

Variation in Information Needs and Quality: Implications for Public Health Surveillance and Biomedical Informatics

Brian E. Dixon, MPA, PhD^{a,b,c}, Patrick T. S. Lai, MPH^a, Shaun J. Grannis, MD, MS^{b,d}

^aDepartment of BioHealth Informatics, Indiana University School of Informatics and Computing, Indiana University-Purdue University Indianapolis, Indianapolis, IN;

^bCenter for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN;

^cCenter for Implementing Evidence-Based Practice, Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service, Indianapolis, IN;

^dDepartment of Family Medicine, School of Medicine, Indiana University; ^e Indiana Health Information Exchange, Indianapolis, IN

Abstract

Understanding variation among users' information needs and the quality of information in an electronic system is important for informaticians to ensure data are fit-for-use in answering important questions in clinical and public health. To measure variation in satisfaction with currently reported data, as well as perceived importance and need with respect to completeness and timeliness, we surveyed epidemiologists and other public health professionals across multiple jurisdictions. We observed consensus for some data elements, such as county of residence, which respondents perceived as important and felt should always be reported. However information needs differed for many data elements, especially when comparing notifiable diseases such as chlamydia to seasonal (influenza) and chronic (diabetes) diseases. Given the trend towards greater volume and variety of data as inputs to surveillance systems, variation of information needs impacts system design and practice. Systems must be flexible and highly configurable to accommodate variation, and informaticians must measure and improve systems and business processes to accommodate for variation of both users and information.

Background and Significance

At their core, clinical and public health are information enterprises, and data are their lifeblood. Nearly all clinical and public health activities involve generating, collecting, storing, analyzing, or sharing data about individual patients or populations. Data are particularly vital to joint activities including disease surveillance (1), population health assessments (2), and health care policy. Effective clinical and public health practice therefore requires access to representative, complete, and timely data from multiple sources (3, 4). Attributes of data such as representativeness, completeness, and timeliness are characterized in the information sciences literature as dimensions of data quality (5).

The quality of data is most often measured with respect to its fit-for-use in a given context (8), and unfortunately the quality of data across the myriad sources from which they are captured, stored, analyzed, or reported is heterogeneous and often poor (6, 7). Therefore various contexts in clinical and public health necessitate different levels of data quality. For example, emergency department chief complaint data are very timely but often lack specificity, making these data useful for early detection of unexplained increases in health care utilization but increased activity does not always signal a true outbreak of disease (9).

Understanding the characteristics of data and their sources, including quality, is important for informaticians to ensure data are fit-for-use to support important processes in clinical and public health (10). Knowing the quality of data is particularly important given that data are increasingly available to clinical and public health professionals through a variety of information systems, including electronic health record (EHR), decision support, and surveillance systems (11).

To ensure data are fit-for-use, users (or data consumers) must first articulate their information needs within specified contexts of use. Gathering and documenting the data requirements for a given context of use is often referred to as an information needs assessment, which is often recommended for business process analysis, workflow redesign, and the design and deployment of user-centered information systems in an organization, including immunization and surveillance systems in public health departments (12-14). Unfortunately there is limited published information available on the information needs in public health with respect to surveillance systems. While several surveillance systems have been deployed and a variety of jurisdictions utilize such systems, it is unclear whether the information

needs across jurisdictions vary with respect to completeness, timeliness, and other key factors identified as barriers to traditional surveillance prior to the deployment of e-surveillance systems.

To measure 1) satisfaction with currently reported data; 2) perceived importance and need with respect to completeness and timeliness of specific data elements; and 3) consensus in responses across multiple jurisdictions, we surveyed epidemiologists and other public health professionals. In this article we describe the development and administration of the survey as well as the results. We then discuss the findings in the context of advancing public health surveillance in an era of increasingly electronic data sources available for use by health departments. Our methods and findings are further applicable to other disciplines within the field of biomedical informatics.

Materials and Methods

Study participants

Health officials and surveillance program managers in eight states were contacted directly by the researchers and invited to participate in the study. Following permission from health department officials, the researchers then invited front-line health department surveillance staff via email to complete the survey. Most public health professionals have access to email and the Internet at their place of work. Furthermore, state health department employees were given permission to use state resources to complete the survey during their normal working hours.

The study originated with state and county health departments participating in the U.S. Centers for Disease Control and Prevention (CDC) funded Indiana Center of Excellence in Public Health Informatics (ICEPHI). However, the study's scope expanded to include other jurisdictions following a discussion of the project at a workshop convened by the Public Health Informatics Institute (PHII), where surveillance officials from several states expressed a desire to participate in the study, so the research team then reached out to an additional six states. The research team ultimately invited 45 individuals at 15 health departments in 8 states to participate in the study.

Survey design

The study used a novel questionnaire designed by the researchers to examine the perceptions and needs of communicable disease program staff with respect to data quality. Parts of the instrument used model questions and concepts from previous studies examining public health professionals' information needs (15). However, the majority of the questionnaire was developed by integrating the dimensions of data quality with specific communicable disease use cases. The use cases were iteratively developed by informatics and epidemiology professionals associated with ICEPHI. The group selected a subset of diseases of high priority that represented the spectrum of disease classes encountered by health departments: seasonal illness (e.g., influenza), sexually transmitted infections (e.g., chlamydia), acute illness (e.g., pertussis), high priority illness (e.g., meningitis), and chronic illness (e.g., diabetes). The selected data elements were drawn from those required under Indiana state law to be reported to health departments and aligned with the core data elements outlined by the Council for State and Territorial Epidemiologists (CSTE) (16).

Prior to use by respondents, the questionnaire was pilot tested by several public health professionals in county and state health departments. These participants provided feedback to the study team with respect to unclear and inappropriate items as well as wording. The instrument was modified following the pilot testing.

We invited participants via email to complete a web-based survey. Reminder emails were sent to invited participants to encourage participation. The survey was fielded between September 2011 and February 2012. The extended time was necessary due to changing the scope from Indiana-only to a survey of health department staff in multiple states and jurisdictions.

Skip logic was used to minimize time burden on respondents, advancing respondents beyond sections that did not pertain to their role within the health department. Furthermore, respondents were given the choice to skip unclear and other questions they felt they could not answer. For example, if the respondent indicated that he or she did not work with influenza data, then the survey instrument skipped over the series of questions regarding influenza data quality. Within each section of the survey, respondents could skip individual questions or items related to completeness, timeliness, or satisfaction.

Survey data collection

The survey included questions examining respondent characteristics as well as respondents' perceptions of data quality. Demographic questions sought to assess respondents' place of employment, role, and length of service. All other questions focused on respondent satisfaction with the quality of data submitted to the health department, including the relative importance, completeness, and timeliness of various data elements.

The survey used Likert-style scales for respondents to rate various dimensions of data quality and satisfaction for each of the common data elements included in reports for the selected diseases. For example, respondents were asked to indicate their level of satisfaction overall with existing data for each disease using the following scale: 1 = Very Dissatisfied, 2 = Dissatisfied, 3 = Neither Dissatisfied or Satisfied, 4 = Satisfied, and 5 = Very Satisfied. For each data element reported for a particular disease, respondents were asked to rate the relative importance of the element using the following scale: 1 = Highly Unimportant (data element provides no value to the surveillance process; I never use this data element), 2 = Unimportant (data element provides minimal value to the surveillance process; I rarely use this data element), 3 = Neither Unimportant nor Important (data element provides limited value to the surveillance process; I use this data element, but I use it only because I am required to by policy or law), 4 = Important (data element adds value to the surveillance process; I often use this data element), 5 = Highly Important (the surveillance process would be difficult if this data element was missing; I almost always use this data element).

In addition to importance, respondents were asked to characterize the completeness of the data received in notifiable disease reports. The following scale was used to measure respondents' perception of completeness: 1 = Mostly Incomplete (Defined as 0.00% to 29.99% complete; it is okay for this value to be missing most of the time), 2 = Marginally Complete (Defined as 30.00% to 49.99% complete; it is okay for this value to be missing much of the time), 3 = Mostly Complete (Defined as 50.00% to 79.99% complete; it is okay for this value to be missing some of the time), 4 = Nearly Complete (Defined as 80.00% to 99.99% complete; it is okay for this value to be missing occasionally), and 5 = Always Complete (Defined as 100% complete; it is not okay for this value to ever be missing).

Finally, the survey asked respondents to indicate their desired and minimally acceptable levels of timeliness associated with data reported across diseases. To measure the ratings involved with respondent's perception on timeliness, the following scale was used: 1 = As soon as a clinician determines the diagnosis, 2 = Within 24 hours of suspicion or diagnosis, 3 = Within 48 hours of suspicion or diagnosis, 4 = Within 72 hours of suspicion or diagnosis, 5 = Within 1 week of suspicion or diagnosis, or 9 = No preference or opinion.

Statistical Analysis

The goals of the analysis were to 1) describe respondents' perceptions of notifiable disease report data quality with respect to satisfaction, importance, desired completeness, and desired timeliness; and 2) measure respondents' consensus across jurisdictions. Descriptive statistics were utilized to characterize respondents' demographics, and Chronbach's alpha was used to assess reliability. To analyze perceptions of data quality, we used median values and the mean absolute deviation (MAD) from the median, a robust measure that represents the average of the absolute differences between each score assigned by the respondent and the overall mean (17, 18).

The analysis further examined respondents' consensus regarding perceived importance of data elements and the minimally acceptable levels of completeness. We defined consensus as 70% or greater respondent agreement on the importance or minimum level of completeness for a data element. This *a priori* level of consensus has been used in previous studies (19, 20). All analyses were performed using SPSS Version 20.

Results

Demographics

Table 1 depicts information on the respondents' characteristics. A total of 33 individuals from more than six states responded to the survey. Among respondents, a majority (60.6%) worked at a local health department, and nearly one-third (30.3%) worked at a state health departments. Two-thirds of the respondents (66.7%) stated they served as an epidemiologist within their organization. Three-quarters (75.8%) of respondents reported working in public health for at least 10 years. A majority of respondents indicated they were previously asked to describe their information needs and participate in projects that gathered public health workers' information needs.

Table 1. Respondent characteristics and demographic information

	Frequency
Place of employment	
Local Health Department	20 (60.6%)
State Health Department	10 (30.3%)
Academic Institution	2 (6.1%)
Federal Government	1 (3.0%)
Title or role within organization	
Epidemiologist	22 (66.7%)
Program Manager/Coordinator	4 (12.1%)
Public Health Nurse	2 (6.1%)
Informaticist (IT Specialist, Health Informaticist)	2 (6.1%)
Other (Infectious Disease Physician, Researcher, Evaluator)	3 (9.1%)
Years worked with public health	
<1 year	1 (3.0%)
1-3 years	3 (9.1%)
3-5 years	1 (3.0%)
5-10 years	3 (9.1%)
10-15 years	10 (27.3%)
>15 years	16 (48.5%)
Have you ever been asked to outline or describe your needs when it comes to the data and information you use to perform your job?	
Yes	24 (72.7%)
No	7 (21.2%)
No Response	2 (6.1%)
Have you ever participated in a project to gather or analyze the data or information needs of others working in public health?	
Yes	19 (57.6%)
No	14 (42.4%)

Reliability

Calculation of Chronbach's alpha (α) ranged from 0.93 to 0.99 across diseases and dimensions of data quality with the exception of chlamydia for which the sample size was too small. Reliability is considered excellent when α is greater than 0.90.

Satisfaction of data received

Table 2 depicts respondents' satisfaction with the data they currently receive. Many respondents were satisfied with the data they received. However, the levels of satisfaction varied across diseases. For three of the diseases (influenza, meningitis, and pertussis), more than two-thirds of respondents indicated they were satisfied or highly satisfied with existing data. For chlamydia and diabetes, however, respondents were less satisfied. For chlamydia, respondents were evenly split between being satisfied or dissatisfied. For diabetes, five respondents (62.5%) were ambivalent about current data, and two respondents (25%) indicated they were very dissatisfied with existing data.

Table 2. Satisfaction of data currently received by disease

Disease	Very		Neither		Very
	Dissatisfied	Dissatisfied	Dissatisfied or Satisfied	Satisfied	Satisfied
Influenza (<i>N</i> = 18)	0 (0.0%)	2 (11.1%)	3 (16.7%)	12 (66.7%)	1 (5.6%)
Chlamydia (<i>N</i> = 5)	0 (0.0%)	2 (40.0%)	1 (20.0%)	2 (40.0%)	0 (0.0%)
Meningitis (<i>N</i> = 19)	0 (0.0%)	2 (10.5%)	4 (21.1%)	9 (47.4%)	4 (21.1%)
Pertussis (<i>N</i> = 12)	0 (0.0%)	2 (16.7%)	2 (16.7%)	7 (58.3%)	1 (8.3%)
Diabetes (<i>N</i> = 8)	2 (25.0%)	0 (0.0%)	5 (62.5%)	1 (12.5%)	0 (0.0%)

Ratings are Very Dissatisfied (median = 1), Dissatisfied (median = 2), Neither Satisfied or Dissatisfied (median = 3), Satisfied (median = 4), and Very Satisfied (median = 5).

Importance of data element

Table 3 summarizes respondent ratings for reportable data elements across all five diseases with respect to the data elements' relative importance to public health work processes. The values represent median scores from the Likert scale and the mean absolute deviation (MAD). Data elements for which consensus were achieved are bolded for emphasis. Importance ranged from 2.0 (Unimportant) to 5.0 (Highly Important) with MAD values ranging from 0.00 to 1.75. Notably, respondents failed to achieve consensus on any data elements for diabetes.

Table 3. Importance of data element in public health work process by disease

	Influenza (<i>N</i> = 18)	Chlamydia (<i>N</i> = 5)	Meningitis (<i>N</i> = 19)	Pertussis (<i>N</i> = 12)	Diabetes (<i>N</i> = 8)
Data Element					
Patient Name	2.5 (1.11)	5.0 (0.00)	5.0 (0.72)	5.0 (0.18)	2.0 (1.13)
Patient Address	3.5 (1.06)	5.0 (0.25)	5.0 (0.78)	5.0 (0.36)	3.0 (1.13)
Patient City	4.0 (0.89)	5.0 (0.00)	5.0 (0.72)	5.0 (0.36)	3.0 (1.00)
Patient Zip Code	4.0 (0.56)	5.0 (0.25)	5.0 (1.00)	5.0 (0.45)	4.0 (0.38)
Patient County of Residence	5.0 (0.44)	5.0 (0.25)	5.0 (0.78)	5.0 (0.27)	4.0 (1.00)
Patient Primary Phone Number	2.5 (1.11)	4.5 (0.75)	5.0 (0.72)	5.0 (0.27)	2.5 (1.25)
Patient Date of Birth or Age	5.0 (0.44)	5.0 (0.25)	5.0 (0.89)	5.0 (0.18)	5.0 (0.50)
Patient Sex	4.0 (0.78)	5.0 (0.25)	4.0 (1.00)	5.0 (0.64)	5.0 (0.50)
Patient Race	4.0 (0.89)	4.5 (0.75)	4.0 (0.82)	4.0 (0.45)	5.0 (0.63)
Patient Ethnicity	4.0 (0.78)	3.5 (1.00)	4.0 (0.88)	4.0 (0.55)	4.5 (0.88)
Patient Pregnancy Status	4.5 (0.94)	4.5 (1.00)	5.0 (1.00)	4.0 (0.55)	4.0 (0.63)
Patient's Next of Kin Name	2.0 (0.61)	3.0 (0.50)	4.0 (1.11)	4.0 (0.82)	2.0 (0.75)
Patient's Next of Kin Primary Phone Number	2.0 (0.61)	3.0 (0.50)	4.0 (1.17)	4.0 (1.00)	2.0 (0.75)
Provider Name	4.0 (0.61)	5.0 (0.50)	5.0 (0.78)	5.0 (0.36)	3.0 (1.00)
Provider Address	4.0 (0.67)	5.0 (0.50)	4.0 (0.71)	4.0 (0.64)	3.5 (1.25)
Provider Phone Number	4.0 (0.67)	5.0 (0.50)	5.0 (0.83)	5.0 (0.45)	3.0 (0.88)
Date of Onset	5.0 (0.53)	3.5 (1.75)	5.0 (0.78)	5.0 (0.18)	3.5 (0.88)
Reason for Visit	5.0 (0.44)	4.0 (1.50)	5.0 (0.88)	4.0 (0.55)	3.5 (0.75)
Clinical Diagnosis	5.0 (0.56)	4.5 (1.25)	5.0 (0.89)	5.0 (0.36)	4.0 (0.38)
Definitive Diagnostic Test Results	5.0 (0.33)	5.0 (0.00)	5.0 (0.72)	5.0 (0.18)	4.0 (0.38)
Person Completing Report Name	3.0 (0.89)	5.0 (1.00)	4.0 (0.71)	4.0 (0.64)	2.0 (1.38)
Person Completing Report Address	3.0 (0.67)	3.0 (1.00)	4.0 (0.94)	4.0 (0.82)	2.0 (1.38)
Person Completing Report Phone Number	3.5 (1.00)	5.0 (1.00)	4.0 (0.89)	5.0 (0.73)	3.0 (1.29)

Ratings are Highly Unimportant (median = 1), Unimportant (median = 2), Neither Important nor Unimportant (median = 3), Important (median = 4), and Highly Important (median = 5). Values are medians (mean absolute deviation from median). Values with respondent agreement of 70% or greater (consensus) are bold.

Minimally acceptable level of completeness

Table 4 summarizes respondent ratings for the minimally acceptable level of completeness for data elements across all five diseases. Values represent median scores from the Likert scale and the mean absolute deviation (MAD). Data elements for which there was consensus are bolded for emphasis. Minimally acceptable levels of completeness ranged from 1.0 (Mostly Incomplete) to 5.0 (Always Complete) with MAD values ranging from 0.00 to 2.00. Of note, Influenza achieved consensus for the only two data elements.

Table 4. Minimally acceptable level of completeness of data element by disease

	Influenza (N = 18)	Chlamydia (N = 5)	Meningitis (N = 19)	Pertussis (N = 12)	Diabetes (N = 8)
Data Element					
Patient Name	2.5 (1.79)	5.0 (0.00)	5.0 (0.00)	5.0 (0.09)	2.0 (1.71)
Patient Address	4.0 (1.50)	5.0 (0.25)	5.0 (0.38)	5.0 (0.18)	3.0 (1.71)
Patient City	5.0 (0.87)	5.0 (0.25)	5.0 (0.27)	5.0 (0.09)	4.5 (1.50)
Patient Zip Code	4.0 (0.47)	4.5 (0.75)	5.0 (0.50)	5.0 (0.36)	5.0 (0.29)
Patient County of Residence	5.0 (0.27)	4.5 (0.75)	5.0 (0.19)	5.0 (0.00)	5.0 (0.71)
Patient Primary Phone Number	1.0 (1.29)	5.0 (0.25)	5.0 (0.19)	5.0 (0.18)	2.0 (1.43)
Patient Date of Birth or Age	5.0 (0.33)	5.0 (0.25)	5.0 (0.06)	5.0 (0.09)	5.0 (0.29)
Patient Sex	4.0 (0.79)	5.0 (0.25)	5.0 (0.44)	5.0 (0.45)	5.0 (0.29)
Patient Race	3.0 (0.93)	4.0 (0.25)	3.5 (0.94)	4.0 (0.91)	5.0 (0.43)
Patient Ethnicity	3.0 (0.79)	4.0 (0.75)	3.0 (0.88)	4.0 (0.91)	4.0 (0.43)
Patient Pregnancy Status	3.5 (1.07)	3.0 (1.25)	4.0 (1.00)	4.0 (0.82)	4.0 (1.00)
Patient's Next of Kin Name	1.0 (0.69)	2.5 (0.75)	4.0 (0.81)	4.0 (0.82)	1.0 (1.00)
Patient's Next of Kin Primary Phone Number	1.0 (0.69)	2.5 (0.75)	3.5 (0.81)	4.0 (0.91)	1.0 (1.00)
Provider Name	4.0 (0.93)	5.0 (0.25)	5.0 (0.13)	5.0 (0.18)	4.0 (1.29)
Provider Address	3.0 (1.14)	4.0 (0.25)	4.0 (0.63)	4.0 (0.55)	4.0 (1.43)
Provider Phone Number	4.0 (1.14)	5.0 (0.25)	5.0 (0.31)	5.0 (0.18)	4.0 (1.57)
Date of Onset	5.0 (0.60)	3.0 (2.00)	5.0 (0.63)	5.0 (0.36)	5.0 (0.43)
Reason for Visit	5.0 (0.29)	2.5 (1.75)	4.5 (1.06)	5.0 (0.73)	3.0 (1.00)
Clinical Diagnosis	5.0 (0.60)	4.0 (1.50)	5.0 (0.63)	5.0 (0.27)	5.0 (0.29)
Definitive Diagnostic Test Results	5.0 (0.50)	5.0 (0.00)	5.0 (0.19)	5.0 (0.18)	5.0 (0.14)
Person Completing Report Name	4.0 (1.33)	4.5 (0.75)	5.0 (0.63)	4.0 (0.82)	4.0 (1.43)
Person Completing Report Address	2.0 (1.21)	2.5 (1.25)	4.0 (0.94)	4.0 (1.00)	4.0 (1.43)
Person Completing Report Phone Number	4.0 (1.33)	4.5 (0.75)	5.0 (0.73)	5.0 (0.82)	4.0 (1.43)

Ratings are Mostly Incomplete (median = 1), Marginally Complete (median = 2), Mostly Complete (median = 3), Nearly Complete (median = 4), and Always Complete (median = 5). Values are medians (mean absolute deviation from median). Values with consensus are bold.

Desired and minimally acceptable timeliness

Table 5 illustrates respondents' desired timeliness and minimally acceptable timeliness (MAT) for new reported cases by disease. Values represent the number of respondents and percentage of respondents for each disease that indicated a specific desired or minimally acceptable timeliness.

Table 5. Desired and minimally acceptable timeliness (MAT) for new reported cases by disease

Disease	As soon as clinician determines diagnosis	Within 24 Hours	Within 48 Hours	Within 72 Hours	Within 1 Week	No preference or no opinion
Influenza (Desired) (<i>N</i> = 18)	5 (27.8%)	6 (33.3%)	3 (16.7%)	3 (16.7%)	1 (5.6%)	0 (0.0%)
Influenza (MAT) (<i>N</i> = 18)	0 (0.0%)	1 (5.6%)	2 (11.1%)	6 (33.3%)	8 (44.4%)	1 (5.6%)
Chlamydia (Desired) (<i>N</i> = 4)	0 (0.0%)	2 (50.0%)	2 (50.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Chlamydia (MAT) (<i>N</i> = 4)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (25.0%)	3 (75.0%)	0 (0.0%)
Meningitis (Desired) (<i>N</i> = 17)	13 (76.5%)	3 (17.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (5.9%)
Meningitis (MAT) (<i>N</i> = 17)	4 (23.5%)	9 (52.9%)	3 (17.6%)	0 (0.0%)	0 (0.0%)	1 (5.9%)
Pertussis (Desired) (<i>N</i> = 11)	3 (27.3%)	5 (45.5%)	1 (9.1%)	1 (9.1%)	0 (0.0%)	1 (9.1%)
Pertussis (MAT) (<i>N</i> = 11)	2 (18.2%)	4 (36.4%)	1 (9.1%)	3 (27.3%)	0 (0.0%)	1 (9.1%)
Diabetes (Desired) (<i>N</i> = 8)	0 (0.0%)	1 (12.5%)	0 (0.0%)	0 (0.0%)	1 (12.5%)	6 (75.0%)
Diabetes (MAT) (<i>N</i> = 8)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (25.0%)	6 (75.0%)

Discussion

We surveyed epidemiologists and other public health professionals regarding their information needs with respect to reporting processes important to surveillance. A novel but reliable (Chronbach's $\alpha > .93$) survey instrument measured the relative importance, completeness, and timeliness of various commonly reported data elements across a range of diseases that health departments monitor. The results show a wide range of variation in information needs and perceived data quality across jurisdictions and diseases. Such variation has a number of implications for surveillance practice as well as systems designed to support it. The findings further have importance for the field of biomedical informatics given that similar variation is likely to exist in many clinical and public health domains.

Several core data elements reported to public health stood out as being extremely important to public health work processes: clinical diagnosis, definitive diagnostic test result, age, county of residence, zip code, city, and information on the treating provider. These fields received at least a mean score of 4.0 with a low MAD (< 1.0) across at least 4 of the 5 diseases asked about in the survey. As expected, these same fields were associated with higher mean scores and lower MAD values for desired level of completeness, indicating a relationship between perceived importance and desired completeness of a data element. For the other fields, perceived importance or desired completeness varied either within a specific disease (e.g., large variation in scores; a high MAD) or across diseases (e.g., mean of 5.0 for some diseases but mean < 4.0 for other diseases). These results suggest that despite these data elements being recognized by CSTE and other groups as core reporting elements, epidemiologists and others vary in their perceptions of these data importance or necessity to surveillance work processes.

We further observed a pattern in the variation across disease categories in which information needs for chlamydia, meningitis, and pertussis (notifiable diseases) were generally consistent while the needs for influenza and diabetes (seasonal and chronic diseases) varied more dramatically. For example, patient and provider details were more consistently perceived as important and needed to be more complete for the notifiable diseases. The means for patient fields such as name, address, and phone number had means of 5.0, low MAD values, and achieved greater than 70% agreement on the scores for chlamydia, meningitis, and pertussis. Mean scores for influenza varied from 2.5 for patient name to 4.0 for patient city; MAD values were above 1.0 for patient name and patient address; and none of the fields were found to have consensus. For diabetes, scores were generally neutral, and the MAD values were above 1.0.

Greater consensus and consistency in information needs for notifiable diseases is likely explained by the fact that these diseases are reportable in nearly every state, therefore many public health departments have been receiving data on these diseases for a significant number of years (21). Furthermore, public health laws generally require providers to report information on these diseases and require health departments to verify that appropriate treatment is being administered to the patient. This requires follow-up with either the patient or providers and is the case for

meningitis, where the health department may need to provide prophylaxis to contacts in order to limit the spread of the disease. Finally, current data routinely captured by health departments on influenza and diabetes tends to be de-identified as follow-up on these diseases is generally not required, and this may explain the lower levels of consensus for importance and completeness observed among respondents.

Clinical and public health organizations are undergoing radical transformation due to changes in policy and technology that will require greater emphasis on the use of electronic data and information from a greater number of sources. The Health Information Technology for Economic and Clinical Health (HITECH) provisions of the American Recovery and Reinvestment Act of 2009 and mobile devices are accelerating the adoption and use of health information technologies to collect and manage greater volumes of electronic data (22-26). The increasing adoption of clinical information systems by hospitals and providers has prompted the CDC, CSTE, the Association of State and Territorial Health Officials, and National Association of City and County Health Officials, among others, to urge state and local health departments to prepare for a sharp increase in electronic reporting of data (11, 27). In a future state where public health departments are receiving and managing large volumes of electronic data from EHR systems, mobile devices, patient-controlled platforms and social media, understanding and re-evaluating the information needs of front line public health workers will be of critical importance. Variation in needs therefore has a number of implications for surveillance systems and practice as well as the field of biomedical informatics.

Implications for Surveillance Systems and Practice

Our findings complement previous studies that characterize the heterogeneous nature of clinical data across the spectrum of health care delivery systems (6). Many notifiable diseases such as chlamydia and meningitis that require a public health response, including documentation of treatment or administration of prophylaxis, will warrant timely and complete data. On the other hand, chronic diseases such as hypertension or diabetes may not demand the same amount of timeliness or level of information. Therefore information systems that support the range of surveillance performed by a public health agency must allow for variability to support the needs of different public health laws and information needs. Surveillance system developers will need to make their systems flexible and customizable to enable localization when implemented within a health department. Yet systems must remain capable of receiving a large array of input data from a variety of sources, requiring the use of standards and interoperability while enabling localization. Platform based approaches like those used in the SMART (smartplatforms.org) and OpenMRS (openmrs.org) efforts are examples of systems that enable application flexibility required by public health.

Our findings further have implications for the business processes involved in receiving and analyzing surveillance data. Given variation in user needs, health departments may wish to work on standardizing work processes involved in surveillance activities. Such redesign will allow for optimization of human resources in conjunction with surveillance information systems. Business process analysis is a technique where user information needs are articulated and translated into requirements for information systems (28). The technique further allows for business process alignment or changes that don't affect or involved information systems. Using business process analysis in conjunction with enhancements to the information systems will support improvements to surveillance practice. The current revolution underway in which surveillance programs and data are being consolidated provides greater incentive for agencies to identify user needs and optimize business processes that meet those needs.

Implications for Biomedical Informatics

Variation is not unique to public health surveillance systems or practice. Therefore platform-based approaches and business process techniques to improve system design and implementation is applicable across the spectrum of biomedical informatics. Further implications of this study include the role of biomedical informatics in evaluating information systems with respect to how well they meet the information needs of end users. Prior literature suggests there are few studies that have evaluated implemented systems (3, 29) and the evidence base for many informatics interventions remains weak (30). Biomedical informatics scholars and professionals should continue to not only development and implement novel systems and functions but also evaluate their ability to meet user needs in addition to their impact on patient and population health outcomes (31).

Limitations

There are several limitations in our methods and results to note. First, the selection of health departments was not random, potentially introducing bias into the study population. Second, the overall sample size was limited (N=33) and some diseases like chlamydia (N=5) had very small data sets. Therefore analysis of the data was limited to descriptive statistics, and causal relationships should not be inferred. Furthermore, the results may not be

representative of jurisdictions across the entire United States. Future studies should consider larger samples of health department personnel as well as the use of power analysis to enable more robust quantitative exploration of data quality across jurisdictions, states, and regions.

Conclusion

Variation remains an important challenge in biomedical informatics. As information systems increase in their adoption and use, the role of informatics professionals in designing systems and information to meet the needs of end users becomes paramount. Systems must be flexible and highly configurable to accommodate variation, and informaticians must measure and improve systems and business processes to accommodate for variation of both users and information.

Acknowledgements

This work was supported by grant number 1P01HK000077 from the Centers for Disease Control and Prevention and grant number R01HS020909 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention, Agency for Healthcare Research and Quality, or the Department of Veterans Affairs.

References

1. Thacker SB, Qualters JR, Lee LM. Public health surveillance in the United States: evolution and challenges. *MMWR Surveill Summ.* 2012;61:3-9.
2. A national research agenda for public health services and systems. *Am J Prev Med.* 2012;42(5 Suppl 1):S72-8.
3. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR Recomm Rep.* 2004;53(RR-5):1-11.
4. Lombardo JS, Buckeridge DL, editors. *Disease Surveillance: A Public Health Informatics Approach.* Hoboken: John Wiley & Sons; 2007.
5. Batini C, Cappiello C, Francalanci C, Maurino A. Methodologies for Data Quality Assessment and Improvement. *Acm Comput Surv.* 2009;41(3).
6. Dixon BE, McGowan JJ, Grannis SJ. Electronic laboratory data quality and the value of a health information exchange to support public health reporting processes. *AMIA Annu Symp Proc.* 2011;2011:322-30.
7. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems.* 1996;12(4):5-34.
8. Liaw ST, Chen HY, Maneze D, Taggart J, Dennis S, Vagholkar S, et al. Health reform: is routinely collected electronic information fit for purpose? *Emergency medicine Australasia : EMA.* 2012;24(1):57-63.
9. Kass-Hout TA, Xu Z, McMurray P, Park S, Buckeridge DL, Brownstein JS, et al. Application of change point analysis to daily influenza-like illness emergency department visits. *J Am Med Inform Assoc.* 2012;19(6):1075-81.
10. Dixon BE, Grannis S. Why "What Data Are Necessary for This Project?" and Other Basic Questions are Important to Address in Public Health Informatics Practice and Research. *Online J Public Health Inform.* 2011;3(3).
11. Smith PF, Hadler JL, Stanbury M, Rolfs RT, Hopkins RS. "Blueprint Version 2.0": Updating Public Health Surveillance for the 21st Century. *J Public Health Manag Pract.* 2013 May-Jun;19(3):231-9.
12. Randolph GD, Stanley C, Rowe B, Massie SE, Cornett A, Harrison LM, et al. Lessons learned from building a culture and infrastructure for continuous quality improvement at cabarrus health alliance. *J Public Health Manag Pract.* 2012;18(1):55-62.
13. Grannis S, Dixon BE, Brand B. Leveraging immunization data in the e-health era: Exploring the value, tradeoffs, and future directions of immunization data exchange. Atlanta: Public Health Informatics Institute, 2010.
14. Public Health Informatics Institute. *Redesigning Public Health Surveillance in an eHealth World.* Decatur, GA: Public Health Informatics Institute, 2012.
15. 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;96(4):335-42.

16. CSTE. Common Core Data Elements for Case Reporting and Laboratory Result Reporting. Atlanta: Council of State and Territorial Epidemiologists, 2009 Contract No.: 09-SI-01.
17. Gorard S. Revisiting a 90-Year-Old Debate: The Advantages of the Mean Deviation. *British Journal of Educational Studies*. 2005;53(4):417-30.
18. Habib EA. Mean Absolute Deviation about Median as a Tool of Explanatory Data Analysis. *International Journal of Research and Reviews in Applied Sciences*. 2012;11(3):517-23.
19. Mahler DA, Selecky PA, Harrod CG, Benditt JO, Carrieri-Kohlman V, Curtis JR, et al. American College of Chest Physicians consensus statement on the management of dyspnea in patients with advanced lung or heart disease. *Chest*. 2010;137(3):674-91.
20. Zafar SY, Currow DC, Cherny N, Strasser F, Fowler R, Abernethy AP. Consensus-based standards for best supportive care in clinical trials in advanced cancer. *Lancet Oncol*. 2012;13(2):e77-82.
21. Sickbert-Bennett EE, Weber DJ, Poole C, MacDonald PD, Maillard JM. Completeness of communicable disease reporting, North Carolina, USA, 1995-1997 and 2000-2006. *Emerg Infect Dis*. 2011;17(1):23-9.
22. Jha AK, Burke MF, DesRoches C, Joshi MS, Kralovec PD, Campbell EG, et al. Progress toward meaningful use: hospitals' adoption of electronic health records. *The American journal of managed care*. 2011;17(12 Spec No.):SP117-24.
23. Centers for Medicare & Medicaid Services. Medicare and Medicaid Programs; Electronic Health Record Incentive Program--Stage 2. *Federal Register* [Internet]. 2012 [cited 2012 August 24]. Available from: http://www.ofr.gov/OFRUpload/OFRData/2012-21050_PL.pdf
24. Levin D. MHealth: promise and pitfalls. *Front Health Serv Manage*. 2012;29(2):33-9; discussion 40-4.
25. Sclafani J, Tirrell TF, Franko OI. Mobile Tablet Use among Academic Physicians and Trainees. *J Med Syst*. 2013;37(1):9903.
26. Comer KF, Grannis S, Dixon BE, Bodenhamer DJ, Wiehe SE. Incorporating geospatial capacity within clinical data systems to address social determinants of health. *Public Health Rep*. 2011;126 Suppl 3:54-61.
27. Kirkwood J, Jarris PE. Aligning health informatics across the public health enterprise. *J Public Health Manag Pract*. 2012;18(3):288-90.
28. Public Health Informatics Institute. Taking Care of Business: A Collaboration to Define Local Health Department Business Processes. Decatur, GA: Public Health Informatics Institute, 2008.
29. Bravata DM, McDonald KM, Smith WM, Rydzak C, Szeto H, Buckeridge DL, et al. Systematic review: Surveillance systems for early detection of bioterrorism-related diseases. *Ann Intern Med*. 2004;140(11):910-22.
30. Dixon BE. A roadmap for the adoption of e-health. *e-Service Journal*. 2007;5(3):3-13.
31. Whipple EC, Dixon BE, McGowan JJ. Linking health information technology to patient safety and quality outcomes: a bibliometric analysis and review. *Inform Health Soc Care*. 2013 Jan;38(1):1-14.