Information Needs and the Characteristics of Population Data Sources:

An Immunization Information System Case Study

Rebecca Anne Hills

A dissertation
submitted in partial fulfillment
of the requirements of the degree of

Doctor of Philosophy

University of Washington

2011

Program Authorized to Offer Degree:
Medical Education and Biomedical Informatics

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Rebecca Anne Hills

and have found that it is completely and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

William Lober

Reading Committee:

William Lober

Debra Revere

Diane Martin

Diane Martin

Date: 10/07/2011

Signature

Date 12/14/2011

University of Washington

Abstract

Information Needs and the Characteristics of Population Data Sources: An Immunization Information System Case Study

Rebecca Anne Hills

Chair of the Supervisory committee: Associate Professor William Lober
School of Medicine and School of Nursing

Data and information are vital to the daily work of public health practitioners, the data they use come from a variety of sources. Examples of these data sources are: vital statistics databases, surveillance data, morbidity data, and Immunization Information Systems (IISs). These IISs are of particular interest because of their near ubiquity in the Unites States, their importance for public health practice, and their most basic function of providing cross-organizational access to immunization-related clinical data for both public and private health care providers. As the infrastructure to connect electronic health record (EHR) systems and public health systems expands, public health practitioners will have the opportunity to access an unprecedented volume of patient level clinical information. The flood of information and data will have the greatest public health impact if understood and organized within the framework of public health practitioners' data and information needs. This work uses qualitative methods to identify and understand the information needs of public health practitioners related to immunization work and the data and information source characteristics that are important in meeting those needs. The study also uses quantitative methods to describe two important data source characteristics in Washington's IIS: timeliness and data element completeness. Results point to three main types of information needs of public health practitioners: individual level, population level and context-specific information (vaccine-specific information in this case). These results further the understanding of information work in public health across local and state public health organizations. The results also provide solid evidence related to the effect of different methods of data transfer on data quality. In addition, synthesis of the qualitative and quantitative components of this work provides evidence to support a set of recommendations presented to state level stakeholders in Washington. This research will help inform the development of technical and non-technical infrastructure to support data sharing between healthcare providers, health information exchanges, and public health organizations.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Thank you to those who answered questions, read drafts, discussed ideas, critiqued presentations, reread drafts, offered support, and encouraged me through this leg of my journey.

Thanks to my advisory committee: Bill Lober, Debra Revere, Diane Martin, Neil Abernethy, Rita Altamore and Will Welton. I'd particularly like to thank Debra Revere, a dear friend and cherished colleague who spent hours upon hours helping me brainstorm, think through, edit, and rehash ideas related to this work.

I would also like to acknowledge the many other University of Washington faculty who helped me along the way, especially Anne Turner, George Demiris, Janet Baseman and Mark Oberle.

I'd like to thank my BHI graduate student colleagues: Blaine Reeder for tremendous help in collecting and analyzing data. I'd also like to thank Jamie Pina, Alicia Guidry, Casey Overby, Mike Galdzicki and CJ Inman.

Thanks to Meg, Alex and Karen Hills and to Kristofer Reed for your unwavering support.

The financial support from the National Library of Medicine expedited the completion of my graduate studies. Thank you, Valerie Florence.

Finally, the public health practitioners across Washington who shared their stories, opinions, thoughts and time with me were instrumental in this work. I'd particularly like to thank Sherry Riddick at CHILD Profile and Chas DeBolt at the Washington Department of Health for their generosity and unending support.

Dedication

To my parents, Meg and Alex Hills, who made all of this possible.

Chapter 1: Introduction

Background

As the “scientific core” of public health (1), epidemiologists rely on the quality and character of data as well as the quality of the information systems used to collect, store, and provide access to these data. The data used by public health professionals come from a variety of sources. Examples of these data sources are birth and death records from vital statistics databases, data from the US Census, morbidity data from national surveys such as the National Health Interview Survey, and local data collected during community health assessments (2). As the infrastructure to connect electronic health record (EHR) systems through Health Information Exchange (HIE) organizations expands, public health practitioners will have opportunities to access and utilize unprecedented amounts of patient level clinical information. This flood of data and information will have its greatest public health impact if organized and understood within the framework of public health practitioners' data and information needs. This dissertation focuses on public health information needs, information sources, and the perceived value of information in one important area of public health work: immunizations.

Immunizations and Immunization Information Systems

Since 1796, when Edward Jenner performed the first vaccination using cowpox inoculation to protect against the related disease, smallpox (3), not only has smallpox been eradicated, but many new vaccines have been developed, tested, and widely used for disease prevention. The incidence of most vaccine-preventable diseases has been reduced 95% since vaccines were first put into common use in the 1900s (4). Although infectious disease (ID) mortality has decreased significantly in the US (5), increases in ID cases (6)(7), ID hospitalizations (8), bioterrorism threats (9)(10), and concerns about vaccine safety (11)(12) have raised public awareness of vaccination recommendations and policies. Immunization Information Systems (IISs), the registries established across the US to store and record immunization data, are of particular value for researchers and epidemiologists interested in investigating these topics and others related to the current immunization landscape.

IISs, sometimes referred to as immunization registries, are “confidential, population-based, computerized information systems that attempt to collect vaccination data about all children within a geographic area” (13). Today IISs are active or under development in 49 states (14) and, although their functionality differs, most systems provide a variety of services beyond simply acting as repositories of immunization information, including:

- facilitating reminder and recall notifications for parents,
- forecasting recommended immunizations for children,
- providing data for immunization coverage assessments,
- managing vaccine inventory,
- maintaining birth-to-death vaccination histories,
- connecting with other health information systems,
- generating official immunization certificates,
- identifying "pockets of need" within communities,
- assisting policymakers in making informed vaccine-related decisions, and
- assisting providers and public health in fulfilling reporting requirements related to immunization and vaccine distribution (14-17).

As evidenced by the list above, IISs have multiple stakeholders: parents and families, health care providers, policymakers, public health practitioners, and pharmacists. Today, systems that track immunizations in the US are among the most mature information systems used regularly in the field of public health (18).

CHILD Profile, Washington State's Immunization Information System

Washington State's IIS and child health promotion system, the Children's Health Immunizations Linkages and Development Profile (CHILD Profile), was developed in 1993 with funding from the Robert Wood Johnson Foundation as a part of the All Kids Count program (19). Today, daily operations are carried out by Public Health-Seattle and King County (PHSKC), working under contract with the State, along with a private software vendor that specializes in IIS development and maintenance. The system's current statewide provider coverage is 93%, with 98% of public providers and 66% of private providers submitting data to CHILD Profile. Methods of connection vary; in 2009, 8% of providers used Health Level Seven (HL7) connections, 14% used batch upload of flat files from provider EHR or billing systems, and 78% entered data manually using the web interface. By contrast, that same year a review of the actual number of records submitted to CHILD Profile showed that 81% came in via batch file, 13% via HL7 message, and only 6% via manual data entry (20). Although HL7 connections are encouraged, no systematic study has compared data quality across these three methods of data transfer.

A formal evaluation of CHILD Profile's health promotion component is conducted every three years by the State Department of Health and the University of Washington (21). The evaluation includes a survey of parents to gather feedback on information needs, health behavior, and the usefulness of the promotional materials. The evaluation results inform recommended changes to the materials and future assessments.

Two of the major lessons learned from the All Kids Count program were to: "involve stakeholders from the beginning" and "define the requirements of the system to support the users' needs" (22). Although all public health practitioners have a stake in IIS development, initial efforts focused almost exclusively on the needs of health care providers. To date, no detailed assessment of an IIS has focused primarily on the information needs or requirements of public health practitioners. Although public and private providers are essential to IIS functioning, to maximize overall system usefulness, the information needs of non-provider stakeholders should also be considered during system design and evaluation activities.

Information Needs and Public Health

Since the first immunization tracking systems were implemented, the fields of information science and health informatics, along with the sub-domain of public health informatics, have evolved significantly. Informaticians now understand much more than they once did about the information needs of health care professionals, especially in specific care contexts or within groups of similar providers (23)(24)(25). Revere et al. (2007) described the information work of public health practitioners broadly as pertaining to three common types of information: synthesized and collated information, content sources, and data (26). Work published by Turner in 2008 reinforced the idea that information work varies depending on the role of the individual practitioner within public health (27). Although general studies of information needs within the field of public health (28)(29)(30) and within some specific populations of public health practitioners (27)(31) have been completed, to date there are no published studies of the information needs of public health practitioners working in the area of immunization.

The research presented here will help public health organizations make the best use of the influx of individual level clinical data brought about through emergent HIEs. This series of studies used various methods to better understand information needs and data source characteristics related to immunization. We used qualitative methods to learn about the information needs of public health practitioners pertaining to immunization data and to assess the importance and significance of population data source characteristics. We also used quantitative methods to describe data source characteristics related to data quality. Finally, we attempted to communicate and verify the results of our work with stakeholders in public health practice.

Research Questions

The goal of this research is to further knowledge of the information needs and practices of public health practitioners as well as to describe the important characteristics of population data sources. It is our intent that this information will help inform the development of technical and non-technical infrastructure to support data sharing between healthcare providers, HIEs, and public health organizations. As means to accomplish this goal, we aim to answer the following research questions:

- What are the information needs of public health practitioners with regard to immunization information?
- What are the characteristics of one public health information system (the CHILD Profile IIS) with respect to information needs related to data quality, specifically, timeliness and completeness?

Overview of the Research

Although IIS data are used consistently by some public health practitioners, anecdotal evidence has suggested that not all public health practitioners are satisfying their immunization related information needs efficiently. One reason for inefficiency in IIS data sources may be data quality, important attributes of which include coverage, timeliness, completeness, and accuracy. The importance to public health practitioners of these properties and other characteristics of immunization data is not well understood.

The overarching goal of this research is to help public health organizations make the best use of the increased availability of individual level clinical data made possible by the emergence of HIEs and similar organizations. IISs are one of the most common sources of clinical data made available through providers to public health, and as providers respond to incentives for Meaningful Use (MU) (32), we can expect an increased number of them to contribute electronic data to IISs. With that in mind, these systems offer a unique opportunity for inquiry. We used qualitative and quantitative methods to advance understanding of the information needs of public health practitioners pertaining to immunization data, and to study some of the important data source characteristics, such as data quality, of the IIS in Washington State.

Specific Aim 1

Use interviews and qualitative analysis to investigate and document the information needs of public health practitioners pertaining to immunization information.

We will use interviews and qualitative data analysis techniques such as thematic analysis to investigate and describe the information needs of public health practitioners pertaining to immunizations. Our analysis will also identify data source characteristics that are important to public health practitioners for completing their immunization related work.

Specific Aim 2

Use quantitative methods to describe the timeliness and data element completeness of data from Washington's IIS and compare these measures across different modes of data transfer.

The goal of this aim is to describe some of the data quality characteristics of Washington's IIS data using accepted measures. The CDC's Updated Guidelines for Evaluating Public Health Surveillance Systems (33) address timeliness and completeness measures. In addition, Saarlas and coauthors offer guidance for measuring data quality characteristics in IISs (34). This aim describes the timeliness and data element completeness characteristics of Washington State's IIS and compares these measures across different modes of data transfer.

Specific Aim 3

Develop and verify recommendations for improving the usefulness of existing data sources for public health practitioners and other stakeholders.

The first two aims of this dissertation involve the collection of empirical evidence to better understand information needs of public health practitioners and the data source characteristics that might affect the IIS's ability to meet those needs. In the third aim we synthesize what we have learned about information needs and data source characteristics in order to develop recommendations for improving the usefulness of the IIS as a data source. We present the final set of recommendations, the process we used for developing those recommendations, as well as the process used for collecting feedback from stakeholders to ensure feasibility and acceptability of the recommendations.

Significance

The inclusion of immunization data transfer to public health in the Meaningful Use definition and rules is evidence that facilitating the exchange of immunization data is perceived as a valuable use of health information technology (32). Understanding the characteristics of the data being made available to the field of public health and how these characteristics affect the usefulness of the data will, we hope, prove valuable as rules are further specified and other public health uses are written into legislation.

The contributions of the research described here will be: [1] the introduction of rigorously collected descriptions of public health practitioners' information needs pertaining to immunizations, [2] the identification of important characteristics of data sources, [3] the systematic characterization of two components of data quality in an IIS, and [4] recommendations for improving the usefulness of IISs data for public health practice.

The broader impacts of this research will be to: [1] foster relationships between the University of Washington and CHILD Profile staff and integrate the research and practice communities, and [2] inform decisions about the use of individual level clinical data to populate aggregate health data sources in the age of HIEs and Meaningful Use of EHR technology.

By describing the information needs of public health practitioners and identifying the data source characteristics that are of the greatest importance to the field of public health, we will encourage decisions leading to the best possible use of new data sources. Wisely taking advantage of the emerging trend toward information exchange has the potential both to inform public health decision-makers and to better protect the health of the public.

Description of Chapters

The current chapter is a general introduction to the dissertation; this chapter states the research questions and briefly describes each of the three specific aims. Chapter 2 is a description of the work related to Aim 1 where we present background information and describe the methods, results, and conclusions from our qualitative inquiry into the information needs and practices of public health practitioners. Chapter 3 describes Aim 2, a quantitative assessment of two data quality characteristics: timeliness and data element completeness. In Chapter 4, we describe the third aim which is a synthesis of the results from Aims 1 and 2. We describe the development of recommendations to be presented to stakeholders. In this final chapter we will report on all of the facets of our process, discuss the changes to our recommendations based on the feedback from stakeholders, and discusses possible implications of our work as well as areas for further study.

Chapter 2: Information needs and data source characteristics related to immunization: a qualitative study

Background

Immunizations are one of the best known and most effective programs of public health; immunization activities focusing on promotion and/or administration take place at nearly every public health organization at all levels of infrastructure. Immunization work has been supported by information systems known as Immunization Registries or Immunization Information Systems (IISs) for many years (35). Today, most public health leaders appreciate the importance of these population based registries (36); all but one of the 50 states in the US have or are developing IISs (14).

These information systems offer enormous potential to public health practice; however, existing systems are far from perfect. Studies examining the quality of data in IISs have shown that these data are often incomplete and inaccurate (37-39). Other documented issues include the timeliness of reporting (40; 41) and coverage, i.e., the percentage of children in the population with a record in the IIS (40; 42). These data quality issues are being addressed; however we still do not fully understand how data quality problems might impact usefulness for public health practitioners. In addition, it has become clear that there may be other barriers to fully realizing the benefits of IIS data for public health practitioners.

The current health information technology (HIT) climate in the United States provides a strong incentive to understand information work in health care, particularly in the realm of immunization. Recent national health policy changes encourage the Meaningful Use (MU) of HIT (32). In order to receive financial incentives for MU, providers and hospitals must meet a set of criteria; one optional criterion on that list requires organizations to test a connection between the Electronic Health Record (EHR) system and a local IIS using established protocols. This elevated level of attention to IIS, immunization information, and interoperability in general, suggests the need to better understand the primary and secondary uses of the data collected and exchanged using these enhanced systems.

Immunization work has many facets, some of which are information and data-intensive. For clinicians, information needs related to immunization exist in part due to the complex guidelines surrounding the immunization schedule. Existing IISs have been designed primarily to meet these clinical immunization information needs, leaving non-clinical information needs— for example, those of public health practitioners concerned with assessment—as a secondary consideration. Although the importance of immunization information for public health work is widely recognized (43-45) the work, tasks, information needs and conditions surrounding this work in public health practice have not, to date, been thoroughly documented.

We conducted a qualitative inquiry to further our understanding of immunization work in public health organizations. The purpose of this study is to describe the information needs of public health practitioners with regard to immunization data and information sources, and identify data source characteristics important to workers when they seek, access, and use this information. This chapter reports on our research findings and presents the emergent themes resulting from our qualitative analysis of interview data.

Methods

This study uses qualitative methods to gain a broad understanding of the information needs and practices of public health practitioners. We chose to use qualitative methods because we began the research without a focused set of questions about information needs, i.e., we did not have an initial list of information needs. Qualitative methods allow for naturally flowing conversation that can be followed in different directions based on the interviewee, and did not constrain our work to one particular instrument, set of questions, or group of respondents.

Setting

Washington has about 6.7 million residents, approximately half of whom reside in the Seattle-Tacoma-Bellevue metropolitan area. The rest of the state, though dotted with smaller urban centers, is predominantly rural. Public health in Washington is run at the state level by the Washington Department of Health, and locally at 35 county or multi-county health departments run by local government. The 2009 National Immunization Survey estimates that 71.2% of children aged 19-35 months in Washington are up-to-date on their modified vaccine series (≥4 doses DTP/DT/DTaP, ≥3 doses of poliovirus, ≥1 dose of any measles-containing vaccine, ≥3 doses of Hepatitis B, ≥1 dose of varicella vaccine, and ≥4 doses of PCV) which is very close to the national estimate of 72.7% coverage (46). Though estimated coverage is average, a recent CDC report found that Washington had the highest rate of non-medical exemptions in the country at 5.7% (47). The state legislature recently passed a law that aims to curb the high number of exemptions that are thought to be made out of convenience (48).

Interviews

Between June 2010 and February 2011 we conducted 12 individual and 3 group interviews with public health practitioners (n=20) working in Washington State. We identified subjects through the recommendations of colleagues and department and program directors within local health jurisdictions. Of those contacted, 100% agreed to participate or provided a referral to more appropriate individuals within their organizations. One or two interviewers (RH, BR, DR) conducted each in-person interview. Sessions lasted between 30 and 90 minutes and were recorded using a digital audio recorder.

An interview guide (see Appendix A) directed the semi-structured interviews. This guide was developed collaboratively, iteratively piloted with colleagues, and revised prior to use in the field. Questions focused on interviewees' daily work related to immunizations, information needs related to data sources, the use of data sources associated with immunization-related tasks, and the data source characteristics that are important to them in seeking and using information for immunization-related work. The University of Washington Institutional Review Board granted approval for this work with human subjects.

Coding and Analysis

Audio recordings of interviews were transcribed verbatim and imported into the NVivo8 qualitative data analysis software package (49). A team of two coders (RH, BR) and one qualitative analysis expert (DR) analyzed the interview data. The codebook development process was based on the approach presented by MacQueen et al.; codebook structure and content was guided by the research questions, and refined by iterative and collaborative evaluation (50). Appendix B contains the final version of the codebook.

Analysis was framed by an inductive approach, allowing the words of the interviewees to guide the coding, rather than using a pre-determined external framework to direct the analysis. We identified and analyzed patterns in the data using thematic analysis techniques. Thematic analysis does not require the imposition of a pre-existing theoretical framework (51)(52). Although our work was informed by the Leckie model of information seeking in professionals (53), we chose not to prescribe this model pre-analysis, and aimed to gain a broad view of the information work of our interviewees.

When the team reached consensus on the codebook, two coders (RH and BR) analyzed the first interview transcript together (Subject PH1), then independently coded a single interview

(Subject PH2). The individual coding sessions were compared for consistency, and inconsistent, ambiguous, or inappropriate codes were reconciled by team discussion. Enhanced definitions and coding rules were subsequently added to the codebook and each coder was able to make changes to their own coding that reflected the group consensus. We calculated inter-rater agreement both before and after code reconciliation: agreement remained 97% or above, and the Kappa coefficient of 0.57 revealed no lower than moderate agreement between the two coders (54). All transcripts were analyzed by the primary coder (RH) and checks of consistency were completed with both coders for 2 of the 15 transcripts. Table 1 details percent agreement and Kappa coefficients.

Table 2.1: Percent Agreement and Kappa Coefficients for Inter-rater Agreement

	Before Reconciliation		After Reconciliation	
Interview	Percent Agreement	Kappa Coefficient	Percent Agreement	Kappa Coefficient
PH2	98	.57	99	.77
PH11	97	.68	99	.95

We identified emergent themes related to information needs using the thematic and content analysis techniques as described by Krippendorff (55). Our team performed analysis tasks simultaneously with ongoing recruiting and interview activities until code saturation was reached, at which time recruitment stopped.

Results

In this section we will first describe the demographic characteristics of the population we interviewed, then describe our findings related to information needs and data source characteristics, and finally present the emergent themes revealed during the analysis.

Demographics

We interviewed a total of 20 individuals working either at a local health jurisdiction in Washington (N=18, 9 rural, 9 urban LHJs) (56) or for the Washington State Department of Health (N=2). Participants held a variety of positions within their organizations including: nurse, administrator, manager, epidemiologist, and director. The majority of interviewees were experienced in public health; only two indicated that they had worked in public health for fewer than five years. Level of training ranged from RN to doctoral degree. Demographic characteristics of participants are detailed in Table 2.

Table 2.2: Characteristics of Interviewees

Interviewee Characteristics		
Job Role	**State (N=2)**	**Local (N=18)**
Nurse/Nursing Supervisor	0	1
Immunization Program Employee	0	5
Immunization Program Director/Manager	0	5
Epidemiologist	2	3
Senior Health Administrator	0	2
Other-Director or Manager	0	2
Supervisory Role		
Supervises other employees	0	11
Does not supervise other employees	1	5
Unassigned	1	2
Time in Public Health		
New to public health (<5 years)	1	1
Experienced in public health (>=5 years)	1	17
Rural vs. Urban Setting		
Rural <100 persons per square mile	NA	9
Urban >=100 persons per square mile	NA	9
Highest Level of Training		
RN	0	4
BSN	0	5
MPH/MSPH	1	4
PhD/ScD/MD/DVM	1	3
BA/BS	0	1
Unassigned	0	1

Information Needs

We coded 16 types of information needs related to immunization work and grouped those codes into 3 categories: individual patient information, population level information (aggregate patient information), and vaccine information. Most of the information needs we coded are closely linked to the tasks they support.

Table 2.3: References to Information Needs of Public Health Practitioners Related to Immunization Information and Data Sources

Information needs	Need category*	Individuals with this need
Information and data for answering population questions	P	13
Information about a patient	I	12
Information for performing regular reporting	P	10
Vaccine inventory and VFC ordering information	V	9
Information from and about schools	I, P	8
Vaccine schedule information	V	7
Information to support mass vaccination efforts	I, P, V	6
Information to support AFIX/provider assessment	I, P, V	6
Customizable reports	P	6
Information to support patient reminder and recall	I	5
Child care center immunization records	I, P	4
Information to support provider training	P	4
Consent Information	I	3
Billing information	I	3
Travel immunization information	V	1
Laboratory data and information	I, P	1

**I=individual patient information needs, P=population level information needs, V=vaccine information needs*

Individual patient information

Interviewees in positions involving direct contact with patients often reported the need for individual patient information while administering vaccine during regular office visits or mass immunization events; during phone calls with a provider, parent, school or employer; or during outbreak investigations. Other activities requiring individual patient information included: provider assessments, i.e., quality assurance activities such as the Assessment Feedback Incentives Exchange (AFIX) program visits, billing duties, tracking consent information, looking up laboratory data, and recall activities (using patient data to generate lists of individuals due for immunizations). Ability to access needed individual patient information varied, and barriers to fulfilling these information needs existed in the form of accessibility issues, data quality problems, and the lack of linkages between systems.

Population level information or aggregate patient information

Most participants also reported that information on the population level or aggregate patient information was a significant immunization-related need. Activities associated with this information need included: routine reporting duties such as quality assurance/provider assessment (AFIX) visits, assessing school exemption rates, reporting for the state and the Centers for Disease Control and Prevention (CDC), reporting on vaccines administered and coverage rates, reports related to inventory tracking, and assessment of vaccine coverage in the community with both geographic and demographic stratification. Participants reported the need for population level descriptions of immunization rates in the context of routine reporting, local tracking activities, and special activities such as grant writing. Though most of the aggregate patient data and information needs were for local and state level data, some participants described needing and using national level data, primarily from the National Immunization Survey. In some cases these information needs were being met, but in others the information available was not sufficient to meet the need for a variety of reasons including: data quality concerns, data format issues, user interface limitations, and time and resource constraints.

Vaccine Information

The final category, vaccine information, is an information need independent of patients, clients, or the population but very specific to the context of immunization programs. This vaccine information need was apparent when practitioners were working on activities related to scheduling and logistical issues regarding vaccine administration, transport and storage. Official CDC or other manufacturer vaccine information was referenced by approximately half of the interviewees. Vaccine information includes contraindications, regular schedule and catch-up schedule for a vaccine as well as information about storage. This vaccine information is needed for both routine childhood immunizations as well as adult and travel vaccines. Public health practitioners reported needing up-to-date information regarding childhood immunization schedules, catch-up schedules, and contraindications in order to answer questions from providers, parents or the public. Travel immunization information refers to recommendations for immunizations for travel to specific countries. Many participants also reported the need for information related to inventory tracking. Discussing the tracking of a vaccine order, PH10 reported: "it's in the system somewhere but it takes a number of phone calls right now to find that out and usually it hasn't been shipped yet because there is some sort of delay and so they can tell me and then in turn I call the provider." Interviewees identified barriers related to inventory and expressed concern about the timeliness of the vaccine information contained in the IIS that is used for vaccine forecasting, that is the immunization schedule guidelines used for making recommendations about which vaccines a child should receive.

Interviewees reported using more than 30 information and data sources and information systems to meet their needs. As participants discussed their use of the systems and data sources, they also described the barriers they encountered and the attributes of systems that facilitated their use.

Data and Information Source Characteristics

Participants reported using several types of information systems in their regular immunization-related work, both locally administered systems and external systems whose use was mandated by the state or federal government. The state IIS and local Electronic Medical Record (EMR) systems were mentioned most frequently.

We coded 9 data and information source characteristics that were important to public health practitioners: data quality, bias, data format, data linkages, metadata availability, granularity, prior success, reputation and usefulness. The coded characteristics can be sorted into three categories: primarily technical characteristics, primarily administrative characteristics and hybrid characteristics that can be viewed as both technical and administrative. Data source characteristics and corresponding categories are represented in Figure 1.

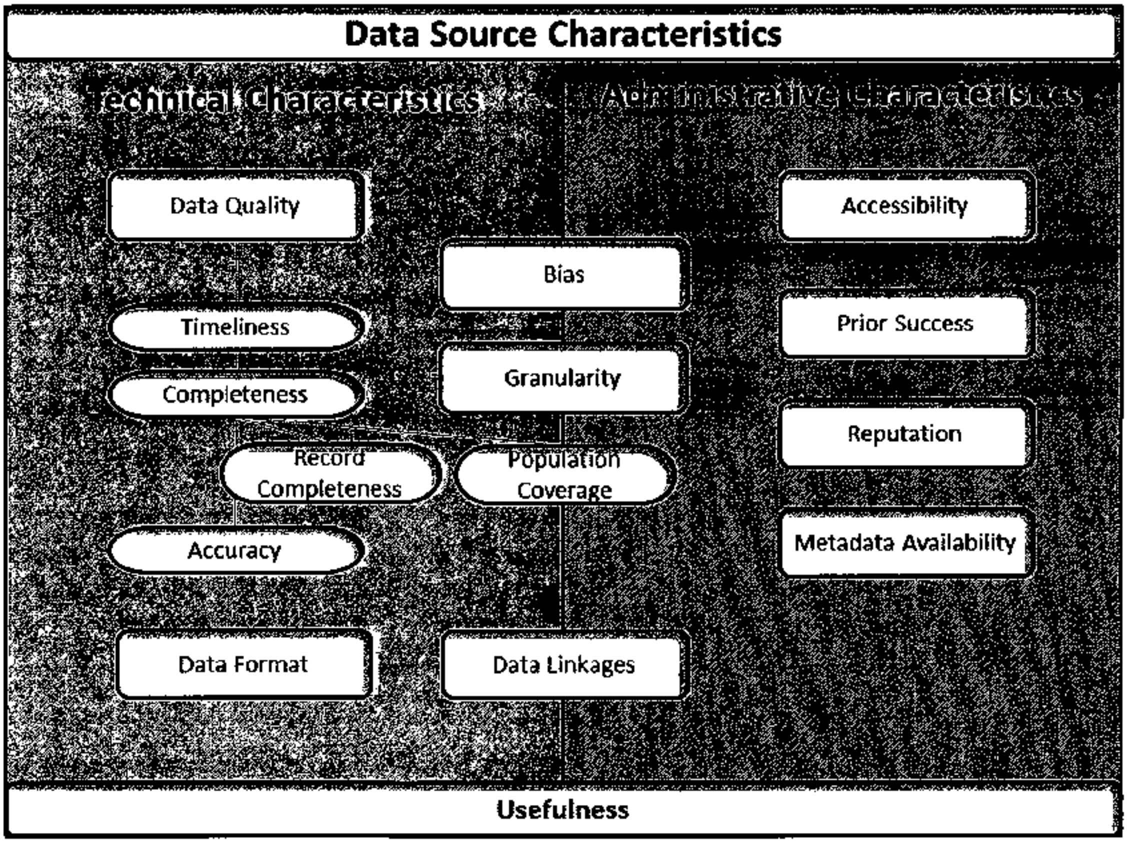


Figure 2.1: Data Source Characteristics

Of the technical characteristics, data quality was most important to the interviewees. Three essential and commonly recognized components of data quality were discussed during the interviews: timeliness, completeness and accuracy. Of the three, accuracy was discussed more often than any other characteristic and was most often mentioned in the context of the immunization information system or the local EHR. Accuracy was important to interviewees with both individual and aggregate information needs. Two types of completeness emerged from the analysis: population coverage, or completeness on the population level, and individual record completeness, the number of completed fields in an individual patient record. While both were important, population coverage was discussed more often. Participants also reported that timeliness was an important data quality attribute, particularly in the context of individual level needs, such as looking up whether an individual had received 1 or 2 doses of H1N1 vaccine.

Data format was the next technical characteristic that emerged during the analysis. Data format can refer to both the file type of a dataset-for example, a SAS dataset or comma-delimited file and the variable definitions, for example, age groups defined in 10-year intervals or age, represented as an integer or date of birth. Format was important to participants as it could affect the immediate usability of data or information for a specific purpose.

Administrative characteristics were also identified as important in the interviews. Accessibility, metadata availability, prior success, and reputation of the data source were mentioned most often. Accessibility was important to almost all of the practitioners we interviewed. Some described accessibility as a barrier (PH11): "we at the local level only have the capacity to look at individual practices." Limited or incomplete access to datasets or information systems caused problems for practitioners. Some attempted to address these issues by developing workarounds such as sharing passwords and using multiple logins. Metadata availability was also very important, primarily to individuals with population-level dataset needs. Several participants pointed to data dictionaries and other "data about the data" as providing context and other critical information for data analysis. One public health practitioner said that some datasets she has used include helpful documentation:

> *Documentation, data dictionaries, coding dictionaries, and then the State Department of Health always with their vital statistics releases...they also have a 'technical notes' so if questions have changed or there are discontinuities or if there are known issues ...so those are all useful to know as an analyst, as the caveats of the data. – PH2*

Generally, public health practitioners explained that in their data-centric work environment they value having information about why and how data are collected as it aids in analysis and helps them understand potential limitations and biases of the data. One interviewee (PH10a) reported that "...you've got to have people that kind of know the history... to really interpret the data."

Data source characteristics that share technical and administrative influences included bias, granularity, data linkages and usefulness. We categorized bias as both a technical and administrative characteristic because the causes of bias can be administrative or technical. For example, biased data may occur because of differences between provider organizations with regard to data sharing policies (administrative), or communication and connectivity limitations may have an effect on data or data quality (technical). Granularity, or the level of detail included in the data, was mentioned most often in the context of geographic specificity. The most common example of this was the lack of acceptable geographic specificity in the National Immunization Survey (NIS) data for purposes of assessing the immunization status of local populations; the finest detail available at this time in Washington is a division of the state into two geographic regions: Western Washington and Eastern Washington. Data linkages refer to one or more "keys" in each record of a dataset that allow users of the data to link records to other data sources. Three participants identified the linkage between data sources as being important in their work. One example is linkages between IIS records and communicable disease reporting systems, which increase the potential usefulness of both systems and help public health practitioners do their work more efficiently.

Finally, most participants discussed the usefulness of data and information sources. Usefulness is a cross-cutting characteristic of data sources, and it may be dependent on other characteristics of the data source such as data quality, granularity and data linkages.

Emergent Themes

Four primary themes related to the immunization work of public health practitioners emerged from our analysis:

1. *Public health practitioners value local data and information*

Interviewees emphasized the importance of the local data for public health workers and described problems they had experienced with its relative unavailability. Practitioners indicated that large national and state information sources are generally unsuitable for local use because of their lack of locality and granularity. The scale and sampling techniques offered by large data collection operations facilitated their usefulness for measuring population characteristics in some cases; however, as one participant stated in reference to the National Immunization Survey, "...the fact that we can't even get [our county's] number anymore from the last couple of years really minimizes its usefulness for us here..." (PH2). In contrast, locally collected data from community health surveys, local patient information systems or other local sources were often described favorably, though the resources needed to gather those data often restricted their availability.

2. *Public health practitioners have two levels of aggregate information needs: reporting needs and data-analytic needs*

Most of the practitioners we interviewed expressed the need for information that comes from large data sets, but the needs can be divided into two distinct categories: reporting needs and data-analytic needs. We defined reporting needs as ad hoc or regularly run data summaries or reports. One example of a report is the simple calculation of immunization coverage rates stratified by age. One practitioner described this need as being unmet because of incomplete population coverage in the IIS (PH7): "our only source of data on immunization rates is the information in CHILD Profile. And of course it's not complete, so we do not have any data that is complete for immunization rates. And it would be wonderful to have but we don't have it." Data-analytic needs were often expressed by epidemiologists and others with sophisticated analytic skills and duties; these practitioners needed datasets to answer complex questions and perform exploratory analyses. Practitioners expressing data-analytic needs also described the availability of metadata and the cleanliness of the data as essential components of a useable data source. One epidemiologist expressed frustration over time put into formatting and cleaning a frequently used state dataset: "Everything is an absolute disaster; I've had to clean it every year." (PH4)

3. *Aggregate population views, e.g., reporting functions, are not well supported in the state IIS*

As a major source for patient immunization data for public health practitioners, the local IIS is the logical place to turn when their work requires reports for state programs, mass immunization events, and local ad hoc information reporting needs. Though some reporting functions are available, practitioners at local health jurisdictions described compiling their own data both from the IIS and from other sources, often using Excel or paper, to meet many of their reporting requirements. Interviewees expressed the need for built-in reports to cover a broad population, such as all providers or patients in the health jurisdiction, as well as customizable reports where parameters can be set dynamically.

4. *Perceived usefulness of the IIS resource differed by the primary information need experienced*

Nearly all of the interviewees described the IIS as an essential tool for their work. Although the system was almost universally praised, interviewees also discussed its limitations. Those limitations differed for practitioners with aggregate-level information needs as compared with patient-level information needs. Practitioners using the information system to access individual patient records found the tool essential for administering vaccine (PH9):

> *...we don't like to poke them. I mean we poke them they cry, we don't like that. Flipside is we really don't want them to get sick and die of disease, so if we don't have a record we're going to poke them. But I can't count the number of times that we have saved shots for children by using CHILD Profile here. Over and over and over.*

Practitioners interested primarily in population views had specific concerns about population coverage, biases, data linkages, and duplication errors at the patient and immunization levels. Those interested primarily in accessing individual patient immunization information expressed concerns and mistrust regarding two methods of data transfer: HL7 connections (HL7 is an increasingly popular messaging standard often used to transfer immunization data between EHRs and an IIS) and batch file uploads. In addition, they expressed concerns about the timeliness of data, completeness of individual records, data accuracy issues resulting from use of billing data, accuracy of influenza vaccination data, and timeliness of the immunization forecasting information. One practitioner pointed out that timeliness is not always important, depending on the primary use of the data (PH7): "...frankly, I don't know how big of a problem it is if the data is a month too old. I mean to me that's pretty good data. I mean I think the concern is if you use it for ordering it may be problematic." Other practitioners, with an interest in vaccine-related information had concerns about the efficiency of inventory tracking systems and the timeliness of updates to vaccine recommendations in the IIS.

To summarize, the information needs of public health practitioners working in the area of immunizations are primarily centered on patient or record level information, and aggregate or population level data and information. In addressing their information needs, practitioners consider technical and administrative characteristics of an information source. Analysis revealed that for immunization work, data quality, accessibility, and the availability of metadata are important in assessing suitability of a data source for a practitioner's work.

Discussion

Purposeful development and strategic use of information systems in clinical and public health practice offers the potential to have a major effect on the quality and efficiency of public health work. In order to design, develop, and evaluate these systems, it is vital that we understand the information needs and the information work done in the public health setting.

Information behavior in the realm of clinical care and health services has been the focus of numerous inquiries. Many of these studies concentrate on the information seeking behavior and information needs of clinicians (57)(53)(58) often in the context of a particular tool such as the Infobutton manager (23), decision support systems (59), or a specific clinical setting such as general practice (60) or emergency departments (61).

Though clinical activities are a component of public health work, the diversity of professionals working within public health organizations and the unique work done by these organizations suggests the need to understand public health work in context. Studies from outside of public health often lack the context that is essential to explain the information behavior and information needs of these practitioners. Relatively few studies of the information needs of public health workers have been completed, but published work has found that this group's information needs are as diverse as their job roles (27)(26)(62).

Though other studies of public health practitioners examined information needs more generally, (27)(63)(64)(62)(65) this study focused on information needs related to immunization work. We found that public health practitioners working in a variety of roles within the area of immunization have a clearly focused set of information needs. The information needs we uncovered fell into one of three categories: patient level information, population level information, or vaccine information. The overwhelming need was for *local* patient and population level information.

Although we did not set out specifically to assess met and unmet information needs, our analysis provided insight into how well needs are currently being met. Overall, patient level information needs were being met satisfactorily, often by local patient data systems such as the local EMR, patient-tracking system or the state's IIS. In contrast, population level or aggregate data needs were not being satisfactorily met. Local data and information were often difficult if not impossible to obtain because of access issues. When local data were obtained in the process of completing population-reporting or related tasks, those data were often plagued with data quality issues. National data such as the National Immunization Survey, and United States Census were sometimes used successfully, but often, these data were not sufficient to meet the interviewee's information needs because of a lack of geographic specificity. When needs were not met directly, practitioners worked with the information they were able to obtain or in some cases developed work-arounds, such as obtaining multiple user IDs for a single system, in order to access all of the information necessary to complete a task.

Our analysis attempted to capture all data characteristics for immunization-related information work. Our analysis identified data quality, accessibility and format as the characteristics most valued by practitioners. Poor quality of patient and population data was an important concern for many of the practitioners; however in most cases the quality of data and information in available systems, though not ideal, was sufficient to at least partially meet general information needs. Previous studies of data quality in IIS have focused on timeliness, population level completeness, and accuracy (41)(38)(66)(67); our analysis identified these three aspects of data quality as well as data element completeness.

Accessibility was an important issue for many of our interviewees and was raised in the context of accessing patient, provider, and population level immunization information. Because immunization data are a part of an individual's health care record and contains personal health information, the sharing of these data is restricted by law, and policies for determining who can access information are often conservative. Public health practice is in the unique position of being both a service provider in need of individual-level information, and a government health agency in need of individual and population-level information. Reconciling these two very different needs remains a challenge for developers, implementers and public health managers, as practitioners continue to struggle with information access in their daily work.

Data format is sometimes considered a trivial and easily adjusted attribute of a data or information source; however differences in formatting can significantly impact the usefulness of information for public health practitioners. Analysis identified variable coding, file type, and reporting formats as important attributes of information or data sources for immunization-related work. Format limitations can sometimes be overcome, however today in the United States public health practitioners have limited resources and may lack the tools, time, or skills necessary to transform the data into a usable format.

The primary local system and opportunities for improvement

The state IIS is an important data and information source for local and state public health practitioners. Because of its widespread availability and current coverage, and its potential to cover nearly 100% of the state's population, we will offer several suggestions for the IIS, informed by this study.

Like many state-level immunization registry systems, Washington's IIS was originally developed to be a repository of immunization information for use by health care providers. As one interviewee stated, it was "…for the use by providers to check their patients' immunization histories and it was not to be used for assessment, for evaluation, for research…" (PH14). The current system has evolved, but the expanded group of stakeholders has evolved further, representing much broader interest in the data, for purposes other than direct patient care.

Practitioners who used the IIS had specific ideas about enriching the information resource and making it more useable. Population coverage was a significant concern. Suggestions for improvements include:

- Include more child care centers as IIS users
- Require use of the system for state-controlled inventory management
- Allow providers to view aggregate immunization rates and the immunization rates of regional provider peers to leverage inherent competitiveness
- Improve inventory management capabilities
- Facilitate data and information sharing between states
- Allow and facilitate linkages between IIS and other public health systems (school systems, reportable condition systems, lab systems)
- Improve public health access to different levels of information

As the HIE movement gains traction nationally, the need will increase to build an infrastructure that supports the aggregation of patient-level information and provides providers and public health practitioners access to those data. This study confirms that a principal challenge for these systems will be the diversity of the stakeholder groups, whose needs and context-specific constraints must be considered simultaneously in design. Designers of public health information systems will be well served by considering the information needs and data source characteristics that are important to public health practitioners. While this study describes the information needs and information source characteristics important to public health practitioners in one specific context, it is a roadmap for conducting similarly detailed inquiries of other stakeholder groups to inform system design projects.

Participatory design methods (68) include representatives from all stakeholder groups in the design, testing and evaluation of systems. Incorporating the information needs, data source characteristics and themes from this study into participatory and scenario-based design activities will help ensure that all public health stakeholders are satisfied with the final system design and that their critical information needs are being met.

While the findings from this study can be used to inform the design of systems to meet immunization-related information needs of public health practitioners, they may also be applied more broadly, as there are parallels between immunization work and other types of public health work that make use of individual and population level health information. One example is laboratory reports of reportable conditions, which are used both at the individual patient level for case investigation as well as at an aggregate level for surveillance activities. These similarities suggest that design and evaluation efforts in other areas of public health practice might leverage the work done in the area of immunization as new systems and information exchange capabilities become available.

Limitations

This study has several limitations. Participants were drawn from a convenience sample and the individuals we interviewed may be systematically different from individuals in similar roles who were not interviewed. This could introduce selection bias, the impact of which is unknown. In addition, although our sample covered a large geographic area and included organizations of varying sizes, caution should be exercised when attempting to generalize the results to practitioners in other jurisdictions and states because the practice of public health may vary among organizations.

Qualitative research has also been criticized for the subjective nature of the analysis. Indeed, coding interview data includes a highly subjective element; however attempts to mitigate this include inter-rater reliability measures (69) for several coded transcripts. These reliability measures (Kappa scores and agreement) suggest a sustained high level of agreement between coders.

Other studies have used participant observation to assess information needs and provide a complete description of work processes in settings outside of public health. The timeframe of this study was a limitation, but future work may benefit from an observational phase. Future work would also benefit from the use of mixed methods to gather and analyze both qualitative and quantitative measures of information needs. This would allow triangulation and validation of measures of information needs and perceptions related to different information source characteristics.

Conclusion

Information systems to support public health practice often present a unique set of challenges for system designers who are confronted with a diverse group of users and stakeholders with complex relationships. Stakeholders include clinicians, administrators, funders and payers, who perform a variety of roles within public health practice. This diversity requires a broad understanding of information work, information needs and the importance of information source characteristics across organizations and roles. With this study, we have described the information work of public health practitioners with respect to immunizations, and built the evidence base for informing design, development and evaluation of information systems to support public health, clinical practice, as well as the exchange of information among organizations.

We found several important categories of information needs among public health practitioners including: individual patient-level information, population level information and vaccine-specific information. In current practice, information needs are met with both local systems and a small number of widely available systems. While some systems may be very useful at the patient level, they may be less than ideal for practitioners seeking population-level information. System designers will continue to face challenges in balancing these two levels of information needs, and in making the trade-offs involved in maximizing data quality attributes and other technical and administrative characteristics. However, making use of user-centered design methods (70) and engaging all stakeholders to understand their information needs should lead to more usable and useful systems.

Chapter 3: Timeliness and data element completeness of immunization data in Washington State in 2010: a comparison of data transfer methods

Background

While Immunization Information Systems (IISs) were originally intended to simply record vaccination data for the residents of a geographic area, their functionality has been extended to include sending caregiver reminders, forecasting recommended immunizations, running reports, assessing coverage, managing inventory, and generating immunization certificates (15-17). In 2010 the Taskforce on Community Preventive Services made a statement supporting the use of IISs, citing a large body of evidence backing the assertion that IISs are effective in increasing immunization rates (71). Today 77% of children under six in the United States have 2 or more vaccinations records stored in a fully functional IIS (14), but the Healthy People 2020 objectives set a goal of 95% IIS coverage rate for this age group (72).

Expanding coverage may increase the utility of IISs (66), but utility of any information system or data source depends on a combination of the administrative and technical characteristics and the data and information contained therein. Data quality, most often measured by characterizing the timeliness, completeness and accuracy of data, describes an important and multi-faceted predictor of system usefulness (73). Data quality has been well-studied inside and outside the field of public health. Much of the relevant literature on this topic is published in journals of business management, manufacturing, and computing (74-76). Across the health and informatics literature, many additional dimensions of data quality have been identified: clarity, comprehensiveness, conciseness ,consistency, content, coverage, credibility, efficiency, flexibility, format, freedom from bias, importance, level of detail, precision, relevance, reliability, scope, specificity, sufficiency, usability, usefulness, and validity (74; 77-79).

In public health, as in other fields, information systems are important resources that are essential for daily work. In the healthcare literature, several studies list quality as an important factor in choosing an information source (53; 58; 80; 81). Revere et al. (2007) found in a review of the literature that quality was a consistently expressed need for public health information seekers (26).

Technical and sociotechnical barriers as well as resource constraints continue to impede the pursuit of high quality data in IISs. Studies have shown that IIS data are often incomplete and inaccurate (37-39). The timeliness of immunization data is also a concern; the CDC's 2006 Immunization Information System Annual Report (16) and a study conducted using data from Arizona and Minnesota (41) found that less than 75% of all immunizations were recorded in an IIS within 30 days of administration. However, IISs and the environment in which they exist are evolving rapidly, and electronic submission of registry data is becoming more common.

In 2006 Kolasa and coauthors found that, in a high risk area of Philadelphia, providers who used electronic submission of Electronic Health Record (EHR) data had significantly more children in the registry and higher registry-reported immunization coverage rates than providers using billing records or log forms for data submission. Since the Kolasa study, the CDC has suggested that an increase in uptake of standards such as Health Level Seven (HL7) messaging in both IISs and EHR systems and increased connections between IISs and EHRs may benefit the general quality, timeliness (14) and completeness of IIS data (44). Data from the 2009 Immunization Information Systems Annual Report (IISAR) survey showed that 59% of CDC IIS grantees were capable of using HL7 messaging standards to send and receive data (14). Creating, processing and receiving HL7 messages has been a part of the National Immunization Program's Minimum Functional Standards for Registries since 2001(82). However, the proportion of providers using HL7 is expected to continue to increase as a result of Health Information Technology for Clinical and Economic Health (HITECH) Act incentives meant to encourage the adoption of certified EHR systems (32). Under HITECH, Medicare and Medicaid provide financial incentives for the Meaningful Use (MU) of EHR technology. HITECH's Meaningful Use criteria emphasize clinical- and quality-focused uses of EHR data but also include several criteria related to connections between EHRs and public health systems. One IIS-related criterion requires a provider's EHR systems be capable of submitting immunization data to a registry using HL7 V2 messaging standards. Though some have speculated that MU incentives will lead to an improvement in data quality, it is not clear exactly what the impact of an increased proportion of HL7 connections will be on the quality of IIS data. The immunization community has long recognized the importance of data quality for the function of programs and systems, and, in 2007, the National Vaccine Advisory Committee recommended "...the adoption of a guidebook and best practices for IIS...to adopt uniform operational guidance and quality control procedures that ensure good data quality " (83). This guidebook has since been developed, and is one of five operational guidelines containing best practices for IIS functional areas (84). The data quality assurance guidelines were adopted by several state immunization programs (84). For example, state officials in Washington undertook a review of established methods for data loading and quality checks in 2009 and, as a result, adopted 21 of the 32 recommended business rules (20).

Healthcare providers across Washington State currently contribute data to the Children's Health Immunizations Linkages and Development (CHILD) Profile system. The CHILD Profile system has been in existence since 1993 (19) and currently covers 96% of children under 6; it is used by 100% of public providers and 81% of private providers in the state (85). CHILD Profile uses several types of data exchange: HL7 connections between EHR systems and the IIS, batch uploads of data, and web data entry and record retrieval. Generally, vital records (birth certificates) are considered a definitive source for patient date of birth. Provider health records (electronic or paper) are accepted as the definitive source for immunization information and serve to update demographic data. On the other hand, electronic billing records (usually transmitted via batch upload) are considered an inferior source of immunization and demographic information (86). The IIS community has thus far focused on evaluating incoming data being entered or imported into an IIS. In fact, the recommendations published by the American Immunization Registry Association (AIRA) Modeling of Immunization Registry Operations Workgroup (MIROW) (86) focuses its data quality chapter on "incoming data." The IIS in Washington is now in use by providers, schools and public health, and it is important to assess and report on the quality of these data to inform the focus of future data quality efforts and the interpretation of secondary uses of the data.

Not only are IISs immensely useful for providers and public health professionals administering immunizations, but they offer great potential as a population data source for other uses such as community health planning, vital records and epidemiologic studies (87). Because of the prevalence of IISs, their relative maturity, and success in linking clinical practice with public health, it has been suggested that they may be important building blocks for national health information systems (18). As such, IISs offer a unique opportunity to study data quality across different methods of data transfer in a system used by both clinical and public health practitioners. In coming years, MU incentives have the potential to change the quantity and quality of provider-IIS connections across the United States. However, evidence of data quality differences across different types of information exchange is sparse. With this research, we hope to gain a better understanding of the possible impact on IIS data quality of different types of provider-IIS exchanges. Specifically, we used a de-identified dataset from 2010 to describe the timeliness and data element completeness of immunization data in Washington's IIS. We also evaluated whether or not the different types of data exchange between provider and IIS were associated with different levels of timeliness and/or data element completeness of production-level IIS data.

Methods

We used one year of data from Washington State's IIS to determine the timeliness and data element completeness of immunization data. In our analysis we paid special attention to differences among types of data exchange.

Sample Dataset, Inclusion and Exclusion Criteria

Our sample included childhood immunization transactions, defined as a vaccine administered and/or recorded by any participating organization in Washington State for which the recorded service date was between 1/1/2010 and 12/31/2010 (inclusive). The sample included all transactions existing in the Washington State IIS database on July 27, 2011 where at least one immunization was administered during 2010, and the demographic record for the individual indicated an age between 0 and 18 sometime during the year 2010 (1/1/1991<DOB<12/31/2010). We did not include "historical" immunizations entered from paper records or provider report after the immunization was administered, e.g., the parent brought a copy of the vaccination record with some immunizations that were not previously entered in the IIS. Although birth certificate data, uploaded weekly as a batch file from Washington State Vital Records, is an important source for demographic data, as well as hepatitis birth dose records, we excluded vaccination records derived from information included on the birth certificate as they made up a very small proportion of the recorded vaccinations. The birth certificate data are different from the other vaccination data transferred by providers to the IIS because the only vaccination consistently reported this way is the Hepatitis B birth dose, and because the birth certificate files are transferred to CHILD Profile not by providers, but by the Center for Health Statistics at the Washington State Department of Health.

We obtained a dataset that included individual person records, but that excluded most Protected Health Information (PHI). In all fields for which PHI was redacted, data completeness was indicated: "1" - complete and "0" - incomplete. We were granted approval for the study and a waiver of informed consent from the institutional review boards of both the Washington State Department of Health and the University of Washington.

Evaluation methods

Timeliness

Preliminary work and a literature review indicate that the appropriate immunization-level timeliness measure (ImTM) is the number of days between immunization administration (Service Date) and submission of the data to the IIS (Entry Date) (34). Saarlas specifies three categories: ImTM ≤ 7 days, ImTM = 8-30 days, and ImTM ≥ 31 days (34). We calculated this measure for each vaccine record in our sample and examined the data aggregated by visits as well, that is, we aggregated all records where all visit-level fields (Patient Identifier, Service Date, Entry Date, Medical Organization and Medical Facility) were equal, thus consolidating all immunizations that were administered during the same visit, into a single record.

Completeness

The MIROW data quality assurance guidelines specify two types of completeness: comprehensive reporting of vaccination events (population coverage) and complete recording of vaccination data elements (record or data element completeness) (86). Comprehensive reporting is an important characteristic of any population registry. However, our study focused on the complete recording of data elements. We do this because our inquiry is limited to data existing in the IIS and because chart reviews, which would be necessary to assess comprehensive reporting, would be impractical given the number and distribution of providers statewide.

We measured the completeness of both demographic and vaccination data elements. We also calculated the number of complete fields per record for both demographic records and vaccination records. The following measures were calculated as proportions for each record in the sample:

> DC = Number of complete demographic fields /Total number of demographic fields (13 fields)
> IC = Number of immunization-level fields complete/Total number of immunization-level fields (6 fields)

We examined the following fields to calculate demographic completeness: First Name, Middle Name, Last Name, Address, City, State, Zip, Phone, Social Security Number (SSN), Birthdate, Medical Organization and Facility. To calculate immunization level completeness and visit-level completeness we examined: Date of Service, Date of Entry, Current Procedural Terminology (CPT) Code, Immunization Description, Medical Organization, and Facility.

In addition to overall completeness measures, we examined the completeness of the minimum/mandatory data items as defined in the MIROW Data Quality Assurance document (86). As defined, this minimum dataset comprises: First Name, Last Name and Birthdate, as well as Provider Organization Information (coded as the presence of Medical Organization or Facility Organization fields), Vaccine Encounter Date (Service Date) and Vaccine Type (coded as the presence of CPT Code and/or Immunization Description).

The timeliness and completeness measures described above were compared across data exchange types: HL7, Batch, Manual Entry (Web Entry) and in the case of demographic records, Birth Certificate.

All statistical analyses were performed using SAS ® software, Version 9.2 (88). SAS code is included in Appendix C.

Results

After excluding 482 vaccinations records that came from the birth certificate batch file, and three immunization records with negative timeliness measures (entry date was before the service date), a total of 757,476 individual patient records and 2,634,101 vaccination records were included in our sample.

Of the patients included in the sample, 28% were born between 2008 and 2010 (approximately ages 3 and under during 2010), 26% between 2004 and 2007 (approximately ages 4-6 during 2010) and 46% between 1991 and 2003 (approximately ages 7-18 during 2010). Each patient included in the sample was linked to between 1 and 29 vaccination records; the mean number of vaccination records per patient record was 3.5. Within the 2,634,101 vaccination records, 1,269,347 unique visit records were identified for which Patient ID, Service Date, Entry Date, Medical Organization and Facility were equal. For example, if a child received three immunizations during one visit to their provider, these three immunizations were consolidated into a single visit record.

All patient demographic records had a complete data exchange type field: 54.76% were listed as Batch, 14.79% as Birth Certificate (vaccination records with transmission type Birth Certificate were excluded, however demographic records remain), 16.88% as HL7 and 13.56% as Manual

Data Entry. The data exchange type field was also complete for all vaccination records in our sample: 52.78% of vaccination records were listed as Batch, 30.08% as Manual Data Entry and 18.13% as HL7 transmission type.

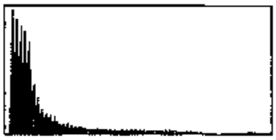
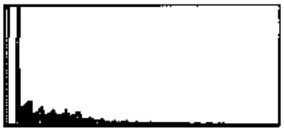
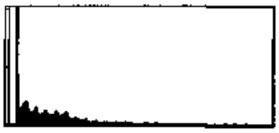
Timeliness

The median timeliness measure for all vaccination records was 11 days. When immunizations were aggregated by visit, the median timeliness measure for the sample was 12 days. Mean, standard deviation (SD), median, and interquartile ranges (IQR) for timeliness were similar when data were analyzed using a dataset aggregated by vaccination record and by visit. In both cases, the greatest median timeliness measure was 31 days for the Batch category (mean 66.02 and 63.87 days for vaccination level and visit level, respectively). Both ways of grouping the data showed the mean timeliness for Manual Data Entry much lower, at around 9 days and the median at 0 (Entry Date = Service Date). Table 3.1 shows sample size, mean, SD, median, IQR and a histogram to demonstrate skew for each data exchange type.

First, we categorized timeliness using the Saarlas categorization described above. Most visit records fell into the <7 days category (44.13%) while 24.78% fell in the 7-30 days category and 31.09% fell into the over 30 days category. Overall, Manual Data Entry (88.52%) and HL7 (80.40%) had a far greater proportion of records in the <7 days category than Batch records (8.43%).

The CDC's IIS Technical Working Group established a limit of 30 days post vaccine administration for the arrival of records into the IIS (82); we categorized timeliness measures in this way (≤30 days or >30 days) and explored the effect of data exchange type on this binary variable. Relative to Manual Entry, HL7 records had more than twice the likelihood (OR=2.79; 95% CI 2.74-2.84) of being unacceptably late, i.e., having a timeliness measure over 30 days. Batch records were over 18 times more likely than Manual records to have a timeliness measure over 30 days (OR=18.77; 95% CI 18.49-19.07). Details and proportion of records in each category are shown in Table 3.2.

Table 3.1: Descriptive statistics and histograms* comparing timeliness of vaccination records and visit records by exchange type**

		Dataset			
		Vaccination Record		**Visit Record**	
Manual Entry					
	N		792,402		361,712
	Mean		8.31		9.63
	SD*		40.26		43.74
	Median		0		0
	IQR**		0-1		0-2
HL7					
	N		477,652		227,101
	Mean		29.23		28.51
	SD		84.69		82.42
	Median		0		0
	IQR		0-3		0-4
Batch					
	N		1,364,047		680,534
	Mean		66.02		63.87
	SD		88.95		85.84
	Median		31.00		31.00
	IQR		16-67		16-65
Total					
	N		2,634,101		1,269,347
	Mean		41.99		42.09
	SD		80.97		79.36
	Median		11.00		12.00
	IQR		0-37		0-38

*SD = Standard Deviation

**IQR = Interquartile Range

***Histograms are of identical scale across rows but not columns (for comparison and to demonstrate skewness)

Table 3.2: Association of Data Exchange Type with Timeliness Measure >30 Days

Exchange	Timeliness Measure (%)				
Type	N	<7 days	7-30 days	>30 days	OR (CI)*
Manual	361,712	88.52	6.27	5.21	1 (ref)
HL7	227,101	80.40	6.31	13.29	2.79 (2.74-2.84)
Batch	680,543	8.43	40.78	50.78	18.77(18.49-19.07)
Total	1,269,347	44.13	24.78	31.09	-

*OR indicates odds ratio, CI indicates confidence interval calculated for a timeliness measure categorized as ≤30 days or >30 days, test of null hypothesis (regression coefficient equal to zero, Chi-square DF=2) was significant, P <.0001

Completeness

The mean completeness measure for all vaccination records was 99.28%. Date of Service, Immunization Description and Medical Organization fields were complete for all vaccination records in the sample. Mean completeness differed only slightly by exchange type. For demographic information, Batch records had the highest mean completeness at 90.76%. HL7 records had the highest mean completeness (99.50%) for vaccination records. Mean completeness measures for both demographic and vaccination records across data transmission methods are shown in Table 3.3.

Table 3.3: Mean record completeness for demographic and vaccination records by method of data exchange

	Percent Completeness – Mean				
Demographic Records	Manual	HL7	Batch	Birth Cert	Total
	87.66	89.98	90.76	89.84	90.07
Vaccination Records	Manual	HL7	Batch	--	Total
	99.76	99.50	98.93	--	99.28

We also calculated the percent completeness for each demographic field. Several fields were 100% complete for each exchange type: First Name, Last Name and Birth Date. Several fields were close to complete: Address, City, State, and Zip Code (all 98% complete or higher). Middle Name was incomplete in many cases, ranging from 48.92% (HL7) to 81.10% (Birth Certificate) complete. Using logistic regression we calculated the OR for a complete Middle Name field. HL7 records were less likely to have a complete Middle Name field than Manual Entry records, but both Batch and Birth Certificate records were significantly more likely to have a complete Middle Name field than Manual Entry records (OR=1.45, 95% CI 1.43-1.47 and OR=3.52, 95% CI 3.45-3.58, respectively). Phone number varied in completeness by exchange type; only 80.86% of Manual Entry records were complete for this field, but 95.46% of HL7 records had a completed phone number field, making them nearly 5 times more likely to have a phone number in the IIS (OR= 4.98, 95% CI 4.83-5.14). Social Security Number also varied by transmission method; 9.44% of Birth Certificate records were complete for this field but 32.95% of Batch records carried an entry for SSN. Medical Organization and Facility were both relatively complete fields with the lowest completeness levels in records from Birth Certificates and the highest for both fields in records where the exchange type was an HL7 message. Table 3.4 shows percent demographic data element completeness by exchange type and associated OR and 95% CI.

Table 3.4: Percent Completeness and Association of Data Exchange Type with Complete Data Elements

Data Element	Data Exchange Type	Percent Complete Records	OR (CI)*
First Name	All	100%	**
Middle Name	Manual	54.98%	1 (ref)
	HL7	48.92%	0.78 (0.77-0.80)
	Batch	63.95%	1.45 (1.43-1.47)
	Birth Certificate	81.10%	3.52 (3.45-3.58)
Last Name	All	100%	**
Address	Manual	98.44%	1 (ref)
	HL7	99.94%	27.78 (21.97-35.13)
	Batch	99.95%	30.03 (26.07-34.59)
	Birth Certificate	99.96%	41.32 (30.52-55.95)
City	Manual	98.81%	1 (ref)
	HL7	99.99%	81.30 (51.68-127.94)
	Batch	99.99%	161.58 (113.13-230.78)
	Birth Certificate	99.97%	36.57 (26.37-50.73)
State	Manual	99.21%	1 (ref)
	HL7	99.99%	59.69 (36.92-96.49)
	Batch	99.99%	64.52 (48.62-85.62)
	Birth Certificate	99.96%	20.68 (15.22-28.11)
Zip Code	Manual	98.76%	1 (ref)
	HL7	99.98%	69.60 (46.08-105.13)
	Batch	99.98%	70.17 (55.50-88.71)
	Birth Certificate	99.96%	31.16 (23.14-41.96)
Phone	Manual	80.86%	1 (ref)
	HL7	95.46%	4.98 (4.83-5.14)
	Batch	88.22%	1.77 (1.74-1.81)
	Birth Certificate	88.29%	1.79 (1.74-1.83)
Birthdate	All	100%	**
SSN	Manual	19.32%	1 (ref)
	HL7	29.93%	1.78 (1.75-1.82)
	Batch	32.95%	2.05 (2.02-2.09)
	Birth Certificate	9.444%	0.44 (0.43-0.45)
Medical Organization	Manual	98.22%	1 (ref)
	HL7	99.92%	22.78 (18.28-27.15)
	Batch	99.43%	3.15 (2.96-3.35)
	Birth Certificate	96.07%	0.44 (0.42-0.47)
Facility	Manual	94.60%	1 (ref)
	HL7	95.61%	1.24 (1.20-1.30)
	Batch	95.51%	1.21 (1.18-1.25)
	Birth Certificate	93.20%	0.78 (0.76-0.81)

*OR indicates odds ratio, CI indicates 95% confidence interval calculated for a complete data element, test of null hypothesis(regression coefficient equal to zero, Chi-square DF=3) was significant, P <.0001
**No OR calculated, all observations have the same response (complete)

Total field completeness by record was high, but varied by field and type of record (demographic vs. vaccination records). The mean field completeness measure (number of completed fields/total number of fields) for the demographic records was 90.07%, indicating an average of 11.70 out of 13 fields completed per record. First Name, Last Name and Date of Birth, fields identified in MIROW's minimum dataset (86), were 100% complete.

Field completeness for vaccination data elements was very high. Most fields (Date of Entry, Date of Service, Immunization Description, and Medical Organization) were 100% complete for all exchange types. CPT Code was more than 99% complete for all exchange types. Facility was 98.80% complete for manual entry records but only 93.58% complete for Batch records. A logistic regression to detect an association between exchange type and field completeness showed that HL7 records were 6 times more likely than Manual records to have a complete CPT Code field (95% CI 5.16-7.07). In the case of the Facility field, HL7 and Batch records were both less likely than Manual records to have a value in that field. A report of all field completeness measures and applicable OR and 95% CI for vaccination fields is shown in Table 3.5.

Table 3.5: Percent completeness of immunization fields – one record per vaccination

Data Element	Data Exchange Type	Percent Complete Records	OR (CI)*
Date of Entry	All	100%	**
Date of Service	All	100%	**
CPT Code	Manual	99.79%	1 (ref)
	HL7	99.96%	6.04 (5.16-7.07)
	Batch	100%	485.88 (2.17.98 - >999)
Imm. Desc.	All	100%	**
Med Org	All	100%	**
Facility	Manual	98.80%	1 (ref)
	HL7	97.03%	0.18 (0.17-0.18)
	Batch	93.58%	0.40 (0.39-0.41)

*OR indicates odds ratio, CI indicates confidence interval calculated for complete data element, test of null hypothesis (regression coefficient equal to zero, Chi-square DF=2) was significant, P <.0001
**No OR calculated, all observations have the same response (complete)

We also examined the vaccination data by number of fields complete and data exchange type. We found no vaccination records in the sample with fewer than 4 completed fields, and 98.60% of Manual Entry records had all 6 fields complete. Only 93.58% of records entered using Batch uploads were 100% complete.

Discussion

Administrators of IISs spend significant time and effort ensuring the quality of the data in their systems. Quality checks on data entering the system are standard for all types of data exchange. Data quality checks and de-duplication processes often take place before data are available in the production system. In the system we studied, all incoming records are held as *reserve* records until de-duplication and other quality rules are applied to determine which records will be used to populate the *master* record maintained in the live system. It is important to consider these details of data flow when examining the results of our analysis. We examined *master* records, which are, in general, of higher overall quality than *reserve* records, with fewer duplicate records. The data we analyzed are the data available to providers enrolled to use the IIS and to the epidemiologists who make immunization data requests to the State of Washington.

Timeliness

We compared descriptive statistics (Table 3.1) for the vaccination dataset and the dataset aggregated by visit; the results were similar for the two methods of grouping. Based on the similar results for the totals and within data exchange types, we aggregated data by visit for all subsequent analyses related to timeliness.

Overall, we found that the timeliness of vaccination data was highly variable (See Table 3.1 and Histograms in Appendix D). However, most data entered manually (88.52%) and most HL7 records (80.40%) fell into the category indicating they were relatively timely, with less than 7 days between date of immunization and date entered into the IIS. Vaccination records that arrived via a batch data transfer had a significantly smaller proportion arriving within a week to the IIS: only 8.43% of batch vaccination records fell into the less than 7 days category.

Our dataset was very large; even when reduced to records aggregated by visit, over 1.2 million records remained. Because of the size of the dataset, it is not surprising that we found a statistically significant association between each type of data exchange and acceptable timeliness measures (timeliness measure ≤30 days vs. >30 days. Although statistical significance does not imply practical significance, the summary data and relative ORs and 95% CIs convey differences that could impact the decision to use data for a particular purpose. For example, providers and public health efforts to reach out to parents with reminders for upcoming recommended vaccinations, will be able to have more trust in a dataset with more timely data. Also, epidemiologists considering use of IIS data for outbreak control activities and reportable condition case investigation require a certain level of timeliness in order for the data to be useful. Both the significance test and the raw differences in median timeliness measures point to the need for improvements in batch entry procedures or the conversion of organizations using batch to transition to HL7. Unfortunately, both of these options create more work, at least initially, for IIS administrators. Setting up and testing HL7 field mapping and connections to provider EHRs is still a labor intensive process. Although use of Batch files for exchanging data also involves mapping, anecdotal reports indicate that the HL7 process may require more time and resources to set up than a traditional Batch file. In addition, many of the quality checks run on data before they move into the production system require manual, human-driven analysis of the data.

Completeness

Our examination of data element completeness includes two types of data: demographic data and vaccination level data. For our completeness analysis, we used the vaccination dataset and not the visit-level dataset. Both demographic data and vaccination data were quite complete. All demographic fields were at least 93% complete except: Middle Name (62.73%), SSN (27.11%), and Phone Number (88.45%); completeness percentages are shown in Table 3.4. Because not all patients have a middle name, the field legitimately may be empty. Middle name is nonetheless an important field for use in de-duplication processes (20). Social Security Number brings another set of concerns, and it is not surprising that Birth Certificate Data (newborns may not have an SSN assigned at the time Birth Certificate Data is issued) and Manual Entry (privacy concerns) had low completeness rates for these fields: 9.44% and 19.32% complete, respectively. SSN is not often relied upon in registry work, but is potentially useful for population level studies where linking to other datasets (birth, death, registries from other states) may be necessary. Similarly, Phone Number offers challenges because it may not be consistent. However, it is a useful data element for record matching, for epidemiologic outbreak control activities, and it can be a way to contact patients or parents (via SMS or voice) for recall and reminder outreach. We suspect that it will be of practical significance to those making use of the Phone Number field for outreach activities that the field is almost 5 times more likely to be complete for HL7 records.

For the fields we examined, vaccination level data were even more complete than demographic data: on average vaccination fields were 99.79% complete. The only field with a higher proportion of missing values was the Facility field. Facility was more complete for manually entered records and only 93.58% complete for Batch records. Facility information can be important for allowing providers to find records submitted in the form of billing or Medicaid data (usually Batch submission) (20). MIROW's data quality guidelines specify immunization date, CVX code, manufacturer, lot number and administering provider as important fields for vaccination level data(86). Our dataset included immunization date, and administering provider, and completeness was near 100% for both of those fields.

Most of the CDC's NVAC core data elements (83) are for the most part present in the CHILD Profile IIS. The 2009 Immunization Information System Annual Report shows all data elements of interest are at least "being collected" by the CHILD Profile System in Washington (89), which indicates that all fields can be received by the system, though levels of participation on the part of providers may vary. The differences we found in levels of completeness in the vaccination data, though statistically significant, are likely not practically significant (0.83%); however, the differences we found in completeness of some demographic data elements may be of concern to data consumers.

Our earlier investigation of information needs in public health related to immunizations (see Chapter 2) found that concerns about both population completeness and data element completeness may prevent secondary users of IIS data from accessing the data for population analyses and assessments. Although Batch records from billing systems and Medicaid are sometimes considered a source of data element incompleteness, batch transmission actually had the highest mean completeness for demographic records (90.76%).

Implications

Numerous timeliness, completeness, de-duplication and logic checks take place before data are made available in CHILD Profile. The Data Quality Manager for the IIS recently implemented new business rules resulting in an improvement in the quality of data entering the system (20; 84). The increasing use of standards for data storage and transmission may decrease the need for these types of checks, putting more of the burden of data quality assurance on IIS and EHR system managers and developers.

Although the National Immunization Program, in its Minimum Functional Standards for Registries in 2001 gave an official recommendation to use HL7 messaging (82), the MU incentives of the HITECH Act may be even more effective in encouraging adoption. Financial incentives and the even stronger encouragement in the form of penalties for non-compliance are likely to significantly increase the number of providers submitting data to IISs using HL7 messaging. If the type of HL7 messaging used most often in 2010 (Version 2) remains the same, we can say with some level of confidence that this could move the timeliness and completeness measures toward the descriptions we see in these analyses for HL7 messages. If the migration to HL7 messaging comes from currently unenrolled providers, or those submitting Batch records, this will mean an improvement in timeliness, as seen in Table 3.2. Although our results show that Manual Entry is more timely than HL7 messaging, the details of HL7 are highly dependent on the implementation. Ideally, HL7 messages can be sent in real-time or at least daily.

The quality of data contained in HL7 messages is dependent on the quality and the consistency of source system data, i.e., the data in the EHR, billing system or Medicaid data file. Again, certified EHRs should offer improvements in these areas and could lead to overall improvements in data element completeness within an IIS.

Overall, the migration of providers to HL7 connections with IISs will likely not have an immediate or significant impact on the data element completeness of IIS data in systems similar to the Washington State IIS. Depending on the frequency of the HL7 messaging, timeliness however, could improve significantly. The migration of providers currently sending batch files, to use of HL7 messaging, could decrease the proportion of records transmitted more than 30 days after vaccine administration (Table 3.2). If EHR-IIS connections make use of near-real time HL7 messaging for communication, the overall timeliness could improve even more. Near-real time HL7 connections offer the potential to improve data quality. Bi-directional interfaces provide advantages to providers as the time of vaccine administration by allowing the provider's system to display or import IIS data, which could contain records from other providers, therefore rendering a more complete picture of the individual's current vaccine status. Overall, the potential and likely benefits of migration to HL7 interfaces, especially when many of the existing batch interfaces are migrated, could have a positive impact on data quality.

Limitations

The most significant limitation of this study was that we were not able to examine two highly relevant components of data quality (Chapter 2): vaccination record coverage completeness and accuracy, both of which could be measured with chart review and interview methods. Additionally, conclusions made from our analysis about data quality must be interpreted only with respect to this specific production dataset. Our study examined data in the production version of the Washington IIS which includes master records only. The IIS also maintains reserve records that may include duplicates or other errors, which have been eliminated by the numerous data quality and logic checks run on the data before the reserve record is moved to a master record. An analysis including records in the reserve dataset may find different associations between method of information exchange and timeliness and completeness measures.

We also limited our analysis to records with a "Service Date" in 2010. The distribution of transmission types used by providers and other organizations and those transmission types accepted and encouraged by IISs administrators is changing quickly. It is possible that inclusion of additional years of data may be advantageous in drawing conclusions about the quality of data in the production system.

An additional limitation of our analysis is the lack of information about the source of the data. In the case of Batch records, it is likely to be important to differentiate between Medicaid, other billing, and EHR data. We were unable to distinguish between these sources with the available dataset.

Conclusion

Although data quality is an important characteristic of a data source for providers and those operating immunization programs, it is also important for ensuring that IIS data are a useful and trusted source of population level data for public health practitioners and researchers (see Chapter 2). Our work takes steps to describe the timeliness and data element completeness of Washington's IIS. In doing so, we have attempted to provide information that may be of use for persons or organizations considering the use of IIS data or data exchange with an IIS. Though the transition to HL7 messaging will be a challenging road for many immunization programs and providers, we believe there is demonstrated evidence that while the transition may not immediately impact completeness in systems with a similar demographic makeup as the Washington State IIS, it could lead to improvements in the timeliness of IIS data.

IISs are one of the most mature public health information systems used in the United States. As Hinman and Ross posited in 2010, "registries provide a model for how public health information systems can support clinical decision making, meet public health demands for assessment and assurance, improve quality of care and contain costs" (18). The systems' maturity, the experience that public health and providers have with them, their interfaces, and the resulting population data sources, allow them to serve as exemplars for future connections between the two entities. We hope that in a similar way this work can be useful beyond the world of Immunization Information Systems, to inform and direct the development of other types of provider-public health electronic connections for purposes of assessment, assurance, and ultimately to safeguard the health of the population.

Chapter 4: Enhancing the Usefulness of a Public Health Data Source

Background

Immunization Information Systems (IISs) are one of the most advanced and widely used systems in public health. These systems are used by public and private providers alike to track and submit individual level immunization data, and they are used by public health officials and researchers to assess coverage and answer questions about immunization in the community. Immunization programs are working to enhance their already advanced information systems to better meet the information needs of providers, as well as the needs of other IIS stakeholders. This is particularly important in the context of recent rapid Health Information Technology (HIT) adoption (90)(91), incentives encouraging the Meaningful Use of HIT (32), and increased interest in Health Information Exchange (HIE) (92)(93).

Though IIS functions vary by system, most provide support for public and private provider immunization patient tracking as well as for basic administrative tasks required of an immunization program (14). In addition to these functions, immunization registries are often touted as excellent resources for assurance and assessment activities (87)(43)(94). However, based on our work in Washington, the state's IIS facilitates some of these functions much more efficiently than others. Functions related to the entry, retrieval and use of individual patient level data are much better supported than those related to population level uses of data (Chapter 2).

Today, most geographic areas in the US are covered by at least one active IIS. Although population coverage of IISs in the US is increasing, the CDC estimated in that in 2009 only 77% of children under six actually had data in an IIS (95)(14). Due to policy changes and the evolving health information interoperability climate, there is imminent potential for a rapid increase in provider participation in these systems (14)(96). This increase in the number of providers with electronic connections to IISs will expand the population coverage of the systems and augment the potential value for both individual and population level uses.

The focus on provider-IIS connectivity is a small part of a much larger movement toward increased interoperability among healthcare information systems in the United States. Immunization Information Systems are by no means the most advanced systems used in health and medicine today; however they are vetted, respected, and have a longer history than almost any other system that directly connects providers with public health, and that supports the bidirectional exchange of patient-level data (97)(35). A number of expert reviews using primarily consensus-driven methods have put forward recommendations and functional standards for IIS development (84)(82); however there have been few specific recommendations based on primary data collection from users, presenting concrete suggestions for maximizing the usefulness of the IIS as a data source for individual and population level use.

As provider enrollment improves nationwide, IISs will likely be called upon to fulfill additional functions. This rise in coverage will be beneficial; however an increased number of total records in the system will not, on its own, improve the overall usefulness of IIS data. Other factors must be considered as well (Chapter 2 and3).

This chapter offers guidance for the development of this evolving data source. Two recently completed studies—a qualitative exploration of information needs related to immunization (Chapter 2), and a quantitative analysis of data source characteristics of immunization repositories (Chapter 3)—helped to formulate a set of recommendations for improving the utility of information systems and the data they contain. These two studies take different approaches to examining the same information work and information systems. Using different methods, they attempt to address the following two research questions:

- What are the information needs of public health practitioners with regard to immunization information?
- What are the characteristics of one public health information system (the CHILD Profile IIS) with respect to information needs related to data quality, specifically, timeliness and completeness?

We make use of mixed (qualitative and quantitative) methods, allowing for more complete insights into aspects of the work that would not have been possible using only a single method of analysis. In this chapter, we present our list of specific recommendations resulting from the synthesis of these two studies, as well as a summary of the feedback on these recommendations from state level stakeholders in Washington. The significance of this work goes beyond the immunization domain and can inform the development of other public health systems where individual as well as aggregate or population views of data are important.

Methods

The work presented in this chapter is based on the two studies presented in Chapters 2 and 3. In the first, a qualitative study of the information needs of public health practitioners with respect to immunization information, we identified information needs of public health practitioners that fell into three categories: individual level needs, aggregate or population level needs and context-specific needs (in this case, vaccine-related). Individual level needs were often related to accessing and using the individual patient's immunization record for purposes of vaccine administration or patient lookup. Reporting on coverage of a geographic area and using large samples or entire datasets to answer population level questions are both examples of population level information needs. Vaccine-related information needs were specific to immunization programs and involved access to immunization schedules, vaccine contraindications and storage information about vaccine inventory. Based on those information needs, we also identified the data source characteristics that were relevant to the work done by interviewees. The most important of these were: accessibility, metadata availability, data format, granularity, and data quality.

The second study was a quantitative description of two aspects of data quality—identified as an important data source characteristic in Chapter 2—in Washington State's IIS: timeliness and data element completeness. This study found that timeliness varied significantly by data exchange type (Web data entry, HL7, or Batch), with data coming from providers via batch transfer much more likely to arrive 30 days or more after the vaccine was administered when compared with HL7 messages and Web data entry. Completeness was very high overall, but we found areas for possible improvement in middle name, phone number and facility fields of the demographic records.

We reexamined coded transcripts from the qualitative study of information needs to find unmet information needs that were identified in more than one interview and that were feasible to address. Using this technique we found a number of items that could be formed into recommendations, many of these were consolidated due to similarities. We re-examined results from the quantitative study of data quality and used these to inform the existing list and to form new recommendations. The resulting list, and further synthesis of and reflection upon the two studies led to a list of recommendations for improving the utility of immunization data. Specifically, the recommendations were aimed at helping better meet the information needs of state and local public health practitioners working in the area of immunization. These recommendations were presented to state-level stakeholders. Presentation slides from the stakeholder interviews are contained in Appendix E.

We presented the recommendations during 4 one-on-one interviews and small group meetings with a total of 8 stakeholders working at the state level in Washington. These stakeholders included epidemiologists, IIS program administrators and informatics professionals. At each stakeholder meeting we presented background on the two studies that informed the recommendations and then presented the list of recommendations, the final versions of which are shown in Tables 4.1 and 4.2. Following each presentation, we conducted a question, answer, and discussion session during which feedback was gathered. We collected feedback on general concerns, feasibility, high priority areas, relative importance, and information about which recommendations had already been identified and addressed. The feedback was unstructured. We made field notes during, and immediately after each meeting. Field notes were synthesized, and incorporated into the final list of recommendations presented in the Results section below.

Results

Each recommendation is an outcome of the synthesis of the results of the two studies we describe in Chapters 2 and 3 of this dissertation. The list comprises two categories: system enhancement recommendations, and process and strategic recommendations. In Tables 4.1 and 4.2 we present these two sets of recommendations. Within each table, the recommendations are organized by information need category; recommendations addressing individual level information needs are presented first, followed by population level needs, and vaccine information needs.

System enhancement recommendations, presented in Table 4.1, are primarily focused on technical improvements and modifications to the IIS. The changes we suggest would add to, or enhance system functions, enable access to the IIS for new groups of users, and improve connections between the IIS and other systems such as provider EHRs and IISs in other states.

The process and strategic recommendations presented in Table 4.2 are related to the availability of information but are independent of the information system itself. Process and strategic suggestions address potential modifications to workflow or in some cases suggest higher level organizational and strategic change. These are aimed at improving general organizational performance related to immunization.

Feedback from stakeholders on the suggested enhancements and strategic recommendations was positive overall; in several cases, criticism and questions were used to modify and clarify the recommendations. For instance, one interviewee suggested that additional examples of ad-hoc queries be listed, and another stakeholder asked for more description of the recommendations related to metadata availability. Most stakeholders focused their comments on the recommendations most closely related to their daily work. Feedback led to the consolidation of several recommendations. Clarifying questions also led to rewording of some items; however, none of the recommendations were eliminated completely, and no new recommendations were added to the list. Based on suggestions and questions during the meetings, a number of specific examples were added to the recommendation lists.

Several of the recommendations we presented were identified as priorities for the immunization program prior to the compilation of our list; several of these have already been addressed. These items are indicated in Table 4.1 by check-marks. The recommendations that have been addressed, at least in part, in Washington include: facilitate access to IIS data for schools, enhance the local health department views of IIS data, evaluate and enhance the inventory management tools in the IIS, and enhance IIS functions for mass vaccination campaigns.

Table 4.1: System enhancement recommendations, examples, and information need category addressed

#		Recommendations and examples	Category*
1	✓	**Facilitate access to IIS data for state licensed child care providers**	**I, P**
2	✓	**Facilitate access to IIS data for schools**	**I, P**
3		**Facilitate linkages with other immunization related data sources**	**I, P**
		Link or facilitate side-by-side comparison of immunization coverage and school exemption data Query for both an immunization record and a case-report record or facilitate the linking of records in these two datasets Facilitate linkages between the Public Health Issue Management System (PHIMS) and the IIS	
4		**Enhance the local health department view of individual level IIS data**	**I, P**
		Allow LHJ access to expanded individual level data for all providers in the jurisdiction	
5	✓	**Enhance IIS functions for mass vaccination campaigns**	**I, P**
		Off-site real-time data entry Tracking immunizations and patient look-up Timely access to mass vaccination campaign data Excel spreadsheet upload of mass vaccination campaign data	
6	✓	**Facilitate live sharing of IIS data with neighboring states (OR, ID)**	**I**
		Investigate the use of Integrating the Healthcare Enterprise Profiles for live sharing. Two profiles may be appropriate: Query for Existing Data and Immunization Content	
7	✓	**Expand reporting functions**	**P**
		Immunization coverage rate for a single vaccine for an age group in a geographic region Reports that facilitate comparison of coverage rates between providers Ability to run and save ad-hoc reports Facilitate the generation of reports from the IIS for performance measures	
8		**Provide summary information describing the IIS dataset (metadata)**	**P**
		Related to both system and strategic suggestions, examples in table 4.2	
9	✓	**Evaluate and enhance inventory management tools in the IIS**	**V**
		Allow providers or local health jurisdictions to look up ship dates Streamline vaccine ordering and tracking process (reduce number of steps, intermediaries)	

**Information need categories: I=Individual level, P=Population level, V=Vaccine level*

A check-mark (✓) indicates that the suggestion has been partially or satisfactorily addressed in Washington

Table 4.2: Process and strategic recommendations, examples, and information need category addressed

#	Recommendations and examples	Category*
1	**Encourage timeliness in all types of data and information entry (will serve to increase usefulness and reputation)**	I, P, V
	Encourage live or daily HL7 feeds rather than HL7 messages sent in batches Investigate more timely update of vaccine recommendation data in the IIS	
2	**Provide information about alternate and linkable datasets**	I, P
3	**Streamline the data request process to reduce burden for the IIS administrators**	I, P
4	**Raise awareness among providers of the IIS functions that can save them time and money to encourage connectivity between Electronic Medical Record (EMR) systems and the IIS**	I, P
5	**Examine the processes for the three primary data transfer methods: HL7, Manual Data Entry and Batch to identify potential areas for improvement in timeliness measures.**	I,P
6	**Review local health jurisdiction level access issues for individual level and population level data**	I, P
	Review current business rules Review current data sharing agreements with providers	
7	**Provide public use, de-identified dataset as an export to common file type (xls, csv)**	P
8	**Look to vital records, Washington Tracking Network (WTN) and Community Health Assessment Tool (CHAT) for guidance on examples of providing summary data, interactive queries and standard datasets to researchers**	P
9	**Consider the use of free and open source tools for querying and visualization of public use and/or aggregated IIS data**	P
	Google Charts: http://code.google.com/apis/chart/ (Example – Gossamer Health)	
10	**Create or facilitate the establishment of a Community of Practice (CoP) for IIS data users (and prospective users)**	P
	Provide metadata to prospective users (examples in item 9, below) Encourage structured communication for data requests (implement a structured query language to facilitate communication between data requestors and IIS administrators)	
11	**Provide summary information describing the IIS dataset (metadata)**	P
	Immunization coverage in the state based on IIS data Provider enrollment based on IIS administrative records Average timeliness of records stratified by provider and patient variables of interest Data element completeness based on IIS data Estimated population completeness based on IIS data and population (census) data Comparisons with National Immunization Survey (NIS) data Provide access to metadata, including descriptive (title, author, keywords), structural (order, data structure) and administrative (explain source and data origination) metadata Provide Entity Relationship Diagram (ERD) or other description of data structure Provide description of how data are collected	
12	**Leverage healthy competition between organizations (schools, providers, local health jurisdictions) and use IIS data to make comparisons**	P
13	**Review common information needs of providers with respect to vaccine recommendations**	V

**Information need categories: I=Individual level, P=Population level, V=Vaccine level*

Some of the recommendations were higher priorities for the stakeholders. Interest and discussion was mostly focused on Table 4.2. Three recommendations in this table—all related to population level information needs—received significantly more attention during the stakeholder interviews:

- streamline the data request process to reduce burden for the IIS administrators (Table 4.2, Item # 3),
- create or facilitate the establishment of a Community of Practice (CoP) for IIS data users (and prospective users) (Table 4.2, Item # 10),
- provide summary information describing the IIS dataset (metadata)(Table 4.2, Item # 11).

The interest in these items demonstrates recognition, on the part of these stakeholders, of inadequacies of the immunization information system in providing reliable and efficient access to population level data. These three items and population level data issues are further explored in the Discussion section.

Discussion

We made specific suggestions for improvements that complement existing sets of recommendations and best practice guidelines. Both AIRA and CDC documents (84)(82) have used IIS experts to analyze practices and come to consensus on best practice guidelines. Our process used interviews with practitioners from the field (expert users), combined with a quantitative analysis of actual IIS data to produce a set of recommendations aimed specifically at meeting the needs of the public health practitioners. Integration of the results of our two studies (Chapters 2 and 3) led to evidence-based recommendations, which were reviewed and verified by state level stakeholders (local subject matter experts) in Washington.

In the Methods section we identified three recommendations related to availability of the aggregated patient data, or to a population dataset, that were of importance to stakeholders. In this section we discuss this focus on population-level uses of the data, the generalizability of our recommendations, the limitations of our research, and future work.

Aggregate patient data and population-level information needs

The importance of supporting population level uses of immunization data was a recurring theme that we found during the analyses for Chapters 2 and 3, and as we reflected on our experience as researchers using the IIS dataset. This theme manifests not only in the need for expanded reporting functions for local health jurisdictions, but also in the need to make the IIS dataset more accessible. With respect to the IIS dataset, one immediate need revealed during our analyses is the need for better descriptions of the data. Interviewees relayed needs related to general descriptive statistics from the IIS dataset as well as information about data structure, data definitions and descriptions of how data were collected. As reflected in the recommendations (Table 4.2, Item # 11), the availability of these types of descriptive information about the IIS dataset could significantly increase the usefulness of the population level data for some users.

The attributes described above are best described as metadata, often referred to as "data about data." Metadata is defined by the National Information Standards Organization (NISO) as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource." (98) There are three main types of metadata: descriptive, structural, and administrative. Although metadata can be used to describe any type of resource, the following examples are relevant for information resources such as scientific datasets, as opposed to other non-digital resources such as photographs or works of art. Descriptive metadata provides a description of a resource to enable discovery or identification, for example descriptive statistics, a data dictionary, and general information about population and ages included in a patient dataset. Structural metadata describes how objects within the resource have been combined; an entity-relationship diagram or descriptions of the tables and relationships within a database would provide this type of metadata. Administrative metadata helps with resource management and might include information about how data were collected, who has access to the identified data, who has access to de-identified data, and the process by which data should be requested and accessed.

We have recommended all three types of metadata to enhance the usefulness of immunization related information resources. The availability of metadata would facilitate identification of useful data resources, streamline the data request process and save time. Metadata are sometimes discovered in the process of working with a dataset, whereas providing that same metadata in advance may help data requestors work more quickly and efficiently with less assistance from information resource administrators.

Similarly, providing metadata to users and potential users of the IIS data source could benefit IIS administrators who spend a substantial amount of time and effort answering questions about the dataset via email. We suggest that a central repository for metadata be created and that this repository contain simple annual or semi-annual summary information such as: the total number of patient and vaccination records in the database, number of providers currently contributing data to the system, the number of patients who have "opted out" of the IIS, as well as summary data quality measures. Metadata describing the data structure and definitions within the dataset will also be useful and might include: data elements, available tables, relationships, and a data dictionary. Background information about how data are collected, stored and processed can also be valuable for dataset users, e.g., data collection procedures, data deduplication procedures, and the standard parameters of data request queries used to create datasets for users.

In addition to access to summary metadata, a structured method of communicating data requests could also streamline the process for both data requestors and those responsible for creating and delivering the requested dataset. The current process involves extensive email communication to solidify a data request. Imposing structure on this request process could make the process faster and easier for both the information requestor and the resource administrators. The census bureau (99), the CDC (100), and some states (101) have used web forms to structure data requests; the forms simultaneously constrain the request and expose the underlying structure of the data to the data requestor. Using such a form, parameters are made explicit and the current process (involving repeated email communication and multiple attempts at creating the desired dataset) is made more efficient.

Communities of Practice (CoP) were also discussed at length during our interviews (Table 4.2, Item #10). A CoP is a group of individuals who share a common interest in some activity or practice area; they also share the desire to interact with others interested in the practice area and to learn from those interactions. The CoP model has been used in business and health care sectors to encourage the exchange of experiential information between practitioners doing similar work (102)(103). Because population level use of the IIS dataset to date has been relatively limited, it would be useful to facilitate the exchange of experiential information about requesting, accessing, using, querying and analyzing the IIS dataset. This offers potential benefit not only to the data users, but also to IIS administrators for whom the burden of repeated queries would be relieved by providing other resources for data users through the CoP.

Dataset availability and the process of requesting datasets received significant attention during our discussions with state officials. Work in this area has begun with respect to Institutional Review Board applications, and early indications are that work will soon begin on developing some of the information tools described here.

Generalizability of recommendations

We believe that our recommendations, although developed based on data collected in Washington, will be useful to other immunization programs and to others hoping to make available population data sources made up of aggregated patient data.

The immunization information system in Washington is just one of many IISs in use around the country and the level of sophistication and integration of these systems varies widely (14). For less advanced systems, our recommendations for enhancements (Table 4.1) may be used as a starting place for the development of requirements and scenarios of use to guide development. Where more advanced systems are already being used, the process and strategic recommendations listed in Table 4.2 may prove more useful as they provide ideas about potential areas of improvement and general directions to enhance the performance of an immunization program. The process and strategic recommendations have potential to be used to provide direction for improvement in the area of data collection and provision of data for public health.

Although our recommendations are specific to immunization work, and focused on immunization information systems, we believe that generalized versions of these recommendations may be applied more broadly. Across the United States, HIE organizations and other movements toward increased health information sharing are gaining support. Federal level goals for adoption of electronic health record (72) and incentives for Meaningful Use of certified health information systems (32) encourage the sharing of electronic data and connections with public health. It may soon be possible to leverage HIE data to provide the public health community with access to large, population level datasets, similar to the current model represented by immunization information systems (18)(104). As public health organizations gain and provide access to these newly available information resources, evidence-based recommendations can serve to inform system design, redesign, or program evaluation efforts.

We identified generalizable themes in our final list of recommendations by examining each recommendation and asking the question "is this a context-specific recommendation or one that will be broadly useful?" Among those that we identified as possibly useful outside of the immunization context, we found three recognizable themes:

- access to individual level views and uses of the data have a unique set of scenarios of use that vary by stakeholder or user role,
- access to population level views of the data require a separate and unique set of scenarios of use that vary by stakeholder and user role, and
- access to metadata of all kinds is important to most population level uses of an information system and its data.

These three themes align with two of the three types of information needs we identified in Chapter 2: individual level information needs and population level information needs. During the course of the analyses for Chapter 2, we found that population level information needs have received less attention during the development and requirements gathering process than individual level needs. Overall, the individual level information needs of public health practice are being satisfactorily met, while population level information needs are not. Though it may seem that data meeting individual level information needs will lead to data that satisfactorily meet population level needs, we found that this is not the case. One critical difference between these two types of information is the provision of metadata.

In addition to individual and population level information needs, a third type of need we identified was context-specific information we defined as "vaccine information". This type of information is critical to the functioning of health care providers, who need information about vaccination schedules and contraindications, and to the functioning of local and state health departments who need all of the information about each vaccine, and often track inventory and ordering for the Vaccines for Children (VFC) program. This immunization context-specific information is not directly generalizable to other population data sources. However it is an important part of the work being done by public health, and it is logical that the information system managing other immunization program work would also cover vaccine information. It is necessary to fully understand the work environment and context in which an information system will be used in order to ensure that all information needs are considered as systems are developed.

Limitations

This work has several limitations. The studies that informed these recommendations took place in one state, and in that state only a sample of the local health jurisdictions were included in the qualitative component of the work; therefore, the generalizability of the recommendations may be limited. The time between the start of data collection and the presentation of the recommendations was more than a year, and by that time some of the issues we identified had been resolved. Although this limits the usefulness of the work for this state, it suggests that the recommendations could be useful to other states whose immunization information systems are at earlier stages of development. Some of the recommendations we made are ambitious, and may be beyond the scope of work possible for immunization programs operating in the current fiscal climate.

Future work

Health information exchange is still in its infancy, and the era of fully interoperable information systems has not yet arrived; the work we have described is just beginning. Future work following from these studies will include private providers in an analysis of information needs, a full evaluation of data quality in the IIS including accuracy and population level completeness, and a study of the use of our recommendations and their impact on data source usefulness on individual and population levels. In addition, we believe that applying our recommendations to other areas of public health practice such as case-reporting and surveillance and their associated data sources will be a valuable test of their generalizability.

Conclusion

There is a strong desire among all stakeholders to improve the quality of immunization data sources in Washington and around the United States. It will be wise to take advantage of the current interest and momentum surrounding the exchange of health information to enhance existing IISs and increase their overall utility. To inform these changes, we undertook a qualitative study of information needs related to immunization and a quantitative study of data source characteristics.

As a result of our mixed methods approach, we were able to describe important aspects of information work and characteristics of population data sources. We found that three types of information needs exist in the area of immunization: individual record level needs, population level needs and vaccine information needs. Our quantitative analysis explored several data quality measures important to users of the IIS dataset. The integration of our results using mixed methods techniques allowed us to extract more valuable insight than would have been possible using only the sum of the individual studies; through this synthesis, we developed a set of evidence-based recommendations.

We hope that the recommendations presented here will inform developments in IISs to improve the utility of data sources and take into account the unique needs and perspectives of all users. On a broader scale, this work contributes several general recommendations for making the large population datasets—that may soon be available through interoperable systems and health information exchange organizations—useful to public health practice. Finally, the use of mixed methods in this work offers a new perspective for those exploring information needs and attempting to inform the design of public health information systems.

References

1. Institute of Medicine. The Future of Public Health. Washington, DC: National Academy Press; 1988.

2. O'Carroll PW, Yasnoff WA, Ripp LH, Martin EL. Public Health Informatics and Information Systems. 1st ed. New York: Springer; 2003. 824 p.

3. Baxby D. Vaccination: Jenner's Legacy. Berkeley, U.K.: Jenner Museum, Jenner Educational Trust; 1994. 36 p.

4. CDC. Impact of vaccines universally recommended for children--United States, 1990-1998. MMWR Morb Mortal Wkly Rep. 1999;48(12):243-8.

5. Armstrong G, Conn L, Pinner R. Trends in infectious disease mortality in the United States during the 20th century. JAMA. 1999 Jan;281(1):61-6.

6. CDC. Measles --- United States, January--May 20, 2011. MMWR Morb Mortal Wkly Rep. 2011;60(20):666-8.

7. CDC. Pertussis --- California, Januar--June 2010. MMWR Morb Mortal Wkly Rep. 2010;59(26):817.

8. Christensen KLY, Holman RC, Steiner C, Sejvar JJ, Stoll BJ, Schonberger LB. Infectious disease hospitalizations in the United States. Clin Infect Dis 2009 Oct;49(7):1025-35.

9. Yih WK, Lieu TA, Rego VH, O'Brien MA, Shay DK, Yokoe DS, Platt R. Attitudes of healthcare workers in U.S. hospitals regarding smallpox vaccination. BMC Public Health. 2003;3(20)

10. Schultz DG, Mihelic MR, Mihelic FM. A relationship between US healthcare worker smallpox vaccination rates in 2003 and presidential election results in 2004. Am J Disaster Med. 2007 Jul-Aug;2(4):189-194.

11. Chatterjee A, O'Keefe C. Current controversies in the USA regarding vaccine safety. Expert Rev Vaccines. 2010 May;9(5):497-502.

12. Charo RA. Politics, parents, and prophylaxis--mandating HPV vaccination in the United States. N Engl J Med. 2007 May;356(19):1905-08.

13. Centers for Disease Control and Prevention. What are Immunization Information Systems? [Internet]. 2011;[cited 2011 May 18] Available from: http://www.cdc.gov/vaccines/programs/iis/faq.htm

14. CDC. Progress in Immunization Information Systems --- United States, 2009. MMWR Morb Mortal Wkly Rep. 2011;60(01):10-12.

15. CDC. Immunization registry strategic plan 2002-2007 [Internet]. 2005;[cited 2011] Available from: www.cdc.gov/vaccines/programs/iis/activities/downloads/strat-plan.pdf

16. Canavan BC, Kurilo M, Moss T, McLaren R, Berry K, Thomas C, Rasulnia B, Kelly J, Urquhart G. Immunization information systems progress - United States, 2005. MMWR Morb Mortal Wkly Rep. 2006;55(49):1327-1329.

17. Boom JA, Dragsbaek AC, Nelson CS. The success of an immunization information system in the wake of Hurricane Katrina. Pediatrics. 2007 June;119(6):1213-17.

18. Hinman AR, Ross DA. Immunization registries can be building blocks for national health information systems. Health Aff. 2010 Apr;29:676-82.

19. Baker B. CHILD profile: An immunization registry and more--a new approach to tracking and surveillance. Children's Health, Immunization, Linkages and Development. Vaccine. 1998;16(18):v-vi.

20. Riddick S, Warren J. Data Quality Improvement: An Example of Using Consensus-Based Best Practice Guidelines to Improve IIS Operations [Internet]. 43rd National Immunization Conference 2009;[cited 2011 Aug 28] Available from: http://cdc.confex.com/cdc/nic2009/recordingredirect.cgi/id/4532

21. Washington State Department of Health. CHILD Profile Health Promotion Phone Survey Report 2007 [Internet]. 2009;6.[cited 2011 Jun 5] Available from: http://www.doh.wa.gov/cfh/childprofile/documents/2007phonesurvey.pdf

22. Saarlas KN, Hinman AR, Ross DA, C WW, Wild EL, Hastings TM, Richmond PA. All Kids Count 1991-2004: developing information systems to improve child health and the delivery of immunizations and preventive services. J Public Health Manag Pract 2004;Suppl:S3-15.

23. Collins SA, Currie LM, Bakken S, Cimino JJ. Information needs, Infobutton Manager use, and satisfaction by clinician type: a case study. J Am Med Inform Assoc 2008 Jan;16(1):140-2.

24. Randell R, Mitchell N, Thompson C, McCaughan D, Dowding D. From pull to push: understanding nurses' information needs. Health Informatics J 2009 Jun;15(2):75-85.

25. Borycki EM, Lemieux-Charles L, Nagle L, Eysenbach G. Novice nurse information needs in paper and hybrid electronic-paper environments: a qualitative analysis. Stud Health Technol Inform 2009 Jan;150:913-7.

26. Revere D, Turner AM, Madhavan A, Rambo N, Bugni PF, Kimball A, Fuller SS. Understanding the information needs of public health practitioners: a literature review to inform design of an interactive digital knowledge management system. J Biomed Inform 2007;40(4):410-421.

27. Turner AM, Stavri Z, Revere D, Altamore R. From the ground up: information needs of nurses in a rural public health department in Oregon. J Med Libr Assoc 2008 Oct;96(4):335-42.

28. Forsetlund L, Bjorndal A. The potential for research-based information in public health: identifying unrecognised information needs. BMC Public Health. 2001 Jan;1(1)

29. LaPelle NR, Luckmann R, Simpson EH, Martin ER. Identifying strategies to improve access to credible and relevant information for public health professionals: a qualitative study. BMC Public Health. 2006;6:89-102.

30. Twose C, Swartz P, Bunker E, Roderer NK, Oliver KB. Public health practitioners' information access and use patterns in the Maryland (USA) public health departments of Anne Arundel and Wicomico Counties. Health information and libraries journal 2008 Mar;25(1):13-22.

31. Pham MT, Jones AQ, Sargeant JM, Marshall BJ, Dewey CE. A qualitative exploration of the perceptions and information needs of public health inspectors responsible for food safety. BMC Public Health. 2010 Jan;10:345.

32. "Medicare and Medicaid Programs; Electronic Health Record Incentive Program: Final Rule" 75 Federal Register 144 (28 July 2010). pp. 44314-588. 42 CFR Parts 412, 413, 422 et al. GPO Access: edocket.access.gpo.gov/2010/pdf/2010-17207.pdf. 2010.

33. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Morb Mortal Wkly Rep. 2001 Jul;50(RR-13):1-35.

34. Saarlas KN, Edwards K, Wild E, Richmond P. Developing performance measures for immunization registries. J Publ Health Manag Pract. 2003;9(1):47-57.

35. Loeser H, Zvagulis I, Hercz L, Pless IB. The organization and evaluation of a computer-assisted, centralized immunization registry. Am J Public Health 1983 Nov;73(11):1298-301.

36. National Vaccine Advisory Committee. Development of community and state-based immunization registries: report of the National Vaccine Advisory Committee (NVAC). 1999;

37. Khare M, Piccinino L, Barker LE, Linkins RW. Assessment of immunization registry databases as supplemental sources of data to improve ascertainment of vaccination coverage estimates in the national immunization survey. Arch Pediatr Adolesc Med 2006;160(8):838-842.

38. Kolasa MS, Lutz JP, Cofsky A, Jones T. Provider chart audits and outreach to parents: impact in improving childhood immunization coverage and immunization information system completeness. J Public Health Manag Pract 2009;15(6):459-463.

39. White KE, Anderson J, Stanley M, Ehresmann K. Evaluating hepatitis B universal birth dose vaccination at Minnesota birthing hospitals by utilizing immunization information systems, birth certificates, and chart reviews, 2007-2008. J Public Health Manag Pract 2009;15(6):464-470.

40. CDC. Immunization information systems progress--United States, 2006. MMWR Morb Mortal Wkly Rep. 2008;57(11):289-291.

41. Zimmerman LA, Bartlett DL, Enger KS, Gosney K, Williams WG. Influenza vaccination coverage: findings from immunization information systems. BMC Pediatr. 2007;7:28.

42. US Department of Health and Human Services. Healthy People 2010: Understanding and Improving Health. 2000;

43. Wood D, Saarlas KN, Inkelas M, Matyas BT. Immunization registries in the United States: implications for the practice of public health in a changing health care system. Annu Rev Public Health. 1999;20:231-255.

44. CDC. Progress in immunization information systems - United States, 2008. MMWR Morb Mortal Wkly Rep. 2010 Feb;59(5):133-5.

45. Hinman AR, Eichwald J, Linzer D, Saarlas KN. Integrating child health information systems. Am J Public Health. 2005 Nov;95(11):1923-7.

46. CDC. National and state vaccination coverage among children aged 19-35 months--United States, 2010. MMWR Morb Mortal Wkly Rep. 2011;60(34):1157-63.

47. CDC. Vaccination Coverage Among Children in Kindergarten --- United States, 2009--10 School Year. MMWR Morb Mortal Wkly Rep. 2011;60(21):700-4.

48. Immunization of child care and school children against certain vaccine-preventable diseases. WAC 246-105. Washington Administrative Code; 2011.

49. QSR International. NVivo 8, Qualitative Data Analysis Software. 2009;

50. MacQueen KM, McLellan E, Kay K, Milstein B. Codebook Development for Team-Based Qualitative Analysis. Field Methods 1998 May;10(2):31-36.

51. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks: Sage Publications, Inc; 1998. 200 p.

52. Glaser B, Strauss A. The Discovery of Grounded Theory. New York: Aldine; 1967. 271 p.

53. Leckie GJ, Pettigrew KE, Sylvain C. Modeling the information seeking of professionals: A general model derived from research on engineers, health care professionals and lawyers. Library Quarterly 1996;66:161-193.

54. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005 May;37(5):360-3.

55. Krippendorff K. Content analysis : an introduction to its methodology. 2nd ed. Thousand Oaks, Calif.: Sage; 2004. 440 p.

56. Washington State Department of Health. Guidelines for Using Rural-Urban Classification Systems for Public Health Assessment [Internet]. [date unknown];[cited 2011 Nov 4] Available from: http://www.doh.wa.gov/data/guidelines/RuralUrban3.htm

57. Hung PW, Johnson SB, Kaufman DR, Mendonça EA. A multi-level model of information seeking in the clinical domain. Journal of biomedical informatics 2008 May;41(2):357-70.

58. Landry CF. Work roles, tasks, and the information behavior of dentists. Journal of the American Society for Information Science and Technology. 2006 Dec;57(14):1896-1908.

59. Weber S. Critical care nurse practitioners and clinical nurse specialists interface patterns with computer-based decision support systems. Journal of the American Academy of Nurse Practitioners 2007 Nov;19(11):580-90.

60. Crowe S, Tully MP, Cantrill JA. Information in general medical practices: the information processing model. Family practice 2010 Apr;27(2):230-6.

61. Reddy M, Spence PR. Finding answers: information needs of a multidisciplinary patient care team in an emergency department. AMIA Annu Symp Proc. 2006 Jan;649-53.

62. Chambers LW, Haynes RB, Pickering R, McKibbon A, Walker-Dilks CJ, Panton L, Goldblatt E. New approaches to addressing information needs in local public health agencies. Can J Public Health 1991;82(2):109-114.

63. Ryan JL. AZHealthInfo: a collaborative model for supporting the health information needs of public health workers, public librarians, consumers, and communities in Arizona. J Med Libr Assoc 2007;95(3):349-351.

64. Rutland JD, Smith AM. Information needs of the "frontline" public health workforce. Public Health. 2010 Nov;124(11):659-63.

65. Dobbins M, Jack S, Thomas H, Kothari A. Public health decision-makers' informational needs and preferences for receiving research evidence. Worldviews Evid Based Nurs. 2007 Jan;4(3):156-63.

66. Goldstein ND, Maiese BA. A Brief Review of Vaccination Coverage in Immunization Registries. Online Journal of Public Health Informatics .. 2011;3(1):1-7.

67. Clarke KR, Lutz JP. The Philadelphia Story. Ambulatory Pediatrics. 2006;6(1):21-24.

68. Carroll JM. Making Use: Scenario-Based Design of Human-Computer Interactions. The MIT Press; 2009.

69. Miles MB, Huberman AM. Qualitative data analysis : an expanded sourcebook. 2nd ed. Thousand Oaks: Sage Publications; 1994.

70. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform. 2004 Feb;37(1):56-76.

71. CDC. Universally recommended vaccinations: immunization information systems. [Internet]. Guide to Community Preventive Services. 2010;[cited 2011 Jul 15] Available from: www.thecommunityguide.org/vaccines/universally/imminfosystems.html

72. US Department of Health and Human Services. Healthy People 2020 Objectives [Internet]. 2010;[cited 2011 Jul 16] Available from: http://www.healthypeople.gov/2020/topicsobjectives2020/pdfs/Immunization.pdf

73. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc. 2002 Nov-Dec;9(6):600-611.

74. English LP. Total Information Quality Management: a complete methodology for IQ management. DM Review 2003;

75. Power D. Data quality and MDM: a data quality mindset increases MDM program success. Information Management 2009;19(2):40-41.

76. Watts S, Shankaranarayanan G, Even A. Data quality assessment in context: A cognitive perspective. Decision Support Systems 2009;48(1):202-211.

77. Case DO. Looking for information : a survey of research on information seeking, needs, and behavior. 2nd ed. Boston: Elsevier/Academic Press; 2007.

78. Lima CR, Schramm JM, Coeli CM, da Silva ME. Review of data quality dimensions and applied methods in the evaluation of health information systems. Cad. Saúde Pública 2009;25(10):2095-109.

79. Wand Y, Wang RY. Anchoring data quality dimensions in ontological foundations. Communications of the ACM. 1996;39(11):86-95.

80. Turner AM, Petrochilos D, Nelson DE, Allen E, Liddy ED. Access and use of the Internet for health information seeking: a survey of local public health professionals in the northwest. J Public Health Manag Pract. 2009;15(1):67-69.

81. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? Ann Intern Med 1985;103(4):596-599.

82. CDC. IIS: 2001 Minimum Functional Standards for Registries [Internet]. 2011;[cited 2011 Aug 18] Available from: www.cdc.gov/vaccines/programs/iis/stds/min-funct-std-2001.htm

83. National Vaccine Advisory Committee. Immunization Information systems: National Vaccine Advisory Committee (NVAC) Progress Report [Internet]. 2007;38.[cited 2011 Aug 16] Available from: www.hhs.gov/nvpo/nvac/NVACIISReport20070911.doc

84. Williams W, Lowery NE, Lyalin D, Lambrecht N, Riddick S, Sutliff C, Papadouka V. Development and Utilization of Best Practice Operational Guidelines for Immunization Information Systems. J Publ Health Manag Pract. 2011;17(5):449-456.

85. CDC. IISAR Data Participation Rates [Internet]. 2009;[cited 2011 Jul 16] Available from: http://www.cdc.gov/vaccines/programs/iis/rates/data-rates.htm

86. AIRA Modeling of Immunization Registry Operations Working (MIROW) Group (eds). Data Quality Assurance in Immunization Information Systems: Incoming Data [Internet]. American Immunization Registry Association (AIRA) 2008;[cited 2011 Jan 8] Available from: http://www.immregistries.org/pdf/AIRA_MIROW_Chap3_DQA_02112008.pdf

87. Freeman V a, DeFriese GH. The challenge and potential of childhood immunization registries. Annu Rev Public Health. 2003 Jan;24(20):227-46.

88. SAS Institute. SAS 9.2 for Windows. Cary, NC: 2011.

89. CDC. IISAR [Internet]. 2009;[cited 2011 Oct 29] Available from: http://www.cdc.gov/vaccines/programs/iis/rates/default.htm

90. Pallin DJ, Sullivan AF, Espinola JA, Landman AB, Camargo CA. Increasing Adoption of Computerized Provider Order Entry, and Persistent Regional Disparities, in US Emergency Departments. Ann Emerg Med. 2011 Jul;58(6):543-550.

91. Hing E, Hsiao C-J. Electronic medical record use by office-based physicians and their practices: United States, 2007. National Health Statistics Reports. 2010 Mar;(23):1-11.

92. Halamka J, Overhage JM, Ricciardi L, Rishel W, Shirky C, Diamond C. Exchanging health information: local distribution, national coordination. Health Aff. 2005;24(5):1170-9.

93. Shapiro JS, Mostashari F, Hripcsak G, Soulakis N, Kuperman G. Using Health Information Exchange to Improve Public Health. Am J Public Health. 2011 Feb;101(4):616-23.

94. Sahni LC, Boom JA, Patel MM, Baker CJ, Rench MA, Parashar UD, Tate JE. Use of an immunization information system to assess the effectiveness of pentavalent rotavirus vaccine in US children. Vaccine. 2010 Aug;28(38):6314-7.

95. Kroger AT, Sumaya CV, Pickering LK, Atkinson WL. General Recommendations on Immunization. MMWR Morb Mortal Wkly Rep. 2011;60(RR02):1-60.

96. CDC. American Recovery and Reinvestment Act funding announcement for enhancing the interoperability of electronic health records (EHR) and immunization information systems (IIS) [Internet]. Funding Announcement 2011;[cited 2011 Mar 9] Available from: http://www.grants.gov/search/search.do?mode=VIEW&oppId=54435

97. Galloway T. The Use of Computers in Vaccination and Immunization Schemes. Copenhagen: Regional Office WHO; 1974.

98. NISO. Understanding Metadata [Internet]. NISO Press 2004;1-20.[cited 2011 Sep 16] Available from: http://www.niso.org/publications/press/UnderstandingMetadata.pdf

99. Bureau UC. American FactFinder [Internet]. Census 2010 2010;[cited 2011] Available from: http://factfinder2.census.gov/

100. CDC. IIS Interactive Data Query Tool [Internet]. 2011;[cited 2011] Available from: http://www2a.cdc.gov/nip/registry/IISAR/IISAR_Query.asp

101. Colorado Department of Public Health and Environment. Colorado Health Information Dataset [Internet]. 2011;[cited 2011] Available from: http://www.cdphe.state.co.us/cohid/

102. Wild EL, Richmond PA, de Merode L, Smith JD. All Kids Count Connections: a community of practice on integrating child health information systems. J Public Health Manag Pract. 2004 Nov;Suppl:S61-5.

103. Li LC, Grimshaw JM, Nielsen C, Judd M, Coyte PC, Graham ID. Use of communities of practice in business and health care sectors: a systematic review. Implement Sci 2009 Jan;4:27.

104. Diamond CC, Mostashari F, Shirky C. Collecting and sharing data for population health: a new paradigm. Health Affairs. 2009;28(2):454-66.

Appendix A: Interview Guide

Introduction

I am conducting this interview as a part of a project to study the information needs related to immunizations of those working in public health. You have been selected because of your role in public health. Your input is extremely valuable and your time is appreciated. If you agree, I'd like to record this interview; the audio tape will be used to transcribe and code the information gathered from the interview. The information you share will not be connected to your name in any reports, publications or presentations of the research results. All information shared during this interview will be confidential and themes emerging from the analysis of multiple interviews will be presented only in aggregate and will not be associated with your name. Your name will be known only to the interviewer(s). I estimate this interview will take between 40 and 60 minutes. Thanks for participating.

Background questions - current position and training

First, I'm going to ask you a few questions about your background and training.

1. What is your current job title?
2. How long have you been in this position?
3. How long have you been working in this organization?
4. How long have you been working in Public Health?
5. A few questions about your training:
    a. What is your degree in?
    b. What was your field of study?
    c. Have you had any training in public health (certificate, courses through employer, etc.)?

How you do your work – tasks, information needs and information systems

Now I'd like to talk with you about your work and the tools you need to do that work.

6. Can you describe the kind of work you do? [What are your main responsibilities?]

7. What systems or tools do you use to do this work? [databases, Excel spreadsheets, web based tools, email, sending information out, retrieving information]

8. Can you talk about how the work you do is related to immunizations?

Now I'm going to ask you some more specific questions about immunization information that you may use as a part of your job.

9. Have you ever had a situation where you needed to access immunization information?

    a. If yes – Can you walk me through a situation where you have a need for immunization information? [If possible, let's go through the steps you take starting with the trigger that makes you aware that there is a need for the data or information all the way down to the time when you either acquire or decide not to pursue the information.]

    b. If no – Could you imagine a time that you might need to access immunization information for your job? [examples based on previous questions]

10. Have you ever used CHILD Profile or Data and Information from CHILD Profile? [Explain if unfamiliar]

    a. If yes – Can you please describe your experience accessing and using CHILD Profile? [Were there any particular issues you encountered? How did you end up using the data?]

    b. If no – What would you like to know about a system like CHILD Profile to decide if it would be worthwhile for you to access its data/information? [use one of the systems that they do use as an example]

11. How does immunization information help inform decisions in your daily work? [Is CHILD Profile ever a part of these decisions, could it ever be a part of the decision making process?]

12. What is your favorite system or tool, and why? [What makes it a trusted source? A useful source?]

13. Now, thinking about your favorite system, how does it compare with CHILD Profile?

14. Have you ever experienced difficulties when using or trying to use immunization data? [administrative, technical, data quality, etc. specifically, fields missing, up-to-date status, completeness, consistency, timeliness]

Thank you!

Appendix B: Codebook

Project Title: Understanding the Relationship between Information Needs and the Characteristics of Population Data Sources: An Immunization Information System Case Study

Version 3 - Date: 2/2/2011

Document Description: The following is a code book for the thematic and descriptive coding of interviews to identify the information needs and perceptions about data source characteristics of public health practitioners related to immunization data and information. Unstructured interviews focused on the subjects' daily work, how they interact with immunization data and information and how their information needs are met or not met by the current systems in use.

From project proposal:

[We] will use content analysis techniques as described by Krippendorff to identify emergent themes related to immunization information needs of the interviewees. We will focus on the individual tasks performed and the identification and prioritization of information needs related to data and characteristics of data sources such as: access, availability, quality, reliability, and timeliness, as suggested by Revere's review article. We will also consider the factors identified as important by Leckie et al.: accessibility, cost, familiarity, packaging, and trustworthiness. In addition, we will identify and code data uses such as: program evaluation, case-investigation, and vaccine administration. This analytic component will enable us to characterize the perceived and realized uses for immunization data within each of the work-roles or positions of the individuals interviewed.

These codes were developed by Rebecca Hills, Debra Revere and Blaine Reeder in the summer of 2010.

Primary coding questions

What are the information needs of public health practitioners related to immunizations?

Secondary coding questions

What are the tasks performed by public health professionals where immunization information is important?

Which data or information source characteristics are important to public health professionals when seeking information related to immunization?

Instructions for the application of free nodes or tree nodes in NVIVO

1. Start at the beginning of a sentence if possible.
2. Code full sentences or the largest possible block of text.
3. Exclusive coding is not necessary – more than one code can be applied to the same block of text/utterance.
4. Apply codes as best as you are able according to these guidelines. If problems with existing codes are detected, please make a note using "memo" in NVIVO for use during reconciliation.
5. During individual coding, if a new code is necessary, code the utterance using a generic "new code" code. The new code can be added during reconciliation.
6. If codes seem to need renaming, please make a note of the code, the change you suggest to the code name, and the reasons you think renaming is necessary. Codes can be renamed during reconciliation.

Case Attributes (represented as attributes of cases in NVIVO)

1. **Interviewee characteristics**
    a. **Role**
        i. **Code: Nurse/Nursing Supervisor**
            Definition: a public health practitioner trained in nursing who has direct patient or client contact and provides services consistent with the nursing profession. Also, public health practitioners trained in nursing who supervise a staff of other nurses
        ii. **Code: Immunization program director/manager**
            Definition: a public health practitioner working in an administrative or hybrid role, closely involved with or directing the immunization program. The majority of this person's duties involve non-clinical task, although some clinical tasks may be performed
        iii. **Code: Epidemiologist**
            Definition: a public health practitioner working primarily in the field of epidemiology, tracking and studying the health of populations and investigating outbreaks of disease
        iv. **Code: Senior Health Administrator**
            Definition: a public health practitioner in a high-level administrative role
        v. **Code: Immunization program employee**
            Definition: a public health practitioner working under an Immunization Program Director or Manager
        vi. **Code: Other - Manager**
            Definition: a public health practitioner in a managerial role such as "Public Health Services Manager" that is not dealing primarily with the immunization program

    b. **Time in Public Health**
        i. **Code: New to public health (worked in public health for less than 5 years)**
        ii. **Code: Experienced in public health (worked in public health for 5 years or more)**
    c. **Supervisory role**
        i. **Code: Yes – supervising one or more employees**
        ii. **Code: No – not supervising any employees**
    d. **Training**
        Highest level of training completed
        i. **Code: LPN – Licensed Practical Nurse**
        ii. **Code: RN – Registered Nurse**

iii. Code: BSN – Bachelor of Science in Nursing

iv. Code: MPH/MSPH – Master's in Public Health/Master's of Science in Public Health

v. Code: PhD – Doctoral Degree

vi. Code: BA/BS – Bachelor of Arts or Bachelor of Science degree

vii. Code: Other training

e. Organization Type

i. Code: Individual employed by a Local Health Jurisdiction

ii. Code: Individual employed by the State Health department

f. Child Profile Connection

i. Code: HL7 Unidirectional

ii. Code: HL7 Bidirectional

iii. Code: Batch – regular data transfers of sets of records from one system to another (CHILD Profile)

iv. Code: Web entry

v. Code: Connected-other

vi. Code: No connection

g. Rural_Urban – code for the county in which the interviewee is currently working (for LHDs only)

i. Code: Rural (population density < 100/square mile)

ii. Code: Urban (population density >=100/square mile)

Codes (represented as tree nodes in NVIVO)

1. **Barriers and Facilitators**
    a. **Code: Barriers**
        1. Definition: Any situation or artifact that prevents work from being done or goals from being accomplished. Code all barriers that are related to or barriers to completing immunization work.
        2. Cues: Participant uses words such as CAN'T, WON'T, COULDN'T, and other negators when referring to tasks or work goals.
        3. Example: "I can't get the reports I need"
        4. Do not code: *none available*

    b. **Code: Facilitator**
        1. Definition: Any situation or artifact that helps, allows, or is necessary for work activities and tasks. Code all facilitators that are related immunization work.
        2. Cues: Participant uses phrases such as: "it's nice that... "we couldn't do it without..." "we have to have..."
        3. Example: *none available*
        4. Do not code: *none available*

    c. **Code: Workaround**
        1. Definition: A participant describes dealing with a barrier in a creative way to complete their work.
        2. Cues: Participant uses phrases to describe a barrier, but also the way around the barrier.
        3. Example: "We can't get at it directly ourselves so we just use a different login"
        4. Do not code: *none available*

    d. **Code: Wouldn't it be Nice (WIBN)**
        1. Definition: A participant describes something they would like to be able to do, but are not currently able to do.
        2. Cues: Participant uses phrases such as: "it would be nice if..." "I wish I could..." "if only we could just..."
        3. Example: "It would be really cool if..."
        4. Do not code: *none available*

2. **Data or Information Source Characteristics**
    i. **Code: Accessibility**
        1. Definition: The degree to which the data or information in a system are accessible to the user, i.e., difficulty of retrieval and use of information.

   2. Cues: Participant describes difficulty or ease of gaining access to the data source.
   3. Example: I don't have a username for that system.
   4. Do not code: *none available*

ii. **Code: Bias**
   1. Definition: A systematic error or difference in measurement or collection of data.
   2. Cues: Participant describes on observed or potential bias in the data, participant expresses concern over biases
   3. Example: "If you have a certain group of providers who are connected, and then another group who are not, what if they are different in some way?"
   4. Do not code: *none available*

iii. **Code: Data Linkages**
   1. Definition: Data linked or capable of being linked between datasets
   2. Cues: Participant describes difficulty or ease of gaining access to the linked data. Participant describes linkages or lack of linkages
   3. Example: "They aren't linked, right now."
   4. Do not code: *none available*

iv. **Code: Description of Data Collection**
   1. Definition: A description of how the data were collected and put into the system
   2. Cues: Participant refers to the process used to collect and input data either by the provider, public health or other stakeholder
   3. Example: "I don't know if they are just going with what the child says, or if they have some backup data source they are looking at."
   4. Do not code as: *none available*

v. ***Code: Granularity***
   1. Definition: The degree of detail or precision contained in the data. Most often referring to the precision of the geographic location of a record.
   2. Cues: Participant refers to a lack of precision or mismatch in level of precision including inability to answer questions with the data because of this lack of precision.
   3. Example: "We can't say because it doesn't get that fine grained"

4. Do not code as: field data granularity, i.e, how the field is broken up: address vs. street, city, state, zip

vi. **Code: Reputation**
    1. Definition: The reputation of a dataset. Unofficial communication between practitioners about use of a dataset or system.
    2. Cues: Participant indicates that he or she has heard via unofficial channels about a dataset or system.
    3. Example: none available
    4. Do not code: *none available*

vii. **Code: Usefulness**
    1. Definition: The perceived usefulness of the immunization data source. This is a broad code and could apply to any aspect of then interviewees work or to perceived usefulness to persons other than the interviewee.
    2. Cues: Participant refers to usefulness of the data or discusses ways the data could be used to facilitate work.
    3. Example: "could be used for allocation of funds"
    4. Do not code: *none available*

viii. **Code: Prior success**
    1. Definition: Previous experience with a dataset or system
    2. Cues: Participant refers to a past experience that is influencing their current information seeking behavior or perceptions of the system or dataset.
    3. Example:
    4. Do not code: *none available*

ix. **Code: Data format**
    1. Definition: Data format includes data type, file format and representation of the data on a dataset and individual variable level.
    2. Cues: Participant refers to coding of variables or data type as a significant issue.
    3. Example:
    4. Do not code: *none available*

x. **Data quality**
    1. **Code: Timeliness**

1. Definition: The degree to which data are available when needed.
2. Cues: Participant refers to data either being available in a timely manner or unavailable when needed.
3. Example: "the data are too old" "takes too long to get into the system"
4. Do not code: *none available*

2. **Completeness**
   a. **Code: Population coverage**
      1. Definition: The degree to which the data or information in the system represent the population of interest, i.e., how well does the data in this system represent the population? (a complete dataset would contain a record for every individual in the population)
      2. Cues: Participant refers to data being not representative of the population.
      3. Example: "coverage rate" "missing children" "provider participation" "participation rate"
      4. Do not code: variable completeness

   b. **Code: Individual record completeness**
      1. Definition: The degree to which a record contains all variables necessary for a given use. (a complete record would have no missing values)
      2. Cues: Participant refers to either missing demographic or immunization information within an existing record
      3. Example: "gender is not there" "lot number is missing"
      4. Do not code: registration completeness

3. **Code: Accuracy**
      1. Definition: The degree to which data represent true information about the individuals in the population.
      2. Cues: Participant refers to data either being wrong, inaccurate or having poor quality.
      3. Example:
      4. Do not code: *none available*

3. **Data Sources and Systems – Code only the first time the participant mentions an information system by name.**

   NOTE: For most data sources, Cues and Examples are left blank except in exceptional cases. Cues are always "Participant describes the use of the data from or interaction with SYSTEM". Examples of these statements are self-explanatory and easy to identify.

   i. **Code: Behavioral Risk Factor Surveillance System**
      1. Definition: BRFSS
      2. Cues: *NA*
      3. Example: *NA*
      4. Do not code: *none available*

   ii. **Code: Community Health Assessment Tool**
      1. Definition*: CHAT*
      2. Cues: *NA*
      3. Example: *NA*
      4. Do not code*: none available*

   iii. **Code: Child Care Records**
      1. Definition: Use of records from child care providers
      2. Cues: *NA*
      3. Example: *NA*
      4. Do not code: *none available*

   iv. **Code: CHILD Profile**
      1. Definition: Washington's Immunization Registry or Immunization Information System. Either the data coming from the system or the system interface.
      2. Cues: *NA*
      3. Example: "Data from CHILD Profile," "I looked it up in the registry," "Immunization Information System," "IIS"
      4. Do not code: CHILD Profile's educational outreach program (Immunization information sent to parents through the mail.

   v. **Code: CIMS**
      1. Definition: Use of the Client Information Management System (CIMS)
      2. Cues: Participant describes the use of data from or direct use of CIMS.
      3. Example: "CIMS"
      4. Do not code: *none available*

vi. **Code: Healthy Youth Survey**
   1. Definition: Data or information from the Health Youth Survey
   2. Cues: *NA*
   3. Example: *NA*
   4. Do not code: *none available*

vii. **Code: Infolinks**
   1. Definition: Use of the Infolinks System
   2. Cues: Participant describes the use of data from or direct use of the Infolinks System.
   3. Example: "Infolinks"
   4. Do not code: *none available*

viii. **Code: INSIGHT**
   1. Definition: Use of the INSIGHT System
   2. Cues: Participant describes the use of data from or direct use of the INSIGHT System.
   3. Example: "INSIGHT"
   4. Do not code: *none available*

ix. **Code: KIPHS**
   1. Definition: Use of the Kansas Integrated Public Health System (KIPHS)
   2. Cues: *NA*
   3. Example: *NA*
   4. Do not code: *none available*

x. **Code: Local Data**
   1. Definition: Use of locally collected data, survey data, community health assessment data collected by an LHJ
   2. Cues: NA
   3. Example: NA
   4. Do not code: *none available*

xi. **Code: National Immunization Survey**
   1. Definition: Use of data collected by the National Immunization Survey.
   2. Cues: Participant describes the use of data or issues surrounding use of data from the National Immunization Survey
   3. Example: "National Immunization Survey dataset" "NIS Data"
   4. Do not code: *none available*

xii. **Code: OAS Gold**

1. Definition: Use of the OAS Gold System
2. Cues: Participant describes the use of data from or direct use of the OAS Gold System.
3. Example: "OAS Gold"
4. Do not code: *none available*

xiii. **Code: Paper**

1. Definition: Use of paper for recording and tracking information. A system based wholly or partially upon the use of paper.
2. Cues: Participant describes the use of a paper-based system of any kind.
3. Example: "We record that on paper"
4. Do not code: *none available*

xiv. **Code: PHIMS**

1. Definition: Use of the Public Health Issue Management System (PHIMS)
2. Cues: *NA*
3. Example: *NA*
4. Do not code: *none available*

xv. **Code: PHRED**

1. Definition: Use of the Public Health Reporting of Electronic Data (PHRED)
2. Cues: *NA*
3. Example: *NA*
4. Do not code: *none available*

xvi. **Code: PRAMS**

1. Definition: Use of the Pregnancy Risk Assessment Monitoring Survey (PRAMS)
2. Cues: *NA*
3. Example: *NA*
4. Do not code: *none available*

xvii. **Code: School absenteeism data**

1. Definition: Use of the data from schools indicating absenteeism rates
2. Cues: NA
3. Example: *NA*
4. Do not code: *none available*

xviii. **Code: School Exemption Data**
  1. Definition: Use of the data from schools indicating exemption rates for vaccination upon student registration. Data collected by schools and compiled by the state
  2. Cues: NA
  3. Example: *NA*
  4. Do not code: *none available*

xix. **Code: Shoreland TRAVAX**
  1. Definition: Use of the Shoreland TRAVAX system for recording, retrieving and tracking information.
  2. Cues: Participant describes the use of the TRAVAX system.
  3. Example: "We use Shoreland TRAVAX for travel clinic shots."
  4. Do not code: *none available*

xx. **Code: Signature**
  1. Definition: Use of the Signature System
  2. Cues: Participant describes the use of data from or direct use of the Signature System.
  3. Example: "Signature"
  4. Do not code: *none available*

xxi. **Code: US Census Data**
  1. Definition: Use of the data from the United States Census
  2. Cues: NA
  3. Example: NA
  4. Do not code: *none available*

xxii. **Code: VISTA**
  1. Definition: Use of VISTA
  2. Cues: NA
  3. Example: NA
  4. Do not code: *none available*

xxiii. **Code: Vital Statistics Data**
  1. Definition: Use of the Vital Statistics Data (Birth Certificates, Death Certificates)
  2. Cues: NA
  3. Example: NA
  4. Do not code: *none available*

xxiv. **Code: Youth Risk Behavioral Surveillance System**
  1. Definition: Youth Risk Behavioral Surveillance System

        2. Cues: NA
        3. Example: NA
        4. Do not code: *none available*

4. **Information needs**
    a. **Code: AFIX**
        1. Definition: Performing AFIX visits with providers
        2. Cues: *NA*
        3. Example: *NA*
        4. Do not code: *none available*

    b. **Code: Answering Ad-hoc population level questions**
        1. Definition: Examining datasets to answer questions about the population. One-off or ad-hoc questions.
        2. Cues: Participant refers population research, epidemiology, answering general questions or providing rates.
        3. Example: "Calculating coverage rates by county"
        4. Do not code: *Scheduled or unscheduled lookup of patient-level information should not be coded as "Answering Ad-hoc population level questions". Reports or population level views of data used regularly should not be coded here, rather they should be coded as "Regular Reporting"*

    c. **Code: Billing Information**
        1. Definition: Information related to billing
        2. Cues: Participant refers to using, accessing or needing information related to billing
        3. Example: *none available*
        4. Do not code: *none available*

    d. **Code: Consent information**
        1. Definition: Looking up a consent form for immunizations
        2. Cues: Participant refers to a finding a consent form.
        3. Example:
        4. Do not code: *none available*

    e. **Code: Customizable reports**
        1. Definition: Participant expresses the need for unique, Ad Hoc reports
        2. Cues: Participant refers to designing one-off reports, unique reports that have not been not pre-defined in the system.
        3. Example: *none available*

4. Do not code: *Regularly run reports*

f. **Code: Information from and about Schools**
    1. Definition: Information about schools. Exemption data, absenteeism data or other school-related data or information
    2. Cues: *NA*
    3. Example: *NA*
    4. Do not code: *none available*

g. **Code: Mass vaccination information**
    1. Definition: Information needs related to mass vaccination activities such as those taking place for H1N1 and Seasonal Influenza vaccine
    2. Cues: Participant refers to mass vaccination clinic or a special immunization event. H1N1 vaccination events.
    3. Example:
    4. Do not code: *none available*

h. **Code: Patient information**
    1. Definition: Looking up information about an individual patient
    2. Cues: Participant refers to a clinic visit, patient visit, administering immunizations or looking up a patient's information.
    3. Example:
    4. Do not code: *none available*

i. **Code: Provider Assessment/CoCASA information**
    1. Definition: Information needs related to the Comprehensive Clinic Assessment Software Application
    2. Cues: CoCASA, Provider Assessment visits
    3. Example:
    4. Do not code: *none available*

j. **Code: Recall**
    1. Definition: Recall refers to calling a patient in when he or she is due for vaccinations according to the accepted schedule. Recall may take the form of phone calls or mailings
    2. Cues: Participant refers to using the recall feature of CHILD Profile, or of their EMR or another system
    3. Example:
    4. Do not code: *none available*

k. **Code: Regular Reporting**

1. Definition: Creating regular reports related to immunizations. Reports needed and run on a regular basis. Reporting on more than one client, aggregating client records.
2. Cues: Participant refers to reports, running reports, or reporting to internal or external stakeholders who request reports
3. Example: *none available*
4. Do not code: *Ad Hoc reporting should not be coded as Regular Reporting since it is not done on a regular basis, or has not yet been established as a regular part of a practitioner's work*

l. **Code: Travel Immunization Information**

1. Definition: Looking up information besides patient information for a travel visit. This could be country-specific recommendations or other information about the client's travel.
2. Cues:
3. Example:
4. Do not code: *none available*

m. **Code: Vaccine Information**

1. Definition: Information specific to a vaccine including dosing, schedule or side-effects.
2. Cues: Participant refers to looking up vaccine-specific information.
3. Example: NA
4. Do not code: *none available*

n. **Code: Vaccine Inventory Tracking**

1. Definition: Recording or tracking information about inventory in the public health department or in area provider offices.
2. Cues: Participant refers to inventory information, inventory-related tasks or inventory information needs. Especially related to VFC ordering.
3. Example: NA
4. Do not code: *none available*

o. **Code: Laboratory Information**

1. Definition: Information from laboratories, test results on individual patients and aggregate information about groups of patients or the laboratory itself.
2. Cues: Participant describes accessing information from a laboratory or from another source to find out laboratory test results for a patient or population level data from or about laboratories

3. Example: NA
4. Do not code: *none available*

p. **Code: Patient lookup – other**
   1. Definition: Looking up information about an individual patient outside of the time the patient is in the clinic.
   2. Cues: Participant refers to looking for, needing or looking up individual level patient data when the patient is not present in the clinic.
   3. Example:
   4. Do not code: *none available*

5. **Perceptions**
   a. **Code: Perceptions**
      1. Definition: Perceptions or feelings about immunization, a job role, work, the state of public health.
      2. Cues: Participant refers to a situation and describes their perception of that situation. May trigger strong feelings that something needs to change.
      3. Example: "This situation is REALLY bad"
      4. Do not code: *none available*

6. **Tasks**
   a. **Code: Tasks**
      1. Definition: Performing a task as a part of one's work duties. Code all tasks related to immunization work.
      2. Cues: Participant refers to completion of a task
      3. Example: Then I have to make all of the phone calls for the next day's vaccinations
      4. Do not code: *none available*
   ii. **Code: Provider Training**
      1. Definition: Training providers to use a system
      2. Cues: Participant refers to visits or phone calls to train providers
      3. Example: "We tag team the provider training"
      4. Do not code: *none available*

Memos and annotations - Use memos or annotations to document exceptional quotes, either coded or un-coded that may be useful for describing the work. Please add a brief description or reason for the annotation in the annotation or memo text field.

Appendix C: SAS Code

```
/* PROGRAM 1 */
/* This program imports all data from two pipe-delimited files provided by CPIR programmers
into SAS datasets.
Updates some records with a new dataset (birth records only) but not complete dataset
provided by CPIR.
Performs several data consistncy checks and allows for visual checking using PROC FREQ */

/* ASSIGN LIBRARY*/
libname SASData 'C:\Users\hillsr\Documents\BHI by Project\Dissertation\Aim2\Data\SASData\';

/* IMPORT PATIENT RECORDS FROM PATIENT_NOHIST (THIRD TRY WITH HISTORICALS DELETED) PIPE
DELIMITED FILE ALLOWING FORM ISSINGS IF "||" */
data SASData.patient_import_nohist_pre;
infile 'C:\Users\hillsr\Documents\BHI by Project\Dissertation\Aim
2\Data\Dataset_1\NoHist\patient_nohist.txt' dlm='|' DSD;
input SIISID :$10. FirstName MiddleName LastName Address City State Zip Phone SSN DOB Sex
ZipThree
DOB_Year MedOrg :$100. Facility :$100. Trans :$20.;
run;

/* IMPORT BIRTH ONLY PATIENT RECORDS FROM PATIENT_NOHIST(Fourth try with only BIRTH records)
PIPE DELIMITED FILE ALLOWING FOR MISSING IF "||" */
data SASData.patient_import_birth_08092011 ;
infile 'C:\Users\hillsr\Documents\BHI by Project\Dissertation\Aim
2\Data\patient_birth_only_08092011.txt' dlm='|' DSD ;
input SIISID :$10. FirstName MiddleName LastName Address City State Zip Phone SSN DOB Sex
ZipThree
DOB_Year MedOrg :$100. Facility :$100. Trans :$20.;
run;

/*SORT DATASETS BEFORE UPDATE*/
proc sort data = SASData.patient_import_nohist_pre;
by SIISID;
run;

data patient_import_birth_08092011;
set SASData.patient_import_birth_08092011;
run;

proc sort data = patient_import_birth_08092011;
by SIISID;
run;

/* USE UPDATE STATEMENT TO UPDATE THE PATIENT RECORD FIELDS (ESP TRANS) WITH VALUES FROM THE
BIRTH DATASET */
data SASData.patient_import_nohist;
update SASData.patient_import_nohist_pre patient_import_birth_08092011;
by SIISID;
run;

/* IMPORT VACCINATION RECORDS (THRID TRY WITH HISTORICALS DELETED) PIPE DELIMITED FILE
ALLOWING FOR MISSINGS IF "||" */
data SASData.vaccine_import_nohist;
infile 'C:\Users\hillsr\Documents\BHI by Project\Dissertation\Aim
2\Data\Dataset_1\NoHist\vaccines_nohist.txt' dlm='|' DSD;
input SIISID :$10. EntryDate :mmddyy10. ServiceDate :mmddyy10. CPTCode ImmDesc :$50. MedOrg
:$50. Facility :$50. Trans :$20.;
run;

/* WORK IN TEMP DATASET*/
/* REDUCE THE VACCINE DATASET TO ONLY SERVICEDATE IN 2010*/
/* DIFFERENCE HERE WILL BE THE NUMBER OF RECORDS EXLUDED BECAUSE SERVICE DATE WAS NOT IN
2010*/
data vaccine_import_nohist_2010;
set SASData.vaccine_import_nohist;
if Year(ServiceDate) = 2010;
run;
```

```
/*REDUCE THE VACCINE DATASET, REMOVE TRANS = 'BIRTH' RECORDS*/
/*DIFFERENCE HERE WILL BE THE NUMBER OF RECORDS EXLUDED BECAUSE TRANS=BIRTH*/
data vaccine_import_nohist_2010;
set vaccine_import_nohist_2010;
if trans NE 'BIRTH';
run;
/*REDUCE THE VACCINE DATSET - REMOVE RECORDS WHERE ENTRYDATE WAS BEFORE SERVICEDATE
(ENTRYDATE<SERVICEDATE) */
/*DIFFERENCE HERE WILL BE THE NUMBER OF RECORDS EXLUDED BCS ENTRYDATE WAS BEFORE
SERVICEDATE*/
data SASData.vaccine_import_nohist_2010;
set vaccine_import_nohist_2010;
if EntryDate GE ServiceDate;
run;

proc contents data = SASData.vaccine_import_nohist_2010;
run;
/* PRINT THE FIRST 10 RECORDS OF VACCINE IMPORT NOHIST */
proc print data=SASData.vaccine_import_nohist_2010 (obs=10);
run;

/*RUN THESE TO GET THE DIFFERENCE IN NUMBERS OF RECORDS WHEN HISTORICALS ARE REMOVED*/
proc contents data= SASData.patient_import;
run;
proc contents data= SASData.patient_import_nohist;
run;
proc contents data= SASData.vaccine_import;
run;
proc contents data= SASData.vaccine_import_nohist_2010;
run;

/*CHECK TO MAKE SURE THE DATA MATCH THE CONSTRAINTS OF THE DATA REQUEST*/
/*MAKE SURE DOB IS BETWEEN 1991 AND 2010*/
proc means data= SASData.patient_import_nohist;
var DOB_Year;
run;

proc tabulate data=SASData.vaccine_import_nohist_2010;
var ServiceDate;
table min max, ServiceDate * f=monyy. ;
run;

proc tabulate data=SASData.vaccine_import_nohist_2010;
var EntryDate;
table min max, EntryDate * f=monyy. ;
run;

PROC TABULATE data=SASData.vaccine_import_nohist_2010;
CLASS ServiceDate;
FORMAT ServiceDate monyy.;
TABLE ServiceDate;
RUN;

PROC TABULATE data=SASData.vaccine_import_nohist_2010;
CLASS EntryDate;
FORMAT EntryDate monyy.;
TABLE EntryDate;
RUN;

/*PRINT THE FIRST 10 RECORDS OF VACCINE IMPORT NOHIST 2010*/
proc print data=SASData.vaccine_import_nohist_2010 (obs=10);
var SIISID EntryDate ServiceDate ImmDesc ;
format EntryDate MMDDYY8. ServiceDate MMDDYY8.;
run;

quit;
/*USE  SASData.vaccine_import_nohist_2010 and SASData.patient_import_nohist*/
```

```
/* PROGRAM 2 */

/* ASSIGN LIBRARY */
libname SASData 'C:\Users\hillsr\Documents\BHI by Project\Dissertation\Aim2\Data\SASData\';


/*The following code manipulates the existing patient and vaccine*/
*/datasets, merges them, adds some  and produces some descriptive statistics*/
/*working with two datasets from SASData directory: patient_import_nohist.sas and
vaccine_import_nohist_2010.sas */
/*bring these two datasets into work directories*/

/* READ IN SAMPLE */
data patient_work;
set SASData.patient_import_nohist;
run;

/*ADD "_V" TO FACILITY, MEDORG, AND TRANS, DROP ORIGINAL VARIABLES*/
data vaccine_work;
set SASData.vaccine_import_nohist_2010;
Facility_V = Facility;
MedOrg_V=MedOrg;
Trans_V=Trans;
Drop Facility Medorg Trans;
run;

/*DATA CHECKS PROC PRINT PROC CONTENTS*/
proc print data = patient_work (obs=10);
run;
proc contents data=patient_work;
run;
proc contents data=vaccine_work;
run;

/*Work with the patient dataset -- add fields 0/1 for MedOrg and Facility for completeness
calculations*/
data patient_work2;
set patient_work;

/*Create MedOrgBool for 0/1 indication for MedOrg field*/
if MedOrg="" then MedOrgBool=0;
else if MedOrg=" " then MedOrgBool=0;
else MedOrgBool=1;

/*Create FacilityBool for 0/1 indication for Facility field*/
if Facility="" then FacilityBool=0;
else if Facility=" " then FacilityBool=0;
else FacilityBool=1;

/*CREATE TRANSBOOL FOR 0/1 INDICATION FOR TRANS FIELD*/
if Trans="" then TransBool=0;
else if Trans=" " then TransBool=0;
else TransBool=1;
run;

/*SORT THE PATIENT DATASET*/
proc sort data = patient_work2;
by Trans;
run;

/*CREATE A SERIES OF BOOLEAN INDICATOR VARIABLES FOR ENTRYDATE, SERVICEDATE, FACILITY, ETC.
USE " V" FOR FIELDS REPEATED FROM PATIENT RECORDS*/
data vaccine_work2;
set vaccine_work;

/*CREATE 0/1 INDICATION FOR ALL FIELDS*/
if EntryDate=. then EntryDateBool=0;
else EntryDateBool=1;

if ServiceDate=. then ServiceDateBool=0;
else ServiceDateBool=1;
```

```
if MedOrg_V="" then MedOrgBool_V=0;
else if MedOrg_V=" " then MedOrgBool_V=0;
else MedOrgBool_V=1;

if Facility_V="" then FacilityBool_V=0;
else if Facility_V=" " then FacilityBool_V=0;
else FacilityBool_V=1;

if ImmDesc="" then ImmDescBool=0;
else if ImmDesc=" " then ImmDescBool=0;
else ImmDescBool=1;

if Trans_V="" then TransBool_V=0;
else if Trans_V=" " then TransBool_V=0;
else TransBool_V=1;

if CPTCode=. then CPTCodeBool=0;
else CPTCodeBool=1;

run;

/*create table for proc freq VACCINES (this one will have hl7, batch, and manual)*/
/*make sure trans doesn't have any missing values*/
proc freq data=vaccine_work2;
title ' check to see if there are any missings in the trans field';
tables trans_V;
run;

/*link patients to vaccines with a merge*/
/*first sort both by SIISID, then merge and separate records from one but not both input
datasets*/
proc sort data=patient_work2;
by SIISID;
run;
proc sort data=vaccine_work2;
by SIISID;
run;

data CP_merged from_patient from_vaccine;
merge patient_work2 (IN=patient) vaccine_work2 (IN=vaccine);
by SIISID;
if patient=1 and vaccine=1 then output CP_merged;
else if patient=1 and vaccine=0 then output from_patient;
else if patient=0 and vaccine=1 then output from_vaccine;
run;

proc print data=cp_merged (obs=10);
   title1 'TEST PRINT: 10 records from cp_merged (merge)';
   title2 'records in both';
run;
proc print data=from_patient (obs=10);
   title1 'TEST PRINT: 10 records from from_patient (merge)';
   title2 'Records in patient only';
run;
proc print data=from_vaccine (obs=10);
   title1 'TEST PRINT: 10 records from from_vaccine (merge)';
   title2 'Records in vaccine only';
run;
/*end merge*/

/*Create permanent SAS dataset with final merged data (final_merged)*/
data SASData.final_merged;
set cp_merged;
run;

/*create a dataset with just the demographic record for each SIISID that has a record in the
vaccine file*/

data SasData.CP_DISTINCT_ID ;
merge patient_work2 (IN=patient) vaccine_work2 (IN=vaccine);
```

```
by SIISID;
if patient=1 and vaccine=1 and (SIISID NE SIISID_Last) then output SasData.CP_DISTINCT_ID;
SIISID_Last = SIISID;
RETAIN SIISID_Last ;
run;

proc print data=SasData.CP_DISTINCT_ID (obs=10);
   title 'TEST PRINT: 10 records from CP_DISTINCT_ID - one record per merged SIISID';
run;

PROC CONTENTS data = SasData.CP_DISTINCT_ID;
title 'check this against the proc contents below for CP_Distinct_ID2 is it the same?'
run;

/*calculate two additional completeness fields: demog_compl_tot (total demog completeness)*/
/*and demog_compl_min (completeness for the fields contributing to the minimum dataset)*/
DATA sasdata.cp_distinct_id;
set sasdata.cp_distinct_id;
demog_compl_tot = FirstName + MiddleName + LastName + Address + City + State + Zip + Phone +
SSN + DOB + Sex+ MedOrgBool + FacilityBool;
pct_demog_flds= demog_compl_tot/13;
demog_compl_min = FirstName + LastName + DOB;
if demog_compl_tot in(4 5 6 7 8 )then completecat=1;
        else
                if demog_compl_tot in (9 10 11) then completecat=2;
                else
                        if demog_compl_tot in (12 13) then completecat=3;
                                else completecat=0;
run;

/*calculate  additional completeness fields for the imms part of the records: immdesc */
/*calculate  additional completeness fields for the TOTAL record:  */
quit;
DATA Final_merged;
set SASData.Final_merged;
run;
DATA Final_merged;
set Final_merged;
/*MedOrgBool_V is always 1 so I commented out this calculation*/
/*If MedOrgBool_V or FacilityBool_V then Org = 1;*/
ImmDesc_txt=ImmDesc;
If ImmDescBool or CPTCodeBool then ImmDesc = 1;
imm_compl_min = ServiceDateBool + MedOrgBool + ImmDesc;
imm_compl_tot = EntryDateBool + ServiceDateBool + MedOrgBool_V + FacilityBool_V + ImmDescBool
+ CPTCodeBool;
total_compl = FirstName + MiddleName + LastName + Address + City + State + Zip + Phone + SSN
+ DOB + Sex+ MedOrgBool + FacilityBool +
                                EntryDateBool + ServiceDateBool + MedOrgBool_V + FacilityBool_V
+ ImmDescBool + CPTCodeBool;
total_compl_min = ServiceDateBool + MedOrgBool + ImmDesc + FirstName + LastName + DOB;
num_dem_flds = FirstName + MiddleName + LastName + Address + City + State + Zip + Phone + SSN
+ DOB + Sex+ MedOrgBool + FacilityBool;
num_imm_flds = EntryDateBool + ServiceDateBool + MedOrgBool_V + FacilityBool_V + ImmDescBool
+ CPTCodeBool;
prct_imm_flds = num_imm_flds/6;
run;
quit;
DATA SASData.Final_merged;
set Final_merged;
run;

data SASData.final_merged;
        set SASData.final_merged;
        if num_dem_flds in(4 5 6 7 8 )then completecat=1;
        else
                if num_dem_flds in (9 10 11) then completecat=2;
                else
                        if num_dem_flds in (12 13) then completecat=3;
                                else completecat=0;
run;
```

```
/*NODUPKEY WILL ELIMATE ANY OPSERVATIONS THAT HAVE THE SAME VALUES FOR THE BY VARIABLES*/
/*CONFIRMED THAT BOTH CP_DISTINCT_ID AND CP_DISTINCT_ID2 HAVE 757290 RECORDS*/

proc sort data=SASData.final_merged out=SASData.providers NODUPKEY;
BY MedOrg;
RUN;

proc sort data=SASData.final_merged out=SASData.facility NODUPKEY;
BY Facility;
RUN;

proc sort data=SASData.final_merged out=SASData.facility NODUPKEY;
BY Facility trans_v;
RUN;

proc sort data=SASData.final_merged out=SASData.facility_1 NODUPKEY;
BY Facility;
RUN;

proc print data = SASData.facility;
var facility trans_v serviceDate;
run;

proc print data = sasdata.facility ;
var MedOrg Facility trans;
run;

proc print data = sasdata.providers ;
var SIISID MedOrg Facility trans;
run;
proc contents data= sasdata.providers;
run;
```

```
/* PROGRAM 3 */
/*THIS PROGRAM RUNS DESCRIPTIVE STATS ON THE MERGED SAS DATASETS*/

/*ASSIGN LIBRARY*/
libname SASData 'C:\Users\hillsr\Documents\BHI by Project\Dissertation\Aim2\Data\SASData\';
proc format;
value age 1991-2003 = '1991-2003'
                       2004-2007 = '2004-2007'
                       2008-2010 = '2008-2010';
run;

/*USE SASData.Final_merged and SasData.CP_DISTINCT_ID(2) /
proc contents data=SASData.final_merged;
title 'number of observations here is the number of vaccinations';
run;
proc contents data=SasData.CP_DISTINCT_ID;
title 'number of observations here is the number of demog records';
run;

/*proc freq to determine  min max and mean number of imms records per patient record*/
/*used 'ODS TRACE ON' AND 'ODS TRACE OFF' to determine output data name: OneWayFreqs*/
/*run proc FREQ on the merged data, output to FreqOutput*/
/*close output window*/
ods _all_ close;
proc freq data=sasdata.final_merged ;
tables SIISID;

/*OPEN FREQOUTPUT FOR OUTPUT FROM PROC FREQ*/
ODS OUTPUT OneWayFreqs = FreqOutput;
run;

/*RE-OPEN OUTPUT WINDOW*/
ods listing;

/*run proc means on frequency output to get the min max and mean number of imms per patient
record*/
proc means data=freqoutput;
var Frequency;
   title 'Means on the Frequency Output data to get Min, Max and Mean number of imms per
patient record';
run;

proc freq data = sasdata.cp_distinct_id;
format dob_year age.;
tables  dob_year ;
title1 'Age Distribution of total sample';
run;

proc freq data = sasdata.final_merged;
tables  trans_v;
title1 'Transmission type Distribution of total sample (all vaccination records)';
run;
/* PROGRAM 4 */

/* ASSIGN LIBRARY */
libname SASData 'C:\Users\hillsr\Documents\BHI by Project\Dissertation\Aim2\Data\SASData\';


/*create the format for immunization timeliness categories*/
proc format;
value ImmCatF 0 = 'negative days'
                         1 = '<7 days'
                         2= '7-30 days'
                         3='>30 days';
value ImmLogF 0 = '<=30 days'
                         1 = '>30 days';
run;

/*USE SASData.Final_merged and SasData.CP_DISTINCT_ID */
/*bring these datasets into work directory*/
```

```
data final_merged;
set SASData.Final_merged;
run;

data CP_DISTINCT_ID;
set SASData.CP_DISTINCT_ID;
run;

/*WORK WITH THE MAIN DATASET -- CALCULATE IMTM (IMMUNIZATION LEVEL TIMELINESS) AND RELATED
CATEGORICAL FIELDS*/
data final_merged;
set final_merged;
/*CREATE IMTM AND IMTMCAT*/
ImTM=EntryDate - ServiceDate;
If ImTM<0 then ImTMCat=0;
Else If ImTM<=7 then ImTMCat=1;
Else If ImTM<=30 then ImTMCat=2;
Else ImTMCat=3;

/*CREATE ImTMLog AND TRANS FOR THE LOGISTIC REGRESSION*/
/*IMTMLOG = 0 WILL INDICATE A RECORD WHERE IMTM<=30 DAYS
  IMTMLOG = 1 WILL INDICATE A RECORD WHERE IMTM >30 DAYS */

If ImTM<=30 then ImTMLog = 0;
Else ImTMLog = 1;

If trans_v = "BATCH" THEN do;
               batch = 1;
               trans_l=3;
       end;
       else batch = 0;
If trans_v = "MANUAL ENTRY" then do;
               manual = 1;
               trans_l=1;
       end;
       else manual = 0;
If trans_v = "HL7" then do;
               HL7=1;
               trans_l=2;
       end;
       else HL7=0;

run;

proc freq data=final_merged;
tables ImTMCat;
title 'Make sure there are no IMTMCat = 0/negative dates';
run;

/*create a new dataset with one record per visit per patient */
proc sort data=final_merged;
by SIISID ServiceDate EntryDate MedOrg Facility;
run;

data visit_level;
set final_merged;
If SIISID = SIISID_L and ServiceDate = ServiceDate_L and EntryDate = EntryDate_L and
MedOrg=MedOrg_L and Facility = Facility_L
then delete;
else do;
       output visit_level;
       SIISID_L = SIISID;
       ServiceDate_L = ServiceDate;
       EntryDate_L = EntryDate;
       MedOrg_L=MedOrg;
       Facility_L = Facility;
end;
retain SIISID_L ServiceDate_L EntryDate_L MedOrg_L Facility_L;
run;
quit;
```

```
proc contents data=visit_level;
title 'number of observations here is the number of visit-level records';
run;
quit;

proc sort data=final_merged;
by trans_V;
run;

/*Reports for All immunization records*/
/*run proc means on the ImTM field*/
proc means data=final_merged;
var ImTM;
title 'Display means of timeliness measures total - FOR INDIVIDUAL LEVEL RECORDS';
run;
quit;

/*examine the ImTM field*/
proc means data=final_merged;
var ImTM;
by trans_V;
title 'Display means of timeliness measures by transfer method - FOR INDIVIDUAL LEVEL
RECORDS';
run;

proc freq data=final_merged;
format ImTMCat ImmCatF.;
tables ImTMCat;
title 'Displays frequencies for the timeliness categories for the entire sample- FOR
INDIVIDUAL LEVEL RECORDS';
run;

/*run the same reports on visit level (visit_level) as were run on the individual level*/
proc sort data=visit_level;
by trans_V;
run;

/*run proc means on the ImTM field*/
proc means data=visit_level;
var ImTM;
title 'Display means of timeliness measures TOTAL - FOR VISIT LEVEL RECORDS';
run;

proc means data=visit_level;
var ImTM;
by trans_V;
title 'Display means of timeliness measures BY DATA TRANSFER METHOD - FOR VISIT LEVEL
RECORDS';
run;

proc freq data=visit_level;
format ImTMCat ImmCatF.;
tables ImTMCat;
title 'Displays frequencies for the timeliness categories for the entire sample - FOR VISIT
LEVEL RECORDS';
run;

proc freq data=visit_level;
format ImTMCat ImmCatF.;
tables ImTMCat;
title 'Display percentages for each of the Timeliness categories for all records BY TRANS
Category- FOR VISIT LEVEL RECORDS';
by trans_V;
run;

proc freq data=final_merged;
format ImTMCat ImmCatF.;
tables ImTMCat;
title 'Display percentages for each of the Timeliness categories for all records- FOR
INDIVIDUAL LEVEL RECORDS';
by trans_V;
```

```
run;

/*CHISQ using only visit level data, perform chisquare test on trans x timeliness category*/
proc freq data=visit_level;
format ImTMCat ImmCatF.;
tables ImTMCat * TRans_V/CHISQ;
title 'chi square test for timeliness category * transmission mode - FOR VISIT LEVEL
RECORDS';
run;

/*SORT BOTH WORK DATASETS*/
proc sort data=final_merged;
by Trans_V;
run;

proc sort data=visit_level;
by Trans_V;
run;

/*HISTOGRAMS*/
/*VACCINATION LEVEL*/
proc univariate data=final_merged noprint;
      var ImTM;
      histogram / cfill=ltgray
               endpoints=0 to 570 by 1 noframe
               vscale=percent
               vaxisLabel="Percent";
               Label ImTM = "Timeliness Measure";
               title 'All records - Vaccination';
 run;

proc univariate data=final_merged noprint;
      var ImTM;
      histogram / cfill=ltgray
               endpoints=0 to 570 by 1 noframe
               vscale=percent
               vaxisLabel="Percent";
               where Trans_V="MANUAL ENTRY";
               Label ImTM = "Timeliness Measure";
               title 'Manual Entry - Vaccination';
 run;
proc univariate data=final_merged noprint;
      var ImTM;
      histogram / cfill=ltgray
               endpoints=0 to 570 by 1 noframe
               vscale=percent
               vaxisLabel="Percent";
               where Trans_V="BATCH";
               Label ImTM = "Timeliness Measure";
               title 'Batch - Vaccination';
 run;

 proc univariate data=final_merged noprint;
      var ImTM;
      histogram / cfill=ltgray
               endpoints=0 to 570 by 1 noframe
               vscale=percent
               vaxisLabel="Percent";
               where Trans_V="HL7";
               Label ImTM = "Timeliness Measure - Vaccination";
               title 'HL7';
 run;

 /*VISIT LEVEL HISTOGRAMS*/
proc univariate data=visit_level noprint;
      var ImTM;
      histogram / cfill=ltgray
               endpoints=0 to 570 by 1 noframe
               vscale=percent
               vaxisLabel="Percent";
               Label ImTM = "Timeliness Measure";
```

```
               title 'All records - Visit';
 run;

proc univariate data=visit_level noprint;
      var ImTM;
      histogram / cfill=ltgray
               endpoints=0 to 570 by 1 noframe
               vscale=percent
               vaxisLabel="Percent";
               where Trans_V="MANUAL ENTRY";
               Label ImTM = "Timeliness Measure";
               title 'Manual Entry - Visit';
 run;
proc univariate data=visit_level noprint;
      var ImTM;
      histogram / cfill=ltgray
               endpoints=0 to 570 by 1 noframe
               vscale=percent
               vaxisLabel="Percent";
               where Trans_V="BATCH";
               Label ImTM = "Timeliness Measure";
               title 'Batch - Visit';
 run;

 proc univariate data=visit_level noprint;
      var ImTM;
      histogram / cfill=ltgray
               endpoints=0 to 570 by 1 noframe
               vscale=percent
               vaxisLabel="Percent";
               where Trans_V="HL7";
               Label ImTM = "Timeliness Measure";
               title 'HL7 - Visit';
 run;
/*END HISTOGRAMS*/

/*plots and histograms of timeliness for all immunization records grouped by vaccination*/
proc sort data=final_merged;
by Trans_V;
run;
 proc univariate data=final_merged;
      var ImTM;
      histogram;
          title 'Histogram of Timeliness - Vaccination Level ';
 run;

 proc univariate data=final_merged;
      var ImTM;
      histogram;
          By Trans_V;
          title 'Histogram of Timeliness - Vaccination Level - by data transfer method';
 run;

proc freq data = visit_level;
tables trans_v;
run;

/*plots and histograms of timeliness for all immunization records grouped by visit*/
proc sort data=visit_level;
by Trans_V;
run;
 proc univariate data=visit_level;
      var ImTM;
      histogram;
          title 'Histogram of Timeliness - Visit Level Only';
 run;

 proc univariate data=visit_level;
      var ImTM;
      histogram;
          By Trans_V;
```

```
          title 'Histogram of Timeliness - Visit Level Only - by data transfer method';
 run;
/*boxplots*/
 proc univariate data=visit_level PLOT;
      var ImTM;
          title 'Box Plot - ImTM - Visit Level Only';
 run;

 proc univariate data=visit_level PLOT;
      var ImTM;
          By Trans_V;
          title 'Box Plot - ImTM - Visit Level Only - by data transfer method';
 run;

/* test cochran mantel-haenzel to account for effect of provider*/
proc freq data=visit_level;
format ImTMCat ImmCatF.;
tables MedOrg_V*ImTMCat*Trans_V/cmh;
title 'Cochran mantel-haenzel controlling for MedOrg';
run;

proc freq data=visit_level;
format ImTMCat ImmCatF.;
tables Facility_V*ImTMCat*Trans_V/cmh;
title 'Cochran mantel-haenzel controlling for Facility';
run;

proc logistic data = visit_level DESCENDING;
        class trans_v(param = ref ref="MANUAL ENTRY");
        model ImTMLog = trans_v;
        title 'LOGISTIC REGRESSION MODEL: ImTMLog=trans_v';
run;

/* create permanent sas datasets*/
data SASData.visit_level1;
set visit_level;
run;
data SASData.final_merged;
set final_merged;
run;
```

```
/* PROGRAM 5 */

/* ASSIGN LIBRARY */
libname SASData 'C:\Users\hillsr\Documents\BHI by Project\Dissertation\Aim2\Data\SASData\';

/*create the format for immunization timeliness categories*/
proc format;
value complete 0 = 'incomplete'
                         1 = 'complete';
run;
/*working with two datasets from SASData directory: SASData.final_merged AND
SASData.visit_level1 AND SasData.CP_DISTINCT_ID */
/*bring these three datasets into work directories*/

data final_merged;
set SASData.final_merged;
run;

/*read in only first 500000*/
data visit_level;
set SASData.visit_level1;
run;

/*read in only first 500000*/
data distinct_pt;
set SasData.CP_DISTINCT_ID;
run;

proc contents data = final_merged order=varnum;
run;
proc contents data = visit_level order=varnum;
run;
proc contents data = distinct_pt order=varnum;
run;

/*BOOL FIELDS HAVE BEEN ADDED FOR ALL PATIENT AND VACCINE FIELDS IN MERGE PROGRAM*/

/*Calculations for 'Completeness of demographic and immunization fields table*/
proc freq data = distinct_pt;
format FirstName MiddleName LastName City State Zip Phone DOB SSN MedOrgBool FacilityBool
complete.;
tables
        FirstName
        MiddleName
        LastName
        Address
        City
        State
        Zip
        Phone
        DOB
        SSN
        MedOrgBool
        FacilityBool
        Trans;
title 'COMPLETENESS FOR ALL DEMOGRAPHIC FIELDS FOR ALL PATIENTS IN THE SAMPLE';
run;

/*SORT THE DISTINCT_PT DATASET*/
proc sort data = distinct_pt;
by Trans;
run;

/*PROC FREQ FOR COMPLETENESS BY TRANSMISSION TYPE*/
proc freq data = distinct_pt;
format FirstName MiddleName LastName City State Zip Phone DOB SSN MedOrgBool FacilityBool
complete.;
tables FirstName*Trans MiddleName*Trans LastName*Trans City*Trans
State*Trans Zip*Trans Phone*Trans DOB*Trans SSN*Trans
MedOrgBool*Trans FacilityBool*Trans /NOROW;
title 'COMPLETENESS FOR ALL DEMOGRAPHIC FIELDS BY TRANSMISSION METHOD';
```

```
run;


/*PROC FREQ FOR 'Completeness of IMMUNIZATION FIELDS*/
proc freq data=final_merged;
format CPTCodeBool EntryDateBool ServiceDateBool ImmDescBool MedOrgBool_V FacilityBool_V
complete.;
tables CPTCodeBool
        EntryDateBool
        ServiceDateBool
        ImmDescBool
        MedOrgBool_V
        FacilityBool_V
        Trans_V;
title 'COMPLETENESS FOR ALL IMMUNIZATION FIELDS FOR ALL RECORDS IN THE SAMPLE';
run;

/*sort the final merged dataset by trans_v*/
proc sort data = final_merged;
by Trans_v;
run;
/*PROC FREQ FOR 'Completeness of IMMUNIZATION FIELDS BY TRANSMISSION TYPE*/
proc freq data=final_merged;
format CPTCodeBool EntryDateBool ServiceDateBool ImmDescBool MedOrgBool_V FacilityBool_V
complete.;
tables CPTCodeBool*Trans_V
        EntryDateBool*Trans_V
        ServiceDateBool*Trans_V
        ImmDescBool*Trans_V
        MedOrgBool_V*Trans_V
        FacilityBool_V*Trans_V /NOROW;
title 'COMPLETENESS FOR ALL IMMUNIZATION FIELDS FOR ALL RECORDS IN THE SAMPLE BY TRANSMISSION
TYPE';
run;

/*PROC FREQ FOR COMPLETENESS NUMBER FOR DEMOG FOR TOTAL SAMPLE*/
proc freq data=distinct_pt;
tables  demog_compl_tot demog_compl_min;
title 'frequency of number of complete demographic fields - TOTAL DEMOGRAPHIC SAMPLE';
run;
/*PROC FREQ FOR 'COMPLETENESS NUMBER FOR DEMOG BY TRANS*/
proc freq data=distinct_pt;
tables  demog_compl_tot*trans demog_compl_min*trans;
title 'frequency of number of complete demographic fields - TOTAL DEMOGRAPHIC SAMPLE BY
TRANS';
run;


/*PROC FREQ FOR 'Completeness percentage TOTAL SAMPLE - ALL VACCINATIONS*/
proc freq data=final_merged;
tables  imm_compl_tot
                imm_compl_min
                total_compl
                total_compl_min
                num_dem_flds
                num_imm_flds;
title 'COMPLETENESS PERCENTAGE FOR ALL IMMUNIZATION FIELDS FOR ALL RECORDS - TOTAL SAMPLE';
run;

/*CHISQUARE TEST FOR COMPLETENESS CATEGORIES (3)*/
proc freq data=distinct_pt;
tables  completecat*trans/CHISQ;
title 'CHISQUARE TEST FOR COMPLETENESS CATEGORIES - PATIENT';
run;

proc freq data=final_merged;
tables  num_imm_flds*trans_v/CHISQ;
title 'CHISQUARE TEST FOR COMPLETENESS CATEGORIES - VACCINE';
run;

/*PROC FREQ FOR Completeness percentage BY TRANSMISSION TYPE*/
```

```
proc freq data=final_merged;
tables imm_compl_tot*TRANS_V
               imm_compl_min*trans_V
               total_compl*trans_V
               total_compl_min*trans_v
               num_dem_flds*trans_v
               num_imm_flds*trans_v;
title 'COMPLETENESS PERCENTAGE FOR ALL IMMUNIZATION FIELDS FOR ALL RECORDS IN THE SAMPLE BY
TRANSMISSION TYPE';
run;

/*PROC MEANS FOR PERCENTAGE COMPLETE CALCULATIONS*/
proc means data=distinct_pt;
var pct_demog_flds;
title 'MEAN COMPLETENESS PERCENTAGE FOR ALL DEMOGRAPHIC FIELDS - ALL RECORDS IN THE SAMPLE
(DISTINCT)';
run;
/*SORT DISTINCT BY TRANS*/
proc sort data = distinct_pt;
by trans;
run;
proc means data=distinct_pt;
var pct_demog_flds;
by trans;
title 'MEAN COMPLETENESS PERCENTAGE FOR ALL DEMOGRAPHIC FIELDS - BY TRANSMISSION METHOD
(DISTINCT)';
run;

proc means data=final_merged;
var prct_imm_flds;
run;

/*SORT MERGED BY TRANS V*/
proc sort data = final_merged;
by trans_v;
run;

proc means data=final_merged;
var prct_imm_flds;
by trans_v;
title 'MEAN COMPLETENESS PERCENTAGE FOR ALL IMMUNIZATION FIELDS - BY TRANSMISSION METHOD
(VACCINATIONS)';
run;

/*LOGISTIC REGRESSION*/
/*      DESCENDING - MODELS "1" OR COMPLETE AS "EVENT"
        MANUAL ENTRY AS REFERENCE
        THIS WILL GIVE THE OR FOR BATCH VS MANUAL AND HL7 VS MANUAL AND BIRTH CERT VS MANUAL

        NOT COMPUTED FOR: FIRSTNAME, LASTNAME, BIRTHDATE, Entry, Service Imm Desc MedOrg (ALL
OBSERVATION HAVE THE SAME RESPONSE)*/

/*DEMOG FIELDS*/
proc logistic data = CP_DISTINCT_ID DESCENDING;
        class trans(param = ref ref="MANUAL ENTRY");
        model MiddleName = trans;
        title 'LOGISTIC REGRESSION MODEL: MiddleName=trans';
run;

proc logistic data = CP_DISTINCT_ID DESCENDING;
        class trans(param = ref ref="MANUAL ENTRY");
        model Address = trans;
        title 'LOGISTIC REGRESSION MODEL: Address=trans';
run;

proc logistic data = CP_DISTINCT_ID DESCENDING;
        class trans(param = ref ref="MANUAL ENTRY");
        model City = trans;
        title 'LOGISTIC REGRESSION MODEL: City=trans';
run;
```

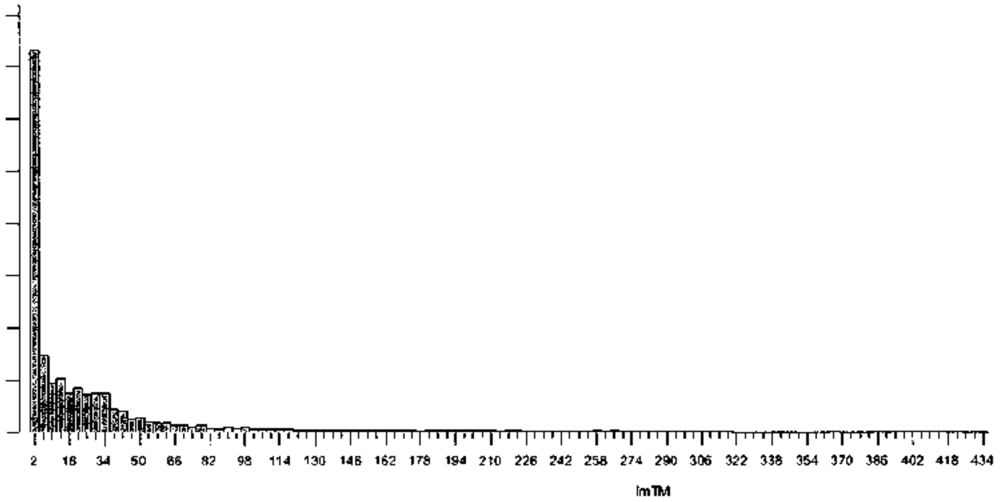
```
proc logistic data = CP_DISTINCT_ID DESCENDING;
       class trans(param = ref ref="MANUAL ENTRY");
       model State = trans;
       title 'LOGISTIC REGRESSION MODEL: State=trans';
run;

proc logistic data = CP_DISTINCT_ID DESCENDING;
       class trans(param = ref ref="MANUAL ENTRY");
       model Zip = trans;
       title 'LOGISTIC REGRESSION MODEL: Zip=trans';
run;

proc logistic data = CP_DISTINCT_ID DESCENDING;
       class trans(param = ref ref="MANUAL ENTRY");
       model Phone = trans;
       title 'LOGISTIC REGRESSION MODEL: Phone=trans';
run;

proc logistic data = CP_DISTINCT_ID DESCENDING;
       class trans(param = ref ref="MANUAL ENTRY");
       model SSN = trans;
       title 'LOGISTIC REGRESSION MODEL: SSN=trans';
run;

proc logistic data = CP_DISTINCT_ID DESCENDING;
       class trans(param = ref ref="MANUAL ENTRY");
       model MedOrgBool = trans;
       title 'LOGISTIC REGRESSION MODEL: MedOrgBool=trans';
run;

proc logistic data = CP_DISTINCT_ID DESCENDING;
       class trans(param = ref ref="MANUAL ENTRY");
       model FacilityBool = trans;
       title 'LOGISTIC REGRESSION MODEL: FacilityBool=trans';
run;

/*VACCINATION FIELDS*/
proc logistic data = final_merged DESCENDING;
       class trans_v(param = ref ref="MANUAL ENTRY");
       model CPTCodeBool = trans_v;
       title 'LOGISTIC REGRESSION MODEL: CPTCodeBool=trans_v';
run;

proc logistic data = final_merged DESCENDING;
       class trans_v(param = ref ref="MANUAL ENTRY");
       model FacilityBool_V = trans_v;
       title 'LOGISTIC REGRESSION MODEL: FacilityBool_V=trans_v';
run;
```

```
/* PROGRAM 6 */
/*THIS PROGRAM EXPLORES INFLUENZA IMMUNIZATIONS*/

/*ASSIGN LIBRARY*/
libname SASData 'C:\Users\hillsr\Documents\BHI by Project\Dissertation\Aim2\Data\SASData\';

/*create the format for immunization timeliness categories*/
proc format;
value ImmCatF 0 = 'negative days'
                        1 = '<7 days'
                        2= '7-30 days'
                        3='>30 days';
value ImmLogF 0 = '<=30 days'
                        1 = '>30 days';
value Flu        0 = 'Non-Flu'
                        1 = 'Flu';
value compl      0 = '0/6'
                        1 = '1/6'
                        2 = '2/6'
                        3 = '3/6'
                        4 = '4/6'
                        5 = '5/6'
                        6 = '6/6';
run;

data tempvacc;
set SASData.final_merged;
run;

/*influenza field as 1 if the imunization administred was an influenza seasonal vaccine*/
data tempvacc;
set tempvacc;
if CPTCode in(90654 90655 90656 90657 90658 90659 90660 906662 90663 90664) then influenza=1;
else influenza=0;
run;

proc sort data=tempvacc;
by trans_v;
run;

proc freq data = tempvacc;
title 'Records in transmission and influenza categories';
format influenza Flu.;
tables influenza;
by trans_V;
run;

proc freq data=tempvacc;
format ImTMCat ImmCatF.;
tables ImTMCat;
title 'Displays frequencies for the timeliness categories for flu vaccinations';
by trans_V;
where influenza=1;
run;


PROC TABULATE data=tempvacc;
format ImTMCat ImmCatF. influenza Flu.;
title 'Frequencies for the timeliness categories by transfer method and flu/nonflu';
CLASS Trans_V ImTMCat Influenza;
TABLE Influenza * Trans_V * ROWPCTN,
ImTMCat ;
RUN;


/*completeness*/
proc freq data=tempvacc;
tables CPTCodeBool FacilityBool_V;
by trans_V;
title 'Displays frequencies for the timeliness categories by transfer method';
run;
```

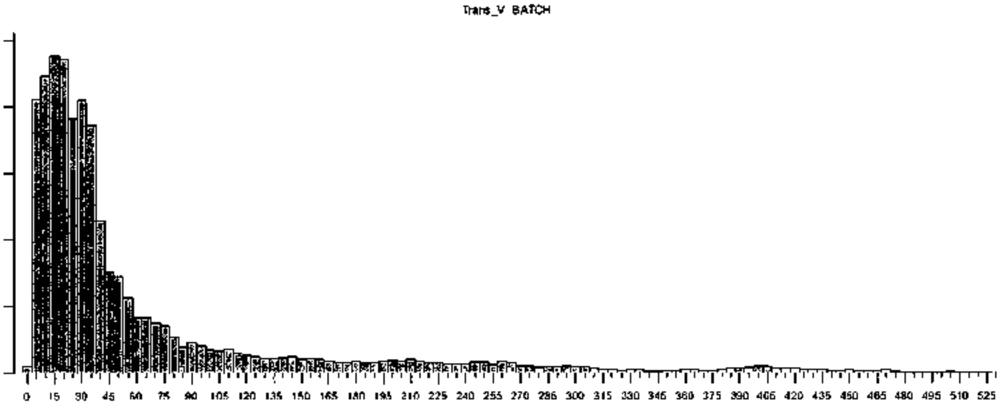
```
PROC TABULATE data=tempvacc;
format  influenza Flu. imm_compl_tot compl.;
title 'Frequencies for the completeness numbers by transfer method and flu/nonflu';
CLASS Trans_V imm_compl_tot Influenza;
TABLE Influenza * Trans_V * ROWPCTN,
imm_compl_tot;
    RUN;
```

Appendix D: Selected SAS Output

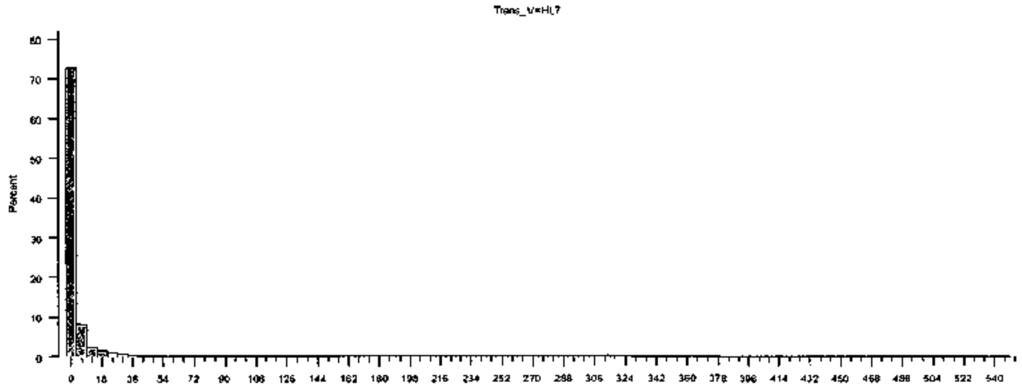
Histogram of Timeliness Measure – Visit Level – All records



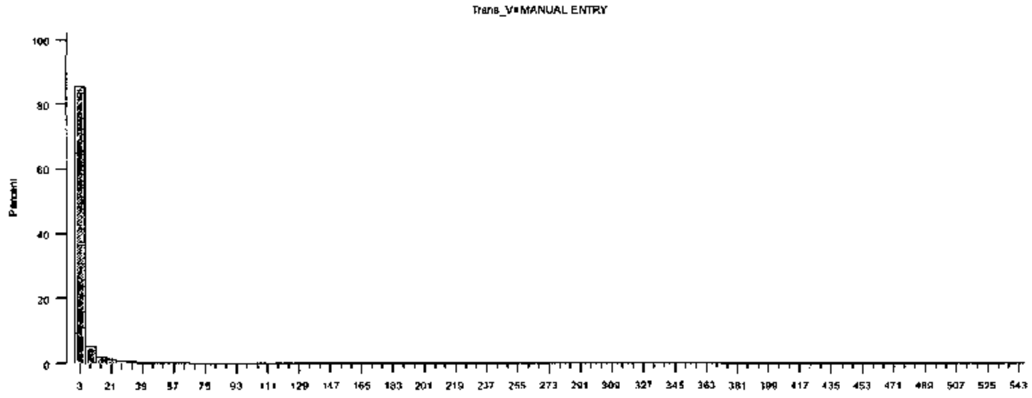
Histogram of Timeliness Measure – Visit Level – Transmission = Batch



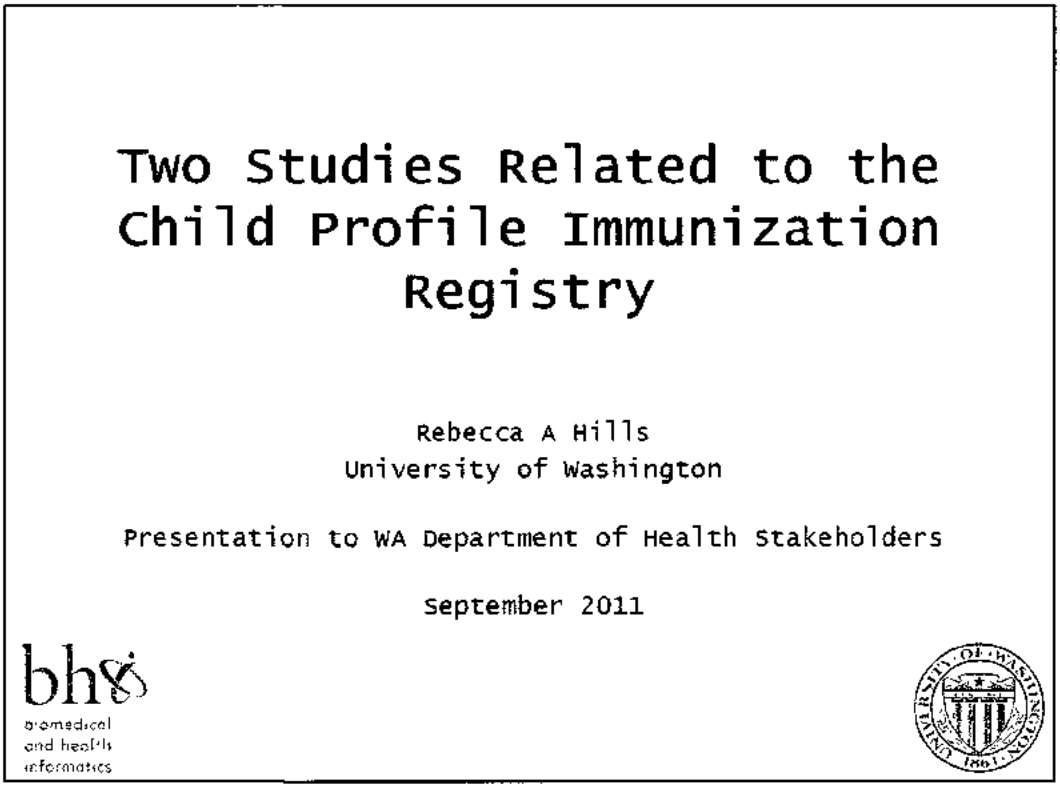
Histogram of Timeliness Measure – Visit Level – Transmission = HL7



Histogram of Timeliness Measure – Visit Level – Transmission = Manual Entry



Appendix E: Presentation Made to Washington Department of Health Stakeholders



Two Studies Related to the Child Profile Immunization Registry

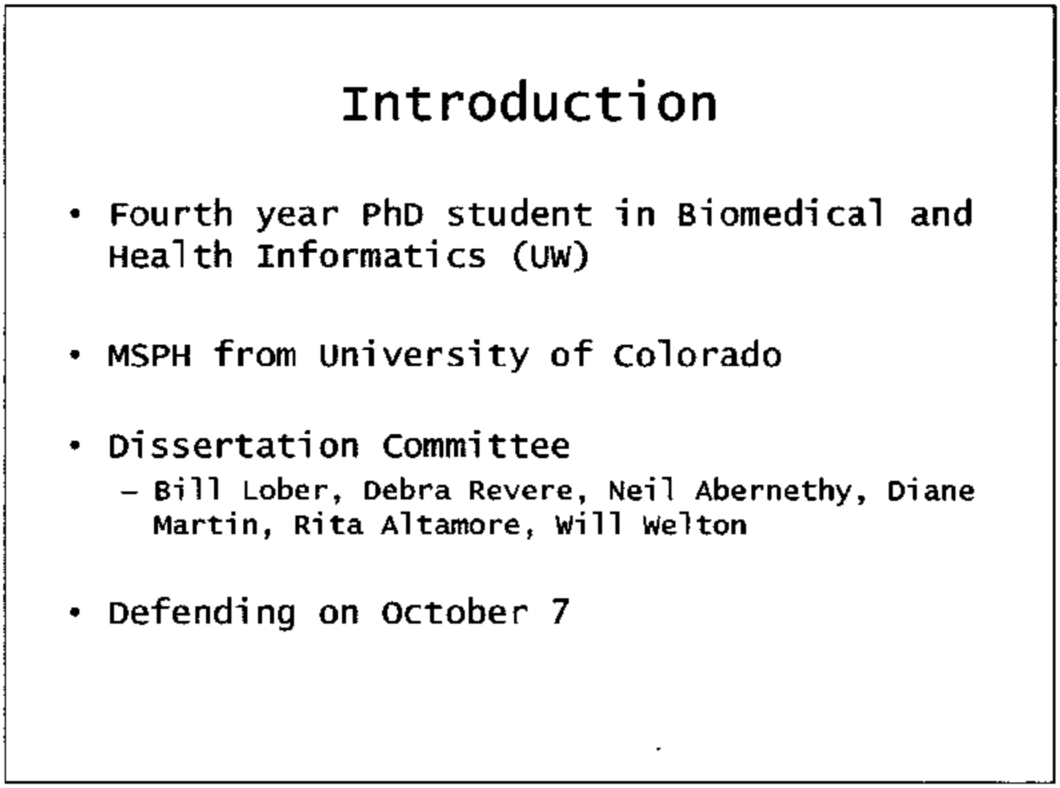
Rebecca A Hills
University of Washington

Presentation to WA Department of Health Stakeholders

September 2011

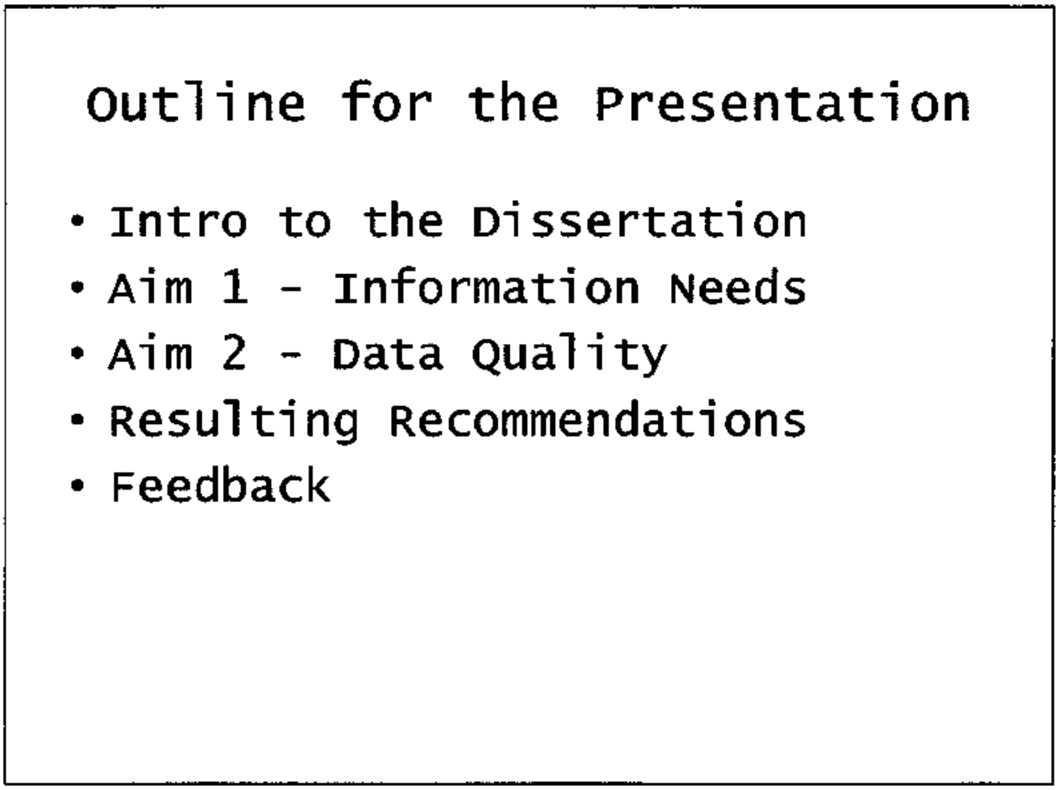
biomedical
and health
informatics

1861



Introduction

- Fourth year PhD student in Biomedical and Health Informatics (UW)
- MSPH from University of Colorado
- Dissertation Committee
  - Bill Lober, Debra Revere, Neil Abernethy, Diane Martin, Rita Altamore, Will Welton
- Defending on October 7



Outline for the Presentation

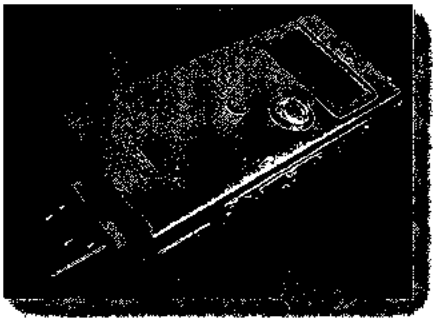
- Intro to the Dissertation
- Aim 1 - Information Needs
- Aim 2 - Data Quality
- Resulting Recommendations
- Feedback

Aim 1 – Information Needs in PH Practice

- Information Needs in PH have been studied broadly
- Motivation:
  - Little available on information needs related to immunization work
  - Questions about epi use of IIS data
- Research Question:
  What are the immunization related information needs of individuals working in public health practice?

Aim 1 - Methods

- Interviews with public health practitioners across WA
- Semi-structured interviews
- Interviews with 20 PH practitioners
- Digitally recorded
- Transcribed verbatim
- Coded in NVivo 8

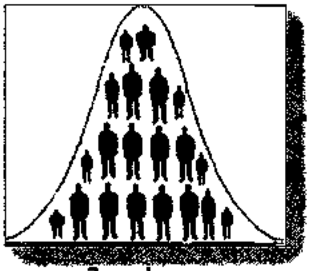


Aim 1 - Results

Three main types of information needs

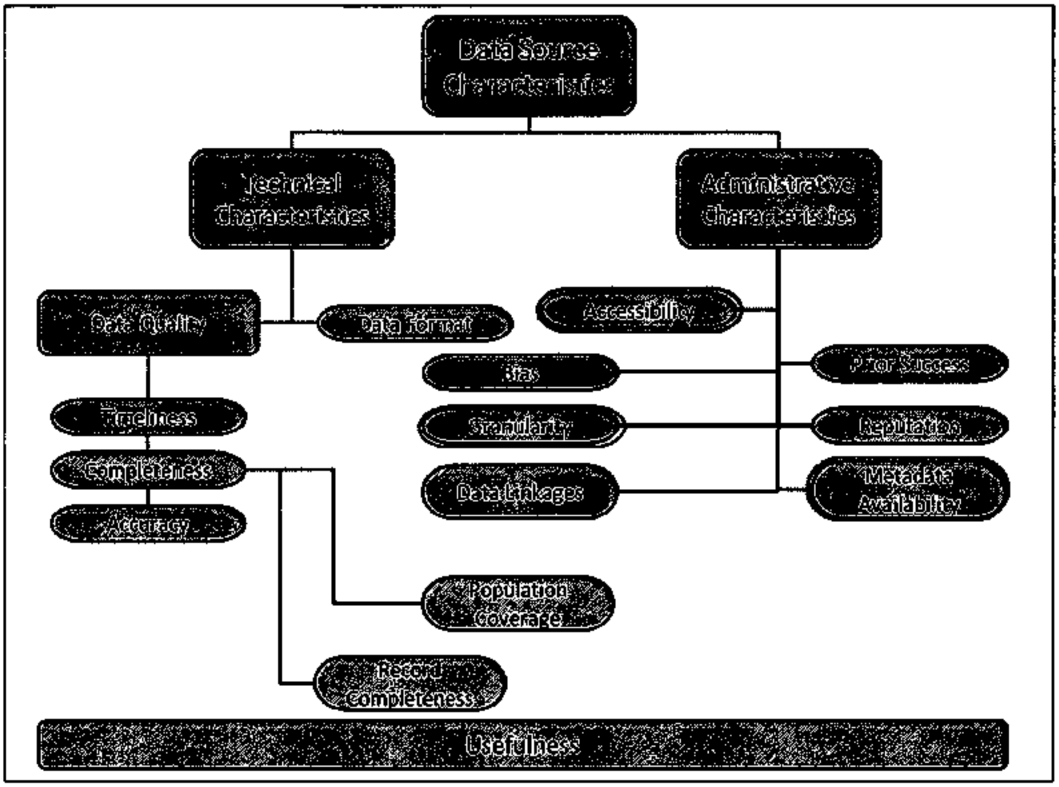
Individual/Record

Context specific
(Vaccine)



Population

Data Source Characteristics



Data Source Characteristics

Technical Characteristics

Administrative Characteristics

Data Quality

Data Format

Accessibility

Bias

Prior Success

Timeliness

Granularity

Reputation

Completeness

Data Linkages

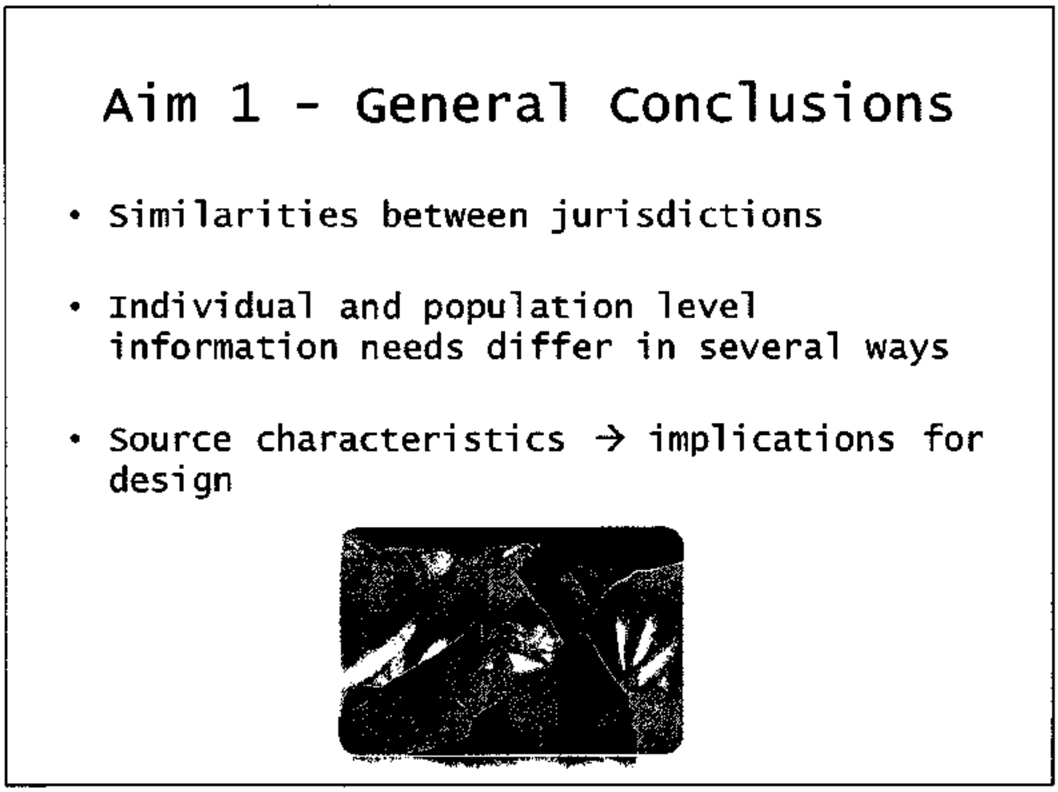
Metadata Availability

Accuracy

Population Coverage

Record Completeness

Usefulness



Aim 1 - General Conclusions

- Similarities between jurisdictions
- Individual and population level information needs differ in several ways
- Source characteristics → implications for design

Switch Gears!

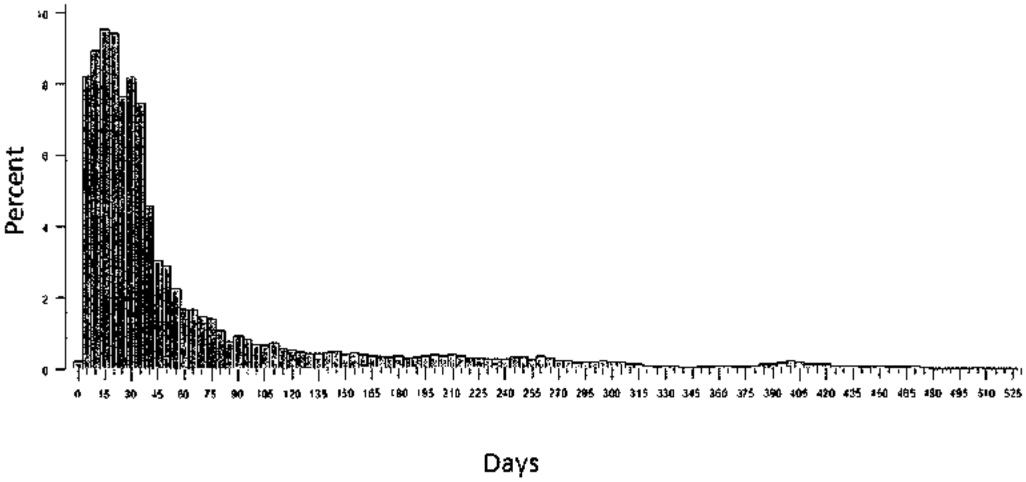
Aim 2 – Timeliness and Data Element Completeness in an IIS

- One of the important characteristics from Aim 1 → Data Quality
- Motivation
  - It is likely that the number of HL7 connections will increase
  - How will that affect data quality?
- Research Question
  - Do timeliness and data element completeness differ across the different modes of data transmission?

Aim 2 - Methods

- Dataset from Child Profile's 2010 transactions
- Children where 1/1/1991 < *DOB* < 12/31/2010
- Excluded historical records
- 757,476 individual patient records
- 2,634,101 vaccination records
- Measured timeliness: entry date-service date
- Measured completeness
- Compared across data transmission methods

Timeliness measure - histogram



Association of Data Transmission Method with Timeliness Measure >30 Days

Data Transmission Mode	N	Timeliness Measure (%) *<7 days*	*7-30 days*	>30 days	OR (CI)*
Manual	361,712	*88.52*	*6.27*	5.21	1 (ref)
HL7	227,101	*80.40*	*6.31*	13.29	2.79 (2.74-2.84)
Batch	680,543	*8.43*	*40.78*	50.78	18.77(18.49-19.07)
Total	1,269,347	*44.13*	*24.78*	31.09	-

*OR indicates odds ratio, CI indicates confidence interval calculated for a timeliness measure categorized as ≤30 days or >30 days, test of null hypothesis(regression coefficient equal to zero, Chi-square DF=2) was significant, P <.0001

Aim 2 - Completeness Results

Percentage Field Completeness

Demographics	Manual Entry	HL7	Batch	Birth Cert	Total
First Name*(%)	100.00	100.00	100.00	100.00	100.00
Middle Name(%)	54.98	48.92	63.95	81.10	62.73
Last Name*(%)	100.00	100.00	100.00	100.00	100.00
Address(%)	98.44	99.94	99.95	99.96	99.74
City(%)	98.81	99.99	99.99	99.97	99.83
State(%)	99.21	99.99	99.99	99.96	99.88
Zip Code(%)	98.76	99.98	99.98	99.96	99.81
Phone(%)	86.86	95.46	88.22	88.29	88.45
Birthdate*(%)	100.00	100.00	100.00	100.00	100.00
SSN(%)	19.32	29.93	32.95	9.44	27.11
Med Org(%)	98.22	99.92	99.43	96.07	98.85
Facility(%)	94.60	96.61	95.51	90.20	95.06
Total Completeness	86.93	88.04	90.00	89.56	89.29

Percent completeness of demographic fields

*minimum dataset as defined by MIROW

Percent Completeness and Association of Data Transmission Method with Complete Data Element

Data Element	Data Transmission Mode	Percent Complete Records	OR (CI)*
Middle Name	Manual	54.98%	1 (ref)
	HL7	48.92%	0.78 (0.77-0.80)
	Batch	63.95%	1.45 (1.43-1.47)
	Birth Certificate	81.10%	3.52 (3.45-3.58)
Phone	Manual	80.86%	1 (ref)
	HL7	95.46%	4.98 (4.83-5.14)
	Batch	88.22%	1.77 (1.74-1.81)
	Birth Certificate	88.29%	1.79 (1.74-1.83)
SSN	Manual	19.32%	1 (ref)
	HL7	29.93%	1.78 (1.75-1.82)
	Batch	32.95%	2.05 (2.02-2.09)
	Birth Certificate	9.44%	0.44 (0.43-0.45)
Facility	Manual	94.60%	1 (ref)
	HL7	95.61%	1.24 (1.20-1.30)
	Batch	95.51%	1.21 (1.18-1.25)
	Birth Certificate	93.20%	0.78 (0.76-0.81)

Aim 2 - Completeness Results

	Field Completeness			
Immunization	Manual Entry	HL7	Batch	Total
Date of Entry (%)	100.00	100.00	100.00	100.00
Date of Service* (%)	100.00	100.00	100.00	100.00
CPT Code (%)	99.79	99.96	100.00	99.93
Imm Description*(%)	100.00	100.00	100.00	100.00
Med Org*(%)	100.00	100.00	100.00	100.00
Facility (%)	98.80	97.03	93.58	98.73
Total Completeness	99.76	99.50	98.93	99.79

Percent completeness of immunization fields – one record per vaccination

*fields contributing to minimum dataset as defined by MIROW

Percent completeness of immunization fields

Data Element	Data Transmission Mode	Percent Complete Records	OR (CI)*
Date of Entry	All	100%	**
Date of Service	All	100%	**
CPT Code	Manual	99.79%	1 (ref)
	HL7	99.96%	6.04 (5.16-7.07)
	Batch	100%	485.88 (217.98->999)
Imm. Desc.	All	100%	**
Med Org	All	100%	**
Facility	Manual	98.80%	1 (ref)
	HL7	97.03%	0.18 (0.17-0.18)
	Batch	93.58%	0.40 (0.39-0.41)

Aim 2 – General Conclusions

- Timeliness was variable but Manual entry and HL7 records were more timely
  - Manual Entry 88.52% in < 7 days category
  - HL7 80.40% in <7 days category
- Statistically significant difference between groups (chi-square, p<.0001)
- Practical significance depends on purpose of using data
  - Outbreak investigation and control
  - Largest problems with batch

Aim 2 - General Conclusions

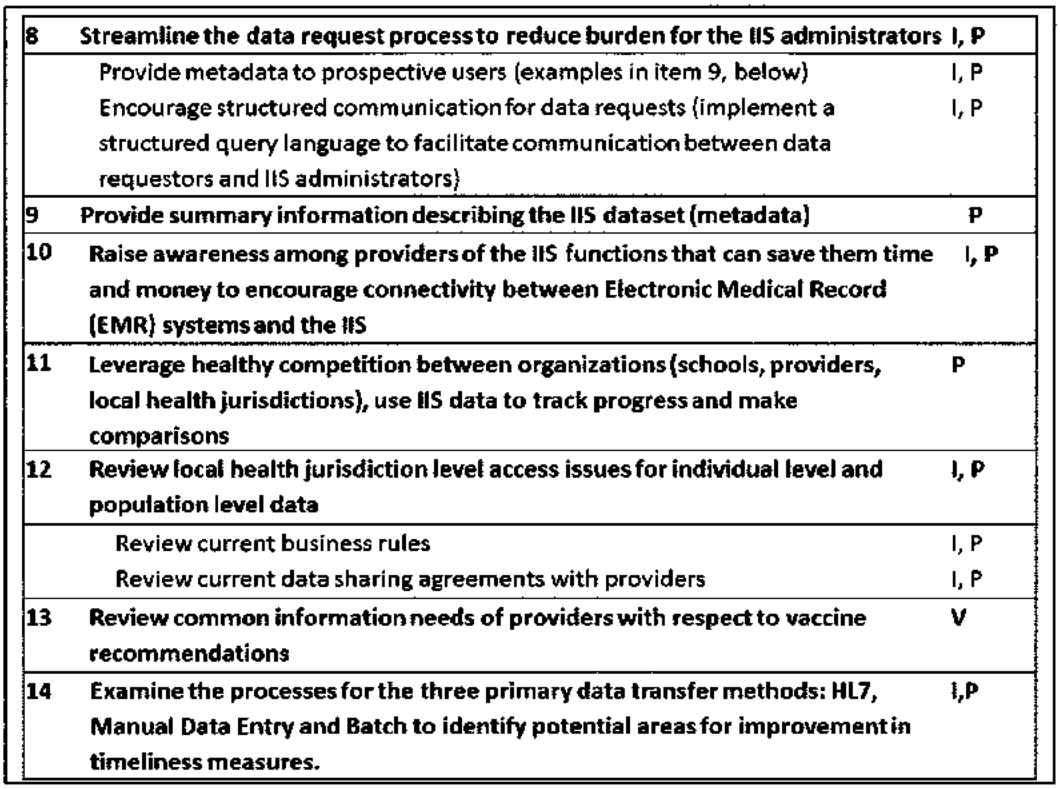
- Overall completeness was very high
- Problem demographic fields (Man/HL7/Batch)
  - Middle Name (55/49/64%)
  - Phone (81/95/88%)
  - SSN (19/30/33%)
  - Facility (95/96/96%)
- Overall demographic records
  - Batch performed better than HL7 and Manual
- Vaccination fields
  - All ~100% complete except Facility for Batch records (93.58%)

System Suggestions

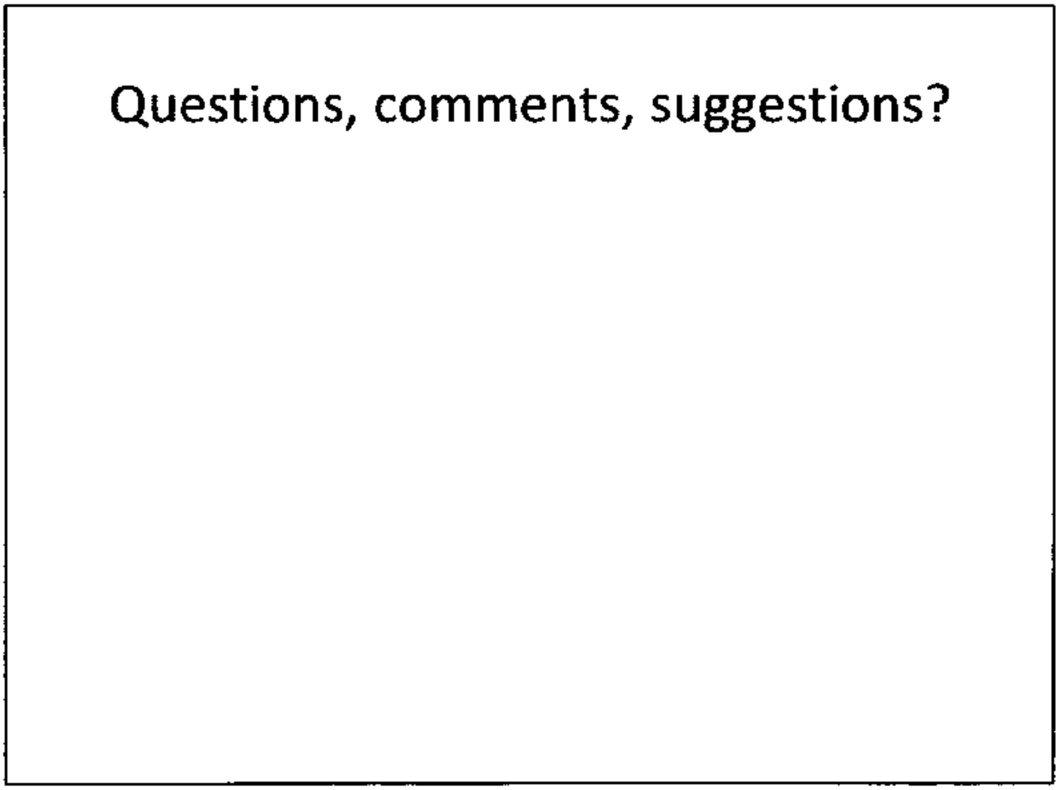
1	✓	**Facilitate live sharing of IIS data between neighboring states (OR, ID)**	**I**
		Investigate the use of Integrating the Healthcare Enterprise Profiles for live sharing. Two profiles may be appropriate: Query for Existing Data and Immunization Content	I
2		**Facilitate access to IIS data for state licensed child care providers**	**I, P**
3	✓	**Facilitate access to IIS data for schools**	**I, P**
4		**Expand reporting functions**	**P**
		Immunization coverage rate for a single vaccine for an age group in a geographic region Reports that facilitate comparison of coverage rates between providers Ability to run and save ad-hoc reports Facilitate the generation of reports from the IIS for performance measures	P P P P
5		**Facilitate linkages with other immunization related data sources**	**I, P**
		Link or facilitate side-by-side comparison of imms coverage and school exemption data Link IIS data to PHIMS data	P I, P
6		**Provide summary information describing the IIS dataset (metadata)**	**P**
		Related to both system and strategic suggestions, examples in table 4.2	
7		**Enhance the local health department view of individual level IIS data**	**I, P**
		Allow LHJ access to expanded individual level data for all providers in the jurisdiction	I
8	✓	**Evaluate and enhance inventory management tools in the IIS**	**V**
		Allow providers or local health jurisdictions to look up ship dates Streamline vaccine ordering and tracking process (reduce number of steps, intermediaries)	V V
9		**Enhance IIS functions for mass vaccination campaigns**	**P, I**
		Off-site real-time data entry Timely access to mass vaccination campaign data Excel spreadsheet upload of mass vaccination campaign data	I P, I I

Process and Strategic Suggestions

1	Encourage more timely data entry (will serve to increase usefulness and reputation)	I, P, V
2	Provide public use dataset (de-identified) as an export to common file type such as Excel or Comma-Separated Values (CSV) file	P
3	Look to vital records, Washington Tracking Network and Community Health Assessment Tool (CHAT) for guidance on examples of providing summary data, interactive queries and standard datasets to researchers	P
4	Consider the use of free and open source tools for querying and visualization of public use and/or aggregated IIS data	P
	Google Charts: http://code.google.com/apis/chart/	P
5	Create or facilitate the establishment of a Community of Practice for IIS data users (and prospective users)	P
6	Provide information about alternate and linkable datasets	I, P
7	Provide description of how data are collected (for prospective non immunization users of the data and information)	I, P



8	**Streamline the data request process to reduce burden for the IIS administrators**	**I, P**
	Provide metadata to prospective users (examples in item 9, below)	I, P
	Encourage structured communication for data requests (implement a structured query language to facilitate communication between data requestors and IIS administrators)	I, P
9	**Provide summary information describing the IIS dataset (metadata)**	**P**
10	**Raise awareness among providers of the IIS functions that can save them time and money to encourage connectivity between Electronic Medical Record (EMR) systems and the IIS**	**I, P**
11	**Leverage healthy competition between organizations (schools, providers, local health jurisdictions), use IIS data to track progress and make comparisons**	**P**
12	**Review local health jurisdiction level access issues for individual level and population level data**	**I, P**
	Review current business rules	I, P
	Review current data sharing agreements with providers	I, P
13	**Review common information needs of providers with respect to vaccine recommendations**	**V**
14	**Examine the processes for the three primary data transfer methods: HL7, Manual Data Entry and Batch to identify potential areas for improvement in timeliness measures.**	**I,P**



Questions, comments, suggestions?

Acknowledgements

A huge thank you to:
Sherry Riddick
Chas DeBolt
Jodi Warren
Yousif Hozail
Pat DeHart
Janna Bardi

And my advisory committee

And of course, all of the public health professionals across Washington who welcomed me into their offices to talk about their work!

bhi
biomedical
and health
informatics

UNIVERSITY OF WASHINGTON 1861

Thanks for your time!

Please contact me for more information or to share thoughts, feedback, or ideas for future work

email: hillsr@uw.edu or rebeccahills@gmail.com
phone: 206.972.3998

bhi
biomedical
and health
informatics

UNIVERSITY OF WASHINGTON 1861

Vita

Rebecca Hills was born in Kotzebue, Alaska. She has also lived and worked in Pennsylvania, Georgia, Colorado and the state of Washington. Rebecca earned a Bachelor of Arts in Biology and German at Bucknell University and a Master of Science in Public Health at the Colorado School of Public Health. She also has non-degree technical training in information system programming from Denver Technical College. In 2011, she earned a Doctor of Philosophy in Biomedical and Health Informatics from the School of Medicine at the University of Washington in Seattle, Washington.