AUDITING THE USE OF LOINC® TO SUPPORT INTEROPERABILITY ACROSS THREE LARGE INSTITUTIONS

by

Ming-Chin Lin

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

Doctor of Philosophy

Department of Biomedical Informatics

The University of Utah

May 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The following faculty members served as the supervisory committee chair and members for the dissertation of Ming-Chin Lin.

Dates at right indicate the members' approval of the dissertation.

Stanley M. Huff,	Chair	**March 13, 2012**
		Date Approved
Clement J. McDonald,	Member	
		Date Approved
Peter J. Haug,	Member	**March 13, 2012**
		Date Approved
Bruce Bray,	Member	**March 13, 2012**
		Date Approved
Lewis J. Frey,	Member	**March 13, 2012**
		Date Approved

The dissertation has also been approved by **Wendy Chapman**, Chair of the Department of **Biomedical Informatics**
and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Logical Observation Identifiers Names and Codes (LOINC®) was developed in 1994 to provide a universal vocabulary for reporting laboratory and clinical observations. This dissertation was aimed at determining whether LOINC is meeting its goal when it is used in the real world.

Three institutions, Associated and Regional University Pathologist (ARUP), Intermountain Healthcare, and Regenstrief Institute, were invited to participate in this research. These institutions represented three of the seven institutions that provided their catalogue of laboratory test names for creating the first version of laboratory LOINC codes. We used the EDs to evaluate the **coverage, correctness, consistency, and competence** of LOINC. For coverage, we analyzed how many laboratory tests being routinely tested in daily operations could be assigned a correct LOINC code. For correctness, we verified the accuracy of LOINC mappings to local codes. For consistency and usefulness, we detected any inconsistencies in LOINC design and measured the degree of semantic interoperability that could be achieved using LOINC.

Besides auditing LOINC code use, we also analyzed the result values that were associated with the LOINC results (i.e., characteristics like the type of result (number, coded value), units of measure, answer set (positive/negative) etc.). We also found that consistent use of result values was important in achieving semantic interoperability when exchanging laboratory data.

Our analysis produced the following results: 1. **Completeness**: LOINC can provide 99% coverage rate for the results in two typical health care institutions and 79% coverage for results from a reference laboratory. 2. **Correctness:** An error rate of 4.5% existed in mappings at the three institutions. 3. C**onsistency and usefulness**: Several complicated or inconsistent designs for LOINC usage were found, which reduced the semantic interoperability of LOINC.

In this research, we developed a systematic approach for auditing LOINC usage in three institutions. We learned that LOINC is not yet perfect, and it needs continued improvement to increase the level of interoperability that can be achieved in exchange of laboratory results. Only auditing standardized terminologies as they are used in real applications can measure their degree of semantic interoperability.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

ACKNOWLEDGEMENTS

With the help of many people, I have finally completed this dissertation, which was a long-long journey. In this journey, I would like to thank many people, professors, colleagues, friends, and families, who supported me in completing this PhD work.

First of all, I want to thank my PhD committee members, who are the greatest combination of people for supporting me in completing this research. The chair of my committee, Dr. Stanley M. Huff, is the most important person in my studies. I extend a special “thanks” for giving me this opportunity to study in his department. He is really a great visionary as demonstrated by proposing this research idea in his LOINC paper published in 1998. He met with me weekly to discuss the progress of my studies and to guide me in the right direction. I not only gained knowledge but also developed a certain attitude about research by working with him.

Dr. Clement J. McDonald, who invented LOINC, is an amazing person. I learned to work harder based on his energetic and enthusiastic attitude. He was especially generous to encourage me in conducting LOINC research even though he was very

busy as the director of National Library of Medicine (NLM) Lister Hill Center. Dr. Peter J. Haug was a member of my committee, but was also the teacher of my master's thesis advisor (Dr. Yu-Chuan Li). I attended Dr. Haug's student meetings and received a lot of good advice from his group, including Jeffery Lin, Chia-Cheng Lee, and Senthil Nachimuthu. Dr. Bruce Bray is a cardiologist with great interest in terminology, especially in the area of cardiology and public health research. He provided me many points of view from real practice. Dr. Lewis Frey, who is an expert concerning caBig, has also actively participated in the caBig vocabulary committee. With his introduction, I worked with his student, Issac Kunz, to develop a very useful string matching algorithm for common data element mapping. I spent lots of time studying with his group, including Issac Kunz, Martin Cryer, and Vasse Nathan, when studying in my cubical at the BMI department.

Outside the PhD committee, I need to thank Daniel Vreeman, who is the director of the LOINC committee. He was also my research collaborator from Regenstrief Institute. He helped in the testing of my parsing program when a number of files were sent back and forth during the early development, and he was also a coauthor on several of my publications.

Finally, I need to thank the continuous support from my parents, my wife (Yi-Wen) and lovely two babies (Yu-Chen, Yu-Lan). They always provided the warmest and strongest support, which enabled me to complete this study.

CHAPTER 1

INTRODUCTION

1.1 Concerns about Using LOINC®

The following examples illustrate several concerns regarding LOINC®.

1.1.1 Case 1

Institution A decided to adopt LOINC® in their internal laboratory system. They wondered if the number of LOINC® codes was sufficient for mapping their internal codes.

1.1.2 Case 2

Institution B mapped their laboratory tests to LOINC® but did not know whether their mappings were correct or not.

1.1.3 Case 3

Dr. Markovski and colleagues wanted to conduct multi-institutional research related to diabetes, with analysis of patients' blood glucose levels. To facilitate the sharing of laboratory test data, they decided to all use LOINC® for data submission, but questioned whether LOINC® could properly handle the combination of laboratory tests between institutions.

The goal of LOINC® is to answer these concerns while promoting confidence in using LOINC® in a variety of laboratory systems.

1.2 Motivations of This Research

In the early stages of developing technology systems (TSs), we did not know how to build a "good" TS. After much trial and error, we arrived at several central principles for building TSs. James Cimino's desiderata (1) summarizes these central principles, including vocabulary content, concept orientation, concept permanence, nonsemantic concept identifiers, polyhierarchy, formal definitions, rejection of "not elsewhere classification" terms, and others. Soon, researchers noted that even these principles did not cover all issues related to implementing TSs. For example, Baorto et al. (2) pointed out that the multi-axial nature of LOINC® definitions did not lead to easy interoperability of laboratory tests. They evaluated LOINC® performance when it was used to combine laboratory tests from three teaching hospitals and found that the complex design of LOINC® could lead users to choose different, conflicting coding strategies. Observing LOINC® usage in real applications enables us to explore those issues arising from real-world use. In this study, we developed systematic methods to

audit LOINC® usage and learned how to use the information to improve the LOINC code system.

1.3 Outline of Dissertation

Chapter 2 of this dissertation provides the background for this study and contains three parts. Part one summarizes LOINC® development. Part two discusses four requirements—completeness, correctness, consistency, and usefulness—for building a functional knowledge system. Part three describes how to utilize extensional definitions (EDs) to audit LOINC®. Chapters 3 through 5 show the use of four approaches to auditing LOINC®: Chapter 3 analyzes LOINC® coverage (completeness), Chapter 4 examines the correctness of LOINC® mapping (correctness), and Chapter 5 evaluates how effectively laboratory test data can be exchanged between institutions using LOINC® (consistency and usefulness). Chapter 6 investigates semantic interoperability of laboratory results reported by LOINC®. Finally, Chapter 7 summarizes our results and discusses the future of LOINC® auditing and usage.

1.4 References

1. CIMINO, J. J. (1998) Desiderata for controlled medical vocabularies in the twenty-first century, *Methods Inf Med*, 37, 394-403.

2. BAORTO, D. M., CIMINO, J. J., PARVIN, C. A. & KAHN, M. G. (1998) Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (LOINC), *Int J Med Inform*, 51, 29-37.

CHAPTER 2

BACKGROUND

2.1 Development of LOINC®

International Statistical Classification of Diseases and Related Health Problems (ICD) was first developed by a French physician, Jacques Bertillon, and was soon adopted by many countries. Its use was (and is) for classifying diseases, symptoms, abnormal findings, social circumstances, etc. Health conditions could be easily categorized by use of a six-digit code unique to each disease. ICD became the "Lingua franca" for communicating disease names. Its codes have since been used worldwide to report information in electronic health records (EHR), billing systems, and health insurance claim systems.

The LOINC® committee first met in February of 1994 with the goal of developing a code system for naming clinical observations and laboratory tests in common use (1). As ICD had done for disease classification, the goal was to have LOINC® become the "Lingua franca" for identifying observations in interoperable data exchange in health care. This exchange makes up a large portion of health care data, such as measurements of sodium, chloride, hemoglobin, red blood cell count, and white blood cell count.

The LOINC® committee wanted to develop a "fully specified name" for each code that would convey all necessary information for differentiating any two different

laboratory tests. To achieve that level of granularity, first, LOINC® incorporated a multi-axial approach which was designed as a comprehensive nomenclature for creating names for clinical findings. Second, LOINC® was incorporated in the work of IUPAC (International Union of Pure and Applied Chemistry) as published in the *Compendium Terminology and Nomenclature of Properties in Clinical Laboratory Science* (The Silver book) (2). IUPAC created standardized nomenclature in chemistry, which was frequently used for naming chemical measurements. It included three axes: 1. The system in which a measurement is made. The system in laboratory settings is usually the specimen type, e.g., urine, blood, ascites, etc. 2. The component, specifying the chemical compound or cell type that is being evaluated, e.g., protein, white blood cell or antibody. 3. The property of the component: further specifying which kind-of-property of the component was being measured, e.g., weight, volume, concentration, etc. Third, the LOINC® committee included experience from EUCLIDES (3), Open Labs, and CEN TC251/PT3-008 (4). Fourth, the LOINC® committee collected laboratory files from seven participating laboratories. The test names in these files were analyzed and used to create the first set of LOINC® names and codes.

2.2 Review of Auditing TSs

Terminology systems such as ICD 9-CM, SNOMED CT, and LOINC® are widely used for improving the interoperability of healthcare data exchange. The high quality of TSs is the foundation of high-quality healthcare data. To achieve quality TSs, in the beginning, researchers focused on making principles for building reliable TSs. Cimino's desiderata (5) provides a set of principles to guide the creation of TSs.

As TSs matured, more institutions were willing to adopt them to represent biomedical informatics knowledge with the goal of improving semantic interoperability. With more content coded by standardized TSs, interest shifted to evaluating how TSs perform in real applications. For example, Bodenreider et al. evaluated the coverage of the Unified Medical Language System (UMLS) for Gene Ontology (6), Andrews et al. investigated the coding consistency of SNOMED CT in reporting rare diseases among three commercial coding companies (7), and Baorto et al. analyzed the coding consistency of LOINC® in three teaching hospitals (8). Evaluating TS use in real applications can provide different perspectives for examining and improving those systems. Min et al. concluded auditing should be an integral part of the terminology design cycle (9). Application-based auditing could include many

perspectives, such as correctness, coverage, and consistency. Arts (10) and Zhu (11) et al. did a thorough review of different approaches in different applications.

Devanbu' et al. (12) defined four requirements of a good knowledge system, including: 1) Completeness: it should have all the necessary knowledge for the intended domain, 2) Correctness: the knowledge should be faithful to the real world, 3) Consistency: the knowledge should not be self-contradictory and 4) Competence: the system should have efficient algorithms to perform the inferences needed for use in clinical applications. These four requirements can be used for evaluating TSs. Therefore, in this research, we want to use these four requirements to evaluate LOINC®.

2.3 Characterizing LOINC® Usage by Generating Extensional Definitions (EDs)

Extensional definitions (EDs) of a group of data, such as mean, standard deviation, or a histogram of data frequencies can reflect the actual meaning of data in production systems. In Table 2.1, we list several elements of EDs that characterize how LOINC® is used in actual laboratory systems. For example, a local description "Creatinine, 24hr urine," tells us this test is a measurement of creatinine in urine after a

Table 2.1 Extensional Definitions (EDs) included local description, mean, standard deviation, units of measure, coded variables and frequency.

Extensional definitions	**Example**	**Containing information for review**
Local description	"Creatinine, 24 hr urine","Sodium urine"	Local description - mainly provides analyte information. In some cases, it also provides method (e.g., EIA), scale/property (e.g., titer), time (e.g., 24 hr) or system information (e.g., urine)
Mean	1.46,137	Mean - provides scale/property information (e.g., SCnc/Qn). This is mainly useful for numeric tests.
Standard deviation	0.54, 7.02	Standard deviation - provides scale/property information (e.g., SCnc/Qn). This is mainly useful for numeric tests.

Table 2.1 Continued

Extensional definitions	**Example**	**Containing information for review**
Units of measure	g/24 h, mmol/L, mg/dl	Units of measure - provides scale/property information. This is mainly useful for numeric tests.
Coded variables and their frequencies	1:8 (109), Negative (900), Positive (899), M1M1 (75)	Coded variables - provides scale/property (e.g., Titr/Qn or ACnc/Qn). For example, M1M1 is a reported value for the genetics test 'ALPHA-1-ANTITRYPSIN PHENOTYPE' and its frequency was 75.
Frequency	50, 184	Frequency - implies whether tests are rare or common

24 hour specimen collection. Another example—the units of measure of a test (mg/dl)—can denote that the test measures a numeric quantity, such as a substance concentration. The frequency tests are performed can help distinguish rare tests from common tests. Zollo et al. successfully used EDs to match local tests among three different institutions (13).

Using a similar approach to that of Zollo et al., we extracted EDs of LOINC® codes as they existed in laboratory systems using a program that did not look at the identity of the patients during the extraction process. We developed this program to obviate any issues in completing our research related to patient privacy concerns.

To extract EDs from different institutions, first, we developed several programs written using Python and Java to parse HL7 messages. Second, we pre-installed the programs in a virtual machine (Linux-based operating system) and distributed the virtual machines containing the extraction algorithms to the participating institutions. Third, collaborating researchers loaded patient data into the virtual machines and used the pre-installed programs to extract EDs. Fourth, the processed EDs were sent back to us for further analysis.

2.4 References

1. FORREY, A. W., MCDONALD, C. J., DEMOOR, G. et al. (1996) 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*, 42, 81-90.

2. (1995) *International Union of Pure and Applied Chemistry/International Federation of Clinical Chemistry. The silver book: compendium of terminology and nomenclature of properties in clinical laboratory science* (Oxford: Blackwell Scientific Publishers).

3. Euclides Foundation International. EUCLIDES [Software] (Zaventem, Belgium: Euclides Foundation International.).

4. ROSSI MORI, A. & CONSORTI, F. (1998) Exploiting the terminological approach from CEN/TC251 and GALEN to support semantic interoperability of healthcare record systems., *Int J Med Inform*, 48, 111-24.

5. CIMINO, J. J. (1998) Desiderata for controlled medical vocabularies in the twenty-first century, *Methods Inf Med*, 37, 394-403.

6. BODENREIDER, O., MITCHELL, J. A. & MCCRAY, A. T. (2002) Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics, *Proc AMIA Symp*, 61-5.

7. ANDREWS, J. E., RICHESSON, R. L. & KRISCHER, J. (2007) Variation of SNOMED CT coding of clinical research concepts among coding experts, *J Am Med Inform Assoc*, 14, 497-506.

8. BAORTO, D. M., CIMINO, J. J., PARVIN, C. A. & KAHN, M. G. (1998) Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (LOINC), *Int J Med Inform*, 51, 29-37.

9. MIN, H., PERL, Y., CHEN, Y. et al. (2006) Auditing as part of the terminology design life cycle, *Journal of the American Medical Informatics Association :*

JAMIA, 13, 676-90.

10. ARTS, D. G., CORNET, R., DE JONGE, E. & DE KEIZER, N. F. (2005) Methods for evaluation of medical terminological systems -- a literature review and a case study, *Methods Inf Med*, 44, 616-25.

11. ZHU, X., FAN, J. W., BAORTO, D. M., WENG, C. & CIMINO, J. J. (2009) A review of auditing methods applied to the content of controlled biomedical terminologies, *Journal of Biomedical Informatics*, 42, 413-25.

12. DEVANBU, P. T. & JONES, M. A. (1994) The use of description logics in KBSE systems: experience report.

13. ZOLLO, K. A. & HUFF, S. M. (2000) Automated mapping of observation codes using extensional definitions, *J Am Med Inform Assoc*, 7, 586-92.

CHAPTER 3

A CHARACTERIZATION OF LOCAL LOINC® MAPPING FOR LABORATORY TESTS IN THREE LARGE INSTITUTIONS

Coauthors:

Daniel J. Vreeman, Clement J. McDonald, Stanely M. Huff

Reprinted with permission from Lin MC, Vreeman DJ, McDonald CJ, Huff SM. A characterization of local LOINC mapping for laboratory tests in three large institutions. Methods of Information in Medicine. 2011;50(2):105-14.

Original Articles

A Characterization of Local LOINC Mapping for Laboratory Tests in Three Large Institutions

M. C. Lin[1]; D. J. Vreeman[2, 3]; C. J. McDonald[4]; S. M. Huff[1, 5]

[1]The Department of Biomedical Informatics, The University of Utah, Salt Lake City, UT, USA;
[2]Regenstrief Institute, Inc., Indianapolis, IN, USA;
[3]Indiana University School of Medicine, Indianapolis, IN, USA;
[4]Lister Hill Center, National Library of Medicine, Washington, D.C., USA;
[5]Intermountain Healthcare, Salt Lake City, UT, USA

Keywords

Controlled vocabulary, LOINC, evaluation research, clinical laboratory information systems

Summary

Objectives: We characterized the use of laboratory LOINC® codes in three large institutions, focused on the following questions: 1) How many local codes had been voluntarily mapped to LOINC codes by each institution? 2) Could additional mappings be found by expert manual review for any local codes that were not initially mapped to LOINC codes by the local institution? and 3) Are there any common characteristics of unmapped local codes that might explain why some local codes were not mapped to LOINC codes by the local institution?

Methods: With Institutional Review Board (IRB) approval, we obtained deidentified data from three large institutions. We calculated the percentage of local codes that have been mapped to LOINC by personnel at each of the institutions. We also analyzed a sample of unmapped local codes to determine whether any additional LOINC mappings could be made and identify common characteristics that might explain why some local codes did not have mappings.

Results: Concept type coverage and concept token coverage (volume of instance data covered) of local codes mapped to LOINC codes were 0.44/0.59, 0.78/0.78 and 0.79/-0.88 for ARUP, Intermountain, and Regenstrief, respectively. After additional expert manual mapping, the results showed mapping rates of 0.63/0.72, 0.83/0.80 and 0.88/0.90, respectively. After excluding local codes which were not useful for inter-institutional data exchange, the mapping rates became 0.73/0.79, 0.90/0.99 and 0.93/0.997, respectively.

Conclusions: Local codes for two institutions could be mapped to LOINC codes with 99% or better concept token coverage, but mapping for a third institution (a reference laboratory) only achieved 79% concept token coverage. Our research supports the conclusions of others that not all local codes should be assigned LOINC codes. There should also be public discussions to develop more precise rules for when LOINC codes should be assigned

Correspondence to:
Stanley M. Huff, MD
South Office Building
5171 South Cottonwood Street
Murray, UT 84107
USA
E-mail: stan.huff@imail.org

Methods Inf Med 2011; 50: 105–114
doi: 10.3414/ME09-01-0072
received: August 18, 2009
accepted: June 13, 2010
prepublished: August 20, 2010

1. Introduction

With the development of electronic health records, there is a strong need to establish standard vocabularies to record patient-related data, especially in reporting laboratory results. Most laboratories use their local codes internally and use LOINC® codes or other standardized codes when there is a need to communicate outside of their own enterprise, e.g. returning results to an ordering physician, the submission of laboratory results to an insurance company, data sharing in a regional clinical data exchange network, or reporting required information to a public health department. Huff et al. noted that when LOINC achieved widespread use, it would be important that sufficient LOINC codes existed to cover the needs of reporting patient data [1]. Researchers have reported that mapping local codes to LOINC codes can be complex [2–4]. Therefore, we were interested in learning:

1. to what extent local codes have been mapped to LOINC codes;
2. what volume of patient test result instances is covered by the mapped codes;
3. how many more local codes could be mapped by expert manual review;
4. how fast the number of local codes is increasing;
5. how fast the number of LOINC codes is increasing;
6. whether there were any common patterns or characteristics of local codes that were not mapped to LOINC that might identify systematic problems in using LOINC.

We did not evaluate the correctness of the local LOINC code mappings in this part of our research.

2. Background

2.1 Development of LOINC

Currently, Health Level Seven (HL7) [5] is the most common electronic message standard used in exchanging clinical data

among hospitals, pharmaceutical manufactures, and public health departments. The observation segment of HL7 messages uses an EAV (entity-attribute-value triplet) [6] strategy to represent clinical data. For example, a serum sodium concentration measurement would be represented conceptually as "Laboratory Test (entity) has Test name = Serum Sodium Concentration (attribute); value =138 mmol/L (value)". Here is an example of the actual syntax of an HL7 Version 2 OBX (observation/result) segment:

OBX|1|NM|2951-2^Serum Sodium Concentration^LN|1|138|mmol/L|||" (1)

In this example, the LOINC code "2951-2" has been used as a standard code to represent the meaning of the serum sodium concentration measurement. LOINC was created to be a universal terminology for the electronic exchange of clinical observations for any kind of data exchange where the EAV approach is used. The intent was that different enterprises would map their local codes to LOINC, and then the LOINC codes would be used as the standard identifiers in data exchange. Essentially, the LOINC codes become the *lingua franca* for identifying observations in interoperable data exchange in health care.

The LOINC committee began to develop a universal vocabulary for reporting laboratory and clinical observations in February of 1994. It released the first version of LOINC codes in the spring of 1995 with about 6000 laboratory test result codes [1, 7]. The LOINC committee releases an updated version of the terminology twice each year. The current LOINC release (version 2.30, Feb 2010) contains 57,693 active codes, including both laboratory and clinical observation codes.

2.2 Current Use of LOINC Codes

Currently, LOINC is widely used in many organizations, including major laboratories (e.g. ARUP, Quest and LabCorp), hospitals, public health departments, health care provider networks (e.g. Indiana Network for Patient Care, INPC) [8], and insurance companies (e.g. United Healthcare) [9]. The National Electronic Disease Surveillance System (NEDSS) of Centers for Disease Control and Prevention (CDC) of the United States recommends using HL7 messages with LOINC codes to submit electronic laboratory reporting and surveillance data to federal agencies and departments [10]. Many studies have also evaluated how well LOINC has been applied to specific domains, such as nursing documents and standardized assessment measures and clinical data in hospital information systems (HIS) [6–8]. Dugas et al. analyzed the coverage of LOINC codes for document types in a German HIS, and reported that more than 93% of the local HIS documents and local document types could be assigned a LOINC code [11].

2.3 Evaluating Terminological Systems

Terminological systems (TSs) can be evaluated from two main perspectives: 1) the content-independent perspective, and 2) the content-dependent perspective [12, 13]. The "content-independent" approach mainly discusses the requirements of terminology systems from a functional, structural, and policy perspective. Examples of content-independent requirements include James Cimino's desiderata for controlled medical vocabularies [14], and the technical specification "Health informatics – Controlled health terminology – Structure and high-level indicators" published by the International Standards Organization (ISO) [15]. The "content-dependent" approach mainly evaluates the use of terminology systems in specific domains. Examples of content-dependent investigations include the evaluation of the coverage of the Unified Medical Language System (UMLS) for coding of concepts in the Gene Ontology (GO) [16], the evaluation of coding consistency of the Systemized Nomenclature of Medicine – Clinical Terms (SNOMED CT) in reporting rare diseases [17], and analyzing the coding consistency of LOINC in three hospitals [2].

By using the content-dependent approach to analyze the coverage of TSs, Cornet et al. defined two types of coverage. 1) **concept type coverage** – the number of concepts in a collection of concepts (e.g. result descriptions in a laboratory test catalog or dictionary) that can be mapped to concepts in a standard terminology. 2) **concept token coverage** – the volume of data instances covered by concepts in a standard terminology. For example if 10 instances (tokens) of hematocrit results are sent on an interface, all 10 instances are covered by the existence of a single hematocrit test code in the standard terminology. Concept token coverage means the percentage of laboratory test instances that have mappings in the standard terminology (▶Table 1) [13]. "Concept type coverage" is calculated by dividing the number of local codes that have been mapped to the reference terminology (i.e. concepts mapped to LOINC in the current study) by the total number of unique local codes. "Concept token coverage" is calculated by assessing instances of laboratory results and is the percentage of laboratory test instances whose code has been mapped to the reference terminology versus the total number of test instances. Compared to concept type coverage, concept token coverage can reflect what percentage of total volume of laboratory tests have LOINC mapping in daily use.

Table 1 The definition of concept type coverage and concept token coverage as used in this article

Definition
Concept type coverage: the number of concepts in a collection of concepts (i.e. result descriptions in a laboratory test catalog or dictionary) that can be mapped to concepts in a standard terminology (number of unique local codes having LOINC mappings/number of unique local codes)
Concept token coverage: the volume of data instances covered by concepts in a standard terminology. For example if 10 instances (tokens) of hematocrit results are sent on an interface, all 10 instances are covered by the existence of a single hematocrit test code in the standard terminology (total number of event IDs for each local code having a LOINC mapping/total number of event IDs for each local code)

2.4 Previous Reports on LOINC Mapping

Two large institutions [3, 4] have reported their LOINC mapping experiences. The common findings from these reports are:

1) The current LOINC database is not yet comprehensive: The LOINC database is still under active development and the number of LOINC codes has increased from about 6300 to 53,000 from 1996 to 2009. Dugas et al. reported that when using the Regenstrief LOINC Mapping Assistant (RELMA®) the LOINC coverage for their hospital information system concepts increased from 77% to 93% between version 3.23 and 3.24 of RELMA [11]. The LOINC committee recommends that any missing concepts be submitted to the LOINC committee for creation of new LOINC codes.

2) The frequency distribution of mapped local codes is highly skewed: concept type coverage was 46% and concept token coverage was 89.9% in the Department of Defense LOINC mapping project [4]. High volume tests are mapped more often than infrequent tests.

3) It is probably not appropriate to assign LOINC codes to all local codes: Some local codes do not carry any clinical information, e.g. an internal "Billed" flag – would not normally be exchanged between institutions. Also, local systems sometimes represent their content in ways that do not conform to HL7 best practices or to the LOINC model, e.g. "See Note", "See Chart" or multiple narrative text results in a field where a single code was expected [3, 4]. Local codes that violate the fundamental principles of unambiguous data exchange would also not be assigned LOINC codes.

3. Methods

3.1 Data Sources

The official LOINC database is stored in Microsoft Access™ 2003 format. We retrieved two fields, "date last changed (Add)" and "class types (laboratory class or clinical class)", of data from the LOINC database between April 1995 and April 2008. The numbers of laboratory and clinical observation codes were cataloged in order to observe the increase in the number of LOINC codes over time.

After obtaining IRB approval, de-identified patient data were collected from three institutions; 1) Associated Regional and University Pathologists, ARUP Laboratories (Salt Lake City, UT), 2) Intermountain Healthcare (Salt Lake City, UT), and 3) Regenstrief Institute, Inc. (Indianapolis, IN). ARUP Laboratories is a national clinical and anatomic pathology reference laboratory and is owned and operated by the Pathology Department of the University of Utah. Intermountain Healthcare is a not-for-profit health care provider organization, with hospitals located in many major cities in Utah. Regenstrief Institute, Inc., is an informatics and health care research organization, that is located on the campus of the Indiana University School of Medicine in Indianapolis.

These three large institutions were founding members of the LOINC committee and have contributed terms and concepts to the LOINC coding system [7]. These institutions represent quite different types of health care organizations. ARUP is a reference laboratory that receives samples from hundreds of clients. Intermountain is a health care provider organization that sends laboratory orders and samples to several different laboratories. Regenstrief is a health care research organization that convened and operates a regional health information exchange called the Indiana Network for Patient Care (INPC). Though ARUP and Intermountain have a similar geographical location, they did not share their resources or dictionaries while performing LOINC mappings. Each of the institutions performed their mappings using internal staff and not by commercial coding service companies. Their experiences provide three independent perspectives of LOINC mapping and usage.

3.2 Data Scope

This research focused on mappings related to laboratory LOINC codes. We chose laboratory test results because laboratory data is one of the most important kinds of data in the medical record and it has been

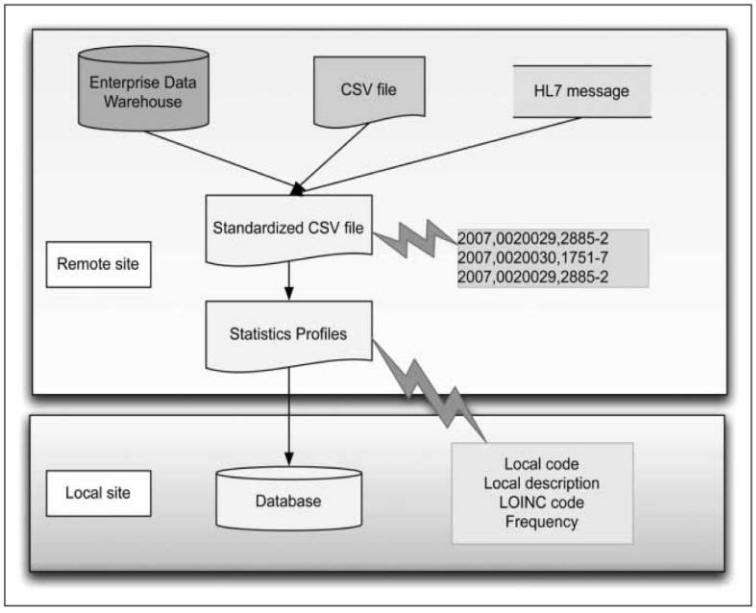

Fig. 1 The steps in data processing. The patient data were initially stored in the source institutions in various formats, with data being stored in an Enterprise Data Warehouse, comma separated values (CSV) files, or HL7 messages. The data was transformed into standardized CSV files at each site. The CSV files were then scanned to generate statistical profiles of each local code. Only the statistical profiles were sent to the authors for analysis.

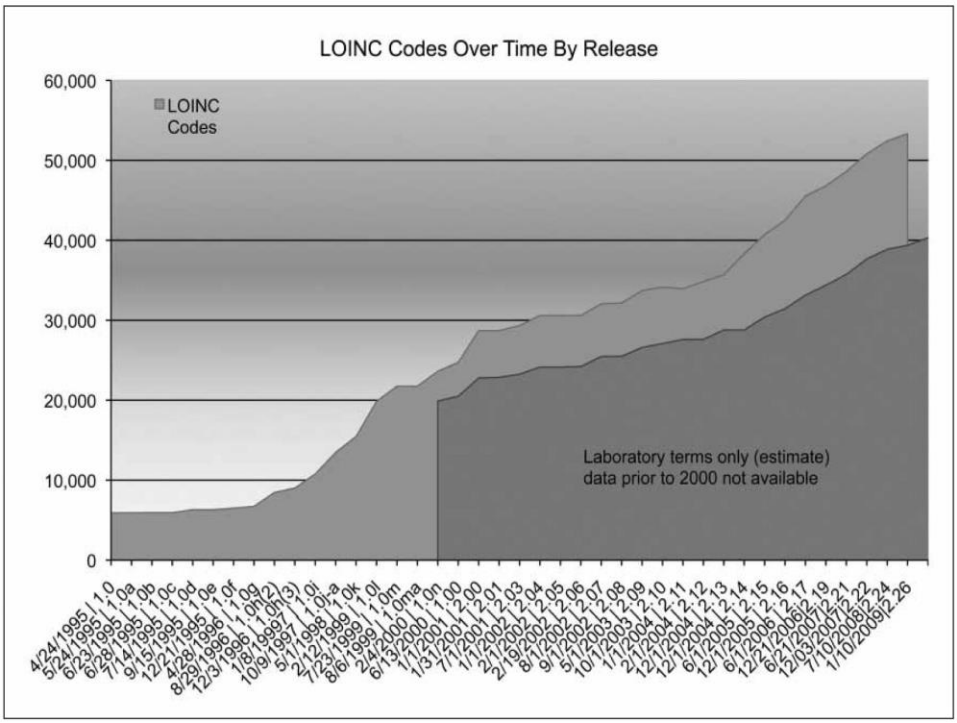

Fig. 2 The number of LOINC codes over time (May 1998 to Jan 2009)

mapped to LOINC codes more frequently than any other kind of data.

At ARUP and Intermountain, the de-identified patient data were collected for the month of April for five consecutive years (each April, from 2003 to 2007).

The data from Regenstrief came from the INPC, which presently includes data from more than 200 source systems and 18 different health systems. Regenstrief maps local system observation codes to terms in the INPC master dictionary, whose terms are also mapped to LOINC [3]. De-identified patient data for a 13-month period (August 2007–August 2008) and the mappings of local codes to LOINC codes (via the INPC master dictionary terms) were extracted from the five founding INPC health systems.

In these three institutions, the mappings were done incrementally and stored in reference tables, which only contain the mappings between local codes and LOINC codes. The version of the LOINC database used and the timestamps of the mappings were not available in these three institutions.

3.3 Data Collection and Processing

The patient data were retrieved by administrative staff at each institution. Each individual test result included the following database elements: 1) event ID, 2) observation ID (local code), and 3) observation description. No identifying information was included. To transform different formats of patient data of each institution to a common format, individual parsing programs were customized for each institution to generate standardized comma separated values (CSV) files (▶Fig. 1). LOINC mappings for local codes were added as a new column in the CSV files, with the LOINC mappings being provided from the reference file supplied by each institution. The CSV files were then scanned to calculate the following numbers: 1) numbers of unique local codes, 2) numbers of unique local codes having a LOINC code mapping, 3) total numbers of event IDs for each local code, and 4) total numbers of event IDs of each local code that was mapped to a LOINC code. Parsing programs were executed at each institution for processing patient data and only final statistical data was sent to the authors for analysis. After obtaining the primitive data as described above, concept type coverage and concept token coverage were calculated. In order to determine if the locally mapped tests were the most frequently resulted tests, cumulative concept token coverage of mapped and unmapped tests were calculated taking into consideration the frequency of the test.

3.4 Manual Review of Unmapped Codes

We wanted to estimate the number of local codes that were not mapped to LOINC codes that could theoretically be mapped by expert manual review of a sample of unmapped local codes.

We used Version 2.22 (released 12/03/2007) of the LOINC database as the target for mapping. To review those unmapped local codes, a 10% sample (concept type coverage) of all local codes from each institution was generated and the identical sample was given to two reviewers for

manual mapping. After manual mapping, reviewers rated results in two categories: 1) "Yes" – locally unmapped codes could be mapped manually, and 2) "NO" – locally unmapped codes could not be mapped manually. To evaluate the inter-rater agreement between two reviewers, the reviewed results were analyzed by using Fleiss' kappa [18], which can handle fixed numbers of reviewers and categorical ratings. Disagreements of manual mapping results from the first two experts were reviewed by a third expert to establish the gold standard. Also, each unmapped code was grouped into one of five categories according to the possible reason that the local code was not mapped: 1) no analyte – no suitable analyte was found in LOINC, 2) ambiguous meaning – the meaning of the local code was not clear and could not be determined by the information available to the reviewer, 3) internal use only – the local code may represent internal laboratory processing status rather than patient data, 4) overly specific methods – the local test name may have an overly specific measurement method, and 5) narrative results – the local code may represent a comment that is context-specific to a single result. After assigning categories to each code, we calculated concept type coverage and concept token coverage for each category of unmapped codes.

After manual review, we recalculated concept type coverage and concept token coverage by two approaches: 1) Adding all newly mapped local codes from the manual review sample to the original mapped local codes: This approach addresses the question of the extent to which current local codes

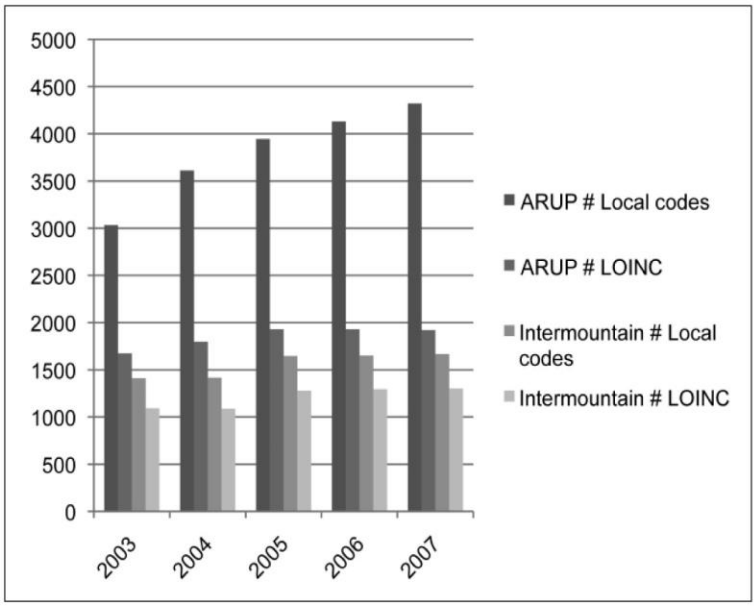

Fig. 3 The number of local codes and LOINC codes used at ARUP and Intermountain (every April, 2003–2007)

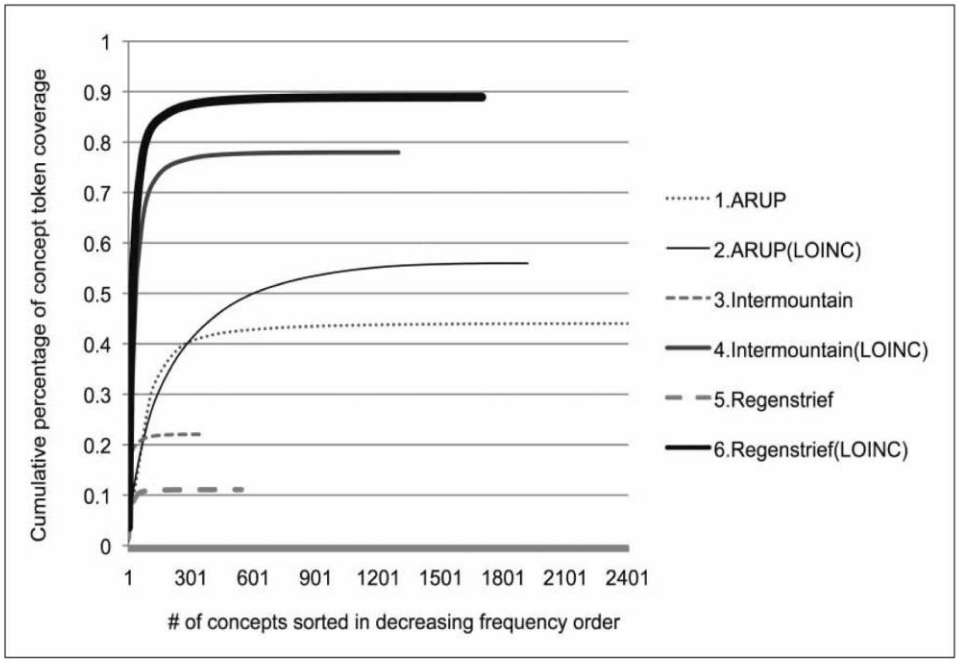

Fig. 4
The cumulative percentage of concept token coverage of mapped and unmapped tests at Intermountain, ARUP and Regenstrief (*) in 2007. The three solid lines represent the cumulative concept token coverage of mapped tests and the three dotted lines represent the percentage of unmapped tests. (*Of the five Regenstrief institutions, only the institution having the biggest volume was used to create this figure.) The results are NOT adjusted for manually mapped concepts.

Table 2 The level of local mappings from each institution. The data sets of Regenstrief consist of local codes collected from five institutions. The numbers (concept type) from the individual institutions are: 1311, 1176, 1471, 1187, and 2242.

	# of local codes	# of local codes mapped to LOINC	Concept type coverage	Concept token coverage
ARUP	4321	1918	44%	59%
Intermountain	1667	1297	78%	78%
Regenstrief	7387	5803	79%	88%

Table 3 The results of mappings before and after manual review of unmapped codes at each institution. After review, the number of new mappings found were 91, 8, and 75, respectively.

	Sample	Mapped	No mapping
ARUP	432	181 + 91	160
Intermountain	167	130 + 8	29
Regenstrief	739	575 + 75	89

Table 4 The percentage of local codes that had LOINC mappings in the original submissions and after manual mapping and review. (*)After excluding two types of local codes: "narrative results" and "internal use only"

	Before review		After review	
	Concept type coverage	Concept token coverage	Concept type coverage	Concept token coverage
ARUP	0.44	0.59	0.63 (0.73)*	0.72 (0.79)*
Intermountain	0.78	0.78	0.83 (0.90)*	0.80 (0.99)*
Regenstrief	0.79	0.88	0.88 (0.93)*	0.90 (0.997)*

can be mapped to LOINC codes by expert manual review. 2) Excluding two types of local codes ("internal use only" and "narrative result"), where assigning LOINC codes is not needed for clinical data exchange. This approach can reveal how well LOINC codes cover just the set of concepts that are useful for clinical data exchange.

4. Results

4.1 The Growth of Local Codes and LOINC Codes

Since May 1998, the number of LOINC codes has grown steadily from 15,464 to 53,345 and the majority of LOINC codes are laboratory terms (▶Fig. 2). At the same time, the number of local codes has also increased continuously. In 2003, at Intermountain, there were 1409 local codes which were mapped to 1092 LOINC codes; in 2007, there were 1667 local codes mapped to 1302 LOINC codes (▶Fig. 3).

4.2 The Cumulative Concept Token Coverage of Mapped and Unmapped Tests

▶Figure 4 shows the cumulative percentage of concept token coverage of mapped and unmapped tests at each institution in 2007. More than 70% of concept token coverage was accounted for by 200 locally mapped tests at Intermountain and Regenstrief.

4.3 The Concept Type Coverage and Concept Token Coverage before and after Manual Review

Agreement among the two reviewers was calculated by using Fleiss' kappa. The kappa value was 0.92 and interpreted as "almost perfect agreement" [19]. The disagreement of results was reviewed by a third expert for generation of the gold standard.

The number (concept type) of local codes in samples from ARUP, Intermountain and Regenstrief were 4321, 1667, and 7387 (▶Table 2). Before sampling for manual review of unmapped codes, the concept type coverage and concept token coverage were 0.44/0.59, 0.78/0.78 and 0.79/0.88 for ARUP, Intermountain, and Regenstrief, respectively.

The one-tenth sample of these data sets contains 432, 167, and 739 codes, respectively (▶Table 3). An attempt was made to manually map all unmapped codes from the samples. After adding the new mappings to the originally mapped codes, concept type coverage and concept token coverage were 0.63/0.72, 0.83/0.80 and 0.88/0.90, respectively (▶Table 4).

4.4 The Analysis of Mapped and Unmapped Codes after Review

▶Figure 5 shows the frequency of initially unmapped local codes which could be mapped after manual review. The most frequently mapped and unmapped codes were listed and ordered based on their frequency in instance data (▶Tables 5 and 6). After categorizing unmapped codes into the five categories of unmapped reasons, concept type coverage and concept token coverage for all unmapped codes in each category were calculated (▶Table 7). The largest concept token coverage (0.64 and 0.92) of unmapped codes at Intermountain and Regenstrief was due to "narrative result", e.g. "Comments Result, Qualitative for GFR", "Interp Gliadin/Gluten IgA"; at ARUP the largest concept token coverage of unmapped codes (0.57) was due to "no analyte", e.g. "NB C12-OH". Across the three institutions, "internal use only", e.g. "Report

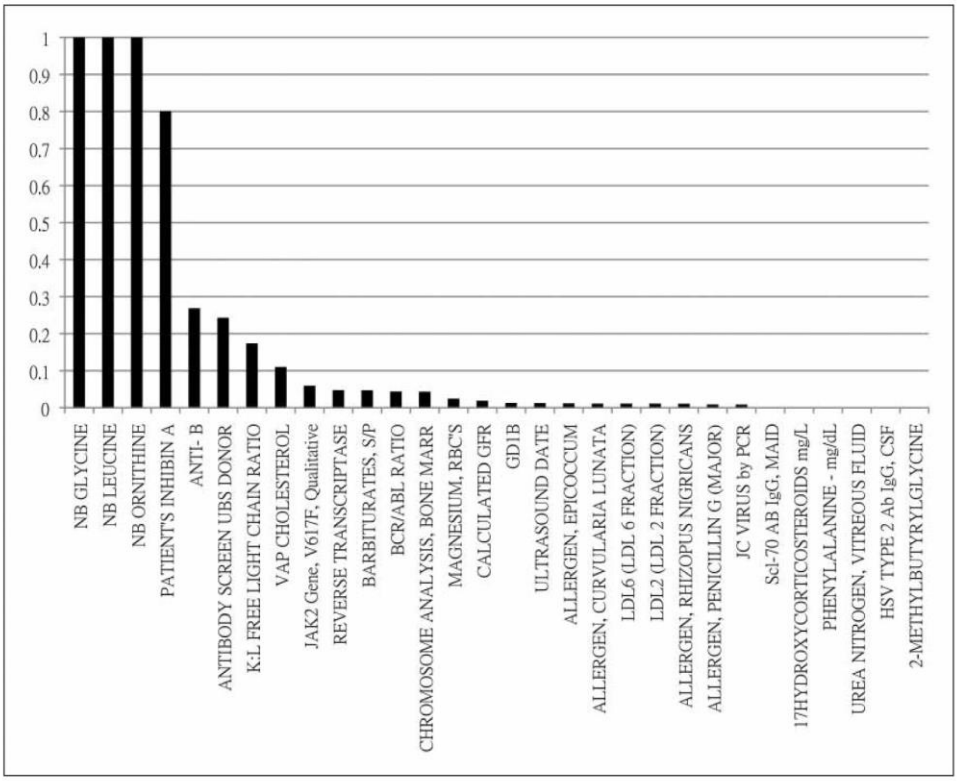

Fig. 5
The histogram of concept token coverage of originally unmapped codes which were manually mapped to LOINC at ARUP. The frequency is normalized by the biggest frequency of the test (NB Glycine).

Status, Qualitative", is a common reason for unmapped codes. After excluding two types of local codes ("narrative results" and "internal use only") from the dataset, concept type coverage and concept token coverage were 0.73/0.79, 0.90/0.99 and 0.93/0.997, respectively (▶Table 4).

5. Discussion

5.1 Local Mapping Is Incomplete

Concept type coverage of mapping increases from 0.44 to 0.63, 0.78 to 0.83 and 0.79 to 0.88 at ARUP, Intermountain and Regenstrief, respectively, which means the local mappings were incomplete in each institution. Some possible reasons were: 1) mapping is a labor-intensive job, so mapping is not performed on all local codes. ▶Figure 4 also shows that frequent tests are more commonly mapped. 2) New local codes and LOINC codes continue to be created and the mapping process does not keep up. It is hard to keep local mappings up to date on the latest LOINC version. 3) Not everyone is using LOINC codes to exchange data yet, therefore there is no urgency to do the LOINC mappings. Although concept type coverage is not 100% yet, these institutions can still report patient data using internal codes.

Table 5 The top 10 newly mapped local terms after manual review are listed by their ranks (based on use in instances of data) in the three institutions. In the Intermountain sample, the number of mapped codes is less than 10.

ARUP	Intermountain	Regenstrief
NB GLYCINE	Cefdinir	MPV
NB LEUCINE	Oxacillin	ALLERGY HX
NB ORNITHINE	5-Hydroxyindoleacetate, Urine Qualitative Sendout	APPEARANCE-UR
PATIENT'S INHIBIN A	Cefotaxime (meningitis)	Gentamicin
ANTIBODY SCREEN UBS DONOR	Oxycodone	Piperacillin
K:L FREE LIGHT CHAIN RATIO	ABO Type	Ceftazidime
VAP CHOLESTEROL	Herpes Simplex Virus 1+2 Ab IgM, Cerebrospinal Fluid Quantitative	INR
JAK2 Gene, V617F, Qualitative		Vancomycin
REVERSE TRANSCRIPTASE		Base Excess
BARBITURATES, S/P		Oxacillin

Table 6 A sample of unmapped concepts showing the categorization of reasons that the codes were not mapped. There are five categories: 1) A – no analyte, 2) M – meaning is not clear, 3) I – internal use, 4) O – overly specific method, and 5) N – narrative result.

ARUP	Reason	Intermountain	Reason	Regenstrief	Reason
ACYLCARNITINE PROFILE	A	Comments Lab Result, Qualitative~for GFR	N	SPECIMEN DESCRIPTION	N
NBC14:1_C16 NBS RATIO	A	Report Status, Qualitative (RPT)	I	Final	N
NB GLU_CIT NBS RATIO	A	Comments	N	LDL MESSAGE	M
NB METHIONE	A	Cerebrospinal Screen, Cerebrospinal Fluid Qualitative	A	Initial Specimen?	I
NB C12-OH	A	Method of Release	I	Other	M
HIRLU	M	Comments Lab Result, Qualitative (CMRSS)	N	Xanthochromic	A
ESTIMATED DUE DATE	I	Specimen Number, Serum Quantitative	I	Engraft Study Post TX	N
DETERMINED BY:	I	Hold Clot (order only)	A	SCL T&B lymph.	A
VT FINAL DIAGNOSIS	I	Comments Lab Result, Qualitative (CVAR)	N	BB Physician	M
ENDOCERVICAL COMPONENT	A	Numbers/Type of Containers:	I	CSF-XANTHCHROMIA	A
UA CULTURE IF ?	I	Comments Lab Result, Qualitative (CPSAF)	N	DETERMINED BY:	I
ANTI- B	A	Comments Lab Result, Qualitative (CMNT)	N	Allergen Scoring Chart	I
META UF INTERP	N	Antigen Type	A	DIABETIC	M
CS ADD REQUEST I RAST	I	RAST Interpretation, Serum Narrative	N	TRICH SOURCE SCREEN	A
DOCTOR REVIEW – PT PCR	N	Comments Lab Result, Qualitative (CFVL)	N	Miscellaneous CPT	I
HEP B CORE AB S/C RATIO	A	Result Date, Quantitative	A	Interp Gliadin/Gluten IgA	N
SP CLINICAL HISTORY	A	MoM for Nuchal Translucency	A	HSV 1,2 DNA Specimen Type	A
OPIATES, NUMERIC INSTRUMENT	O	Phone orders	I	Interpretation	N
BARBITURATE, NUMERIC INSTRMNT	O	Comments Lab Result, Qualitative (CFTA)	N	LS Interpretation	N
INTERPRETATION/SPECIAL CHEM	N	Chronic Lymphocytic Leukemia Panel, Blood Qualitative Flow Cytometry~USE CODE FLOWLL	A	PRE TRANS B/P	I
VT TISSUE DESCRIP-CYTOLOGY	N	Insulin Sensitivity Index, Serum or Plasma Quantitative	A	HLA-DR DQ low res	O
VT MINI DIAGNOSIS	N	Comments Lab Result, Qualitative~Used with CLSW	N	PHOSPHATIDLSER IGG	A
ANATOMIC PATHOLOGY TRACKING T	I	Alpha-Beta %	A	Seq. HLA-B Interp	N
SP COMMENTS	N	Pathologist Interpretation, Qualitative~INACTIVE 8/14/2007	N	Cryptococcus AG BLD Interp	N

5.2 Not All Local Codes Should Be Assigned a LOINC Code

Assigning LOINC codes to local codes like "narrative results" does not help create interoperable data exchange. For example, local observations like "Seq. HLA-B Interp" and "DOCTOR REVIEW – PT PCR", usually have values that are comments or directions to a human reader like "See Note" or "See Chart". LOINC is designed to carry clinical data using the EAV strategy, but narrative results sometimes contain a mix of different kinds of information: analyte names, actions, people's names, and date and time information. A real example of a narrative example is "Colony Bacillus species. Results called to and read back by John 10/02/2008 14:41:56". This result value does not follow the EAV style. It is probably not useful to try to assign LOINC codes that could capture the context of this statement. These kinds of local codes carry important information, but it can only be read and understood by human users. A better strategy is to break the information into discrete data elements so it can be used by automated decision support processes. Terminologists and system developers should avoid using narrative text to encode clinical data for medical exchange and follow the style of discrete EAV data [20].

Assigning LOINC codes to "internal use" codes like "RETICRTR BILL", which has values of "Billed" and "Confirmed", would not typically be useful for inter-enterprise data exchange because they do not carry any clinical data.

At Intermountain and Regenstrief, the main two reasons for unmapped codes are "narrative results" and "internal use only". Assuming that these local codes are not appropriate for inter-enterprise data exchange, a flag could be added to the lab reference table to indicate a "Do not map" status for those items [3, 4]. After excluding "narrative" and "internal use" codes, coverage increased to 0.73/0.79, 0.90/0.99 and 0.93/0.997, respectively. At Intermountain and Regenstrief, the current LOINC database contains codes that could cover about 99% of volume of laboratory tests. New LOINC codes will need to be created for ARUP content if concept token coverage for ARUP is to reach the same level of coverage as currently exists for Regenstrief and Intermountain.

Table 7 The concept type coverage and concept token coverage of unmapped codes in each category. A – no analyte, M – meaning is not clear, I – internal use, O –overly specific method, and N – narrative result. The bold number indicates the largest number in each category of coverage.

	Concept type coverage					Concept token coverage				
	A	M	I	O	N	A	M	I	O	N
ARUP	**0.52**	0.08	0.20	0.04	0.16	**0.57**	0.09	0.22	0.03	0.09
Intermountain	**0.39**	0.0	0.23	0.0	**0.39**	0.05	0.0	0.31	0.0	**0.64**
Regenstrief	**0.40**	0.08	0.22	0.04	0.26	0.01	0.05	0.02	0.002	**0.92**

5.3 Creation of New LOINC Codes

The unmapped local codes in the "no analyte" category should be submitted to the LOINC committee for the creation of new LOINC codes. The unmapped tests which are due to "overly specific method", e.g. "HLA-DR DQ Hi Res Amp2" or "HLA-DR DQ Hi Res Amp1" pose a different problem. These local codes include very specific information about the method. We would propose that if it is desirable to include highly specific method information with the patient result, then the method be sent as coded data in a special "method type" field in the result message, rather than precoordinating the method name into the test code. We also noted inconsistency across institutions regarding specificity of mappings as they relate to methods. It appears that sometimes mappers link the method-specific codes to a more general LOINC code, and at other times they link to a method-specific LOINC code. This causes inconsistency in mappings across institutions. A comprehensive analysis of these inconsistencies is beyond the scope of this paper, but we would like to examine this issue in future work.

The current process of submitting requests for new LOINC codes asks users to provide information for the five primary axes of the LOINC code definition [21]. However, the creation of local codes is often a separate process from mapping to LOINC codes or submitting requests for new LOINC codes, and different people are usually responsible for these separate activities. Therefore, it is often the case that it requires extra effort to gather the information to submit new local codes for the assignment of LOINC codes. People do not always go to the extra effort to submit requests for new LOINC codes to match new local codes. At Regenstrief, they have deployed an Exception Browser [3] to monitor all of the INPC data streams. If there is a new local code which cannot be found in their master dictionary, the Exception Browser generates an exception and requires further actions by a human to deal with the new codes. They can either request new LOINC codes or make a notation in the mapping file that the new local code is to be ignored. This kind of automation can facilitate the appropriate creation of new LOINC codes.

5.4 Version Control of LOINC Mappings

The version of the LOINC database used for mapping was not available from the three institutions. Newer versions of the LOINC database have the possibility of affecting the calculation of concept type coverage and concept token coverage following manual review of initially unmapped local codes. Because the new database has more codes, it could be that an unmapped code can now be mapped whereas at the time of initial mapping no matching concept existed in the older version of the LOINC database. Use of the newer version of the LOINC database could change the number of unmapped local codes in the "no analyte" and of the "overly specific method" categories, but these changes would only make small differences in our overall statistics. Our goal was to estimate the maximal level of LOINC mapping that could reasonably be achieved, and we believe our method leads to a good estimate of the maximum mapping that can be achieved in the current database.

5.5 The Frequency Distribution of Local Codes that Are Mapped to LOINC Is Highly Skewed

In a previous study of INPC laboratory data, it was concluded that 244 to 517 local codes represented 99% of the volume from all institutions and there were 97 local codes that were common to all five institutions [22]. This conclusion also coincides with our observation that only a small number of tests account for a large portion of the volume at Intermountain and Regenstrief, and that about 200 locally mapped tests account for more than 70% of test volume. At ARUP, it takes a larger number of tests to account for the same total volume. A possible reason is that Intermountain and Regenstrief, which are general health care provider organizations, use more common tests, e.g. general biochemistry, but ARUP, which is a reference laboratory, has a greater preponderance of rare tests, e.g. allergen tests, as compared to the other two institutions. Based on these observations, we would predict that general health care organizations, mapping a relatively small number of tests (less than 500), will cover a large volume of the common laboratory tests. Since concept token coverage is higher than concept type coverage, we can infer that on average mapped

local codes occur more often in instances of patient data than the unmapped local codes. To extend this research, we plan to pool all frequent tests and their LOINC mappings from the reference tables of each institution to generate a master index file containing the most frequent local codes and their mappings. This file could then be used by institutions as they begin to map their local codes, and they would initially only need to map the codes which are listed in the master index file. They should be able to reach a high concept token coverage without spending a lot of time mapping all local codes [22].

6. Limitation

The three organizations examined in this study have been intimately involved in LOINC development, and they may be more likely to have local names that match LOINC content and have a better understanding of how to do LOINC mappings. Thus, the three institutions are not representative of institutions in the US or worldwide. The implication is that the percentage of locally mapped local codes and the coverage of local codes in these three institutions is probably higher than would be expected in other institutions. Finally, we did not verify the accuracy and consistency of the mappings of local codes to LOINC codes in this phase of our research, and more work is needed to gain insight into these aspects of mapping across institutions.

7. Conclusions

The number of local codes and LOINC codes continues to grow, which means that each institution needs a process to maintain their local LOINC mappings. For general health care providers, concept token coverage can reach about 99% for daily use. The reference laboratory has a greater number of rare tests, which will require creation of new LOINC codes to reach the same level of concept token coverage. Our research also supports the conclusions of others that not all local codes should be assigned LOINC codes. There should be public discussions about how laboratory processes could be further standardized so that the results produced are more consistent and interoperable. There should also be public discussions to develop more precise rules for when LOINC codes should be assigned. Extending this research to examine the consistency and accuracy of local mappings across institutions will be an important next step in evaluating whether LOINC is meeting its goal of being a universal coding system for observation identifiers.

Acknowledgments

The authors would like to thank Brian Jackson, MD and Alan Terry, MS from ARUP for retrieving and processing the ARUP data set for this research and Chia-Cheng Lee, MD from the Biomedical Informatics Department of the University of Utah for reviewing the sample data.

References

1. Huff SM, Rocha RA, McDonald CJ, De Moor GJ, Fiers T, Bidgood WD, Jr., et al. Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary. J Am Med Inform Assoc1998; 5 (3): 276–292.
2. 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.
3. Vreeman DJ, Stark M, Tomashefski GL, Phillips DR, Dexter PR. Embracing change in a health information exchange. AMIA Annu Symp Proc 2008. pp 768–772.
4. Lau LM, Banning PD, Monson K, Knight E, Wilson PS, Shakib SC. Mapping Department of Defense laboratory results to Logical Observation Identifier Names and Codes (LOINC). AMIA Annu Symp Proc 2005. pp 430–434.
5. Health Level Seven International (cited 2009). Reference Information Model, Health Level Seven, International. Available from: http://www.hl7.org.
6. Nadkarni PM, Marenco L, Chen R, Skoufos E, Shepherd G, Miller P. Organization of heterogeneous scientific data using the EAV/CR representation. J Am Med Inform Assoc 1999; 6 (6): 478–493.
7. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leavelle D, Leland D, 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 Chem1996; 42 (1): 81–90.
8. McDonald CJ, Overhage JM, Barnes M, Schadow G, Blevins L, Dexter PR, et al. 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.
9. McDonald C, Huff S, Suico J, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clinical Chemistry 2003; 49 (4): 624.
10. Potential effects of electronic laboratory reporting on improving timeliness of infectious disease notification – Florida, 2002–2006. MMWR Morb Mortal Wkly Rep 2008; 57 (49): 1325–1328.
11. Dugas M, Thun S, Frankewitsch T, Heitmann KU. LOINC codes for hospital information systems documents: a case study. J Am Med Inform Assoc 2009; 16 (3): 400–403.
12. Arts DG, Cornet R, de Jonge E, de Keizer NF. Methods for evaluation of medical terminological systems – a literature review and a case study. Methods Inf Med 2005; 44 (5): 616–625.
13. Cornet R, de Keizer NF, Abu-Hanna A. A framework for characterizing terminological systems. Methods Inf Med 2006; 45 (3): 253–266.
14. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998; 37 (4–5): 394–403.
15. ISO/TC215. Health informatics – Controlled health terminology – Structure and high-level indicators. 2002. Report NO.: 17117.
16. Bodenreider O, Mitchell JA, McCray AT. Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics. Proc AMIA Symp 2002. pp 61–65.
17. Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. J Am Med Inform Assoc 2007; 14 (4): 497–506.
18. Fleiss J. Measuring nominal scale agreement among many raters. Psychological Bulletin 1971; 76 (5): 378–382.
19. Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977; 33 (1): 159–174.
20. Coyle JF, Mori AR, Huff SM. Standards for detailed clinical models as the basis for medical data exchange and decision support. Int J Med Inform 2003; 69 (2–3): 157–174.
21. LOINC Committee. LOINC Submissions. Indianapolis, IN 2008 (updated 2008 Dec 23). Available from: http://loinc.org/submissions.
22. Vreeman DJ, Finnell JT, Overhage JM. A rationale for parsimonious laboratory term mapping by frequency. AMIA Annu Symp Proc 2007. pp 771–775.

CHAPTER 4

CORRECTNESS OF VOLUNTARY LOINC® MAPPING FOR LABORATORY TESTS IN THREE LARGE INSTITUTIONS

Coauthors:

Daniel J. Vreeman, Clement J. McDonald, Stanely M. Huff

Reprinted with permission from Lin MC, Vreeman DJ, McDonald CJ, Huff SM. Correctness of voluntary LOINC mapping for laboratory tests in three large institutions. American Medical Informatics Association Annual Symposium proceedings. 2010;2010:447-51.

Correctness of Voluntary LOINC Mapping for Laboratory Tests in Three Large Institutions

Ming –Chin Lin, MD[1], Daniel J. Vreeman, PT, DPT[2,3], Clement J. McDonald, MD[4], Stanley M. Huff MD[1,5]
[1]The University of Utah, Salt Lake City, UT
[2]Regenstrief Institute, Inc., Indianapolis, IN
[3]Indiana University School of Medicine, Indianapolis, IN
[4]Lister Hill Center, National Library of Medicine, Washington, D.C.
[5]Intermountain Healthcare, Salt Lake City, UT

Abstract

With IRB approval, we obtained de-identified laboratory test data from 3 large institutions (ARUP, Intermountain, and Regenstrief). In this study we evaluated correctness of mapping local laboratory result codes to Logical Observation Identifier Names and Codes (LOINC®). We received 9,027 laboratory tests mapped to 3,669 unique LOINC codes. A one tenth sample (884 tests) was manually reviewed for correctness of the mappings. After review, there were 4 tests mapped to totally unrelated LOINC codes and there were 36 tests containing at least one error in mapping to the 6 axes of LOINC. The errors of LOINC mapping could be categorized into 4 systematic errors: 1) human errors, 2) mapping to different granularity, 3) lack of knowledge of the meaning of laboratory tests and 4) lack of knowledge of LOINC naming rules. Finally, we discuss how these systematic mapping errors might be avoided in the future.

Introduction

Evaluation of Terminological Systems

Many terminological systems (TSs) are used in Electronic Medical Records (EMR) to enable interoperability in health care. International Classification of Disease (ICD), the Systemized Nomenclature of Medicine (SNOMED), and the Logical Observation Identifier Names and Codes (LOINCR®)[1] are examples of widely used TSs. To improve the development of TSs, it is important to evaluate them from two main perspectives: 1) content independent (functional) evaluations, and 2) content dependent evaluations. Content independent evaluation of TSs discusses the requirements of TSs from a functional, structural and policy perspective. Examples of functional criteria are James Cimino's desiderata[2] for controlled medical vocabularies. Content dependent evaluations focus on concept coverage, term coverage, synonym completeness, etc. Until TSs are in widespread use in health care systems, the usage of TSs can be pooled for analysis. Two examples include the evaluation of the coverage of the Unified Medical Language System (UMLS) for coding of concepts in the Gene Ontology (GO)[3] and analysis of the coding consistency of LOINC in three hospitals[4].

Current LOINC usage and evaluation

The LOINC committee began to develop a universal code system for reporting laboratory and clinical observations in February of 1994. The current LOINC release (version 2.30, February 26, 2010) contains 57,693 active codes, including both laboratory and clinical observation codes. LOINC is widely used in many domains, including major laboratories, hospitals, public health departments, health care provider networks and insurance companies[5].

Since LOINC is in widespread use, Huff et al. proposed that there were two main perspectives for evaluating LOINC: 1) Coverage 2) Correctness[6]. The goal of LOINC is to provide standard codes to improve interoperability when sharing clinical data. In pursuit of that goal, the LOINC database is designed to support greater accuracy and to decrease the time and cost when mapping from local codes to standard codes[6]. Manual mapping is not usually an easy task. Without a good understanding of content and the design of LOINC codes, using LOINC could have two possible types of errors: 1) human errors: simple typographic or selection errors, 2) semantic errors, where there is a difficulty in choosing the correct LOINC code. The second kind of error can occur if LOINC is too complicated for the average mapper to understand, or if the codes have ambiguous meaning. Users could have trouble in aligning local information with the six axis model of

LOINC codes. Lau et al. at 3M Health Care reported that in a large scale mapping project LOINC mapping was time consuming and laborious, and that human variation caused mapping inconsistencies and errors[7].

Evaluating LOINC mappings using extensional definitions

One challenge of evaluating LOINC mapping is how to determine the actual meaning of local codes. Most institutions document very vague descriptions, which contain[5]: 1) Idiosyncratic abbreviations, (e.g. EPI-Cell), 2) No specific type of analyte (e.g. HSV TYPE 1/2), 3) Incomplete information: No description of method (e.g. EIA), scale (e.g. quantitative or ordinal), property (e.g. titer), time (e.g. 24 hour) and specimen type (e.g. Serum). Clarifying the meaning of local codes is very time consuming. One approach to determine the meaning of local codes is to observe test values stored in clinical systems, such as frequency of testing, mean value, standard deviation of the value, units of measure, value type (coded vs numeric). Those profiles or extensional definitions reflect the actual meaning of tests in the system and are called extensional definitions (EDs)[8]. By matching EDs of laboratory tests, Zollo et al. automatically cross mapped local laboratory codes from 3 institutions with an accuracy of 81%[9].

Problem Statement

The accuracy of mappings from local codes to LOINC codes influences the quality of interoperability in exchanging clinical observations. We wanted to evaluate mapping accuracy in existing systems, so we collected voluntary LOINC mappings of laboratory tests from three large institutions and created extensional definitions associated with these tests. We analyzed the correctness of the mappings, identified systematic errors, and then formulated some suggestions that might improve the LOINC mapping process.

Methods

Data sources

With IRB approval, de-identified patient data were collected from three institutions: 1. Associated Regional and University Pathologists, ARUP Laboratories (Salt Lake City, UT) 2. Intermountain Healthcare, Intermountain (Salt Lake City, UT) 3. Regenstrief Institute, Inc. (Indianapolis, IN). ARUP Laboratories is a national clinical and anatomic pathology reference laboratory and is owned and operated by the Pathology Department of the University of Utah. Intermountain Healthcare is a not-for-profit health care provider organization, with hospitals located in many major cities in Utah. Regenstrief Institute, Inc., an informatics and healthcare research organization, that is located on the campus of the Indiana University School of Medicine in Indianapolis. Regenstrief operates a regional health information exchange in central Indiana called the Indiana Network for Patient Care[10] that includes data from more than a hundred source systems and five major hospital systems.

Data scope

This research focused only on mappings related to standard laboratory LOINC codes, i.e. we excluded anatomic pathology and microbiology tests. We chose to focus on laboratory test results because laboratory data is one of the most important kinds of data in the medical record and it has been mapped to LOINC codes more frequently than any other kind of data. At ARUP and Intermountain, the de-identified patient data were collected for the month of April (April, 2007). At Regenstrief, this data was retrieved for a 13 month period (August, 2007 – August, 2008). The mappings of local codes to LOINC codes were also collected from the three participating institutions. In this study, we utilized data collected in 2007 from all three institutions.

Collect data and generate extensional definitions

The patient data were initially stored in the source institutions in various formats, with data being stored in an Enterprise Data Warehouse, comma separated values (CSV) files, or HL7 messages. The patient data were retrieved by administrative staff at each institution. Each individual test result instance included the following database elements: 1.Local code 2.Local description 3.Numeric value 4.Coded variables 5.Units of measure (UOM) 6.LOINC mapping. The retrieved data were transformed into standardized CSV files at each site. The CSV files were then processed to generate extensional definitions (EDs) (Table 1) of each local code. Only EDs were sent to the authors for analysis; no patient identifying information was included. After preparing summary reports, a one tenth sample was examined for LOINC mapping accuracy following explicit review criteria.

Review Criteria

We evaluated the accuracy of the mappings based on instantiating the six axis LOINC model for each local code and comparing that local instance to the definition of the LOINC code to which the local code was mapped (Table 1). For each axis we reviewed,

Extensional definitions	Example	Containing information for review
Local description	"Creatinine, 24 hr urine", "Sodium urine"	Local description - mainly provides analyte information. In some cases, it also provides method (e.g. EIA), scale/property (e.g. titer) , time (e.g. 24 hr) or system information (e.g. urine)
Mean	1.46,137	Mean - provides scale/property information (e.g. SCnc/Qn). It also implies the type of analyte, e.g. The mean of "Sodium" should be around 135-145.
Standard deviation	0.54, 7.02	Standard deviation - provides scale/property information (e.g. SCnc/Qn). It also implies the system, e.g. The same analyte in the body fluid could have greater standard deviation than in the blood.
Coded variables	1:8, Negative, Positive	Coded variables - provides scale/property (e.g. Titr/Qn or ACnc/Qn)
Units of measure	g/24 h, mmol/L, mg/dl	Units of measure - provides scale/property information. Sometimes it also provides time information (e.g. g/24h implies 24 hour)
Frequency	50, 184	Frequency - implies whether tests are frequent (e.g. biochemistry tests) or rare (e.g. allergen test)

Table 1. The example of Extensional Definitions (EDs) including local description, mean, standard deviation, units of measure, coded variables and frequency.

we defined 3 categories of review results: 1) **Correct:** The mapping of a particular axis is correct, e.g. for "Creatinine, 24hr urine" that was mapped to "2162-6:Creatinine:None:MRat:24H:Qn:Urine", the mapping of "Creatinine" for the Analyte is correct. 2) **Error:** The mapping of the axis is incorrect, e.g. a test, "ISLET CELL Ab, IgG" was mapped to "33563-8:Pancreatic islet cell Ab.IgG:None:ACnc:Pt:Qn:Ser", but the test result values are "1:4;1:8;1:16", which are "Titer(s)" and are not "ACnc"; so the mapping is in error. We also considered it an error if a more specific test is mapped to a more general concept, e.g. "Ab.IgG" is mapped to just "Ab". This is considered and error because it represents a loss of meaning when going from the specific code to the more general code. 3) **Unknown:** The mapping of the axis could not be verified due to insufficient information, e.g. a test, "Succinic acid" was mapped to "Succinate/Creatinine (Ratio)". Because the test description only contains "Succinic acid", we cannot confirm the association to "Creatinine". It is often the case that there is only very general information contained in local test code descriptions.

Results

After collecting the data from all three institutions, 9,027 local laboratory tests mapped to 3,669 unique LOINC codes. A one tenth sample of these 3,669 unique LOINC codes contained 884 laboratory tests that were manually reviewed for correctness of the LONC mappings.

Review result	Tests number
Totally unrelated mapping	4
Containing at least one error in mapping to the 6 axes of LOINC	36
Total sample number	884

Table 2. The accuracy of mappings for 884 tests by examining each axis of LOINC mapping.

	A	M	P	T	Sc	S
C	755	733	860	869	877	310
U	111	140	10	11	3	562
E	14	7	10	0	0	8

Table 3. The review results of 880 tests by examining each axis of LOINC mapping. The review results were categorized into 3 categories: 1) Correct **(C)** - The mapping is correct, 2) Unknown **(U)** – The information is insufficient for review, and 3) Error **(E)** – The mapping is an error. (**A**) – Analyte, (**M**) – Method, (**P**) – Property, (**T**) – Time, (**Sc**) –Scale, (**S**) - System

We found 4 tests mapped to totally unrelated LOINC codes (Table 2): 1) "Cannabinoids" was mapped to "Bacteria identified:Culture:Prid:Pt:Nom:Thrt", but "Cannabinoids" is a chemical substance in the nervous and immune systems and not a bacteria. 2) There are two tests having identical test names in one institutions, "EPI CELL-UR" with UOM "/HPF"

Axis	Local codes	Original incorrect mapping	Explanation
Analyte	Glucose 30 Minute **75 g** Glucose PO, Serum or Plasma	Glucose^30M post dose glucose:None:MCnc:Pt:Qn:Ser/ Plas	The correct mapping is "Glucose^30M post **75g** glucose PO:None:MCnc:Pt:Qn:Ser/Plas"
Method	"Bartonella Henselae Ab IgM, Serum Quantitative **EIA**	Bartonella henselae Ab.IgM:**IF**:Titr:Pt:Qn:Ser	The correct method is "**EIA**"
Property	ISLET CELL Ab, IgG	Pancreatic islet cell Ab.IgG:None:**ACnc**:Pt:Qn:Ser	The test has tests value "1:4,1:8,1:16,1". The correct property should be "**Titr**", not "ACnc".
System	**CSF**-MONOS	26486-1:Monocytes/100 leukocytes:None:NFr:Pt:Qn:**Bld**	The correct system is "**CSF**"

Table 4. Example errors of each axis from the manual review.

that was mapped to "Epinephrine:None: MCnc:Pt:Qn:Urine". Based on the local description

and UOM, "EPI CELL-UR" should mean "Epithelia Cell of Urine" per high power field (HPF) using light microscopy. and 3) "Estrogen receptor IP" was mapped to "Basement membrane Ab.IgG:IF:ACnc:Pt:Ord:Ser" and "Estrogen receptor IP" should mean a estrogen receptor, not a basement membrane, which is a thin sheet of fibers underlying the epithelium.

After excluding the above 4 tests, 880 tests were then reviewed by each axis of LOINC mapping (Table 3). Among 880 tests, there are 36 tests containing at least one error in one of the six axes. The examples of errors of each axis are shown in Table 4.

Discussion

Mapping is not yet perfect

After reviewing voluntary LOINC mapping of three institutions, we found that the mappings are not yet perfect. There are four types of systematic errors identified so far: **1) Human errors**: There were 4 tests mapped to totally unrelated LOINC codes. The reasons might be the misunderstanding of the meaning of an acronym or simply picking the wrong LOINC code. **2) Mapping to different granularity:** These types of errors mainly happen when there is a group of LOINC codes having a similar analyte but varying degrees of specificity. For example, LOINC codes were not chosen correctly according to the specific subtype of analyte e.g. Ab.IgG was mapped to Ab, instead of Ab.IgG. Currently the LOINC database provides a multi-axial hierarchy of LOINC codes. When exploring LOINC codes, the parent-child relationship of LOINC codes could be displayed for choosing the correct granularity of mapping. **3) Lack of knowledge of the meaning of laboratory tests:** Laboratory test methods are still being actively developed, e.g. there are enzyme immunoassay (EIA), Western blot, and indirect immunofluorescene (IF) for detecting antibodies. Without understanding the detail of the test itself, it is hard to distinguish the various methods and choose the correct LOINC code. **4) Lack of knowledge of LOINC naming rules:** The same analyte measurement could be mapped to different LOINC codes depending on the context, e.g. "BENZODIAZEPINES, URINE" has three candidate LOINC codes "19283-1:Benzodiazepines cutoff:Screen:MCnc:Pt:Qn:Urine", "19284-9:Benzodiazepines-cutoff:Confirm: MCnc:Pt:Qn:Urine";"19064-5:Bezodiazepines cutoff:null:MCnc:Pt:Qn:Urine" one is for "Screen", one is for "Confirm", and another is for "Null". The methods for "Screen" and "Confirm" could be different, so the sensitivity and precision of tests might be different; therefore two LOINC codes have been created. Without knowing the LOINC naming rules, users cannot choose the correct method among "Screen", "Confirm" or "Null" appropriately.

Improving the correctness of LOINC mapping

Medline indexing consistency can be used as a model for evaluating LOINC mappings. Funk et al. concluded that high quality MEDLINE indexing requires an excellent controlled vocabulary, exemplary quality control and highly trained indexers[11]. Based on the above conclusions, LOINC mapping could be improved by five possible approaches: **1) Use more specific naming conventions for local descriptions:** Based on the above results, many tests do not have a specific description, e.g. 562 tests out of 880 tests do not contain information about "System" (Specimen types) and it is hard to guess the correct specimen type from EDs. Including more specific information for each axis could save time and improve the quality of mapping. **2) Develop automated mapping tools:**

Mapping is a labor intensive job and human errors can cause careless mapping errors. Automated mapping tools can facilitate the mapping process and reduce human errors. For example, local mappers should use the Regenstrief LOINC Mapping Assistant (RELMA) for semi-automated mappings. **3) Use extensional definitions to validate LOINC mapping:** Some parts of EDs explicitly identify the LOINC axis, e.g. a test with UOM, "mcg/24 h" indicates the property, time aspect, and scale type of the LOINC code should be "MRat","24H" and "Qn" respectively. By implementing validation rules in mapping tools, we can detect an invalid mapping when choosing inappropriate codes according to EDs of tests. Furthermore, as more EDs are collected from different places, a reference ED for each LOINC code, e.g. reference mean and standard deviation, UOM, etc. could be distributed with the LOINC database to help build validation rules. **4) Develop enhanced quality control:** Since current LOINC mappings still contain errors, manually evaluating the correctness of LOINC mappings to detect any possible systematic errors and use this information to benchmark the performance of mappings and **5) Provide better training:** Provide better training to teach users to avoid making common errors when mapping. Many errors are caused by a lack of knowledge of LOINC code usage. Evaluating voluntary LOINC mappings across institutions can identify common errors users make, e.g. failure to choose the correct granularity of Antibody IgG or inappropriately choose "Screen" or "Confirm" for the method.

Limitations

We only utilized local names and EDs to evaluate the correctness of the LOINC mappings in this study. To conduct a more thorough evaluation would require complete information, such as "subtype of analyte", "method", "timing" and "system" for all tests. To accomplish this would require cooperation with laboratory technicians and administrative staff who would create local test definitions that describe the complete meaning of all test codes.

Conclusion

By using EDs of laboratory tests, we evaluated the correctness of voluntary LOINC mappings of three large institutions and identified several common errors that occur in current mappings. Understanding those errors can help in the development of better LOINC codes, automated tools, evaluation methods, and training courses to reduce systematic errors in LOINC mapping.

References

1. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leavelle D, Leland D, 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 Chem1996 Jan;42(1):81-90.
2. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med1998 Nov;37(4-5):394-403.
3. Bodenreider O, Mitchell JA, McCray AT. Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics. Proc AMIA Symp2002:61-5.
4. 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 Inform1998 Jul;51(1):29-37.
5. McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clin Chem2003 Apr;49(4):624-33.
6. Huff S, Rocha R, McDonald C, De Moor G, Fiers T, Bidgood Jr W, et al. Development of the logical observation identifier names and codes (LOINC) vocabulary. J Am Med Inf Assoc1998;5(3):276.
7. Lau LM, Johnson K, Monson K, Lam SH, Huff SM. A method for the automated mapping of laboratory results to LOINC. Proc AMIA Symp2000:472-6.
8. Sowa J. Knowledge representation: logical, philosophical, and computational foundations: MIT Press; 2000.
9. Zollo KA, Huff SM. Automated mapping of observation codes using extensional definitions. J Am Med Inform Assoc2000 Nov-Dec;7(6):586-92.
10. McDonald CJ, Overhage JM, Barnes M, Schadow G, Blevins L, Dexter PR, et al. 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 Sep-Oct;24(5):1214-20.
11. Funk ME, Reid CA. Indexing consistency in MEDLINE. Bull Med Libr Assoc1983 Apr;71(2):176-83.

CHAPTER 5

AUDITING CONSISTENCY AND COMPETENCY OF LOINC® AMONG THREE LARGE INSTITUTIONS - USING VERSION SPACES FOR GROUPING LOINC® CODES

Coauthors:

Daniel J. Vreeman, Clement J. McDonald, Stanely M. Huff

Reprinted with permission from Lin MC, Vreeman DJ, McDonald CJ, Huff SM. Auditing consistency and usefulness of LOINC use among three large institutions - using version spaces for grouping LOINC codes. Journal of Biomedical Informatics. 2012;45(4):658-66.

Journal of Biomedical Informatics 45 (2012) 658–666

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Auditing consistency and usefulness of LOINC use among three large institutions – Using version spaces for grouping LOINC codes

M.C. Lin [a], D.J. Vreeman [b,c], Clement J. McDonald [e], S.M. Huff [a,d,*]

[a] The Department of Biomedical Informatics, The University of Utah, Salt Lake City, UT, United States
[b] Regenstrief Institute, Inc., Indianapolis, IN, United States
[c] Indiana University School of Medicine, Indianapolis, IN, United States
[d] Intermountain Healthcare, Salt Lake City, UT, United States
[e] Lister Hill Center, National Library of Medicine, Washington, D.C., United States

ARTICLE INFO

Article history:
Received 3 August 2011
Accepted 18 January 2012
Available online 28 January 2012

Keywords:
Controlled vocabulary
LOINC
Evaluation research
Clinical laboratory information systems
LOINC usage
Consistency
Usefulness
Semantic interoperability
LOINC usage
Data exchange standards

ABSTRACT

Objectives: We wanted to develop a method for evaluating the consistency and usefulness of LOINC code use across different institutions, and to evaluate the degree of interoperability that can be attained when using LOINC codes for laboratory data exchange. Our specific goals were to: (1) Determine if any contradictory knowledge exists in LOINC. (2) Determine how many LOINC codes were used in a truly interoperable fashion between systems. (3) Provide suggestions for improving the semantic interoperability of LOINC.

Methods: We collected Extensional Definitions (EDs) of LOINC usage from three institutions. The version space approach was used to divide LOINC codes into small sets, which made auditing of LOINC use across the institutions feasible. We then compared pairings of LOINC codes from the three institutions for consistency and usefulness.

Results: The number of LOINC codes evaluated were 1917, 1267 and 1693 as obtained from ARUP, Intermountain and Regenstrief respectively. There were 2022, 2030, and 2301 version spaces among ARUP and Intermountain, Intermountain and Regenstrief and ARUP and Regenstrief respectively. Using the EDs as the gold standard, there were 104, 109 and 112 pairs containing contradictory knowledge and there were 1165, 765 and 1121 semantically interoperable pairs. The interoperable pairs were classified into three levels: (1) Level I – No loss of meaning, complete information was exchanged by identical codes. (2) Level II – No loss of meaning, but processing of data was needed to make the data completely comparable. (3) Level III – Some loss of meaning. For example, tests with a specific 'method' could be rolled-up with tests that were 'methodless'.

Conclusions: There are variations in the way LOINC is used for data exchange that result in some data not being truly interoperable across different enterprises. To improve its semantic interoperability, we need to detect and correct any contradictory knowledge within LOINC and add computable relationships that can be used for making reliable inferences about the data. The LOINC committee should also provide detailed guidance on best practices for mapping from local codes to LOINC codes and for using LOINC codes in data exchange.

1. Introduction

Consistency and usefulness are two important characteristics of good terminological systems (TSs), especially for information exchange. *As we use the terms in this article, consistency* means that between any two terms within TSs there is no contradictory knowledge (as represented by the implicit or explicit relationships between concepts), and *usefulness* means that there is knowledge in the terminology that allows for creation of an efficient algorithm for making inferences using the relationships in the terminology for supporting different kinds of use, e.g. information retrieval, data integration or clinical decision support [4]. Auditing TSs can be a difficult task because of the huge number of concepts, e.g. LOINC has more than 65,000 codes. In order to reduce this task to a manageable size, researchers have used semantic methods to search for similar concepts in the UMLS [9] or used semantic structures to partition SNOMED into smaller groups [28]. Previous reports have shown that most inconsistencies in LOINC mapping result from choosing codes that vary in the 'method', 'scale' and 'property' characteristics of the codes. [3,20,26]. The use of version spaces is a common technique used in machine learning for concept discovery [24]. Version spaces are used to divide all hypotheses into smaller

* Corresponding author. Address: South Office Building, 5171 South Cottonwood Street, Murrary, UT 84107, United States. Fax: +1 801 507 9400.
E-mail address: stan.huff@imail.org (S.M. Huff).

1532-0464/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2012.01.008

subspaces to make it possible to search similar concepts by a given set of constraints. This paper describes a systematic method for auditing the consistency and usefulness of LOINC use and discusses potential strategies to approach best practices in the use of LOINC for interoperable data exchange.

1.1. Auditing TSs on policy vs. use

Early papers on TSs development were focused on functional, structural and policy perspectives. These papers include Cimino's desiderata for creating controlled medical vocabularies [10], Chute et al.'s study of functional characteristics of comprehensive health terminology systems in the United States [8], and the technical specification published by International Standard Organization (ISO) – "Health informatics – Controlled health terminology – Structure and high-level indicators" [16]. As TS usage increased, discussions shifted to descriptions of practical use. These studies included analyzing coverage of the UMLS for coding of concepts in the Gene Ontology (GO) [6], comparing coding consistency of SNOMED CT among three commercial coding companies [2] and evaluating the performance of LOINC when comparing laboratory data among three hospitals [3]. To summarize all auditing methods for TSs, Zhu et al. have done a thorough literature review on different auditing methods, including manual, systematic and heuristic methods [29].

1.2. The development of LOINC

1.2.1. Rapid evolution of LOINC model

LOINC provides a universal terminology for reporting laboratory tests and other clinical observations. Since 1994, LOINC has grown from about 6000 codes to more than 65,000 in the current version. As Cimino noted in his desiderata [10], an important characteristic of TSs is to "Evolve Gracefully", and LOINC tries to adhere to this principle [15]. The LOINC committee has emphasized practical experience in using LOINC to improve its design. Whenever the original design of LOINC is not sufficient, the design is enhanced or a new model is created. Before migrating to the current six-axis model, at least four different earlier models were created (Table 1). For example, the first design of LOINC was a four-axis model, but with more implementation, the original model was insufficient for specifying some tests, e.g. it lacked the ability to specify timing (24 h, 12 h, or 4 h, etc.). Therefore, "⟨timing⟩" was added to create a new model.

1.2.2. LOINC in action

Many places adopted LOINC in their daily operations, including large commercial laboratories, hospitals, health care provider networks, insurance companies, and public health departments [21]. Recently, LOINC was adopted as the terminology standard for certification of laboratory orders and results, including electronic reporting of lab results to public health agencies as part of the Centers for Medicare and Medicaid Services (CMS) Electronic Health Record (EHR) "Meaningful Use" incentive program. LOINC was also used in a German Hospital Information System (HIS) to identify the document type of reports sent as Clinical Document Architecture (CDA) documents [14] and to retrieve laboratory data of adverse events automatically from clinical trial databases [7]. LOINC is also frequently used in computerized clinical decision support systems [1]. Although the scope of LOINC covers both clinical and laboratory observations, for the purpose of this paper we focus exclusively on laboratory content.

1.2.3. Evaluations of LOINC

Evaluating LOINC performance in actual practice can help to improve LOINC design. McDonald et al. summarized LOINC development and worldwide use [21]. Lau et al. reported LOINC coverage for the laboratory test dictionary in the US Department of Defense (DoD) [18] and Vreeman et al. reported LOINC coverage for tests in the Indiana Network for Patient Care (INPC) [27]. We also conducted a series of studies about LOINC usage among three large institutions. First, we reported that LOINC codes can cover more than 99% of the volume of every day laboratory tests among two institutions and 79% of tests in a reference laboratory [19]. Second, we evaluated the correctness of LOINC mapping and reported that there were 0.45% (4/884) tests mapped to totally unrelated LOINC codes and 4% (36/884) tests containing at least one error in mapping to the 6 axis model of LOINC [20]. An earlier study by Baorto et al. also evaluated LOINC performance when combining laboratory results associated with congestive heart failure patients among three teaching hospitals [3].

1.2.4. Requirements for ideal LOINC use

According to Devanbu et al.'s defintion of a good knowledge system [13], the best practice of LOINC should have the following characteristics: (1) *Completeness*: it should have all the necessary LOINC codes to cover the domain of interest, (2) *Correctness*: mapped LOINC codes should be faithful to the original meaning of the tests, (3) *Consistency*: the knowledge implied by different LOINC codes should be consistent, e.g. if two different codes have identical meanings, the codes are duplicates and the consistency principle is violated, and (4) *Competence*: usefulness is the fundamental goal of LOINC for supporting use of laboratory data in different fields. Support for the use of ontologic relationships is one of the important competencies of TSs [25]. LOINC should define the relations between codes and combinations of codes that allow users to infer equivalence, if their meanings in data instance representation are interoperable. That is, if the combination of two codes has the same meaning as a single code (a difference in the use of pre- or post-coordination), relationships should exist between the codes that support the assertion of equivalence. Previous evaluations have described LOINC with respect to the first two characteristics, completeness [18,19] and correctness [20]. The focus of this paper is on the evaluation of consistency and usefulness.

1.3. Definition of consistency and usefulness of TSs

1.3.1. Consistency

Consistency in a system implies that the system does not contain contradictory knowledge. Consistency of TSs could be discussed from two perspectives: (1) *Internal consistency:* Inconsistency can

Table 1
Evolution of LOINC model.

	LOINC model	Explanation
1	⟨analyte⟩:⟨specimen⟩:⟨precision⟩:⟨method⟩	Initial model
2	⟨analyte⟩:⟨timing⟩:⟨specimen⟩:⟨precision⟩:⟨method⟩	Adding 'timing' axis
3	⟨analyte⟩.⟨subspecies⟩:⟨property⟩:⟨timing⟩:⟨system⟩:⟨precision⟩:⟨method⟩	Adding chemical subspecies and kind of property
4	⟨analyte⟩.⟨subspecies⟩^⟨chall⟩:⟨property⟩:⟨timing⟩:⟨system⟩:⟨precision⟩:⟨method⟩	Adding 'challenge' information

result if there is a failure to uniformly employ design principles throughout the entire terminology. For example, the hierarchies, semantic rules, mapping polices, etc., can be incomplete or inconsistent. One study investigated inconsistencies in the usage of the 'parent–child' and 'is-a relationship' in the Unified Medical Language System (UMLS) [11]. Another study investigated inconsistent use of semantic or linguistic rules for representing similar terms for SNOMED [5]. They found that two common modifier terms, "acquired" and "congenital", used for the disease "porphyria", are sibling relationships; but for another disease, "acquired keratoderma (D0-22310)" and "congenital keratoderma (D4-40130)" were not sibling relationships but were found in two separate branches in SNOMED [5]. (2) *External consistency*: The terminology allows for inconsistent and non-interoperable representation of instance data. Baorto et al. reported that the same laboratory tests could be mapped to different LOINC codes, e.g. different coding strategies could be used to choose specimen types, serum, plasma or serum/plasma for the same tests across different institutions [3]. A study comparing three versions of SNOMED coding of the same case report forms (CRF) done by three commercial coding companies showed that there was no significant degree of inter-rater agreement in their coding behaviors, because the three coding companies used different coding strategies concerning pre- vs. post-coordination [2]. Sometimes TSs designs were workable (internal consistency), but TSs were not used consistently across institutions (external consistency). Therefore, ensuring that TSs are both internally and externally consistent is crucial for semantic interoperability.

1.3.2. Usefulness

The use of ontologic relationships to support biomedical inferencing, as exemplified in knowledge management, data integration and decision support, is one of the important characteristics of TSs [4]. Often times the same instance data can be represented by different combinations of terms, and an ontology can provide an ability for finding equivalence of those terms. Steindel et al. reported the ability of building the hierarchy of LOINC codes can facilitate public health reporting [26]. To allow consistent reporting of laboratory data to the Centers for Disease Control (CDC) using LOINC, a "Reportable Condition to LOINC mapping table for National Notifiable Disease (NND)" was created for use by all healthcare institutions [22]. However, new LOINC codes for screening tests for NND were continually created and manually adding them into the above mapping table would be labor intensive. To facilitate the update process, a set of rules was developed to allow new LOINC codes to be added to the table according to their similarity in meaning to existing LOINC codes. For example, a methodless measurement was added as a parent for a measurement with a method 'automatically', all different methods for the same analyte would be placed as children under the same analyte, and ordinal (ORD) measurements were considered to be children of quantitative (QN) measurements [26]. Therefore, new LOINC codes could be added "automatically" to the mapping table based on these rules.

1.4. How to audit TSs efficiently

TSs usually consist of many thousands of terms. It is not an easy task to audit TSs, because scanning all pairs of terms and examining their relationships would create huge numbers of combinations. A key strategy in developing an efficient approach for comparing terms is to generate only the most likely pairs, instead of scanning all pairs. In general, it is not very meaningful to compare two things which are not related, e.g. comparing a bacterial screening test to a sodium measurement is less interesting than comparing two different bacterial screening tests. There are two common approaches for searching similar things to generate pairs: (1) *Searching equivalent concepts*: When examining one term, only search for other terms which have similar meaning. Cimino used semantic methods to search for terms having similar meaning to audit the UMLS, e.g. using synonyms and semantic types of concepts. For example, he neglected different *word order* in searching for equivalent concepts, and he found duplicate concepts as shown below [9].

C0000760: ABNORMAL PAP SMEAR
C0240660: PAP SMEAR ABNORMAL

Cornet et al. used a similar approach to detect duplicates in DICE TS and they successfully found four duplicate concepts in a set of 2500 concepts [12].

(2) *Segmentation*: This approach divides TSs into several smaller groups based on their characteristics. It has been applied to evaluating the National Cancer Institute Thesaurus (NCIT) and SNOMED by Perl's group and their colleagues [23,28]. Their method has two main phases for auditing terminologies: the automated preparation phase and (2) the manually guided-discovery phase. The first phase is composed of four steps: (1) divide all concepts into several groups called an area-taxonomy based on concepts having the same roles, (2) construct a compact abstraction network, (3) refine each division into groups of concepts, called P-area's and (4) finally, construct an enhanced abstraction network, called the P-area taxonomy [23]. Basically they segmented NCIT's Biological Process hierarchy, by using a general to specific (divide and conquer) approach using roles (e.g. biological process role, chromosomal location role) to place terms into groups (areas) for constructing the area-taxonomy. The terms within each area could be categorized into partial areas by their semantics as captured by the root role of the partial area. Using this approach, a partial area taxonomy could be created that segmented NCIT contents based on structure and semantics for support of manual auditing, by identifying partial areas with a high likelihood of errors [28].

1.4.1. Concept learning

Concept learning is a common approach in machine learning where searching for similar concepts is based on a given set of attributes [24]. LOINC utilizes six attributes to specify the meaning of laboratory tests, which is very similar to how a specific hypothesis is defined in concept learning. One approach to concept learning is to divide all instances into several *version spaces* based on a given set of constraints [24]. Using this approach, all instances within a given version space will share similar attributes.

1.4.2. Notation

In concept learning, all hypotheses can be represented by a set of attributes. LOINC term names consists of a vector of six constraints (Analyte, Method, Time, Scale, Property, System). For each constraint, the attribute could be entered using the following tokens [24]:

- "?", allow any value
- allow a specific value (e.g., Hematocrit)
- 'Ø', empty set, do not allow any value

Therefore, the most general hypothesis, which includes all LOINC codes, could be represented by the expression

⟨?,?,?,?,?,?⟩

And, the most specific concepts (does not allow any LOINC codes), is represented by

⟨Ø, Ø, Ø, Ø, Ø, Ø⟩

We can specify a group of LOINC codes for measuring Sodium from Blood specimens, no matter what their methods, time, scale, or property by this expression:

⟨Sodium,?,?,?,?, Blood⟩

Using the same methodology, the following expression represents a group having the same analyte.

⟨Analyte,?,?,?,?,?⟩

The 'Analyte' could be substituted by any more specific analyte. If we are counting the number of hypotheses ⟨Analyte,?,?,?,?,?⟩ in a dataset, it equals the count of instances that represent analytes. This expression is also equivalent to the structured query language (SQL) statement 'Group by Analyte'. Similarly, the following expression is used to specify the group having the same analyte and system. The analyte and system could each be substituted by any analyte or system. The number of hypotheses that exist is equal to the number of combinations of analyte and system. It is also equivalent to the SQL statement 'Group by Analyte and System'.

⟨Analyte,?,?,?,?, System⟩

1.5. Proposed framework

We wanted to investigate the consistency and usefulness of LOINC concepts by creating a version space. By examining all the relationships between any two LOINC codes within each version space, we wanted to:

(1) Determine whether there is any contradictory knowledge, e.g. duplicate codes, or code ambiguity.
(2) Detect any combinations of pre and post coordinated tests that are equivalent or where relationships can be created that will enable systems to interoperate.
(3) Provide suggestions for best practices for LOINC users that will improve the semantic interoperability of LOINC.

2. Methods

2.1. Collecting Extensional Definitions (EDs) of LOINC from three large institutions

We first sent out invitations to several major institutions, and three institutions agreed to provide their laboratory data for the experiment. These three institutions were: 1. Associated Regional and University Pathologists, ARUP Laboratories (Salt Lake City, UT) 2. Intermountain Healthcare, Intermountain (Salt Lake City, UT) 3. Regenstrief Institute, Inc. (Indianapolis, IN). With IRB approval, de-identified patient data for general laboratory tests for the year of 2007 for each institution were selected for this study. In our previous studies [19,20], we used all data collected from the five institutions that contribute data to Regenstrief. To avoid selection bias, only the institution with the highest volume of tests was included in this study. We developed a parsing program written in JAVA and Python to process patient data to generate EDs, which include local code, local description, numeric value, units of measure, coded variables and LOINC mappings, from each institution shown as Table 2. Then, we distributed installed parsing programs to each institution and asked collaborators to process the de-identified patient data within the virtual machines. Only processed EDs were sent back to us for analysis.

Table 3 shows how EDs can be used to determine appropriate LOINC naming. For example, there were EDs of two genetic tests retrieved from two different institutions which had identical local names, 'ALPHA-1-ANTITRYPSIN PHENOTYPE'. The local codes were mapped to two different LOINC codes that had different 'method', 'property' and 'scale' attributes. Considering only their local names, there was not enough information to determine whether two LOINC codes with different 'property/scale' were needed. But if we examined the values reported for each test as summarized from their EDs, we can determine that these are similar tests. This comparison of EDs can help us determine whether LOINC naming for specific tests was accurate.

2.2. Creating version spaces for searching similar LOINC concepts (Find-S: Finding a maximally specific hypothesis)

As noted previously, making version spaces is an approach used in concept learning to create subspaces of hypotheses using different constraints. One approach to creating version spaces is Find-S: Finding a maximally specific hypothesis, which creates hypotheses from the most specific to least stringent by relaxing constraints one by one. For example, the most specific hypothesis for specifying a unique LOINC concept is:

⟨Analyte, Method, Time, Scale, Property, System⟩

A less stringent example of a hypothesis can be made by relaxing one constraint (property):

⟨Analyte, Method, Time, Scale,?, System⟩

It has been reported that choosing different 'properties' was the most frequent reason that different coding choices were made in a study about LOINC coding behaviors in congestive heart failure patients [3]. In our previous study about correctness of LOINC mapping, choosing different 'Method', 'Scale' and 'Property' attributes was the most common reason for different coding choices among three large institutions [20]. For example, in the 'Method' axis, some institutions usually use a code that specifies the method (when available), whereas other institutions always choose terms that are "methodless". Another example, is that in the 'Scale/Property' axes, LOINC uses two distinct styles (Prid:Nar vs. Prid:Nom) for

Table 2
Extensional Definitions (EDs) included local description, mean, standard deviation, units of measure, coded variables and frequency.

Extensional definitions	Example	Containing information for review
Local description	"Creatinine, 24 h urine", "Sodium urine"	Local description – mainly provides analyte information. In some cases, it also provides method (e.g. EIA), scale/property (e.g. titer), time (e.g. 24 h) or system information (e.g. urine)
Mean	1.46, 137	Mean – provides scale/property information (e.g. SCnc/Qn). This is mainly for numeric tests
Standard deviation	0.54, 7.02	Standard deviation – provides scale/property information (e.g. SCnc/Qn). This is mainly for numeric tests
Units of measure	g/24 h, mmol/L, mg/dl	Units of measure – provides scale/property information. This is mainly for numeric tests
Coded variables and their frequencies	1:8 (109), Negative (900), Positive (899), M1M1 (75)	Coded variables – provides scale/property (e.g. Titr/Qn or ACnc/Qn). For example, M1M1 is a reported value for the genetics test 'ALPHA-1-ANTITRYPSIN PHENOTYPE' and its frequency was 75
Frequency	50, 184	Frequency – implies whether tests are frequent

Table 3
Example EDs from two different institutions. Two genetic tests are shown along with their local names, LOINC code mapping and reported values. By examining the EDs, we can determine how these two LOINC codes were used in each institution.

Source	Examples of EDs
A	[Local name]: Alpha 1 antitrypsin phenotyping [LOINC]: 6770-2: Alpha 1 antitrypsin phenotyping: Immunofixation: Prid: Pt: Nom: Ser/Plas [Coded Variables and their frequencies]: M1M1 (75), M1M2 (31), M1S (12), MM (8), M1Z (6)
B	[Local name]: Alpha 1 antitrypsin phenotyping [LOINC]:32769-2: Alpha 1 antitrypsin phenotyping: none: Imp: Pt : Nom : Ser/Plas [Coded Variables and their frequencies]: M1M1 (887), M1M2 (278), M1S (91), M1Z (88), MM (86), SEE NOTE (60)

reporting the interpretation of laboratory tests (e.g. CFTR gene mutation analysis). The Narrative (Nar) scale is for free text results (sentences, paragraphs, sections), whereas the Nominal (Nom) scale is used for representing coded values, as when selecting an organism found on culture from a coded list of bacteria. The differences between these types are often subtle and require understanding the reporting system. Steindel et al. also concluded that for some purposes, such as finding any code that could be used to indicate the presence of a particular disease, rolling up LOINC codes and ignoring some LOINC axes (e.g. method, scale, or property) can be beneficial [26]. For example, the following LOINC codes have differences in method, scale or property, but they all can be used to diagnose the infectious disease 'BACILLUS ANTHRACIS':

BACILLUS ANTHRACIS AB	EIA	PT	ORD	ACNC	SER
BACILLUS ANTHRACIS AB	CF	PT	ORD	ACNC	SER
BACILLUS ANTHRACIS AB	ID	PT	QN	TITR	SER

Using the version space constraint notation described above, all three of these concepts can be represented by the following expression:

⟨BACILLUS ANTHRACIS AB,?, PT,?,?, SER⟩

Based on the most common kinds of mapping errors, we decided to choose ⟨Analyte,?, Time,?,?, System⟩ as the optimal design for auditing LOINC.

2.3. Constructing the version space for the ⟨Analyte,?, Time,?,?, System⟩ expression

After receiving the data from each institution, all EDs were loaded into the database. We then grouped all LOINC codes into smaller subspaces by creating version spaces matching the expression ⟨Analyte,?, Time,?,?, System⟩ shown as Table 4. Within each version space, all LOINC codes shared the identical three axes "Analyte", "System" and "Time", while having different values for 'Scale', 'Property' and 'Method'. LOINC flags codes that can be used in post-coordinated expressions by using 'XXX' as a value in certain axes, especially "System" and 'Time'. The 'XXX' signifies that the information content of that axis is communicated elsewhere in an instance of data, for example in a different field in an HL7 message. In creating our version spaces, we interpreted 'XXX' to be any value. For example, we added the code ⟨Leukocytes, Automated count, Pt, NCnc, Qn, XXX⟩ into the version space ⟨Leukocytes,?, Pt,?,?, Bld⟩ and ⟨Leukocytes,?, Pt, ?,?, CSF⟩. We also processed those LOINC codes having 'XXX' in the 'Time' axis using the same approach. In the real world, the meaning of 'XXX' in a term is not literally anything; its meaning is constrained by the nature of the 'Analyte'. Since we used existing concepts from LOINC with 'XXX' the broad interpretation of XXX is not a significant problem.

2.4. Semi-automated review

We developed a semi-automated review process, which could be used to systematically discover the relationships of terms and determine the characteristic of the relationships, such as consistency and usefulness. This process consisted of three phases: (1) Discovery phase: A python program was written to discover all patterns of similar pairs in each version space for manual review, (2) Discussion phase: All similar patterns were manually reviewed and discussed by experts, and (3) Analysis phase: We analyzed all patterns to define formal descriptions (a taxonomy) for them, e.g. (Method vs. Methodless) or (Pre vs. Post-coordinated).

2.4.1. Discovery phase

A python program was developed to scan all LOINC pairs from each version space. For example, in the Table 4, version space ⟨Hemoglobin,?, Pt,,?,?, Bld⟩, between ARUP and Intermountain, there were four pairs, ⟨5,1⟩ ⟨5,7⟩ ⟨6,1⟩ ⟨6,7⟩ chosen for review.

One important principle was that we focused on general issues for analytes having similar patterns. For example, the following patterns, ⟨Imp:Nar⟩ and ⟨Imp:Nom⟩ were considered contradictory designs (ambiguous), because LOINC codes with the same analyte but using these two different patterns were found to report similar things when their full EDs were considered. All analytes having these designs were found to be contradictory.

⟨13514-5: Hemoglobin pattern: Electrophoresis: Pt: **Imp**: **Nar**: Bld⟩
⟨12710-0: Hemoglobin pattern: Electrophoresis: Pt: **Imp**: **Nom**: Bld⟩

or

⟨49291-8: Prophyrins: None: Pt: **Imp**: **Nar**: Urine⟩
⟨44014-9: Prophyrins: None: Pt: **Imp**: **Nom**: Urine⟩

Thus, the patterns, ⟨Imp:Nar⟩ and ⟨Imp:Nom⟩, were flagged as 'contradictory', and automatic checking for these patterns became part of the logic in the review program. Similarly when two patterns (Method:Scale:Property) are semantically interoperable, all LOINC codes having these two patterns are considered semantically interoperable. In the following two examples, one code has a specified method and the other has not. Their meanings are not exactly the same, but they can support a degree of semantic interoperability.

⟨14336-2: Ethanol: **GC**: Pt: MCnc: Qn: Ser/Plas⟩
⟨5643-2: Ethanol: **Null**: Pt: MCnc: Qn: Ser/Plas⟩

or

⟨20405-7:Urobilinogen: **Test strip**: Pt: MCnc: Qn: Urine⟩
⟨3107-0: Urobilinogen: **Null**: Pt: MCnc: Qn: Urine⟩

Table 4
Example of version spaces. Multiple local tests can be mapped to a single LOINC code, because even though the local tests are different, the difference is not significant according to LOINC naming rules.

Index	Source/local name	LOINC	Analyte	Method	Time	Scale	Property	System
Version space: ⟨Hemoglobin,?, Pt,?,?, Bld⟩								
1	Intermountain/ Hemoglobin, Blood Quantitative Calculated	20509-6	Hemoglobin	Calculated	Pt	Qn	MCnc	Bld
2	Regenstrief/ Hemoglobin Bld QN	20509-6	Hemoglobin	Calculated	Pt	Qn	MCnc	Bld
3	Regenstrief/ Hemoglobin, POC	20509-6	Hemoglobin	Calculated	Pt	Qn	MCnc	Bld
4	Regenstrief/Hgb	718-7	Hemoglobin	None	Pt	Qn	MCnc	Bld
5	ARUP/Hemoglobin, inst.	718-7	Hemoglobin	None	Pt	Qn	MCnc	Bld
6	ARUP/Hemoglobin	718-7	Hemoglobin	None	Pt	Qn	MCnc	Bld
7	Intermountain/ Hemoglobin, Blood Quantitative	718-7	Hemoglobin	None	Pt	Qn	MCnc	Bld
Version space: ⟨Alpha-1-fetoprotein,?, Pt, Ser/Plas,?, Ser/Plas,⟩								
1	ARUP/PATIENT∼S AFP	1834-1	Alpha-1-fetoprotein	None	Pt	Qn	MCnc	Ser/Plas
2	ARUP/AFP (TUMOR MARKER)	1834-1	Alpha-1-fetoprotein	None	Pt	Qn	MCnc	Ser/Plas
3	ARUP/MEDIAN	1834-1	Alpha-1-fetoprotein	None	Pt	Qn	MCnc	Ser/Plas
4	Intermountain/Alpha-1-Fetoprotein, Serum Quantitative∼AFP TUMOR MARKER ONLY	1834-1	Alpha-1-fetoprotein	None	Pt	Qn	MCnc	Ser/Plas
5	Intermountain/Alpha-1-Fetoprotein, Serum Quantitative	1834-1	Alpha-1-fetoprotein	None	Pt	Qn	MCnc	Ser/Plas
6	Regenstrief/AFP	1834-1	Alpha-1-fetoprotein	None	Pt	Qn	MCnc	Ser/Plas
7	Regenstrief/AFP SerPl QN	1834-1	Alpha-1-fetoprotein	None	Pt	Qn	MCnc	Ser/Plas
Version space: ⟨Acyl carnitine,?, Pt,?,?, Ser/Plas⟩								
1	ARUP/CARNITINE, EASTERIFIED	14282-8	Acyl carnitine	None	Pt	Qn	SCnc	Ser/Plas
2	Intermountain/ Carnitine Esters, Plasma Quantitative	1717-8	Acyl carnitine	None	Pt	Qn	MCnc	Ser/Plas

The program discovered all similar pairs and generated a text report for manual review.

2.4.2. Discussion phase

After the discovery phase, there were two stages of discussion. In the first stage, small groups of experts reviewed the overall findings. Any confusing findings involving LOINC policy were distributed in the LOINC committee for additional discussion. The discussion phase can increase awareness of consistency and usefulness issues in the general public.

2.4.3. Analysis phase

After the discussion phase, we found some patterns that shared similar characteristics e.g. (Method vs. Methodless) or (Pre vs. Post-coordinated). Therefore, we developed formal descriptions (a taxonomy) for each group. Using those formal descriptions we can communicate consistency and usefulness of LOINC more efficiently.

3. Results

Table 5 shows the number of laboratory tests collected from three institutions and the number of version spaces for each institution. Table 6 shows the distributions of the size of pairwise version spaces between institutions. It also shows that many LOINC codes were used by all institutions and most version spaces contain less than 10 items. After reviewing consistency within each version space, reasons causing contradictory knowledge were classified into three categories (Table 7): (1) Deprecated codes: LOINC codes

Table 5
The number of laboratory tests and the number of version spaces for ⟨Analyte,?, Time,?,?, System⟩ in each institution.

Source	# of laboratory tests	# of version spaces
ARUP	1917	1601
Intermountain	1267	1089
Regenstrief	1693	1440

Table 6
Pairwise comparison of numbers of tests in each version space between institutions. A–I: ARUP and Intermountain, I–R: Intermountain and Regenstrief, A–R:ARUP and Regenstrief.

# of tests	A–I	I–R	A–R
1	1240	1444	1435
2	589	414	631
3	117	96	134
4	41	42	55
5	20	12	23
6	3	12	9
7	4	5	4
8	5	2	4
9	0	0	2
10	0	1	2
11	0	1	1
13	1	0	1
15	1	0	0
28	0	1	0
37	1	0	0
Total	2022	2030	2301

Table 7
Numbers of pairs having contradictory knowledge and their classifications.

	Deprecated codes	Raw measurement vs. interpretation	Ambiguous codes	Total
ARUP and Intermountain	3	84	17	104
Intermountain and Regenstrief	0	108	1	109
ARUP and Regenstrief	2	106	4	112

Table 9
Numbers of pairs having semantically interoperable knowledge, which were classified into three categories. II.a is 'MCnc vs. SCnc' II.b is 'Pre vs. Post-coordinate' and II.c is 'log'.

	I	II (II.a + II.b + II.c)	III	Total
ARUP and Intermountain	956	92(4 + 84 + 4)	117	1165
Intermountain and Regenstrief	559	17(3 + 14 + 0)	189	765
ARUP and Regenstrief	862	73(6 + 65 + 2)	186	1121

were already deprecated in the LOINC distribution file but were still actively being used in laboratory systems. (2) Raw measurement vs. interpretation: Both raw measurements (30 ng/ml) and their interpretations (e.g. Negative) are usually reported in laboratory systems. The LOINC committee created terms for both raw measurements and interpretations but most institutions only choose one code for mapping both kinds of results. (3) Ambiguous codes: There are LOINC codes for reporting similar things, but the institutions have chosen different styles of pre and post coordination.

Table 8 shows examples for each category. The 'Raw measurement vs. interpretation' category is the most frequent reason for conflicting code use. Table 9 reveals results after the review for usefulness. All semantically interoperable pairs could be classified into three levels: (1) Level I: There is no meaning loss. Information is fully exchanged by identical codes, (2) Level II – there is no meaning loss, and information can be fully exchanged, but it requires conversion of units of measure or other kinds of additional processing. Level II interoperability has been further classified into another three sub-categories: II.a is 'MCnc vs. SCnc', II.b is 'Pre vs. Post-coordinated' and II.c is 'value vs. log value', and (3) Level III – there is some meaning loss and only partial information could be exchanged. We have listed examples and further explanation for each category in Table 10. From Table 9, we also learn that there are 956, 559 and 862 laboratory tests where results are exchanged using identical LOINC codes between ARUP and Intermountain, Intermountain and Regenstrief and ARUP and Regenstrief respectively.

4. Discussion

Using LOINC, or any standard coding system to support interoperable data exchange is not an easy task. We evaluated LOINC specifically on consistency and usefulness of use perspectives which revealed several important findings.

4.1. Creating LOINC codes should be consistent

LOINC is still being actively developed. Creation of new codes and updates could cause inconsistency. Ideally when new codes are created, the new code should follow previous design patterns. Currently, there is no way to automatically check consistency except by using expert guidance from the LOINC committee. For long-term maintenance, there should be a standardized method to detect any contradictory knowledge within LOINC. When inconsistencies in LOINC codes are discovered, it leads to the complicated task of correcting the error. This leads to several suggestions about best practices: First, the LOINC committee should decide which codes should be deprecated or when a new code should be created for replacing existing contradictory codes. Second, after changes have been made to the LOINC codes, each institution should update any deprecated codes. Although the status for each LOINC code is clearly specified in the LOINC distribution and the RELMA program includes functions for finding local codes mapped to deprecated LOINCs, we still observed deprecated codes in use in operational databases. This finding suggests that LOINC users are not always diligent in updating their systems with new LOINC releases.

4.2. Minimalism as a better approach

Complicated designs are hard for users to follow. The LOINC committee designed two styles (raw measurements vs. interpretations) of codes for storing raw measurements and interpretations as distinct observations, which is a workable design. We have observed that sometimes when a new kind of test comes onto the market it is initially reported with interpretive codes, but over time the prevalent reporting pattern changes to sending the quantitative results. Therefore, users who aggregate longitudinal data or interface with many partners need extra efforts to accommodate both LOINC styles. Usually laboratories just choose one style for reporting data. Sometimes even within the same institution, we observed different styles due to different mappers. After reporting these issues to the LOINC committee, the LOINC committee suggested that these two different styles be condensed into one style that reports both pieces of information in a single data instance. For reporting the interpretation of raw measurements, the LOINC committee adopted the Value-Cutoffs-Interpretation style recommended by HL7. This style allows sending the value, cutoffs and interpretation in the same result segment of an HL7 message. This style indicates that results such as "POS" and "NEG" should go in

Table 8
Examples of pairs having contradictory knowledge.

Source	LOINC	Analyte	Method	Time	Scale	Property	System
ARUP	7650-5	Ambrosia elatior Ab.IgE	None	Pt	Qn	ACnc	Ser
Regenstrief	6085-5	Ambrosia elatior Ab.IgE	None	Pt	Qn	ACnc	Ser
Explanation: The LOINC code 7650-5 was already deprecated							
Intermoutnain	44014-9	Porphyrins	None	Pt	Nom	Imp	Urine
Regenstrief	2818-3	Porphyrins	None	Pt	Ord	ACnc	Urine
Explanation: Both codes were used to report 'Interpretation', such as 'Negative, Positive or Normal'							
ARUP	21654-9	CFTR gene mutation analysis	Molgen	Pt	Nom	Prid	Bld/Tiss
Intermountain	38404-0	CFTR gene mutation analysis	Molgen	Pt	Nar	Prid	Bld/Tiss
Explanation: Although 'Nom' and 'Nar' were designed to store different types of information, in practice they were used to store similar information							
Regenstrief	49291-8	Porphyrins	None	Pt	Nar	Imp	Urine
Intermountain	44014-9	Porphyrins	None	Pt	Nom	Imp	Urine
Explanation: Although 'Nom' and 'Nar' were designed to store different types of information, in practice they were used to store similar information							

Table 10
Degree of semantic interoperability between two LOINC codes.

Source	LOINC	Analyte	Method	Time	Scale	Property	System
Intermountain	6085-5	Ambrosia elatior Ab.IgE	None	Pt	Qn	ACnc	Ser
Regenstrief	6085-5	Ambrosia elatior Ab.IgE	None	Pt	Qn	ACnc	Ser
Level I: For all relationships in level I, there is no meaning loss. Information is fully exchanged by identical codes							
ARUP	14282-8	Acyl carnitine	None	Pt	Qn	SCnc	Ser/Plas
Intermountain	1717-8	Acyl carnitine	None	Pt	Qn	MCnc	Ser/Plas
Level II.a: For all relationships in the level II, there is no meaning loss. Information could be fully exchanged, but it requires additional processing. In the above case, 'SCnc' measures numbers of molecular of 'Acyl carnitine' and 'MCnc' measures the weight of 'Acyl carnitine'. Therefore, the value for 'MCnc' equals to 'SCnc' multiplied by the molecular weight of 'Acyl carnitine'							
Intermountain	17096-9	Lymphocytes.kappa/100 lymphocytes	None	Pt	Qn	NFr	Bld
ARUP	20617-7	Lymphocytes.kappa/100 lymphocytes	None	Pt	Qn	NFr	XXX
Level II.b: ARUP uses the code in a post-coordinated way. When the value of 'Bld' is included in a second part of the message to further specify the value of 'XXX', the meanings of these two codes are equivalent							
Intermountain	11011-4	Hepatitis C virus RNA	Probe.amp.tar	Pt	Qn	ACnc	Ser/Plas
ARUP	38180-6	Hepatitis C virus RNA	Probe.amp.tar	Pt	Qn	LaCnc	Ser/Plas
Level II.c: 'LaCnc' is a logarithmic scale. So the value of log (ACnc) is equal to 'LaCnc'.							
Intermountain	20570-8	Hematocrit	None	Pt	Qn	VFr	Bld
ARUP	4545-0	Hematocrit	Spun	Pt	Qn	VFr	Bld
Level III: For all relationships in level III, there is some meaning loss. Partial information could be exchanged. Since one institution is using a methodless term (which could represent a value from a spun hematocrit, impedance, or hematology analyzer) it is not possible to know from the LOINC name whether these results are exactly equivalent. Although these two codes are not identical, but they do share identical 5 other axes. The two codes can be used interoperably in any context where the method is not considered important							

the OBX-8 field for normalcy status in an HL7 message. Thus, LOINC codes with a scale of Qn can be appropriately used in cases where the "values" coming back are coded interpretations of the true numeric result value.

Another common issue we observed was differences in mapping to method-specific or methodless LOINC codes. In current laboratory systems, information about the 'method' is not always specified. For most practical purposes, users only care about the 'method' for billing purposes and only care about meaning of tests for analytic purposes. Using 'methodless' LOINC codes with the Value-Method-Interpretation style is one possible way to solve this problem.

The minimalist style of only keeping necessary information in the LOINC codes, can avoid knowledge overloading and overlap for LOINC users, and also avoids the situation where users choose different coding styles.

4.3. Measuring semantic interoperability of LOINC between two institutions

To determine whether LOINC has achieved its goal of improving semantic interoperability of laboratory tests is not an easy task. Any inconsistency in code definitions or use reduces semantic interoperability. If we can use consistent approaches then as more laboratory tests are exchanged we can achieve better semantic interoperability. By classifying semantic interoperability into three well-defined levels, we can better measure how we are doing in using data to its greatest potential. Overall, Level I statistics can be used to measure the semantic interoperability that can be achieved between two institutions by the use of current LOINC codes. It should also be possible to automate data conversions so that Level II interoperability is achieved in working systems. Using the definitions for the different levels of interoperability enables the measurement of improvements in data representation as we improve LOINC code consistency and as systems implement better mapping strategies.

4.4. Using version space approach and EDs for auditing TSs

We demonstrated that using the version space approach can reduce the complexity of auditing TSs. It consists of three important steps: (1) Prepare a set of attributes for concept learning. The crucial step in concept learning is finding a set of attributes that can be used to partition the concepts. Perl and his colleagues used the "divide and conquer method" for constructing the abstraction network (area-taxonomy and P-area taxonomy) according to the roles of concepts, e.g. [has associated location], which is similar to utilizing attributes for representing concepts. Formal concept analysis (FCA) is another similar approach to concept learning, which decomposes compound medical concepts into several atomic concepts by constructing a concept-attribute table, e.g. tonsillitis could be represented by tonsils and inflammation process [17]. Both Perl's approach and FCA could be used to create hierarchy of TSs. All three approaches, concept learning, Perl's taxonomy analysis and FCA, provide methods for measuring the similarities between two concepts, which can be used for grouping or dividing concepts. (2) Grouping terms by creating version spaces: There are general to specific and specific to general approaches for creating version spaces. For those TSs that already had a set of attributes (e.g. description logics), the above two approaches could be used. Our specific to general approach had one advantage in that it created smaller groups at the beginning compared to the "divide and conquer method", which requires multiple steps to generate smaller groups. One limitation of this study was that we only audited the consistency and usefulness within single analyte version spaces. One possible extension of this study would be to use a semantic method for searching subspecies of analytes for creating version spaces, e.g. ⟨Genetic tests,?, Time,?,?, System⟩ or ⟨Antibody,?, Time,?,?, System⟩ [9]. Therefore, consistency and usefulness across a wider range of terms could be evaluated. (3) Using EDs for auditing: Our results revealed that using EDs alone for auditing may not be practical in the real world. Using EDs and version spaces together can provide a better understanding of the quality of TS implementations.

5. Conclusions

Min et al. concluded that auditing should be part of the terminology design life cycle, and LOINC is no exception [23]. There are variations in the way LOINC is used for data exchange that result in some data not being truly interoperable across different enterprises. To improve its semantic interoperability, we need to

detect and correct any contradictory knowledge within LOINC using audit techniques and add computable relationships that can be used for making reliable inferences about LOINC encoded data. The LOINC committee should also provide detailed guidance on best practices for mapping from local codes to LOINC codes and for using LOINC codes in data exchange.

Acknowledgments

The authors would like to thank Brian Jackson, M.D. and Alan Terry, M.S. from ARUP for retrieving and processing the ARUP data set.

References

[1] Ahmadian L, van Engen-Verheul M, Bakhshi-Raiez F, Peek N, Cornet R, de Keizer NF. The role of standardized data and terminological systems in computerized clinical decision support systems: literature review and survey. Int J Med Inform 2011;80:81–93.
[2] Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. J Am Med Inform Assoc 2007;14:497–506.
[3] 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;51:29–37.
[4] Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. Yearb Med Inform 2008:67–79.
[5] Bodenreider O, Burgun A, Rindflesch TC. Assessing the consistency of a biomedical terminology through lexical knowledge. Int J Med Inform 2002;67:85–95.
[6] Bodenreider O, Mitchell JA, McCray AT. Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics. Proc AMIA Symp 2002:61–5.
[7] Brandt CA, Lu CC, Nadkarni PM. Automating identification of adverse events related to abnormal lab results using standard vocabularies. In: AMIA annual symposium proceedings/AMIA symposium. AMIA symposium; 2005. p. 903.
[8] Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. ANSI healthcare informatics standards board vocabulary working group and the computer-based patient records institute working group on codes and structures. J Am Med Inform Assoc: JAMIA 1998;5:503–10.
[9] Cimino JJ. Auditing the Unified Medical Language System with semantic methods. J Am Med Inform Assoc: JAMIA 1998;5:41–51.
[10] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inform Med 1998;37:394–403.
[11] Cimino JJ, Min H, Perl Y. Consistency across the hierarchies of the UMLS semantic network and metathesaurus. J Biomed Inform 2003;36:450–61.
[12] Cornet R, Abu-Hanna A. Auditing description-logic-based medical terminological systems by detecting equivalent concept definitions. Int J Med Inform 2008;77:336–45.
[13] Devanbu PT, Jones MA. The use of description logics in KBSE systems: experience report. IEEE Computer Society Press; 1994. p. 23–35.
[14] Dugas M, Thun S, Frankewitsch T, Heitmann KU. LOINC codes for hospital information systems documents: a case study. J Am Med Inform Assoc 2009;16:400–3.
[15] Huff SM, Rocha RA, McDonald CJ, De Moor GJ, Fiers T, Bidgood Jr WD, et al. Development of the logical observation identifier names and codes (LOINC) vocabulary. J Am Med Inform Assoc 1998;5:276–92.
[16] ISO/TC215. Health informatics – controlled health terminology – structure and high-level indicators. Report No. 17117; 2002.
[17] Jiang G, Ogasawara K, Endoh A, Sakurai T. Context-based ontology building support in clinical domains using formal concept analysis. Int J Med Inform 2003;71:71–81.
[18] Lau LM, Banning PD, Monson K, Knight E, Wilson PS, Shakib SC. Mapping department of defense laboratory results to logical observation identifiers names and codes (LOINC). AMIA Annu Symp Proc 2005:430–4.
[19] Lin MC, Vreeman DJ, McDonald CJ, Huff SM. A characterization of local LOINC mapping for laboratory tests in three large institutions. Methods Inf Med 2011;50:105–14.
[20] Lin MC, Vreeman DJ, McDonald CJ, Huff SM. Correctness of voluntary LOINC mapping for laboratory tests in three large institutions. AMIA annual symposium proceedings/AMIA symposium. AMIA symposium 2010; 2010. p. 447–51.
[21] McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clin Chem 2003;49:624–33.
[22] McDonald CJ, Overhage JM, Dexter P, Takesue BY, Dwyer DM. A framework for capturing clinical data sets from computerized sources. Ann Intern Med 1997;127:675–82.
[23] Min H, Perl Y, Chen Y, Halper M, Geller J, Wang Y. Auditing as part of the terminology design life cycle. J Am Med Inform Assoc: JAMIA 2006;13:676–90.
[24] Mitchell TM, Michell T. Machine learning. McGraw-Hill Series in Computer Science, WCB McGraw-Hill; 1997. ISBN: 0070428077.
[25] Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. Briefings Bioinform 2008;9:75–90.
[26] Steindel S, Loonsk J, Sim A, Doyle T, Chapman R, Groseclose S. Introduction of a hierarchy to LOINC to facilitate public health reporting. Am Med Inform Assoc 2002:737.
[27] Vreeman DJ, Stark M, Tomashefski GL, Phillips DR, Dexter PR. Embracing change in a health information exchange. AMIA Annu Symp Proc 2008:768–72.
[28] Wang Y, Halper M, Min H, Perl Y, Chen Y, Spackman KA. Structural methodologies for auditing SNOMED. J Biomed Inform 2007;40:561–81.
[29] Zhu X, Fan JW, Baorto DM, Weng C, Cimino JJ. A review of auditing methods applied to the content of controlled biomedical terminologies. J Biomed Inform 2009;42:413–25.

CHAPTER 6

INVESTIGATING THE SEMANTIC INTEROPERABILITY OF LABORATORY DATA EXCHANGED USING LOINC® CODES IN THREE LARGE INSTITUTIONS

Coauthors:

Daniel J. Vreeman, Stanely M. Huff

Reprinted with permission from Lin MC, Vreeman DJ, Huff SM. Investigating the semantic interoperability of laboratory data exchanged using LOINC codes in three large institutions. American Medical Informatics Association Annual Symposium proceedings. 2011;2011:805-14.

Investigating the Semantic Interoperability of Laboratory Data Exchanged Using LOINC Codes in Three Large Institutions

Ming –Chin Lin, MD[1], Daniel J. Vreeman, PT, DPT, MSc[2,3], Stanley M. Huff, MD[1,4]
[1]The University of Utah, Salt Lake City, UT
[2]Regenstrief Institute, Inc., Indianapolis, IN
[3]Indiana University School of Medicine, Indianapolis, IN
[4]Intermountain Healthcare, Salt Lake City, UT

Abstract

LOINC codes are seeing increased use in many organizations. In this study, we examined the barriers to semantic interoperability that still exist in electronic data exchange of laboratory results even when LOINC codes are being used as the observation identifiers. We analyzed semantic interoperability of laboratory data exchanged using LOINC codes in three large institutions. To simplify the analytic process, we divided the laboratory data into quantitative and non-quantitative tests. The analysis revealed many inconsistencies even when LOINC codes are used to exchange laboratory data. For quantitative tests, the most frequent problems were inconsistencies in the use of units of measure: variations in the strings used to represent units (unrecognized synonyms), use of units that result in different magnitudes of the numeric quantity, and missing units of measure. For non-quantitative tests, the most frequent problems were acronyms/synonyms, different classes of elements in enumerated lists, and the use of free text. Our findings highlight the limitations of interoperability in current laboratory reporting.

Introduction

Logical Observation Identifiers Names and Codes (LOINC®) was developed in 1994 to provide a universal vocabulary for reporting laboratory and clinical observations[1-3]. We and others have previously evaluated LOINC usage by analyzing coverage, consistency and the correctness of LOINC usage among 3 large institutions[4-6]. One goal of LOINC is to facilitate the aggregation of laboratory data collected from different institutions to support research other kinds of secondary use of clinical data[6]. Our working definition of interoperability is that data from different institutions are interoperable if they are mutually substitutable, that is, if the data from one institution can be used in the patient care and decision support programs of the second institution without the need to translate or convert the data. For example, "body weight 80 kg" and "body weight 176.4 lb" are not semantically interoperable because the units of measure would need to be converted before these two representations would be mutually substitutable. In our previous study[5], when observing the same laboratory tests (having the same LOINC codes) among different institutions, we discovered that laboratory data contained many heterogeneous data formats. For example, laboratory data for "Patient Body Weight (LOINC: 29463-7)" among three institutions could be "80 kg", "160 lb" or "158 lbs". Another example is for reporting "Aspergillus fumigatus 2 Ab (LOINC: 29334-0)", the report value could be "negative", "NEG", "positive" or "POS" among three institutions. This implies that even with proper LOINC use, aggregating data across different institutions will require the conversion of different units of measure and creating synonym tables. Therefore, we investigated the common problems that occur when aggregating data that has been represented using LOINC codes. We collected laboratory data reported using LOINC codes from three large institutions for our investigations. Our goal was to understand:

1) How many heterogeneous data formats are represented when combining tests based on LOINC codes?
2) What kinds of efforts are needed for converting the heterogeneous data formats so that data can be aggregated?

Background

1) LOINC in Action

Currently, LOINC is used for reporting laboratory data in many health organizations, including major laboratories (e.g. ARUP, Quest and LabCorp), large healthcare providers (e.g. Indiana Network for Patient Care and Intermountain Healthcare) and insurance companies (e.g. United Healthcare)[3]. LOINC is also used in many clinical applications, including reporting public health data[7], retrieving laboratory data for clinical studies[6], providing standardized terminology in supporting sharable clinical decision support logic[8] and reporting adverse events[9].

2) Previous Evaluations of LOINC

Terminological Systems (TSs) are not perfect, but can be improved through cycles of scrutiny to detect errors and omissions and then correction. The early stages of developing TSs focused on building them based on functional, structural, and policy perspectives[10, 11]. As TSs are now in use in many applications, investigators have begun to evaluate TSs based on practical issues[12-14]. In some ways LOINC development has followed a similar path. At first we mainly focused on discussing LOINC design philosophy[1, 2]. As LOINC has become more widely used, we are now more interested in evaluating its performance in real applications[4-6, 15]. We have evaluated LOINC usage among three large institutions based on examining LOINC coverage[4] and correctness of LOINC mapping[5].

3) The Design of LOINC

The Entity-attribute-value (EAV) triplet is a common design used in representing clinical data[16]. For example, a body weight measurement would be represented conceptually as "Observation (entity) has Test name = Body weight (attribute); value =84.7 kg (value)". LOINC codes are designed to be used as observation identifiers in Health Level Seven (HL7) messages. Here are two sets of examples of the actual syntax of HL7 Version 2.X OBX (observation/result) segments:

Example 1:

OBX|1|NM|29463-7^Body Weight^LN|1|84.7|kg |||"

OBX|1|NM|29463-7^Body Weight^LN|1|156|lb |||"

OBX|1|NM|29463-7^Body Weight^LN|1|166|lbs |||"

Example 2:

OBX|1|NM| 29334-0^ Aspergillus fumigatus 2 Ab^LN|1|Pos| |||"

OBX|1|NM| 29334-0^ Aspergillus fumigatus 2 Ab^LN|1|Positive| |||"

In example 1, the LOINC code "29463-7" represents the test, "Body Weight", and the "84.7 kg", "156 lb" and "166 lbs" are the values of "Quantitative" measurements. In the example 2, the LOINC code "29334-0" represents the test, "Aspergillus fumigatus 2 Ab", and "Pos" and "Positive" are "Ordinal" measurement values. LOINC codes are defined using a six-axis model: "component", "property", "timing", "system", "scale", and "method". The scale portion of the LOINC name specifies whether the measurement is "quantitative", "ordinal", "nominal", "narrative" etc. (Table1)[1, 2, 17].

Scale Type	Abbreviation	Descriptions
Quantitative	Qn	This is for reporting continuous numeric scale. Valid values are "-7.4", "0.125","<10","1-10","1:256"
Ordinal	Ord	Ordered categorical responses, e.g, "1+","positive", "negative", "reactive", "indeterminate"
Nominal	Nom	Nominal or categorical response that do not have a natural ordering, e.g. "name of bacteria", "yellow", "clear", "bloody"

Narrative	Nar	Text narrative, e.g. description of microscopic part of a surgical papule test.
Quantitative or ordinal	OrdQn	Test can be reported as either "Qrd" or "Qn" . LOINC committee discourages the use of this type.

Table 1. Examples and descriptions for each type of measurement in "Scale" axis.

4) Current work to standardize laboratory data

With the development of EHRs, transferring clinical data among physicians, laboratories, healthcare organizations, and clinical researchers has become a complicated task where users often have to deal with many different data formats. To address this issue, the California HealthCare Foundation (CHCF) started the EHR-lab Interoperability and Connectivity Specification (ELINCS) in 2005[18]. ELINCS was developed to provide for standardized formatting and coding of electronic messages used to exchange data between clinical laboratories and EHR systems. In 2007, HL7 approved the ELINCS standard for transmitting laboratory data, and it has since been evaluated in pilot implementations [19]. Also as an attempt to reduce the variability of data, the International Standards Organization (ISO) developed the ISO/IEC 11179 standard as a framework for consistent data representation[20]. To conform to ISO/IEC 11179 standard, a data element should contain a data element concept (DEC) and one value domain. The value domain consists of a set of permissible values, which is an expression of a value meaning allowed in a specific value domain[20]. The ISO/IEC 11179 standard has been widely adopted for developing common data elements (CDEs) in cancer research[21].

5) Extensional definitions (EDs) to characterize laboratory data

A systematic method can be used to characterize laboratory data by grouping tests having the same local codes together and describing their usage in the system. A test can be characterized by how frequently it is done, its mean value, the standard deviation of the value, its associated unit of measure, and the frequencies of coded values that it has. These test profiles, which are called EDs, reflect the meaning of tests in the system (Table 2). The approach of generating EDs has been applied to automatically map local laboratory tests from 3 institutions[22] and to verify the correctness of the LOINC mappings[5].

Extensional attribute	**Example**	**Value of information for meaning and mapping**
Local description	"Creatinine, 24 hr urine", "Sodium urine"	Provides a human readable meaning for the test
Mean	1.46,137	Mean - the average value for quantitative tests. Provides information for reviewing quantitative tests.
Standard deviation	0.54, 7.02	Standard deviation - a measure of the physiologic consistency of the values of Quantitative tests
Units of measure	g/24 h, mmol/L, mg/dl	Units of measure - provides scale information. Sometimes it also provides time information (e.g. g/24h implies 24 hour)
Coded variables	Yellow (45), Negative (1345), Rare (697), 1+(143)	After grouping the same tests in each institution by their local codes, we calculated the frequency of each coded variable.
Frequency	50, 184	Frequency - implies whether tests are frequent (e.g. biochemistry tests) or rare (e.g. allergen test)

Table 2. The example of extensional definitions (EDs).

Methods

Data sources and scope

This study is based on two LOINC evaluation studies[4, 5]. After sending out invitations, three institutions agreed to join in this research. They were: 1. Associated Regional and University Pathologists, ARUP Laboratories (Salt Lake City, UT) 2. Intermountain Healthcare, Intermountain (Salt Lake City, UT) 3. Regenstrief Institute, Inc. (Indianapolis, IN). ARUP Laboratory is a national clinical and anatomical pathology reference laboratory, which is owned and operated by the Pathology Department of the University Utah. Intermountain Healthcare is a not-for-profit health care provider organization, which

consists of many hospitals located in major cities in Utah. Regenstrief Institute, Inc., is an informatics and healthcare research organization, which operates a regional health information exchange in central Indiana called the Indiana Network for Patient Care (INPC)[23] that includes data from more than a hundred source facilities and thirteen health systems. Regenstrief Institute is located on the campus of the Indiana University School of Medicine in Indianapolis. In our previous study, the Regenstrief dataset was retrieved from five institutions, which share similar resources on their laboratory systems. To avoid selection bias, we only used data collected from the largest institution. With IRB approval, de-identified patient data for the year of 2007 as reported by general laboratory systems for each institution were selected for this study.

Generate extensional definitions (EDs)

The raw patient data were stored in the source institutions with various formats, e.g. HL7 messages or flat files. First, we customized the individual interface for each institution to transform raw data into standardized comma separated value (CSV) files. In CSV files, we loaded the following data elements:

1) **Local code:** The internal test identifier. We can use these codes to group the same tests together for generating EDs.
2) **Local description**: The local test name, which suggests the meaning of test.
3) **Numeric value:** The numeric result of the test, which was used for calculating mean value and standard deviation of value.
4) **Units of measure:** We kept raw presentations of units of measure, without any normalization process of text string.
5) **Coded variables:** The results of non-quantitative tests, e.g. (positive, pos, 1+, rare, yellow, light pink)
6) **LOINC mappings:** The LOINC code for this test (mapped by the source institution), which we use for grouping the same tests among different institutions.

Then, we developed a parsing program written in JAVA and Python to process CSV files to generate EDs (Table 2). To avoid transferring the patient data out from its original institution, we distributed pre-installed parsing programs to each institution and asked collaborators to process the de-identified patient data within the virtual machines. Only processed statistical results were sent back to us for analysis.

Analysis

There are two ways to count numbers of tests. One is to calculate the number of unique concept and another is to calculate the total frequency of each concept. For example, there are 10 "Body weight" tests in the system. If we count the unique concept of "Body weight", it is one. But if we count the volume of "Body weight" tests, it is ten. In the laboratory system, the distributions of tests are highly skewed[24] and counting concepts by their volume provides the most information about the true usage of the test in the system.

After receiving the EDs for all tests at the three institutions, we:

1) Calculated the numbers of LOINC codes for each different "Scale" by counting their unique concept and their total volume
2) Calculated numbers of UOM by counting their unique concept and their total volume
3) Calculated the number of variables by counting their unique concept and their total volume

Manual Review

In order to characterize the differences in presentation of the laboratory among three institutions, we conducted a more detailed manual review on a subset of the data. We grouped tests from the three source institutions by LOINC codes, and then selected a one-tenth sample. To aid in the analysis we divided tests into classes of quantitative and non-quantitative. In on our last study[5], we sampled one-tenth of the LOINC mappings for manual review for correctness of the LOINC mappings. The LOINC mapping errors identified in that prior review were excluded from the current study. We also excluded LOINC concepts that only existed in a single institution. We then reviewed the result values and characterized the differences in presentation by identifying common patterns. A given test could be

assigned to more than one taxonomy, because it could contain more than two heterogeneous presentations.

Results

After receiving the data from all three institutions, there were 4,876 local laboratory tests, which were mapped to 3,078 unique LOINC codes. Among these 4,876 tests, the frequency and percentage of each scale type was determined as shown in Table 3. The most frequent categories were "Qn" and "Ord". The tests in "Narrative" category were relative small, and consisted of unstructured information and were not considered further in the analysis. The distribution of LOINC 'Class' was shown in Table 4.

	ARUP			*Intermountain*			*Regenstrief*		
Scale	A	B	C	A	B	C	A	B	C
Qn	1473	77%	35%	859	68%	81%	1194	71%	95%
Ord	323	17%	46%	233	18%	6%	397	23%	3%
Nominal	96	5%	17%	102	8%	1%	80	5%	2%
OrdQn	21	1%	2%	66	5%	2%	12	1%	<1%
Narrative	4	<1%	1%	7	1%	<1%	9	1%	<1%
Total	1917			1267			1692		

Table 3. The frequency and percentage of each scale. **A**: Number of unique LOINC codes, **B**: Percentage of each LOINC code by counting unique LOINC codes, **C**: Percentage of each LOINC code by counting their total volume

ARUP		*Intermountain*		*Regenstrief*	
LOINC class	Percentage	LOINC class	Percentage	LOINC class	Percentage
CHEM	42.87%	HEM/BC	40.04%	CHEM	61.58%
MICRO	15.00%	CHEM	36.65%	HEM/BC	29.78%
HEM/BC	13.41%	MICRO	5.11%	UA	4.50%
SERO	5.34%	UA	4.97%	MICRO	1.55%
DRUG/TOX	4.64%	ABXBACT	2.41%	SPEC	0.87%
ALLERGY	3.77%	SPEC	2.36%	DRUG/TOX	0.50%
SPEC	3.74%	CLIN	2.18%	BLDBK	0.31%
COAG	2.46%	COAG	1.47%	COAG	0.26%
UA	1.54%	BLDBK	1.35%	SERO	0.17%
OB.US	1.23%	DRUG/TOX	1.02%	CELLMARK	0.17%

Table 4. The distributions of LOINC 'Class'. The 'CHEM', 'HEM/BC','MICRO' are the most frequently used LOINC classes.

Characterization of the different presentations of laboratory data

The frequency distribution of UOM was highly skewed. In Intermountain and Regenstrief, about 10 UOM account more than 85% of all test volume. The examples of the top 10 UOM in each institution are listed (Table 5). The total unique UOM used in ARUP, Intermountain and Regenstrief were 103, 80 and 105 respectively.

ARUP		*Intermountain*		*Regenstrief*	
UOM	Percentage	UOM	Percentage	UOM	Percentage
mg/dL	12%	mg/dL	18%	mg/dL	25%
%	10%	%	16%	mmol/L	23%

ng/mL	7%	mmol/L	16%	%	11%
g/dL	5%	10^3/uL	9%	k/cumm	9%
mmol/L	5%	g/dL	6%	GM/dL	8%
IV	5%	U/L	5%	Units/L	6%
K/uL	5%	K/uL	5%	million/cumm	5%
kU/L	4%	fL	5%	fL	2%
pg/mL	3%	/100 WBC@s	3%	pg	2%
U/L	3%	10^6/uL	3%	mL/min/1.73m2	2%
Total number of unique units of measure					
103		**80**		**105**	

Table 5. Example of the top 15 UOM in three institutions. The percentage is the sum of the total test volume.

We calculated the frequency of each coded variable by summing their whole volume. The frequency distribution of non-quantitative tests was also highly skewed (Figure 1). Among three non- quantitative groups (Ord, Nominal and QnOrd), the "Nominal" category contains the most varied formats and frequent "See note" "See Description" and "HIDE" among three institutions. The examples of frequently used coded variables are shown in Table 6.

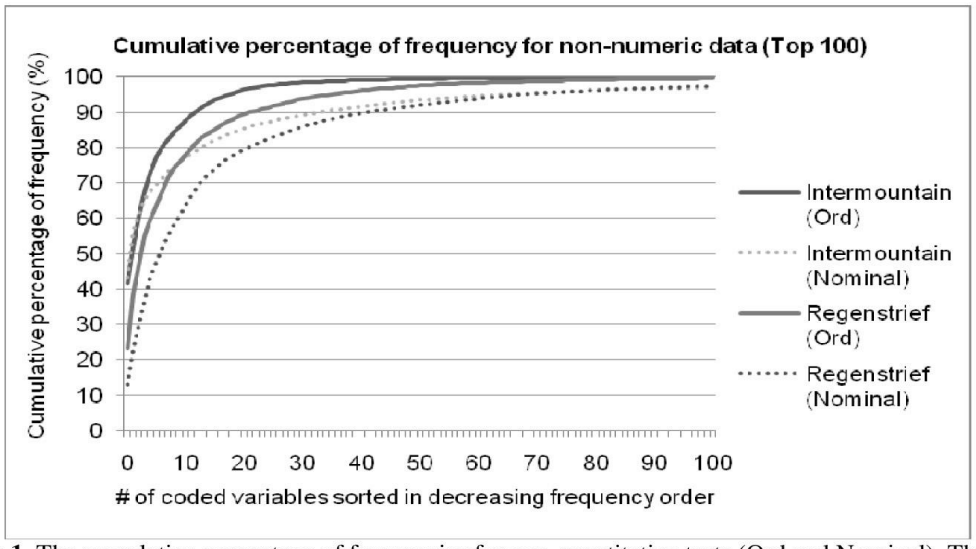

Figure 1. The cumulative percentage of frequencies for non-quantitative tests (Ord and Nominal). The percentage was summed for the total test volume.

Ord			Nominal			OrdQn		
A	**I**	**R**	**A**	**I**	**R**	**A**	**I**	**R**
NEG	NEG	Negative^Negative	SEE NOTE	HIDE	CLEAR^Clear	NONE	SS	Syn-S
NEGATIVE	HIDE,*,	Normal^Normal	NO	NORMAL	YELLOW^Yellow	SEE NOTE	R	Syn-R
NOT APPL	PLTOK	NEG^Negative	NEGATIVE	(null)	310215^Escherichia coli	HIGH	I	space
DETECTED	P1,*,	Few (1+)^Few (1+)	WHITE	CCS	310784^Staphylococcus aureus	POSS	NONE	>=8

POSITIVE	(null)	neg^Negative	CLEAR	SOK	See description^See description	NEGATIVE	DEL	Susceptible
NONE DET	NR	CX7NEG^Negative	US	XT	LTYELL^Light Yellow	<0.1	<=0.25	<=8
SEE NOTE	P2,*,	Neg^Negative	LNMP	COMPAT	SLCLDY^Slightly Cloudy		NDO	Resistant
NON REAC	NDE	NR^Negative	NOT APPL	DEL	760835^Pseudomonas aeruginosa		<=8	>=256
1+	OBSER,*,	RARE^Rare	HISPANIC	LRCA	Normal^Normal		<=0.06	Ceftazidime
NORMAL	TRACE,*,	Positive^Positive	IFE DONE	;BL	No fungal elements seen.		HIDE	

Table 6. Example of coded variables used in ACnc, Nominal and OrdQn category, which were summed to their total volume. They are sorted in decreasing frequency. **A**: ARUP, **I**: Intermoutain, **R**:Regenstrief. "P1" means "one plus" and is a synonym for "1+",

II. Comparing the different presentations of laboratory data

A one-tenth sample of the 3,078 unique LOINC codes contained 479 laboratory tests and 308 unique LOINC codes (Table 7). After removing the mapping errors, there were 445 tests. Only tests appearing in more than one institution were selected for review, which left 229 tests containing 92 LOINC concepts. The most frequent reasons for differences in quantitative test presentation were "missing UOM" and the "Synonymous units" (Table 8). The most frequent reasons for differences in presentation among non-quantitative tests were "Acronym/Synonym" and "Different enumeration" (Table 9).

	Number of tests	Number of LOINC
Total number of tests	4876	3078
After one-tenth sampling	479	308
After removing error mapping	445	293
Appearing in more than one institution (quantitative + non-quantitative)	229 (186+43)	92 (75+17)

Table 7. The number of tests and LOINC codes in the data sets.

Category Type	**Count**	**Examples**	
		Institution A	**Institution B**
Synonym	16	mg/24hr	mg/d
		10^3/uL	K/uL
		mCg/mL	ug/ml
Different magnitude	3	3.98 mol/L	100 umol/L
		23.5 mg/dL	43.2 ug/ml
		84.7 kg	166 lb
Exact match	38	mg/dl	mg/dl
Missing UOM	19		
Number of LOINC	**75**		

Table 8. Comparison of the different presentations of UOM as extracted from three institutions.

Category Type	**Count**	**Examples**	
		Institution A	**Institution B**
Acronym/Synonym	8	NEG	Negative
		SLCLDY	Slight Cloudy
		CLDY	Cloudy
Different enumeration lists	12	Light Pink, Pale Pink, Slightly Yellow	Pink, Yellow

		BLDY, CLEAR, TURB, SLCLDY, CLDY	Turbid, Clear, Hazy, Cloudy
		1+,2+,3+	Rare, Moderate, Many
Free text	3	1+ (few) Acid fast bacilli in concentrated smear	Rare Acid fast bacilli seen .br Results called to and read back by Dr xxx
Perfect match	2	Positive, Negative	Positive, Negative
# of LOINC	**17**		

Table 9. Comparison of the different presentations of non-quantitative test results. The "Acronym/Synonym" and "Different enumeration lists" were the most frequent reasons for inconsistent presentations. These examples were extracted from three institutions.

Discussion

Having a universal observation identifier (e.g. a LOINC code) to aggregate laboratory test data is necessary but not sufficient for full semantic interoperability. Current LOINC codes could cover 99% of the volume of laboratory tests in daily operation in Intermountain and Refenstrief[4]. Yet, this analysis shows that laboratory data have significant variation in their delivered units of measure for quantitative tests and for the answers of non-quantitative tests. Further work (e.g. to define guidelines for units of measure and coded values for laboratory data) for solving those problems is needed. Currently, users combining existing laboratory data might encounter the following issues.

Heterogeneous formats of quantitative tests

1) Missing UOM: Quantitative tests without a UOM is not meaningful;[25] to know the accurate meaning of quantitative tests, users need to know the UOM. The quantitative tests reported in the HL7 OBX segment should always specify UOM. Although not available for this analysis, one problematic messaging pattern we have observed in our experience (and a potential reason why the UOM was missing from our data) is sending the UOM in the NTE segment of the HL7 message.

2) Lack of a standardized code for UOM: 21% (16/75) of the quantitative LOINCs from our sample had variations of synonymous UOM, thus highlighting the potential value of adopting a standardized UOM representation. One standardized UOM developed by Regenstrief is "The Unified Code for Units of Measure (UCUM)"[25]. Other standards are ISO 2955, ANSI X3.501, "ISO+" developed by HL7 and ASTM 1238, and the European Standard ENV 12435[6].

3) There is a need for converting UOM: Inevitability we have to face varied formats of UOM in existing systems, which might have different magnitudes, e.g. we need to convert "lb" to "kg". One approach to solve this problem is to create conversion programs for UOM based on the "dimension" of the measurement. The base system of dimensions consist of length, time, mass, charge, temperature, luminous intensity and angle. For example, "mg/dL" could be represented by "L-3M". UOM within the same dimension can be converted algorithmically. The UCUM project has developed an open-source Java implementation for UOM conversion[25].

Heterogeneous formats of non-quantitative tests

1) Lack of standardized terminology: There are substantial number of synonyms and acronyms used in reporting non-quantitative tests. This variation creates a large burden for those attempting to aggregate data, because it forces users to create mapping tables for all of the synonyms and acronyms appearing as result values. We also observed both "neg" and "negative" used for values of the same test within a single institution, but this type of inconsistency is more commonly seen between different tests and institutions.

2) Lack of standardized enumeration lists (value sets) for reporting encoded data: For example, e.g. reporting urine color as "yellow" in one institution, while another institution could have more fine-grain descriptions, e.g. "light yellow", "yellow" and "dark yellow". Lack of standardized value sets will

hinder data integration. The best way to solve this is to specify the standardized value sets for reporting laboratory data, but for existing systems, a possible solution is to use an ontological approach for grouping different granularities of information under common parents.

3) Lack of permissible value: It is common to observe the same laboratory test containing a quantitative measurement (<1:16) in some instances, and a non-quantitative measurement (negative) in other instances. This happens because of a typical laboratory practice. When measuring the existence or quantity of a "Drug", "Bacteria" or "Antibody" in a sample, users first derive the quantitative measurement from the machine, then compare the measured value to a cut-off value to interpret the quantitative test as being positive or negative. It would be best if both the measured value and the interpretation were sent from the laboratory.

4) Lack of standardized models for reporting complicated data: Sometimes value sets are not sophisticated enough for reporting complicated data, such as reporting genetic tests. For example, reporting "ALPHA-1-ANTITRYPSIN PHENOTYPE', the report value could be "M1M1", "M1M2","MM","M3M3","M1Z" or "M1S". The phenotypes of "ALPHA-1-ANTITRYPSIN" have different alleles variants "M", "S", and "Z", and "M" variants could be divided into six M subtypes. The report value, e.g. "M1M2" implies a model of "types of variants (M, Z, or S)" and "subtype of variants (M1,M2,…,M6)". Standardized models for reporting complicated data are needed to help clinical applications consistently represent the results of complex tests.

5) Lack of standardized strategies for sending some result information: Result reporters often find convenient (but less than optimal) ways of dealing with the complexity of sending both numeric and interpretive data by sending the true test result (and perhaps some additional interpretive or "boilerplate" information) as NTE segments in the HL7 message. We saw evidence of this practice by phrases like "see note" or "see description" appearing in the OBX-5 (observation value) field of the message. Storing results in different strategies, e.g. storing information in NTE segments instead of OBX-5 segment, would hinder data integration.

Limitations

This study only collected data from three institutions. These institutions also provided their laboratory test names for creating the initial set of LOINC codes. Because these three institutions have better knowledge and resources for using LOINC than typical institutions, we cannot necessarily extrapolate our results to other institutions. Another limitation was that we did not compare the consistency of the use of UOM and coded values within the individual institutions. This dataset was collected in 2007 and some newer laboratory tests, e.g. genetic tests, were fewer at that time. For those newer tests, a more recent dataset is needed.

Conclusion

Greater interoperability could be achieved if national standards bodies or the LOINC Committee provided more guidance on best practices in coding of laboratory results. Some possible suggests are: 1) For numeric data: When UOM are appropriate for a given test, reporting UOM should be required and, standard UCUM codes should be used. 2) For enumerated lists, standardized terms and codes should be developed and use in reporting should be required. The differences in enumerated lists could be resolved by creating an ontology for combining different enumerated lists. 3) For complicated data (e.g. genetic tests results), standardized models or patterns for results reporting are needed. It is reasonable to predict that genomic tests will be increase in frequency and will become important tests in clinical decision support systems. We need to standardize how to report these kinds of complex data. Healthcare providers, should be aware of the issues related to coding of laboratory results and adopt best practices in their daily operations. LOINC use in their production databases can provide valuable information for improving LOINC design. Finally, users should profile their LOINC usage periodically to monitor the quality of TSs practice.

References

1. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leavelle D, Leland D, 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 Chem1996 Jan;42(1):81-90.

2. Huff SM, Rocha RA, McDonald CJ, De Moor GJ, Fiers T, Bidgood WD, Jr., et al. Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary. J Am Med Inform Assoc1998 May-Jun;5(3):276-92.
3. McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clin Chem2003 Apr;49(4):624-33.
4. Lin MC, Vreeman DJ, McDonald CJ, Huff SM. A Characterization of Local LOINC Mapping for Laboratory Tests in Three Large Institutions. Methods Inf Med Aug 20;49(5).
5. Lin MC, Vreeman DJ, McDonald CJ, Huff SM. Correctness of Voluntary LOINC Mapping for Laboratory Tests in Three Large Institutions. AMIA Annu Symp Proc;2010:447-51.
6. 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 Inform1998 Jul;51(1):29-37.
7. Centers for Disease Control and Prevention. National Electronic Disease Surveillance System (NEDSS). Centers for Disease Control and Prevention; Available from: http://www.cdc.gov/NEDSS/.
8. Greenes RA. Clinical decision support : the road ahead: Boston: Elsevier Academic Press; 2007.
9. Brandt CA, Lu CC, Nadkarni PM. Automating identification of adverse events related to abnormal lab results using standard vocabularies. AMIA Annu Symp Proc2005:903.
10. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med1998 Nov;37(4-5):394-403.
11. ISO/TC215. Health informatics -- Controlled health terminology -- Structure and high-level indicators. 2002;Report NO.:17117.
12. Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. J Am Med Inform Assoc2007 Jul-Aug;14(4):497-506.
13. Bodenreider O, Mitchell JA, McCray AT. Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics. Proc AMIA Symp2002:61-5.
14. Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. Yearb Med Inform2008:67-79.
15. Dugas M, Thun S, Frankewitsch T, Heitmann KU. LOINC codes for hospital information systems documents: a case study. J Am Med Inform Assoc2009 May-Jun;16(3):400-3.
16. Nadkarni PM, Marenco L, Chen R, Skoufos E, Shepherd G, Miller P. Organization of heterogeneous scientific data using the EAV/CR representation. J Am Med Inform Assoc1999 Nov-Dec;6(6):478-93.
17. LOINC Committee. LOINC Manual. Indianapolis, IN 2011; Available from: http://loinc.org/downloads/files/LOINCManual.pdf.
18. The California HealthCare Foundation (CHCF) ELINCS: The Lab Data Standard for Electronic Health Records. Indianapolis, IN 2011; Available from: http://elincs.chcf.org/.
19. Raths D. ELINCS gets adopted. HL 7 is taking over ELINCS to ensure lab data and EHRs continue towards interoperability. Healthc Inform2007 Mar;24(3):20, 4.
20. ISO/IEC 11179, Information Technology -- Metadata registries (MDR). Indianapolis, IN 2011; Available from: http://metadata-standards.org/11179/.
21. Covitz PA, Hartel F, Schaefer C, De Coronado S, Fragoso G, Sahni H, et al. caCORE: a common infrastructure for cancer informatics. Bioinformatics2003 Dec 12;19(18):2404-12.
22. Zollo KA, Huff SM. Automated mapping of observation codes using extensional definitions. J Am Med Inform Assoc2000 Nov-Dec;7(6):586-92.
23. McDonald CJ, Overhage JM, Barnes M, Schadow G, Blevins L, Dexter PR, et al. 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 Sep-Oct;24(5):1214-20.
24. Vreeman DJ, Finnell JT, Overhage JM. A rationale for parsimonious laboratory term mapping by frequency. AMIA Annu Symp Proc2007:771-5.
25. Schadow G, McDonald CJ, Suico JG, Fohring U, Tolxdorff T. Units of measure in clinical information systems. J Am Med Inform Assoc1999 Mar-Apr;6(2):151-62.

CHAPTER 7

DISCUSSION

7.1 Summary

By characterizing LOINC® usage among three large institutions, we evaluated whether LOINC® is meeting its goal of improving interoperability. We used four different perspectives, including completeness, correctness, consistency, and usefulness, to evaluate LOINC® use.

First, *completeness* of LOINC® can examine the extent to which the LOINC® terminology provides sufficient codes to cover the needs of health care institutions. In the two health care provider institutions, LOINC® covered 99% of test volume. The reference laboratory was about 79%, because it contained a greater number of rare tests. Another finding was that a small number of high frequency tests account for a large percentage of routine test volume; across our three institutions, about 200 tests accounted for more than 70% of test volume. The fastest way to reach a high mapping rate is to map high frequency tests first. The LOINC® committee could develop a starter set of frequently used LOINC® codes for newcomers. It was also found that LOINC® codes were created and deprecated relatively frequently. There is a need to maintain mappings for the updated LOINC® codes, and Regenstrief has deployed an Exception Browser (1) to do so. The Exception Browser monitors all of the INPC data streams and all local codes are classified into two categories: mapped or marked to be

mapped. Therefore, if there are any unrecognized or unclassified codes, the Exception Browser captures those codes. If codes need to be mapped, the following steps are followed. First, check whether there is an existing LOINC® code; if not, submit a new LOINC® code to LOINC® committee. If codes do not need to be mapped, those codes should still be marked as "intentionally not mapped" so that they can be differentiated from those codes that have been submitted for creation of new codes. In laboratory systems, the ability of differentiating local codes from mapped, unmapped, and internal use codes is a necessary mechanism for maintenance.

Correctness of LOINC® mapping is a fundamental requirement of using LOINC codes, because mapping local codes to LOINC codes should preserve the meaning of the original local code. We found that 0.45% (4/884) of tests were mapped to totally unrelated LOINC® codes and 4% (36/884) of tests contained at least one error in mapping to the 6 axis model of LOINC®. The totally unrelated mappings were probably caused by human error and one possible solution to the problem is to use RELMA to perform the LOINC® mappings. The LOINC® committee is still actively developing RELMA, adding new LOINC® codes, improving search algorithms, updating synonym tables, etc. Another common reason for incorrect mapping is the complex and subtle differences in items being mapped to LOINC®. For example, in

reporting toxicology tests, there are codes that distinguish the concept of "screen," "confirm," "threshold," and "cutoff" for different purposes, and users have to fully understand the design of LOINC® to avoid inappropriate mappings. In order to understand the LOINC® design, users are encouraged to read the LOINC® manual and receive training in LOINC® mapping. Sometimes the need for complex LOINC® codes can be obviated by making better use of the structure of the message in which the codes are used. For example, the use of "raw data" and "interpretation" code variants can be avoided by using new fields in the HL7 OBX segment that allow both the raw data and interpretation to be recorded in the same data instance. To address this problem, the LOINC® committee deprecated the "interpretation" style and adopted the Value-Cutoffs-Interpretation style recommended by HL7.

LOINC® also has the need to "Evolve gracefully." LOINC® naming should support *consistency*. We discovered several instances of inconsistent LOINC® naming. One important type of inconsistency is the creation of duplicate concepts, e.g., "Nom/Prid" vs. "Nar/Prid." These two styles have different meanings in LOINC®, but in production use, people tend to use them to store similar information. Based on this example, we learned three things. First, we should resolve knowledge contradictions among existing LOINC® codes. The solution is to deprecate one of the competing

styles of name creation. Second, we should check for internal consistency before creating new LOINC® codes. Third, we should develop a systematic approach to checking the consistency of TSs. We created a method that uses the version spaces approach to reduce the consistency auditing task to a manageable level.

The inclusion of more ontological knowledge in LOINC® will be useful for supporting biomedical applications, e.g., information retrieval, data integration, and clinical decision support systems. We categorized pairs of codes used in different institutions for similar data into three levels: 1) Level I – No loss of meaning, complete information was exchanged by the use of identical codes. 2) Level II – No loss of meaning, but processing of data was needed to make the data completely comparable. 3) Level III – Some loss of meaning. For example, tests with a specific "method" could be rolled-up with tests that were "methodless." It would be useful if the LOINC® committee provided the above three ontological relationships in the LOINC® database for supporting different uses.

7.2 Contribution

In this study, we made several contributions on auditing LOINC as followings.

1. Systematic approach for auditing LOINC

We demonstrated the use of four different perspectives (completeness, correctness, consistency, and usefulness) to audit LOINC and discovered some errors in LOINC usage (Figure 7.1). Understanding these errors can help the LOINC committee to improve LOINC design.

2. Improved terminology characterization methods

The following informative methods were developed as part of this study:

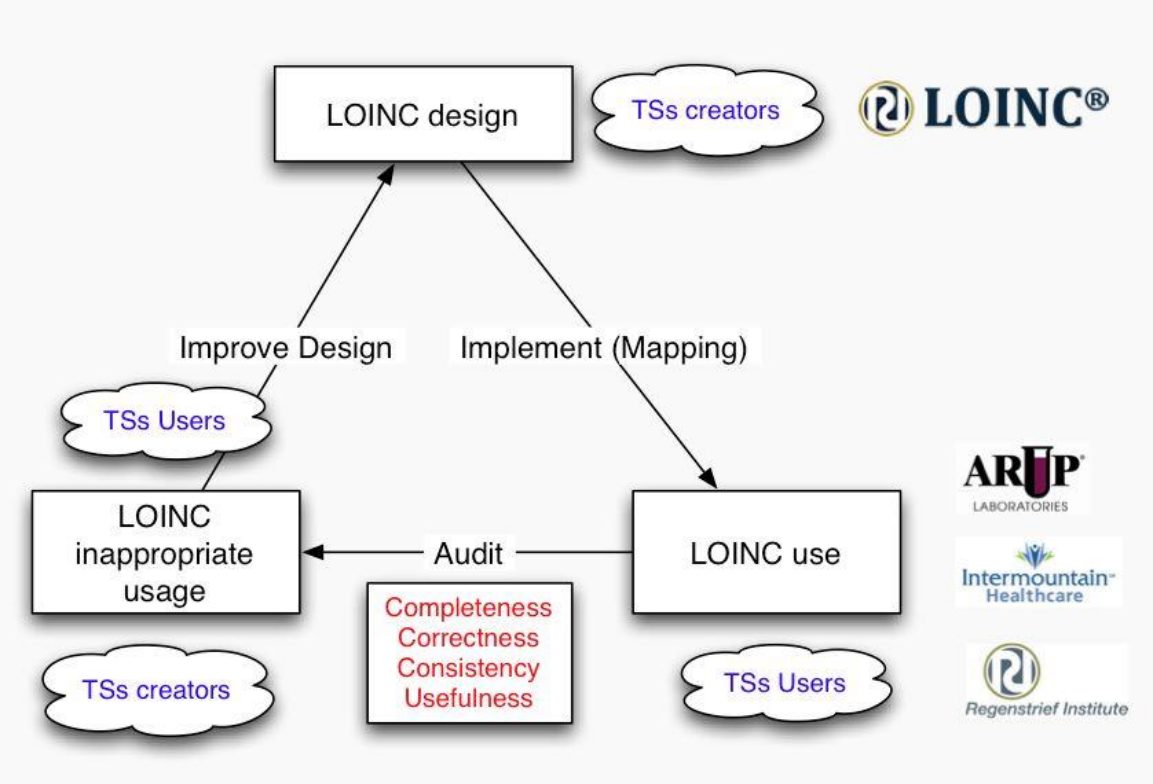

Figure 7.1 The complete design circle of LOINC

2-1. Using virtual machine

Using a Java program that ran in a local virtual machine to generate the extensional definitions for local codes eliminated patient privacy issues for data collection spanning multiple institutions. We sent out pre-installed virtual machine to each institution for processing patient data that reduced patient privacy concerns, since no patient identifiable information needed to leave the local institution.

2-2. Developing EDs for summarizing LOINC usage

The true meaning of codes can only be understood by considering the context of use in actual production systems, but there were only few studies on how to characterize use and meaning of codes in production systems. We transformed LOINC usage to EDs, which allowed us to use informatics methods to characterize LOINC usage, e.g., calculating the similarity between to EDs.

2-3. Using the version space approach for reducing complexity

Auditing LOINC is not an easy task, especially dealing the huge size of LOINC. One way to reduce the complexity is to segment LOINC into several small groups in a meaningful way. We successfully used the version space technique to segment LOINC codes into several smaller groups, which made auditing tasks feasible.

7.3 Future direction

An automated tool could facilitate the auditing process. In our series of studies, it was revealed that auditing LOINC® requires intensive human review and new software developments. Regenstrief's "Exception Browser" is a good example of using an automated tool to reduce human review. EDs of LOINC® can provide information about how LOINC® was used in real systems. Comparing EDs to the LOINC® six axis model could be used to detect errors, e.g., if scale of one LOINC® code in EDs profiles was "titer" and was mapped to "SCnc," the mapping is incorrect. EDs of LOINC® can be used to develop quality control programs.

In the last part of our study, we found that the practical use of LOINC® codes could be enhanced by adding ontological relationships between codes. We identified 3 levels of interoperable links that can be asserted between LOINC® codes, which adds an important piece of ontologic knowledge to the LOINC database.

In auditing consistency, we used one model,<Analyte, System, Time,?,?,?>, to create version spaces, but we did not examine consistency between two different analytes. One possible future direction is to create different models, e.g., <Toxicology, System, Time,?,?,?>, to examine consistency among general classes of analyte

concepts. By changing designs of version spaces, we can evaluate LOINC® more thoroughly.

7.4 References

1. VREEMAN, D. J., STARK, M., TOMASHEFSKI, G. L., PHILLIPS, D. R. & DEXTER, P. R. (2008) Embracing change in a health information exchange, *AMIA Annu Symp Proc*, 768-72.

.