Electronic Laboratory Data Quality and the Value of a Health Information Exchange to Support Public Health Reporting Processes

Brian E. Dixon, MPA, PhD[1,2], Julie J. McGowan, PhD[1,2,3], Shaun J. Grannis, MD, MS[1,2,3]

[1]Indiana University, School of Informatics, Indianapolis, IN; [2]Regenstrief Institute, Indianapolis, IN; [3]Indiana University, School of Medicine, Indianapolis, IN

Abstract

There is increasing interest in leveraging electronic health data across disparate sources for a variety of uses. A fallacy often held by data consumers is that clinical data quality is homogeneous across sources. We examined one attribute of data quality, completeness, in the context of electronic laboratory reporting of notifiable disease information. We evaluated 7.5 million laboratory reports from clinical information systems for their completeness with respect to data needed for public health reporting processes. We also examined the impact of health information exchange (HIE) enhancement methods that attempt to improve completeness. The laboratory data were heterogeneous in their completeness. Fields identifying the patient and test results were usually complete. Fields containing patient demographics, patient contact information, and provider contact information were suboptimal. Data processed by the HIE were often more complete, suggesting that HIEs can support improvements to existing public health reporting processes.

Introduction

Interest in and development of methods for leveraging electronic health record (EHR) data across disparate sources for a variety of use cases (a.k.a. secondary use) is widespread (1). Existing literature discusses the challenges associated with the re-use of EHR data for purposes beyond clinical care, including access to data, privacy protections, identity matching, and the interoperability of data from disparate sources (1, 2). However, often missing from discussions of secondary use is a common, core issue that arguably is more problematic than any other issue relevant to re-use of EHR data: poor data quality.

Poor data quality is common and affects all industries and organizations that employ information systems (3). Typical data quality issues encountered include: inaccurate data, inconsistencies across data sources, and incomplete (or unavailable) data necessary for operations or decisions (4). A large bank found that data in its credit-risk management database were only 60% complete, which necessitated additional scrutiny by anyone using its data (5). In health care, the completeness of data in EHR systems has been found to vary from 30.7 to 100% (6).

Good evidence from the information management literature on the impacts of these issues is sparse, but good estimates of impacts include: increased costs in the range of 8-12% of organizational revenue, and up to 40-60% of a service organization's expenses consumed as a result of poor data; poorer decisions that take longer to make; lower data consumer satisfaction with information systems; and increased difficulty in reengineering work and information flows to improve service delivery (4). Impacts on health care include poorer decisions when humans or machines use poor quality data inputs from EHR systems (7, 8).

Data quality issues have been well examined and documented in the epidemiology literature. For example, spontaneous reporting rates for infectious diseases range from 9% to 99% and have remained relatively unchanged from 1970 – 2000 (9). While some conditions, such as sexually transmitted infections, are reported approximately 80% of the time, many conditions are reported less than half of the time. Timeliness, another attribute of data quality (3), has also been found to be a challenge in public health reporting (10). Delays in the receipt of notifiable disease data (timeliness) and the lack of a complete set of reports (completeness) impact public health agency surveillance processes, including but not limited to the ability of agencies to respond to emerging disease threats.

Electronic laboratory reporting (ELR) was demonstrated just over a decade ago to be an effective method to improve the timeliness of reporting as well as the number of reports submitted to public health agencies (10). Since ELR was shown to be effective, the U.S. government, states, and a number of private foundations have invested millions of dollars into the development, implementation, and adoption of health information, ELR, and surveillance systems (11-15). Despite reported improvements in the timeliness and volume of submitted reports, some studies indicate anecdotally that ELR may not improve the completeness of the data in the submitted reports (16).

Given a paucity of evidence in the literature that ELR does or does not impact the completeness of notifiable disease data, this study examined the completeness of data from clinical information systems. In addition to characterizing the completeness of ELR data, the study further compared raw data directly sent from clinical information systems with data enhanced by a health information exchange (HIE) prior to transmission to a public health agency. If an HIE can improve the completeness of ELR data submitted to public health, it would signify that HIE data enhancement methods are a valuable, effective method for improving notifiable disease data quality. Improving data quality will likely translate into improvements in disease surveillance processes, impacting both clinicians and public health professionals.

Background

Although surveillance methods and practices date back to 1854 when John Snow used reported mortality data and location information to convince authorities to remove a water pump that was the source of a cholera outbreak, modern surveillance activities are aided by computer systems and informatics methods (17). This modernization of surveillance began in earnest during the previous decade following a report commissioned by the U.S. Department of Health and Human Services (HHS) and U.S. Centers for Disease Control and Prevention (CDC) (18) as well as early evidence published on the use of electronic methods to enhance traditionally manual surveillance processes (10). The HHS report and early pioneers illuminated a number of challenges, including but not limited to: 1) a lack of data standards for the exchange of surveillance data between providers and public health and between public health entities; 2) variability in the use of available messaging formats for the exchange of surveillance data between providers and public health and between public health entities; 3) limited decision and analytic support from early computer applications; and 4) a general lack of computer systems in public health laboratories (18, 19).

Policymakers responded to these challenges by funding numerous initiatives which included at least one principal aim to modernize surveillance practices by implementing advanced IT systems and networks that would link health care providers and public health agencies together to better detect and cooperatively address disease outbreaks (15, 20). Funding from the CDC Office of Surveillance, Epidemiology and Laboratory Services is just one example (21).

Many initiatives focused on electronic laboratory reporting (ELR), which involves the electronic submission of laboratory data, following the confirmation of an infectious disease, to a public health agency. Others focused on syndromic surveillance which detects initial manifestations of disease before clinical or laboratory diagnoses are established. All initiatives sought to improve the timeliness, accuracy, and completeness of data needed by public health agencies to perform surveillance activities.

The evidence in recent literature demonstrates that ELR can be effective at improving the timeliness of infectious disease reports (22), and ELR can further increase the number of cases reported to public health agencies (23). However, some researchers anecdotally suggest that ELR may not improve the completeness of the data reported to public health agencies (16). Therefore our study focused on evaluating the completeness of the data output from current clinical information systems. We examined ELR data received directly from laboratory and hospital information systems as well as enhanced data transmitted from an HIE to public health. For the purposes of this study, enhanced data is defined as data that has been validated, including but not limited to correcting units of measure, as well as augmented, including but not limited to mapping local lab test identifiers to standardized vocabulary concepts. Researchers at Regenstrief have previously described the practice of data enhancement and its routine use when exchanging clinical data as a method to improve data quality and interoperability (24, 25).

Methods

The scope of our research included the following aims: 1) the development of a method for evaluating the completeness of laboratory data in the context of public health reporting; 2) measuring the completeness of laboratory data received from clinical information systems and an HIE using the method; and 3) comparing the completeness of the "raw" data from clinical information systems (e.g., unaltered, unedited ELR messages) with the completeness of "enhanced" data from the HIE (e.g., ELR messages having syntax corrected and concepts mapped to standard vocabularies). The study was performed in the context of the lead author's (BED) dissertation and approved by both the Indiana University-Purdue University Indianapolis Institutional Review Board as well as the research council of the Indiana Network for Patient Care (INPC).

The central theme of the study was completeness. Completeness in the context of public health surveillance refers to both the proportion of diagnosed cases reported to public health and the proportion of fields in a case report completed by the submitting hospital or lab (26). As previously described, ELR messages' ability to increase the proportion of diagnosed cases reported to public health has been well established (10, 16, 22). Therefore this study concentrated on measuring the completeness of the data within ELR messages transmitted by a data source (e.g., hospital, laboratory, HIE). The completeness of ELR messages in this context is unknown and only asserted anecdotally by previous research in public health informatics.

A first step in analyzing the completeness of ELR message data involved creating a "minimum data set" for ELR messages that would meet the information needs of public health agencies that receive the data. Public health professionals collate ELR data with data received from other sources to complete a case report. Therefore the aim of ELR should be to provide a set of data that can populate as much of the case report as possible to streamline public health and clinical workflows. For example, when data is missing, public health professionals often call provider organizations to acquire the missing data, which disrupts public health and clinical workflows.

The lead author (BED) created the "minimum data set" by identifying the data elements required under Indiana law for laboratories to report to public health agencies. This initial set was then augmented using data elements reported in the public health literature to be useful to public health agencies for surveillance activities. This refined set was then provided to public health professionals and researchers for review and comment. Feedback was used to create a final list of data elements that would profile a given ELR message set's completeness.

Next a "completeness profile" was calculated for two samples of real-world ELR messages using the minimum data set. Each profile was constructed by dividing the number of values present in a given field by the total possible values that could have been populated in that field. The profiles were then compared to one another.

The study data originated from production information systems utilized by a variety of clinical settings and an HIE. The first sample contained "raw" (unaltered, unedited) Health Level 7 (HL7) messages (Version 2.x) received from 168 distinct hospital and laboratory information system interfaces during a one month period (November 14, 2010 to December 15, 2010) by the INPC, an operational HIE (27) that includes integrated delivery networks, hospitals, independent laboratories, physician practices, radiology centers, and the Indiana State Department of Health (ISDH). These messages were extracted from the INPC's inbound message queue and parsed into a relational database composed of tables that represent logical HL7 segments (e.g., PID, OBR, OBX). Each field within the tables corresponded to an individual field within a HL7 segment (e.g., PID-1, OBR-16, OBX-3). To parse the messages, we employed a clone of the INPC's production methods for deconstructing HL7 messages. These methods are used to receive and process real-world HL7 messages from a variety of clinical information systems, and they have been refined and validated over the INPC's 16 years of operation.

The second sample contained "enhanced" HL7 messages (Version 2.x) representing 49 distinct hospital and laboratory sources processed by the INPC during the same timeframe. These messages were extracted from the outbound message queue which contains reportable messages bound for the state public health agency. The INPC utilizes the Regenstrief Notifiable Condition Detector (NCD) to critically examine HL7 messages from INPC interfaces that potentially contain notifiable disease results. Messages determined to contain reportable results are sent from the outbound queue to the ISDH on behalf of the INPC and its member institutions (e.g., hospitals, labs).

The NCD further enhances the HL7 messages through validation and augmentation methods. For example, local laboratory codes contained within the OBX-3 field are mapped to equivalent Logical Observation Identifiers Names and Codes (LOINC) codes. The LOINC codes are appended to the original messages prior to transmission to the ISDH. The NCD further examines incoming messages for provider information (e.g., National Provider Identifier, address of the hospital or practice, phone number for the department or clinic) and attempts to add any missing provider information found in a table of providers stored in the INPC knowledge repository. Furthermore, labs may improperly place units of measure in a comment field, so the NCD examines comments for key data like units of measure and copies relevant data it finds to the appropriate HL7 field. The messages in this sample were parsed using the same methods as the raw sample into the same relational database for analysis.

Structured query language (SQL) statements were executed to calculate the completeness of each HL7 field within both samples. Each field's "percent complete" was calculated by dividing the count of non-null values by the total number of possible values for that field. The calculated values were input into a completeness profile for each sample, and the difference between the completeness scores across samples was calculated.

Results

Aim 1: Key Fields that Support Notifiable Disease Surveillance Processes

The result of the first aim in this study was the development of a novel method for measuring completeness and comparing the completeness of two or more data sets. We first defined a "minimum data set" that contains the data elements specifically required in state law that are to be reported to public health agencies for notifiable conditions. In Indiana, these data elements are defined in the Indiana Administrative Code (IAC) under 410 IAC 1-2.3-48.

In addition to what is minimally required by law, the minimum data set was constructed to also include those additional elements for which evidence suggests the data aid public health professionals in notifiable disease surveillance processes. A number of peer-reviewed ELR studies (10, 16, 22, 28), as well as white papers published by public health professional organizations such as the International Society for Disease Surveillance (ISDS) and the Council on State and Territorial Epidemiologists (CSTE), discussed useful fields including sex, race, and ethnicity.

The final data set was augmented with "units of measure" as suggested by a group of experts working at and in close proximity to Indiana University's School of Medicine who were consulted for the project. The experts were provided with a draft data set that included the IAC and evidence-based elements. Some of the experts considered units to be helpful since many lab tests are often identical except for the kind of quantity examined in the specimen (29). For example, the concentration of sodium in a urine sample can be measured in terms of its mass concentration (ug/mL) or molar concentration (mmol/L).

Once the minimum data set was defined, the lead author mapped each data element to one or more corresponding fields from the HL7 Version 2 technical specification. The final data elements, their corresponding HL7 fields, and the source of their usefulness are summarized in Table 1.

Key Data Element	Corresponding HL7 Field(s)	Source
Patient's Identifier	Patient Identifier (PID-3)	(30)
Patient's Name	Patient Name (PID-5)	IAC
Patient's Date of Birth	Date of Birth (PID-7)	IAC
Sex (Gender)	Administrative Sex (PID-8)	(16, 28)
Race	Race (PID-10)	(28)
Patient's Address	Patient Address (PID-11)	IAC
Patient's Home Phone Number	Phone Number (PID-13)	(16, 28)
Ethnicity	Ethnic Group (PID-22)	(28)
Name of Attending Physician or Hospital or Clinic or Submitter	Ordering Provider (OBR-16) Ordering Facility Name (ORC-21) Staff Name (STF-3)	IAC
Telephone Number of Attending Physician or Hospital or Clinic or Submitter	Order Callback Phone Number (OBR-17) Ordering Facility Phone Number (ORC-23) Staff Phone (STF-10)	IAC
Address of Attending Physician or Hospital or Clinic or Submitter	Staff Office/Home Address (STF-11) Ordering Provider Address (ORC-24)	IAC
Test Name	Observation Identifier (OBX-3)	IAC
Test Results or Laboratory Interpretation of Test Results	Observation Value (OBX-5)	IAC
Specimen Source	Specimen Source (OBR-15)	(16)
Units of Measure	Units (OBX-6)	Experts
Normal Range	Reference Range (OBX-7)	IAC
Abnormal Flag	Abnormal Flags (OBX-8)	(16)
Status of Test Result	Observation Result Status (OBX-11)	(22)

Table 1 – A Minimum Data Set for Electronic Laboratory Reporting
IAC = Indiana Administrative Code

Aims 2 and 3: Measurement and Comparison of Real-World ELR Data Completeness

The first sample contained 7,592,039 messages from the INPC's "raw" queue for incoming messages. In the raw sample, there were 7,592,039 possible values for fields within the PID segment, 7,471,001 possible values for fields within the OBR segment, and 22,244,305 possible values for fields within the OBX segment.

The second sample contained 16,365 messages from the Regenstrief NCD post-processed queue of reportable results. In the enhanced sample, there were 16,365 possible values within the PID segment, 35,266 possible values for fields within the OBR segment, and 131,665 possible values within the OBX segment.

Table 2 summarizes the calculated completeness for each field in the two samples. The first column contains the key data element name. The second column contains the corresponding HL7 field name. The third column contains the "percent complete" for each field in the raw sample. The fourth column contains the "percent complete" for each field in the enhanced sample. The final column contains the difference between the two "percent complete" values across the samples.

Key Data Element	Corresponding HL7 Field	Percent Complete Raw	Percent Complete Enhanced	Difference
Patient's Identifier	Patient Identifier (PID-3)	99.9%	100%	+0.01%
Patient's Name	Patient Name (PID-5)	99.4%	100%	+0.06%
Patient's Date of Birth	Date of Birth (PID-7)	97.8%	99.8%	+2.0%
Sex (Gender)	Administrative Sex (PID-8)	95.8%	99.9%	+4.1%
Race	Race (PID-10)	38.4%	60.3%	+21.9%
Patient's Address	Patient Address (PID-11)	41.5%	63.3%	+21.8%
Patient's Home Phone Number	Phone Number (PID-13)	38.5%	72.8%	+34.3%
Ethnicity	Ethnic Group (PID-22)	3.5%	18.3%	+14.8%
Name of Attending Physician or Hospital or Clinic or Submitter	Ordering Provider (OBR-16)	57.4%	66.5%	+8.9%
Telephone Number of Attending Physician or Hospital or Clinic or Submitter	Callback Number (OBR-17) Staff Phone (STF-10)	0.15%	73.3%	+73.2%
Address of Attending Physician or Hospital or Clinic or Submitter	Staff Office/Home Address (STF-11)	N/A	84.6%	+84.6%
Test Name	Observation Identifier (OBX-3)	99.3%	100%	+0.07%
Test Results or Laboratory Interpretation of Test Results	Observation Value (OBX-5)	96.3%	98.9%	+2.6%
Specimen Source	Specimen Source (OBR-15)	13.7%	28.7%	+15.0%
Units of Measure	Units (OBX-6)	57.0%	17.5%	-39.5%
Normal Range	Reference Range (OBX-7)	55.8%	18.3%	-37.5%
Abnormal Flag	Abnormal Flags (OBX-8)	33.0%	28.4%	-4.6%
Status of Test Result	Observation Result Status (OBX-11)	92.8%	99.5%	+6.7%

Table 2 – Comparison of INPC Completeness Profiles

The difference between fields across the two samples varied from 0.01% to 84.6%. The completeness for most of the fields increased, although several fields (Units of Measure, Normal Range, Abnormal Flag) decreased. The larger differences (Provider Phone Number, Provider Address) were observed for fields for which the data were directly enhanced by the INPC. Other fields varied in their completeness, although these variations are not attributable to the HIE's enhancement processes.

Discussion

To effectively perform surveillance and their other core functions, public health agencies require access to "timely, accurate, and complete data" (17). The results of this study confirm that laboratory data from clinical information systems are heterogeneous in their completeness. In many cases, data important to public health surveillance processes are missing, indicating suboptimal ELR data quality. The study further demonstrates that HIEs employ methods that can mitigate ELR data deficiencies, improving the completeness of lab data electronically transmitted to public health information systems.

First, the study created a novel method for assessing the completeness of clinical data. While much of the literature on ELR and public health reporting focuses on improving the number of reportable cases submitted to public health agencies, this study measured the completeness of the data within individual reported cases. In previous studies, completeness of a data source is assigned a single value. For example, Effler et al. reported that the electronic communicable disease reporting system accounted for 91% of unique cases of notifiable disease (10). Heterogeneity of data completeness, however, makes it difficult to score an entire information system or data source as being 90% or 40% complete. A single score obscures whether the data source could adequately provide the data elements needed for recipient A versus recipient B. A system tracking spatial-temporal disease spread would benefit from a data source with a more complete address data. More complete address data would not, however, be useful for a statistical service that identifies when the number of reported cases rises above a certain threshold. Therefore this study assigned a percent complete to each of the data elements considered important to notifiable disease surveillance processes. A similar approach should be considered in the future when evaluating data sources to ensure that data consumers (humans or machines) understand the characteristics of the data from those sources.

The study also quantified what many in informatics are likely to encounter routinely: clinical data are heterogeneous in their completeness across and within information systems. Some laboratory information systems almost always transmit the specimen source (e.g., blood, urine), while others almost never provide this data element. Although this concept is not new, it is rarely measured and published.

Unfortunately many public health officials, like data consumers in other health care segments and industries, believe that data is easily and uniformly captured and stored across the spectrum of health care services. Data however are captured for a specific purpose, and the collection of additional data elements is costly. Additional data elements require staff to ask for and then record the information, which translates into additional time and labor. Therefore data consumers must understand the impact of the cost of data collection on the characteristics of data captured in various environments, like their completeness, when making decisions about secondary use. Public health officials, for example, might benefit from understanding that elements like the provider's phone number and address have little clinical relevance to the physician receiving the results of a lab test. These fields are poorly populated by laboratory information systems. Although these fields are required according to state (e.g., IAC) and federal (e.g., meaningful use) regulations, it does not guarantee that they will be complete and available for public health surveillance processes. Few addresses are provided today directly from the labs in the INPC; and very few phone numbers are provided. Thus policies to require additional data elements are unlikely to impact data collection processes unless laboratories and hospitals are incentivized to capture the additional data elements needed for public health surveillance processes.

Comparing the raw ELR messages with messages enhanced by the INPC demonstrates that the INPC employs methods that can improve the completeness of data. The completeness of nearly every field in the enhanced sample was larger than the equivalent fields in the raw sample. The improvements in completeness for provider names, addresses, and phone numbers were a direct result of HIE processes designed to enhance provider information. The INPC identifies all providers present anywhere in the message and resolves their identities using its Master Provider Index. The Master Provider Index is similar to master patient indices given that its function is to store a central list of all providers known to the INPC. The index possesses data elements such as the provider's name, clinic address, phone number, role (e.g., physician, physician assistant), and staff ID number. Using its Master Provider Index, the INPC is able to dramatically increase the amount of provider detail for the messages sent to the state health agency.

The Master Provider Index, however, was not specifically created for the public health reporting use case. The INPC has a more practical use of the index: the accurate delivery of lab results, radiology dictations, and other clinical documents to clinicians. There is intrinsic value to the INPC for knowing who providers are and where they practice to enable results delivery as well as other core HIE functions like access control. Such re-use of core HIE functionality is a benefit beyond improvements in data completeness. Leveraging core functions is one way that HIEs can support public health with little incremental cost. This is important, because in a recent survey of public health officials regarding participation in an HIE financial cost was a major concern (31). If multiple HIE participants are able to benefit from the same core set of technologies, then costs for all participants can be shared and become more reasonable. Sustainability is a top priority for many HIEs, many of which struggle to support themselves when initial grant funding ends.

In addition to improving completeness and leveraging common infrastructure, HIE enhancements to laboratory data will improve public health surveillance and clinical workflow. Missing patient information necessitates a phone call from public health to request, for example, a patient's phone number. This would require the public health nurse to pause the investigation of a new notifiable disease case until the phone number could be identified. It would further require a nurse or other resource at the hospital or clinic to retrieve the voice mail, log into the EHR system (or pull a chart), extract the needed information, and call the public health department to provide the information. Inefficiencies due to data quality issues result in unnecessary costs and disruptions to routine clinical and public health workflows. Therefore any enhancements by HIEs to ELR data will improve public health surveillance processes for both clinicians and public health professionals.

While many fields in the enhanced messages have higher completeness, some fields have lower completeness. The differences between the raw and enhanced messages for these fields, however, are independent of the INPC's internal processes and enhancement methods. The INPC never removes data from a message; the HIE only adds information to or alters the value of a particular field. These differences reflect primarily heterogeneity in the data sources and message types. The raw sample consists of messages from 168 unique data senders, and the data pertain to all types of lab results (e.g., routine tests like white blood cell counts and hemoglobin A1c). The enhanced sample contains messages from 49 unique data senders and pertains only to positive notifiable disease results (e.g., sexually transmitted infections, lead levels in blood). A higher proportion of the enhanced sample contains microbiological cultures or micro results. For micro results, the units and normal range fields should be null as cultures are resolved through interpretation by a human lab technician. Future analyses of message completeness will control for this fact.

Abnormal result flags were also missing more often in the enhanced sample than the raw sample. Approximately five percent fewer abnormal flag values were observed. This outcome can also be explained by the fact that the enhanced sample contained a higher proportion of microbiological cultures. Micro results tend to be reported in a single field within the HL7 message (OBX-5), and some labs place micro results wholly in an NTE segment (a kind of comment field) at the end of the HL7 message. Values of "positive" embedded within an OBX-5 or comment field are challenging to process. Better use of abnormal flags would improve the Regenstrief NCD's ability, and other clinical information systems, to identify and route notifiable cases to clinicians and public health agencies.

Limitations

A limitation of this study is that the impacts upon public health surveillance processes are only estimated. While the literature provides some evidence on the impact of poor data quality, the specific impact of missing data in surveillance processes was not measured. Furthermore, measurable improvements to clinical and public health workflows as a result of INPC data enhancements were not captured in this study. This is work that researchers affiliated with the Indiana Center of Excellence in Public Health Informatics hope to perform in the future.

Additional work for the future includes the development and evaluation of processes to enhance patient-level data. The INPC plans to leverage its Master Patient Index in the same way that it currently leverages the provider index. We hypothesize that this will support a reduction in the number of calls to clinics and hospital wards to obtain additional details about patients who test positive for sexually transmitted infections and other notifiable diseases.

Finally the INPC is arguably one of the most robust and successful HIEs in the U.S. The INPC has partnered with local and state health agencies numerous times for over a decade to improve public health reporting, and the INPC has invested heavily in the development and maintenance of its Master Provider Index. Therefore the results of studies on data within in the INPC may not be generalizable to all HIEs and regions.

Conclusions

Poor quality data exists in clinical information systems, which presents a challenge for those interested in secondary uses of electronic health record data. For public health reporting, a single secondary use case, many data elements necessary to support surveillance processes are missing. Methods employed by HIEs to improve data quality can be leveraged by public health agencies to improve completeness, supporting both local needs to investigate disease outbreaks and federal goals to create meaningful use of EHR systems.

Although there is great opportunity for public health agencies, HIEs, hospitals, and laboratories to collectively improve public health surveillance processes, a number of challenges remain. Financial incentives to stimulate collaboration and data exchange may be necessary in some regions. Better use of existing standards, like the abnormal flag field in HL7, will be necessary to improve identification of notifiable results. Finally, data consumer expectations need to be tempered to recognize not only the possibilities of HIE but also the limitations of certain data sources and systems.

Furthermore, the cost of collecting additional data in EHR and laboratory systems must be better understood by all stakeholders in health information exchange. Financial or other incentives may be required to drive changes to existing data collection workflow. New methods for collecting data *de novo* or leveraging existing data that minimize impact on workflow should be explored.

Ultimately, through research, development, and practice, we can build an information infrastructure capable of supporting secondary uses of electronic clinical data. This infrastructure will enable further improvements in public health surveillance processes not only in Indiana but across many states and regions.

Acknowledgements

The authors acknowledge Roland Gamache and Josette Jones, the other members of Dr. Dixon's doctoral research committee, for their guidance and support on the project and feedback on this paper. The authors further thank Shahid Khokhar of the Regenstrief Institute for his help with the extraction of the HL7 message data into a format suitable for analysis.

This study was supported, in part, by the Indiana Center of Excellence in Public Health Informatics through a grant award (1P01HK000077-01) from the U.S. Centers for Disease Control and Prevention. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention.

References

1. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007 Jan-Feb;14(1):1-9.
2. Elkin PL, Trusko BE, Koppel R, Speroff T, Mohrer D, Sakji S, et al. Secondary use of clinical data. Studies in health technology and informatics. 2010;155:14-29.
3. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems. 1996;12(4):5-34.
4. Redman TC. The Impact of Poor Data Quality on the Typical Enterprise. Communications of the ACM. 1998;41(2):79-82.
5. Bailey JE, Pearson SW. Development of a tool for measuring and analyzing computer user satisfaction. Manaement Science. 1983;29(5):530-45.
6. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997 Sep-Oct;4(5):342-55.
7. Hasan S, Padman R. Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. AMIA Annu Symp Proc. 2006:324-8.
8. Stein HD, Nadkarni P, Erdos J, Miller PL. Exploring the degree of concordance of coded and textual data in answering clinical queries from a clinical data repository. J Am Med Inform Assoc. 2000 Jan-Feb;7(1):42-54.
9. Doyle TJ, Glynn MK, Groseclose SL. Completeness of notifiable infectious disease reporting in the United States: an analytical literature review. Am J Epidemiol. 2002 May 1;155(9):866-74.
10. Effler P, Ching-Lee M, Bogard A, Ieong MC, Nekomoto T, Jernigan D. Statewide system of electronic notifiable disease reporting from clinical laboratories: comparing automated reporting with conventional methods. JAMA. 1999 Nov 17;282(19):1845-50.
11. AHRQ. HHS Awards $139 Million To Drive Adoption of Health Information Technology. Rockville, MD: Agency for Healthcare Research and Quality; [updated October 13, 2004August 3, 2009]; Available from: http://www.ahrq.gov/news/press/pr2004/hhshitpr.htm.
12. Health Resources and Services Administration US. Justification of Estimates for Appropriations Committees. [updated 2007July 20, 2007]; Available from: ftp://ftp.hrsa.gov/about/budgetjustification08.pdf.

13. New York State Department of Health. Health Information Technology (Health IT) Grants - HEAL NY Phase 1. 2006 [updated May; cited 2011 February 18]; Available from: http://www.health.state.ny.us/technology/awards/.

14. Robert Wood Johnson Foundation. Funding. 2010 [cited 2011 February 18]; Available from: http://www.projecthealthdesign.org/about/funding.

15. GAO. EMERGING INFECTIOUS DISEASES: Review of State and Federal Disease Surveillance Efforts. Washington, D.C.: U.S. Government Accountability Office2004. Report No.: GAO-04-877.

16. Overhage JM, Grannis S, McDonald CJ. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. Am J Public Health. 2008 Feb;98(2):344-50.

17. Lombardo JS, Buckeridge DL, editors. Disease Surveillance: A Public Health Informatics Approach. Hoboken: John Wiley & Sons; 2007.

18. Baxter R, Rubin R, Steinberg C, Carroll C, Shapiro J, Yang A. Assessing Core Capacity for Infectious Diseases Surveillance. The Lewin Group, Inc.; 2000 [cited 2010 March 8]; Available from: www.lewin.com/content/publications/808.pdf.

19. Lober WB, Karras BT, Wagner MM, Overhage JM, Davidson AJ, Fraser H, et al. Roundtable on bioterrorism detection: information system-based surveillance. J Am Med Inform Assoc. 2002 Mar-Apr;9(2):105-15.

20. AHRQ. AHRQ Research Relevant to Bioterrorism Preparedness. Rockville, MD: Agency for Healthcare Research and Quality; 2002 [updated March; cited 2010 March 17]; Available from: http://www.ahrq.gov/news/focus/bioterror.htm.

21. Magruder C. Public Health/Health Information Exchange Collaborative: A Model for Advancing Public Health Practice. Online Journal of Public Health Informatics [serial on the Internet]. 2010; 2(2): Available from: http://ojphi.org/htbin/cgiwrap/bin/ojs/index.php/ojphi/article/view/3217/0.

22. Panackal AA, M'Ikanatha N M, Tsui FC, McMahon J, Wagner MM, Dixon BW, et al. Automatic electronic laboratory-based reporting of notifiable infectious diseases at a large health system. Emerg Infect Dis. 2002 Jul;8(7):685-91.

23. Nguyen TQ, Thorpe L, Makki HA, Mostashari F. Benefits and barriers to electronic laboratory results reporting for notifiable diseases: the New York City Department of Health and Mental Hygiene experience. Am J Public Health. 2007 Apr;97 Suppl 1:S142-5.

24. Vreeman DJ, Stark M, Tomashefski GL, Phillips DR, Dexter PR. Embracing change in a health information exchange. AMIA Annu Symp Proc. 2008:768-72.

25. Zafar A, Dixon BE. Pulling back the covers: technical lessons of a real-world health information exchange. Studies in health technology and informatics. 2007;129(Pt 1):488-92.

26. Wurtz R, Cameron BJ. Electronic laboratory reporting for the infectious diseases physician and clinical microbiologist. Clin Infect Dis. 2005 Jun 1;40(11):1638-43.

27. Grannis SJ, Biondich PG, Mamlin BW, Wilson G, Jones L, Overhage JM. How disease surveillance systems can serve as practical building blocks for a health information infrastructure: the Indiana experience. AMIA Annu Symp Proc. 2005:286-90.

28. Vogt RL, Spittle R, Cronquist A, Patnaik JL. Evaluation of the timeliness and completeness of a Web-based notifiable disease reporting system by a local health department. J Public Health Manag Pract. 2006 Nov-Dec;12(6):540-4.

29. McDonald C, Huff SM, Mercer K, Hernandez JA, Vreeman DJ. Logical Observation Identifiers Names and Codes (LOINC®) Users' Guide. [PDF] Indianapolis: Regenstrief Institute; 2010 [cited 2011 January 28]; Available from: http://loinc.org/downloads/files/LOINCManual.pdf.

30. CSTE. Common Core Data Elements for Case Reporting and Laboratory Result Reporting. Atlanta: Council of State and Territorial Epidemiologists2009 Contract No.: 09-SI-01.

31. Hessler BJ, Soper P, Bondy J, Hanes P, Davidson A. Assessing the relationship between health information exchanges and public health agencies. J Public Health Manag Pract. 2009 Sep-Oct;15(5):416-24.