol 2000;96:1003-8.

50. Escobar GJ, Fischer A, Kremers R, Usatin MS, Macedo AM, Gardner MN. Rapid retrieval of neonatal outcomes data: the Kaiser Permanente neonatal minimum data set. Qual Manag Health Care 1997;5(4):19-33.

51. Goossen WT, Jonker MJ, Heitmann KU, Jongeneel-de Haas IC, de Jong T, van der Slikke JW, et al. Electronic patient records: domain message information model perinatology. Int J Med Inf 2003;70:265-76.

52. NCVHS. Information for Health: A strategy for building the National Health Information Infrastructure. Report and recommendation from the NCVHS. Washington D.C.: National Committee on Vital and Health Statistics; 2001 11/15/2001. Available from: http://aspe.hhs.gov/sp/nhii/Documents/NHIIReport2001/default.htm. Accessed Feb 23, 2004.

53. Yasnoff WA.National Health Information Infrastructure: moving toward implementation.2004. Available from: http://aspe.hhs.gov/sp/nhii/. Accessed Feb 23, 2004.

54. Lorenzi N. Strategies for creating successful local health information infrastructure initiatives. Nashville: Vanderbuilt University; 2003 Dec 2003. Report No.: 03EASPE00772. Available from: Accessed Feb 5, 2004.

55. Centers for Medicare & Medicaid Services.The Health Insurance Portability and Accountability Act of 1996 (HIPAA). 2002. Available from: http://www.cms.hhs.gov/hipaa/. Accessed Jan 24, 2004.

56. Hulsey TC. Technical report, an in-house analysis from the perinatal information system and the Division of Pediatric Epidemiology. Charleston: Medical University of South Carolina; 2002 Oct. Report No.: 01PINS00.

57. Simini F. Perinatal information system (SIP): a clinical database in Latin America and the Caribbean. Lancet 1999;354(9172):75.

58. Hripcsak G, Clayton PD, Jenders RA, Cimino JJ, Johnson SB. Design of a clinical event monitor. Comput Biomed Res 1996;29:194-221.

59. Steiner DJ, Coyle JF, Rocha BH, Haug PJ, Huff SM. Medical data abstractionism: fitting an EMR to complex and radically evolving medical information systems. Medinfo 2004;accepted for 2004.

60. 3M HIS. Making sense of the data: using a medical data dictionary to integrate, share, and understand clinical data. White Paper. Salt Lake City, Utah: 3M Health Information Systems; 2002.

61. Niemeck M. Design and implementation of an HL7-interface to transmit data from a clinical data acquisition system to a lifetime data repository. [Masters Thesis]. Vienna: University of Vienna; 2001.

62. Thornton SN, Yu H, Gardner RM. Using point of service clinical documentation to reduce variability in charge capture. Proc AMIA Symp 2002:782-6.

63. Gibson RF, Haug PJ, Horn SD. Lessons from evaluating an automated patient severity index. J Am Med Inform Assoc 1996;3:349-57.

64. Utah Department of Health.About Utah State Immunization Information System (USIIS).2004. Available from: http://www.usiis.org/aboutUSIIS.shtml. Accessed Feb 7, 2004.

65. Nay J.Utah's WebKIDS software saves clinic time.2002. Available from: http://www.immregistries.org/snapshots/april02/utah_webkids.html. Accessed Feb 7, 2004.

66. Friedman CP, Wyatt J. Evaluation methods in medical informatics. New York: Springer; 1997.

67. Ammenwerth E, Iller C, Mansmann U. Can evaluation studies benefit from triangulation? A case study. Int J Med Inf 2003;70:237-48.

68. Kaplan B. Addressing organizational issues into the evaluation of medical systems. J Am Med Inform Assoc 1997;4:94-101.

69. Logan JR, Gorman PN, Middleton B. Measuring the quality of medical records: a method for comparing completeness and correctness of clinical encounter data. Proc AMIA Symp 2001:408-12.

70. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. For the Computer-based Patient Record Institute's work group on codes & structures. J Am Med Inform Assoc 1996;3:224-33.

71. McClay JC, Campbell J. Improved coding of the primary reason for visit to the emergency department using SNOMED. Proc AMIA Symp 2002:499-503.

72. Rocha RA, Huff SM. Coupling vocabularies and data structures: lessons from LOINC. In: Proc AMIA Annu Fall Symp; 1996; 1996. p. 90-4.

73. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. Designing clinical research: an epidemiologic approach. 2nd ed. Baltimore: Williams & Wilkins; 2001.

74. Weiner M, Gress T, Thiemann DR, Jenckes M, Reel SL, Mandell SF, et al. Contrasting views of physicians and nurses about an inpatient computer-based provider order-entry system. J Am Med Inform Assoc 1999;6:234-44.

75. Thomas SM, Overhage JM, Warvel J, McDonald CJ. A comparison of a printed patient summary document with its electronic equivalent: early results. Proc AMIA Symp 2001:701-5.

76. McKnight L, Stetson PD, Bakken S, Curran C, Cimino JJ. Perceived information needs and communication difficulties of inpatient physicians and nurses. Proc AMIA Symp 2001:453-7.

77. Wilson SR, Starr-Schneidkraut N, Cooper MD. Use of the critical incident technique to evaluate the impact of MEDLINE. Final Report. Palo Alto CA: American Institutes for Research; 1989 9/30/1989. Report No.: NLM-OPE-90-01.

78. Costakos DT, Love LA, Kirby RS. The computerized perinatal database: are the data reliable? Am J Perinatol 1998;15:453-9.

79. Goodwin L, Prather J, Schlitz K, Iannacchione MA, Hage M, Hammond WE, et al. Data mining issues for improved birth outcomes. Biomed Sci Instrum 1997;34:291-6.

80. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. N Engl J Med 2003;348:2635-45.

81. Wallace CJ, Stansfield D, Gibb Ellis KA, Clemmer TP. Implementation of an electronic logbook for intensive care units. Proc AMIA Symp 2002:840-4.

82. Wallace CJ, Stansfield D, Clayton PD. Evaluation of electronic documentation quality in acute care. Salt Lake City: Intermountain Health Care Report; 2003.

83. McGraw K, Harbison-Briggs K. Knowledge acquisition: principles and guidelines. Englewood Cliffs, NJ: Prentice Hall; 1989.

84. Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. J Am Med Inform Assoc 1997;4:484-500.

85. Friedman CP, Owens DK, Wyatt JC. Evaluation and technology assessment. In: Shortliffe EH, editor. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001. p. 313.

86. Bakken S, Campbell KE, Cimino JJ, Huff SM, Hammond WE. Toward vocabulary domain specifications for health level 7-coded data elements. J Am Med Inform Assoc 2000;7:333-42.

87. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37:394-403.

88. Intermountain Health Care Women & Newborns Clinical Program.GBS prophylaxis for threatened imminent preterm delivery.2003. Available from: http://ihcweb.co.ihc.com/enterprise/clinedu/womens/women_perinatal_gbs.pdf. Accessed Feb 7, 2004.

89. Sittig DF, Kuperman GJ, Fiskio J. Evaluating physician satisfaction regarding user interactions with an electronic medical record system. Proc AMIA Symp 1999:400-4.

90. Coiera E. When conversation is better than computation. J Am Med Inform Assoc 2000;7:277-86.

91. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system related errors. J Am Med Inform Assoc 2003:2004 Mar-Apr;11:104-12. Epub 2003 Nov 21.

92. Laerum H, Ellingsen G, Faxvaag A. Doctors' use of electronic medical records systems in hospitals: cross sectional survey. Bmj 2001;323:1344-8.

93. Torpy JM. Raising health care quality: process, measures, and system failure. Jama 2002;287:177-8.

94. McNutt RA, Abrams R, Arons DC. Patient safety efforts should focus on medical errors. Jama 2002;287:1997-2001.

95. Morris AH. Treatment algorithms and protocolized care. Curr Opin Crit Care 2003;9(3):236-40.

96. Vermont Oxford Network (VON).Neonatal network.2003. Available from: http://www.vtmednet.org/vtoxford/. Accessed Dec 2, 2003

97. Moyer R. Data warehousing: myths, pitfalls, and the secret weapon. J Ahima 2003;74(10):46-50.

98. McCormick KA. Integrating outcomes into computerized information systems. It's time to get guidelines off the shelf. Outcomes Manag Nurs Pract 2000;4(4):151-4.

99. Huff Murphy H. "Analyzing data dictionary concept usage in a clinical data repository to aid developers" [Masters]. Salt Lake City: University of Utah; 2004.

100. Benson T. Why general practitioners use computers and hospital doctors do not--Part 1: incentives. Bmj 2002;325:1086-9.

101. Tessier C, Waegemann CP. The continuity of care record. Organizations unite to create a portable, electronic record for use at transfer points in patients' care. Healthc Inform 2003;20(10):54-6.

102. Hammond WE, Cimino JJ. Standards in medical informatics. In: Shortliffe EH, editor. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001. p. 224.

103. Connolly TM, Begg CE. Database systems: a practical approach to design, implementation, and management. 3rd ed. Harlow, England ; New York: Addison-Wesley; 2002.

104. Johnson SB. Generic data modeling for clinical repositories. J Am Med Inform Assoc 1996;3:328-39.

105. Institute of Electrical and Electronics Engineers. IEEE standard computer dictionary: a compilation of IEEE standard computer glossaries. New York: IEEE; 1990.