Impact of Selective Mapping Strategies on Automated Laboratory Result Notification to Public Health Authorities

Roland E. Gamache, PhD, MBA[1,2], Brian E. Dixon, MPA, PhD[1,3,4], Shaun Grannis, MD, MS[1,2] , Daniel J. Vreeman, PT, DPT, MSc[1,2]
[1]Regenstrief Institute, Inc., [2]Indiana University School of Medicine, [3]Indiana University School of Informatics, [4]VA HSR&D Center on Implementing Evidence-Based Practice, Roudebush VAMC, Indianapolis, IN

ABSTRACT

Automated electronic laboratory reporting (ELR) for public health has many potential advantages, but requires mapping local laboratory test codes to a standard vocabulary such as LOINC. Mapping only the most frequently reported tests provides one way to prioritize the effort and mitigate the resource burden. We evaluated the implications of selective mapping on ELR for public health by comparing reportable conditions from an operational ELR system with the codes in the LOINC Top 2000. Laboratory result codes in the LOINC Top 2000 accounted for 65.3% of the reportable condition volume. However, by also including the 129 most frequent LOINC codes that identified reportable conditions in our system but were not present in the LOINC Top 2000, this set would cover 98% of the reportable condition volume. Our study highlights the ways that our approach to implementing vocabulary standards impacts secondary data uses such as public health reporting.

INTRODUCTION

Automated electronic laboratory reporting (ELR) for public health requires the detection of specific positive clinical laboratory results.[1-3] However, clinical laboratory results are often identified by idiosyncratic local codes that represent identical concepts in different laboratory systems.[4] Mapping local laboratory test codes to the LOINC® vocabulary standard enables interoperable data sharing and aggregation from many sources.[5] Many forces in the current healthcare environment are accelerating the standardization of laboratory result exchange, including the Meaningful Use (MU) criteria established as part of the U.S. Centers for Medicare and Medicaid Services (CMS) electronic health record (EHR) incentive program.[6] However, the task of mapping each local code in a laboratory test dictionary to a code from a vocabulary standard can be daunting because it is a complex and resource demanding process.[7, 8]

Previous work has shown that as few as 80 distinct laboratory result codes can account for over 80% of the total volume of laboratory results seen in a typical health care system.[4] When faced with the challenge and cost of mapping local codes to a standard vocabulary in resource-limited settings[9], one potential strategy to mitigate this burden is to focus the mapping effort on the highest volume tests. We refer to this frequency-based approach as "selective mapping."

Regenstrief Institute, in collaboration with the Lister Hill National Center for Biomedical Communications at the National Library of Medicine (NLM) have developed an empirically derived list of the 2,017 most commonly reported LOINC codes (which we will refer to as the "LOINC Top 2000").[10] The codes in the LOINC Top 2000 represent about 98% of the test volume carried by three large organizations that mapped all of their laboratory tests to LOINC codes. The LOINC Top 2000 presents a much more manageable target than trying to match all of the codes in a typical laboratory's 2000-5000 term dictionary to one of the 68,000 LOINC codes in current release (December 2011). In addition, Regenstrief and NLM have developed a "Mapper's Guide" [11] that contains assistance about which codes from the LOINC Top 2000 to choose for which purpose. The guide was developed to aid small-to-medium sized organizations achieve the requisite LOINC mapping necessary to achieve MU using the LOINC Top 2000.

Many clinical providers have considered a strategy to first map this selective set of clinical codes to mitigate the burden of mapping the plethora of individual local codes to standard code equivalents. This strategy, referred to as selective mapping, is being considered by providers as an interim step to meet the requirements of the Stage 1

Meaningful Use (MU) criteria established as part of the U.S. Centers for Medicare and Medicaid Services (CMS) electronic health record (EHR) incentive program.[12]

Public health agencies, in response to MU and to a decrease in available resources, are relying more on automated methodologies such as an ELR to collect and analyze information relevant for public health surveillance and outbreak management. ELR has been shown to positively impact notifiable disease reporting and surveillance, improving both the timeliness and the volume of notifiable disease cases reported to public health agencies.[13-15] However, the selective mapping approach may have adverse consequences for public health as it relies more on automated electronic laboratory reporting systems for the reporting of notifiable diseases and population-based surveillance in the community.

To achieve and sustain successful automated ELR for public health reporting, reportable laboratory results must be mapped to a standard code set. This enables the receiving information system at the public health agency to correctly interpret the test and result communicated from the sending information system in the laboratory, hospital, or other health care facility. In other words, mapping enables semantic interoperability between the laboratory, provider, and public health. However, the potential impact of a frequency-based, selective mapping approach on ELR to public health has not yet been evaluated.

The Centers for Disease Control and Prevention (CDC) developed and published the Reportable Condition Mapping Table (RCMT), a resource that provides mappings between reportable conditions and their associated LOINC coded laboratory tests and SNOMED CT result values.[16] A key use of the RCMT is as a filter for identifying which laboratory results should be sent to public health. If local laboratory test codes are not mapped to LOINC, then they cannot easily be automatically reported to or interpreted by the public health agency. Therefore, in addition to the codes in the LOINC Top 2000, the LOINC codes in the RCMT also represent an important target for mapping. If the results that must be reported to public health are not of significant frequency in the relatively small set of large volume clinical laboratory tests, selective mapping based on frequency alone may hinder efforts to increase the usage of ELR to public health.

Given the potential for unintended consequences to public health, the purpose of this study is to evaluate the impact of selective mapping strategies on real-world automated laboratory result notification. Specifically, we evaluated the notifiable disease results from a large operational health information exchange with respect to the codes contained in the LOINC Top 2000. Examples of codes that are included in this dataset are presented in an extract of this list in Table 1.

Table 1. Extract of Observation Codes from the MICRO Class of the LOINC Top 2000 Clinical Observation Codes.

LOINC #	Long Common Name	CLASS	Rank
21441-1	Human papilloma virus 6+11+42+43+44 DNA [Presence] in Cervix by DNA probe	MICRO	1293
42481-2	Human papilloma virus 6+11+42+43+44 DNA [Presence] in Cervix by Probe & signal amplification method	MICRO	557
46082-4	Influenza virus A Ag [Presence] in Nasopharynx by Immunoassay	MICRO	1201
5862-8	Influenza virus A Ag [Presence] in Unspecified specimen by Immunoassay	MICRO	728
5866-9	Influenza virus B Ag [Presence] in Unspecified specimen by Immunoassay	MICRO	796
41499-5	Legionella pneumophila 1 Ag [Presence] in Urine by Immunoassay	MICRO	1169
588-4	Legionella pneumophila Ag [Presence] in Unspecified specimen by Immunofluorescence	MICRO	1360
6448-5	Legionella pneumophila Ag [Presence] in Urine by Radioimmunoassay (RIA)	MICRO	1649
593-4	Legionella sp identified in Unspecified specimen by Organism specific culture	MICRO	1154
12232-5	Measles virus Ag [Presence] in Unspecified specimen by Immunofluorescence	MICRO	467
20479-2	Measles virus IgG Ab [Presence] in Serum	MICRO	1133
35275-7	Measles virus IgG Ab [Presence] in Serum by Immunoassay	MICRO	1134
5244-9	Measles virus IgG Ab [Units/volume] in Serum by Immunoassay	MICRO	627

METHODS

The Indiana Network for Patient Care (INPC) [17, 18] is a 16-year-old health information exchange operated in Indiana. Regenstrief Institute made a pioneering commitment to standards, interoperability, and the interchange of clinical data for clinical, public health, and research purposes. Investigators at Regenstrief created the INPC in 1995 with the goal of providing clinical information at the point of care for the treatment of patients. The INPC now includes clinical data from over 80 hospitals, the public health departments, local laboratories, imaging centers, a few large-group practices closely tied to hospital systems, with plans to continue expanding. The health information exchange data repository carries over 4 billion discrete clinical results, over 79 million text reports, for more than 25 million different patient registrations of over 12 million unique patients.

Investigators at Regenstrief have also developed and implemented an automated ELR and case-notification system called the Notifiable Condition Detector that has been operating within the INPC for over 10 years.[19] The Notifiable Condition Detector uses a standards-based messaging and vocabulary infrastructure (including HL7 and LOINC), and receives more than 350,000 real-time HL7 clinical transactions daily, including laboratory studies, diagnoses, and transcription from more than 24 organizations, national labs and local ancillary service organizations. It translates local proprietary codes into LOINC codes, determines whether the results carried by the message indicate a notifiable condition using a variety of algorithms. It evaluates the presence or absence of an abnormal flags for lab results; if the abnormal flag is not set, the software compares the clinical result to indicator values from the RCMT or to specific organisms listed in the RCMT. Using additional information available in the INPC (including patient and physician data) we augment the original clinical results with more complete patient, physician, and clinical vocabulary, and automatically forward notifiable conditions to appropriate local and state public health personnel. These augmented data serve multiple purposes. First, they satisfy the state reporting mandates for many conditions. Second, they serve as automated triggers to initiate the public health case management process. Third, the lab data can be used to automatically complete case reporting forms, which the provider can then complete. This system, depicted in Figure 1, demonstrated a greater than 4-fold detection rate than traditional physician-based reporting methods.[14]

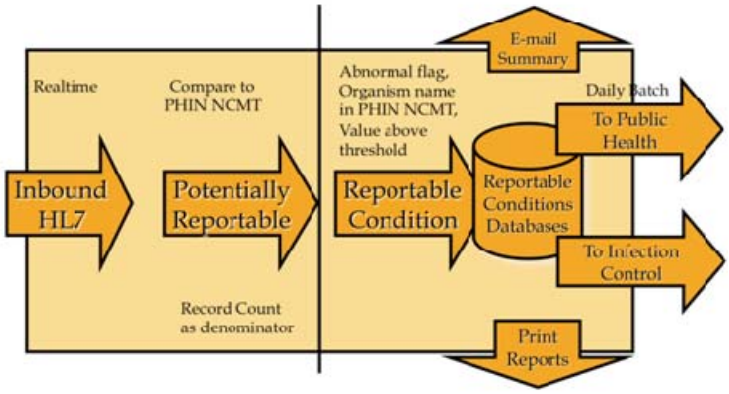

Figure 1: Data flow for the public health condition detector. Inbound HL7 messages are mapped to LOINC and if the test is potentially reportable, it is passed to a second stage where its reportable status is adjudicated.

These results suggest that value sets such as the RCMT will diverge if not maintained on a routine basis in conjunction with its affiliated standards. To evaluate the implications of selective mapping ELR approaches for public health reporting processes, we compared the public health reportable laboratory results identified by the Notifiable Condition Detector (NCD) to codes in the LOINC Top 2000 Version 1.0a. We determined the proportion of laboratory test codes and volume for reportable laboratory results with LOINC codes that were present in the LOINC Top 2000. We also determined the distribution of reportable laboratory tests by volume and performed a sensitivity analysis to gage the impact of the "selective mapping" based on the LOINC Top 2000 on the ability to automatically report these results to public health agencies.

RESULTS

The Regenstrief Institute's Notifiable Condition Detector identified 833,710 potentially reportable conditions to public health from 3/8/97 through 12/5/11. It is not uncommon for a patient to present with multiple reportable conditions, so the number of patients reported to public health would be fewer than the number of cases reported

during this time frame. In total, there were 490 unique observation codes reported over this timeframe from the NCD. At the time of this study there were 6,644 observation codes identified in the RCMT. So the total number of unique codes used in practice is far fewer than the total number of identified public health reportable observation codes listed by the CDC.

Of the 833,710 potentially reportable conditions, 523,783 (65.3%) were identified by LOINC codes that were also included in the LOINC Top 2000, and 277,730 (34.7%) were not included in the LOINC Top 2000 (Figure 2).

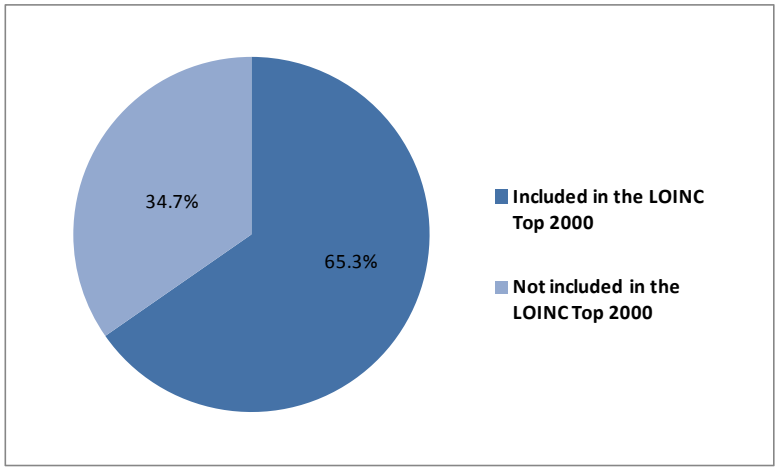

Figure 2. Percent of public health reportable conditions identified by LOINC codes contained in the LOINC Top 2000.

The distribution of 490 observed codes when ordered from most common to least common from the Notifiable Condition Detector is presented in Figure 3. This graph presents the cumulative proportion of observed codes versus the total number of codes reported over this period. Similar to previous results from clinical laboratory systems [4], the shape of the curve appears like a cumulative distribution Pareto with the majority of actual reported public health reportable conditions represented by a relatively few LOINC observation codes. Moreover, the 129 most frequent codes account for 98% of the conditions identified by the NCD and reported to public health.

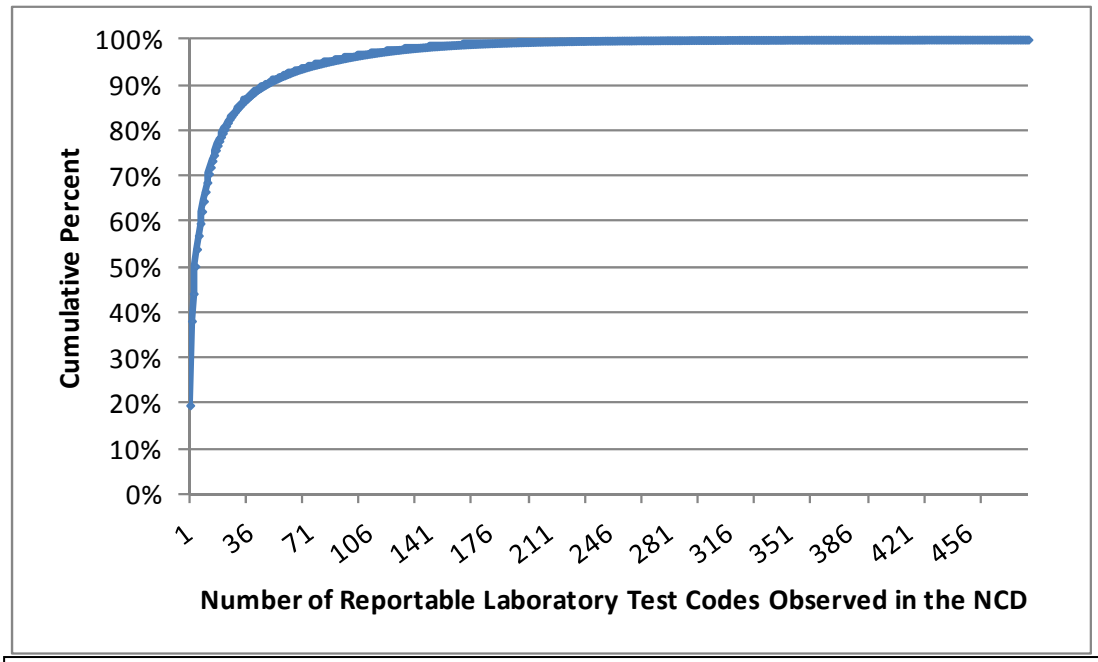

Figure 3. Laboratory test codes ordered by decreasing frequency versus the cumulative frequency of results reported to public health from the NCD.

To determine the potential impact on public health by using a strategy to map tests from the LOINC Top 2000 we performed a sensitivity analysis to determine the relative percent volume of reported public health conditions that would be attainable by mapping an increasing number of LOINC codes from the LOINC Top 2000. This graph is shown in Figure 4. Also, similar to the clinical results, the top 98% of the reported conditions for public health occur in the most frequent 136 codes.

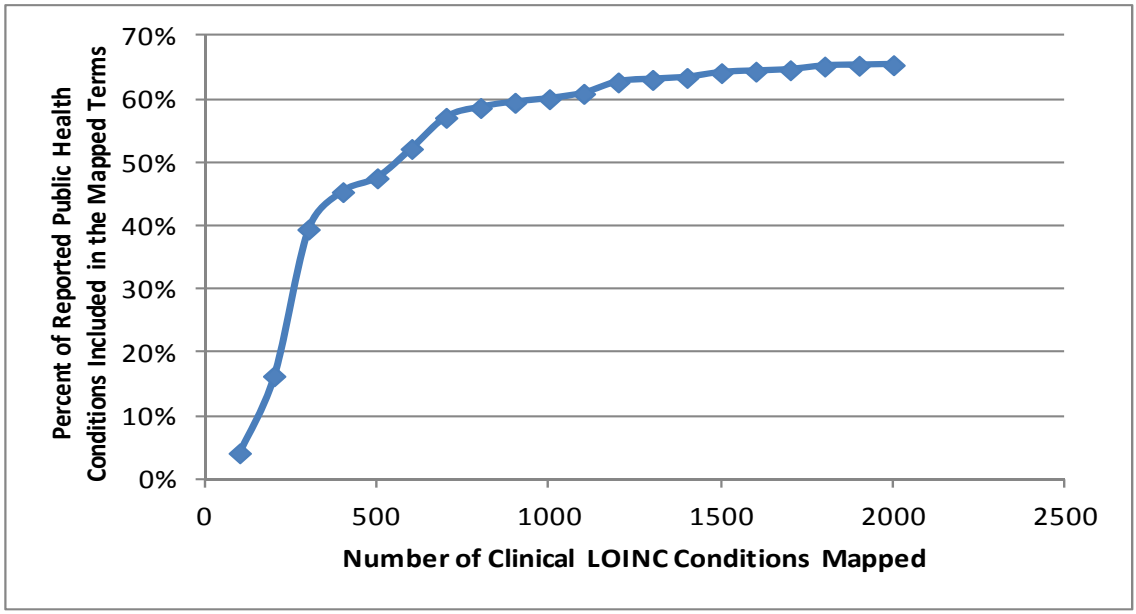

Figure 4. A sensitivity analysis illustrating the rate of notifiable public health results identified as a function of mapping the most common LOINC® codes. By mapping the top 500 most common LOINC® codes, slightly more than 50% of all reportable results would be identified.

Over 50% of the reported conditions to public health could be found by mapping laboratory tests to the 550 most common codes from the LOINC Top 2000. Mapping tests to the next 1,500 most frequently occurring terms from the LOINC Top 2000 would yield only an additional 15% of public health conditions. The last 500 most frequent codes in the LOINC Top 2000 add less than a 2% increase in the number of conditions reported to public health. 35% of the conditions reported to public health were identified by LOINC codes that were not included in the LOINC Top 2000. The most significant laboratory LOINC condition not included in the LOINC Top 2000 was 10368-9, which is capillary blood lead test and is used as a screening test.

In over ten years of automatically reporting reportable conditions to public health jurisdictions in Indiana, only 490 of these LOINC codes were used in our experience. Of the total number of possible mapped codes, just over 13% have been used during the lifetime of our Notifiable Condition Detector.[20]

DISCUSSION

As we contemplate national strategies for health information technology certification and deployment, we need to be cognizant of how our approach to implementing vocabulary standards impacts many stakeholders. This analysis highlights the fact that the laboratory tests most commonly reported for clinical care are not sufficient to cover all of the tests for conditions reportable to public health. Yet, similar to clinical laboratory reporting, few of the tests possible for a reportable condition account for a large proportion of the volume. Indeed, of the more than 3,700 LOINC codes in the CDC RCMT, Regenstrief's Notifiable Condition Detector only used 490 LOINC codes, and 129 of those accounted for 98% of the reportable condition volume.

Our results highlight that we must be mindful of the many potential uses of important health data like laboratory results beyond direct patient care, including public health and quality reporting. As the developers of the LOINC Top 2000 and Vreeman et al [4] were keenly aware, mapping a selective subset based on reporting frequency might

be a good place to start, but is not sufficient to support all health care use cases. The results of this study can help inform those faced with limited resources and the challenge of mapping by illustrating the use case of public health reporting.

We believe that there is value in creating these "top" lists for mapping in the context of clinical care, public health, quality reporting, and potentially other use cases as well. But we also recognize that selective mapping has both advantages and disadvantages. In the context of public health reporting, mapping all laboratory codes has a substantial cost both initially and for the on-going maintenance, and mapping selectively will likely result in certain conditions being missed. For too long however, the position of not using standard vocabularies has hindered the potential positive impact of health information technology.

Similar to the Clean Water Act, we also believe strongly that the mapping and data standardization effort should occur as close to the production source as possible. Physician offices and public health are both downstream recipients of the data from clinical laboratories. While the focus of Meaningful Use's incentive payments is on providers and hospitals that adopt EHRs, the mapping effort is far more effective within the laboratory that has the domain experts who know what the tests really measure. In the Laboratory Interoperability Cooperative initiative (http://www.labinteroperabilitycoop.org/), the CDC has funded an effort that includes helping laboratories that report to public health map their local codes to LOINC. We are also thrilled to see movement within the instrument and test kit manufacturer community such as the IVD Industry Connectivity Consortium (http://www.ivdconnectivity.org) to provide standard codes to their customers and in their interfaces. Standardized coding at the headwaters of the data flow would be a huge efficiency for all the downstream recipients.

The analysis presented in this paper might also serve as a model for public health stakeholders to develop specific strategies to improve case detection processes. Strategic approaches for capturing data that deliver greatest value – such as identifying and mapping high volume public health results first – can help ensure that public health maximally benefits from large-scale efforts such as meaningful use. With public health increasingly receiving electronic transactions from clinical information systems, the analysis presented in this paper is illustrative of the types of strategic analyses that large public health stakeholders should conduct to inform their approach to implementing effective and efficient electronic case detection systems. For example, if a state health department performs analyses similar to this study and they discover that they receive no HIV results electronically, then public health may prioritize obtaining electronic data for that particular condition.

There are several limitations to our analysis. Indiana is a report all state for blood lead testing. This test is also the most frequent observation reported through the NCD. If only the positive blood lead levels were included, then the total percentage of observations reported for public health purposes in the LOINC Top 2000 would be reduced. Currently, lead reporting accounts for 18.5% of the total number of reported results. As mentioned earlier, the mapping table in the NCD and the CDC's RCMT are not in exact agreement. The NCD mapping table is constantly updated by information supplied by our public health users.[19] The CDC's RCMT is not updated at this same frequency so inconsistencies exist. We have shared our table with the CDC when they last updated the RCMT. Additionally, there are conditions that may be monitored by a local jurisdiction that would not be included on a national reportable condition mapping table.

CONCLUSIONS

As communities develop their implementation strategies for Health Information Exchange and Meaningful Use, the addition of a relatively few key public health observation codes would greatly benefit the public health surveillance capabilities of contiguous jurisdictions. Laboratory result codes in the LOINC Top 2000 account for 65.3% of the public health reportable conditions in Indiana. Although the some of the reportable public health conditions may not be represent as high a volume of conditions as other clinical conditions, but these public health cases may have a greater impact on the population health of the community. Using only the Top 2000 clinical codes would potentially miss just under 35% of the reportable public health conditions. However, by reviewing the community public health needs and ensuring that the top 129 LOINC codes identifying public health reportable conditions are also included in this mapped set of local terms would increase the percentage of reported cases to public health to 98% of the reported conditions. Using the same strategy as clinical systems, selectively including public health observations can minimize mapping effort while providing a robust automated reporting system for the community.

ACKNOWLEDGEMENTS

This work was supported, in part, by a contract (HHSN2762008000006C) from the National Library of Medicine, a grant (5R01HS0118553) from the Agency for Healthcare Research and Quality, as well as the Indiana Center of Excellence in Public Health Informatics through a grant (1P01HK000077-01) from the U.S. Centers for Disease Control and Prevention. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the National Library of Medicine, Centers for Disease Control and Prevention, Department of Veterans Affairs, the Agency for Healthcare Research and Quality, or the United States government.

REFERENCES

1. Wurtz R, Cameron BJ: **Electronic laboratory reporting for the infectious diseases physician and clinical microbiologist**. *Clin Infect Dis* 2005, **40**(11):1638-1643.
2. Vogt RL: **Laboratory reporting and disease surveillance**. *J Public Health Manag Pract* 1996, **2**(4):28-30.
3. Silk BJ, Berkelman RL: **A review of strategies for enhancing the completeness of notifiable disease reporting**. *J Public Health Manag Pract* 2005, **11**(3):191-200.
4. Vreeman DJ, Finnell JT, Overhage JM: **A rationale for parsimonious laboratory term mapping by frequency**. *AMIA Annu Symp Proc* 2007:771-775.
5. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leavelle D, Leland D, Fiers T, Charles L, Griffin B, Stalling F *et al*: **Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results**. *Clin Chem* 1996, **42**(1):81-90.
6. CMS: **Medicare and Medicaid Programs; Electronic Health Record Incentive Program; Proposed Rule**. In: *Federal Register.* vol. 75. Washington: Office of the Federal Register, National Archives and Records Administration; 2010: 1844-1892.
7. Baorto DM, Cimino JJ, Parvin CA, Kahn MG: **Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (LOINC)**. *Int J Med Inform* 1998, **51**(1):29-37.
8. Lin MC, Vreeman DJ, McDonald CJ, Huff SM: **Correctness of Voluntary LOINC Mapping for Laboratory Tests in Three Large Institutions**. *AMIA Annu Symp Proc* 2010, **2010**:447-451.
9. Vreeman DJ, Stark M, Tomashefski GL, Phillips DR, Dexter PR: **Embracing change in a health information exchange**. *AMIA Annu Symp Proc* 2008:768-772.
10. **Introduction to the Guide to the Top 2000+ LOINC Laboratory Observations** [http://loinc.org/resolveuid/7e7956dcb7eaa551c354ec0d37b0578c]
11. **Introduction to the Mapper's Guide for the Top 2000 plus LOINC Laboratory Observations** [https://loinc.org/usage/obs/introduction-to-the-mappers-guide-for-the-top-2000-plus-loinc-laboratory-observations.pdf/view]
12. CMS: **Electronic Health Record Incentive Program; Proposed Rule**. In. Edited by Programs MaM, vol. 75; 2010: 1844-1892.
13. Effler P, Ching-Lee M, Bogard A, Ieong M-C, Nekomoto T, Jernigan D: **Statewide System of Electronic Notifiable Disease Reporting From Clinical Laboratories: Comparing Automated Reporting With Conventional Methods**. *JAMA* 1999(19):1845-1850.
14. 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, **98**:344-350.
15. 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**:322-330.
16. **Introduction to the PHIN Notifiable Condition Mapping Tables**. In. Edited by May: Centers for Disease Control; 2004.
17. McDonald CJ, Overhage JM, Barnes M, Schadow G, Blevins L, Dexter PR, Mamlin B, Committee IM: **The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries**. *Health Aff (Millwood)* 2005, **24**(5):1214-1220.

18. Biondich PG, Grannis S: **The Indianapolis Network for Patient Care INPC An Integrated Clinical Information System Informed by Over Thirty Years of Experience**. *Supplement to the J Public Health Manag Pract Suppl* 2004:S81-S86.
19. Grannis S: **An Overview of Succesful Large-Scale Automated Case Detection: Assisting Public Health with the Identification of Reportable Conditions**. In: *Invited Levture for the Centers of Excellence in Public Health Informatics.* Edited by CDC. Atlanta, GA; 2010.
20. 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* 2005:286-290.