Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (LOINC)

David M. Baorto [a,*], James J. Cimino [b], Curtis A. Parvin [a], Michael G. Kahn [c]

[a] *Department of Pathology, Division of Laboratory Medicine, Washington University School of Medicine, Washington University, Box 8118, 660 S. Euclid Avenue, St. Louis, MO 63110, USA*
[b] *Department of Medical Informatics, Columbia University, New York, NY, USA*
[c] *Rodeer Systems, Incorporated, Boulder, CO, USA*

Abstract

A standard set of names and codes for laboratory test results is critical for any endeavor requiring automated data pooling, including multi-institutional research and cross-facility patient care. This need has led to the development of the logical observation identifier names and codes (LOINC) database and its test-naming convention. This study is an expansion of a pilot study using LOINC to exchange laboratory data between Columbia University Medical Center in New York and Barnes Hospital at Washington University in St. Louis, where we described complexities and ambiguities that arose in the LOINC coding process (D.M. Baorto, J.J. Cimino, C.A. Parvin, M.G. Kahn, Proc. Am. Med. Inf. Assoc. 1997). For the present study, we required the same two medical centers to again extract raw laboratory data from their local information system for a defined patient population, translate tests into LOINC and provide aggregate data which could then be used to compare laboratory utilization. Here we examine a larger number of tests from each site which have been recoded using an updated version of the LOINC database. We conclude that the coding of local tests into LOINC can often be complex, especially the 'Kind of Property' field and apparently trivial differences in choices made by individual institutions can result in nonmatches in electronically pooled data. In the present study, 75% of failures to match the same tests between different institutions using LOINC codes were due to differences in local coding choices. LOINC has the potential to eliminate the need for detailed human inspection during the pooling of laboratory data from diverse sites and perhaps even a built-in capability to adjust matching stringency by selecting subsets of LOINC fields required to match. However, a quality standard coding procedure is required and examples highlighted in this paper may require special attention while mapping to LOINC. © 1998 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Laboratory data sets; Multiple institutions; Logical observation identifier names and codes

---

* Corresponding author. Tel.: + 1 314 4547613; e-mail: baorto@pathbox.wustl.edu

1. Introduction

Advocates of electronic medical record systems often propose the aggregation of patient data from multiple clinical databases to support multi-institutional or national health services research studies [2]. To accomplish this goal however, clinical data emerging from different facilities must be comparable. Central to combining clinical information from disparate and independent sources is the use of a standardized set of terms or codes in which contributing organizations agree to encode their data. Once translated into an agreed-upon standard, data from different sources can be combined and analyzed.

In laboratory medicine, an effort to create a standard set of test names has achieved substantial momentum [3]. The logical observation identifier names and codes (LOINC) database was initially motivated by the need to share laboratory test results among disparate clinical systems. The laboratory LOINC database has grown to over 6500 clinical test names or identifiers (version 1.0h). Recently a set of clinical terms, called the Clinical LOINC database, has been added (version 1.0i). LOINC databases are freely available via the Internet (http://www.mcis.duke.edu. standards/termcode/loinc.htm).

LOINC testnames are ASCII strings constructed by combining six component fields separated by a field delimiter (the colon character) [4]. The six fields of a LOINC testname are as follows: < Analyte > : < Kind of property > : < Time aspect > : < Sample type > : < Precision > : < Method > . A detailed description of these can be found in the LOINC manual [4], but an example is 'LEUKOCYTES: NCNC: BLD: QN: AUTOMATED COUNT', where NCNC is 'number concentration', BLD is 'blood' and QN is 'quantitative'. Each name is assigned a

unique LOINC code, the assigned code has no embedded semantics or interpretation. Institutions seeking to exchange laboratory data using the LOINC vocabulary must provide a mapping between each institutional-specific laboratory code or name in their system's term dictionary to LOINC names or codes. Each institution is responsible for representing their tests accurately in the LOINC vocabulary. Thus, as with any standard coding system, the ability to use LOINC as a method to standardize laboratory test names from disparate institutions rests not only on the specificity and completeness of the LOINC vocabulary, but also on the ability of participating departments or institutions to encode their local tests correctly into LOINC identifiers.

The present experiment required combining laboratory test names and results from two independent academic institutions for the purpose of comparing laboratory test utilization in a specific clinical condition (congestive heart failure). Each institution was required to design and execute queries to extract raw data from their local laboratory information system, to provide their own mapping of local test names into LOINC identifiers and to generate aggregate data for comparative analysis. Electronic correspondence between the two research groups defined the patient characteristics for inclusion in the study population and computations for constructing aggregate comparative data, but no attempt was made to restrict or define the process of mapping local codes into LOINC identifiers. Investigators used the same information provided in the same version of the LOINC user's guide [4]. In this follow-up study coding was repeated using an updated version of the LOINC database, but as before done independently at both sites. We focus here only on the complexities and difficulties encountered in the process of

mapping tests to LOINC for the purpose of sharing data between two independent institutions. An analysis of the laboratory test utilization results is reported elsewhere [5].

2. Methods

2.1. Data sources

We assessed the utility of LOINC for sharing laboratory data between two academic hospitals, Barnes Hospital at the Washington University School of Medicine in St. Louis and Presbyterian Hospital at Columbia University in New York (CPMC) by asking the following question, How does laboratory test utilization differ among these hospitals for all patients admitted between January 1 1995 and December 31 1995 with a primary discharge ICD-9 diagnosis of congestive heart failure (428.0)? During the study period, Washington University School of Medicine was affiliated with two different adult teaching hospitals: Barnes Hospital and Jewish Hospital (During 1996, these hospitals were merged as Barnes–Jewish Hospital). Only congestive heart failure patients from Barnes Hospital and CPMC are compared in the present analysis.

2.2. Database queries

Each site was responsible for querying the local laboratory database. At Barnes Hospital, all patient and laboratory data are stored on an IBM mainframe computer in relational database tables for a period of approximately 2 years. Database queries produced ASCII output files containing: (1) a listing of patient registration numbers, admission and discharge dates, for all patients admitted between January 1 1995 and December 31 1995 who had the primary diagnosis of congestive heart failure (ICD9 code 428.0) upon discharge; and (2) all laboratory procedures reported on those patients including test identifier, battery identifier (how test was ordered), test result and test date and time. The files were uploaded into Stata, a commercial statistical software package. For the present study, tests were listed in decreasing order of test volume for each institution. The most frequent 100 tests at each institution were used in subsequent analysis. Similarly, summaries from the CPMC database were obtained in two steps. In the first step, the financial database (IBM's PM/PA) was queried to identify all admissions with the same admission date range and ICD9-CM primary diagnosis as was used for the Barnes Hospital queries. For each admission, the medical record number, admission date and discharge date were obtained. A second query was then made to the clinical database (DB2) to obtain, for each medical record number, all laboratory data falling between the admission and discharge dates.

2.3. LOINC coding

Each site was responsible for mapping local test names into LOINC names with no prior discussion of coding rules, except that LOINC version 1.0i was used. Each site then provided a listing in LOINC of the most common 100 tests ordered on the patient population described above. Codes were also included for each test where a pre-existing code existed in the LOINC database (version 1.0i). Test names and summary data for each test were provided as follows: < local test identifier > | < LOINC code (if it existed in database) > | < Analyte > : < Kind of property measured > : < Time aspect > : < Sample type > : < Precision > | < # patients having test > | < total # ordered >. The fully specified LOINC name is between the

second and third '|'. No patient-specific information was passed between the two institutions. 'True Matches' were clearly the same laboratory test as determined by one of the authors (D.M.B.) who is a laboratory medicine physician.

3. Results

3.1. Reasons for nonmatches with the preexisting LOINC database

The top 100 tests comprised 96% of the total test volume on CHF patients at Barnes and 94% of the total test volume on CHF patients at CPMC. A significant proportion of the LOINC names for the top 100 tests at the two institutions did not have a pre-existing LOINC code in the LOINC database (version 1.0i). Among the top 100 tests at Barnes Hospital 35 tests did not have a matching LOINC code and at CPMC 17 of the top 100 tests did not have a matching LOINC code. We further examined why so many of the most frequent tests did not match test names in the LOINC database.

Of the 35 testnames at Barnes that did not match to a LOINC code, 11 failed based on 'sample type' primarily because the tests at Barnes were done on 'plasma' and LOINC only had a code for 'serum', not 'plasma/serum' or 'plasma'. Of the remaining 24, seven failed based on 'kind of property measured'. In one case, the existing LOINC code for creatine kinase. MB is for catalytic activity, not mass concentration, which is measured at Barnes. In another example, the result for inhaled oxygen concentration at Barnes is reported as either a 'volume ratio' or a 'number fraction' with the units included in the result, so that test could not be coded to a specific LOINC code. Many nonmatches based on 'kind of property' were due to our

assigning properties to those tests which differed from properties assigned by LOINC for the identical test. For example, LOINC assigns the property 'number fraction' for hematocrit while we assigned 'volume fraction', we assigned the property 'range' to erythrocyte size distribution while LOINC assigns 'length' and we assigned 'mass per entity' (entity being erythrocyte) for mean corpuscular hemoglobin while LOINC assigns 'mass concentration'.

Eleven testnames in the Barnes Hospital top 100 list were not present in the LOINC database, including common tests such as 'Carbon dioxide. calculated', 'Troponin I' and 'Indirect antiglobulin test'. One of the missing tests noted in the pilot study, 'Percent neutrophils', has been added to the current LOINC version. Often, results that provide information pertinent to other test results were found to be missing from the LOINC database, such as 'Interval since last dose', 'Time of last dose', 'Time drawn', 'Collection period' and 'Total cells counted'. The other six nonmatches between Barnes and LOINC were due to existing LOINC codes being either too general or too specific to accurately reflect the Barnes test. Two examples were due to a distinction in coagulation testing methodology that LOINC did not consider. The LOINC database codes one prothrombin time (PT) test called 'coagulation tissue factor induced^^^ Patient: Time: PT: PPP' (see LOINC manual for explanation of fields). However, Barnes Hospital has two distinct PT tests which differ by the ISI number of the thromboplastin, one of the reagents used in the assay. While the fully specified LOINC name for these two tests would be identical and both actually match the testname in the LOINC database, these tests can yield vastly different results and should be distinguished.

In the pilot experiment [1], Columbia had many mismatches with the LOINC database

based on 'sample type'. Upon recoding to the newer LOINC version for the present experiment, Columbia allowed matching for most serum/plasma discrepancies. Nevertheless, there were still 17 tests in the Columbia top 100 list which did not have a corresponding entry in the LOINC database and 13 of these were due to 'sample type' discrepancies. Five of these 13 cases were for blood gas analysis on specimens identified as being 'blood' in the CPMC database. The existing LOINC database entries for blood gas specimen type requires a higher specificity, for example, with 'arterial blood', 'venous blood' or 'capillary blood' being the choices. In three additional cases, CPMC coded the specimen type as 'UR' (urine), where LOINC only has entries for 'URNS' (urine sediment). Other examples where CPMC tests did not match entries in the LOINC database were (1) the test for platelet estimate has a measurement

scale of 'SQ' (semi-quantitative), where the existing LOINC entry is 'QN' (quantitative) and (2) there appeared to be no corresponding LOINC entry for 'RBC (red blood cell) morphology'.

3.2. Test matching between the two hospitals

We examined how the top 100 lists at the two hospitals compared with each other. The number of true matches as well as the number of matches obtained by using different fields of the LOINC testname appear in Fig. 1. The 63 true matches were those tests present in the top 100 lists from both Barnes and CPMC. String matching by the LOINC code field alone correctly matched only 27 of those 63 tests. This was not surprising since many LOINC names at each institution had no associated LOINC code. Also note that three tests that were not true matches were
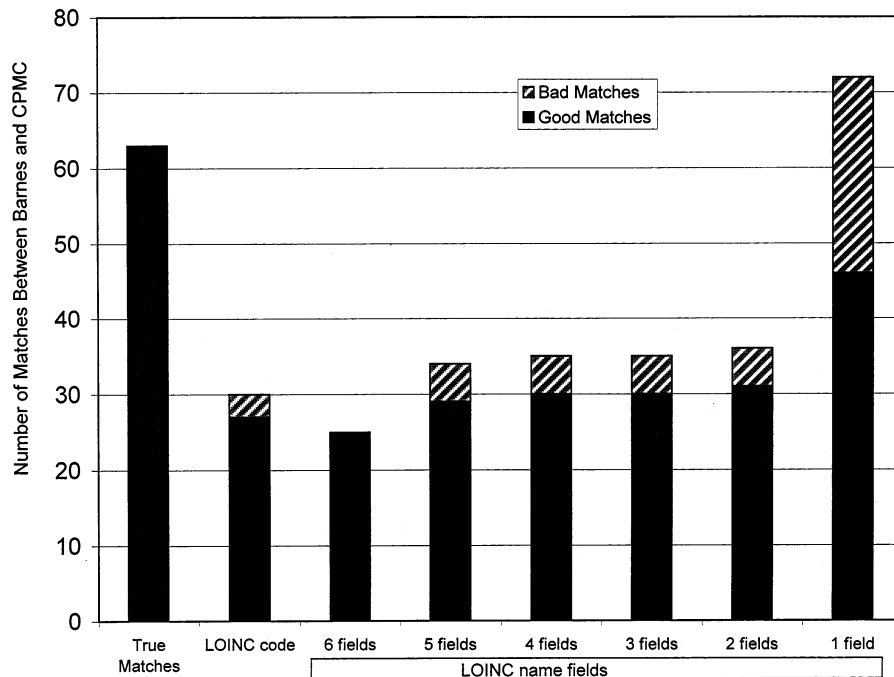

Fig. 1. Test matching between the top 100 lists on CHF patients between Barnes and CPMC.

aligned using the LOINC code. These were three cases where one site compiled their serum and plasma tests separately, yet mapped each to the same LOINC code, while the other site compiled their serum and plasma tests together. String matching using the following six fields of the LOINC test-name matched 25 of the 63 true matches: $<$analyte$>$: $<$sample type$>$: $<$kind of property$>$: $<$time aspect$>$: $<$measurement$>$: $<$precision$>$: $<$method$>$. Note, however, that the matching by the six fields of the fully-specified LOINC name did not align any nonidentical tests. By sequentially removing the last LOINC field from the matching criteria, the matching stringency was reduced in a stepwise manner. Removing $<$method$>$ increased the number of successful matches by four, from 25 to 29 (transition from six to five LOINC fields). However, removing $<$method$>$ from the matching stringency allowed matching of five tests that were nonidentical (primarily by combining leukocyte differential counts done by automated methods with those done by manual methods). Removing $<$measurement$>$ from the matching criteria resulted in one additional true match. Removing $<$time aspect$>$ from the matching stringency added no additional matches, not surprising since the top 100 tests at both institutions included primarily single timepoint measurements, not rates. Removing $<$kind of property$>$ from the matching criteria also had little effect on the number of matches. Removing $<$Sample type$>$ from the matching criteria (transition from two to one LOINC field), had the largest effect on increasing the number of matches between the hospital pairs, increasing the number of true matches successfully attained to 46 of 63. However, clinical relevance is severely impaired, with 26 additional tests being inappropriately matched.

3.3. LOINC matching failures are primarily due to local coding choices

It is notable that the number of true matches is greater than that which could be obtained by LOINC matching even at the lowest stringency. There were 38 cases where string matching based on the six fields of the fully specified LOINC name failed to bring 'true matches' together. We examined the reasons why LOINC failed to match in these cases and divided them into two groups:
1. Failure due to local coding choices.
2. Failure due to another laboratory factor.

Similar to the results of the pilot study, it is notable that the majority of LOINC matching failures are due to local coding choices. We found that 32 of the 38 LOINC matching failures were in this category. Examples are: (a) Analyte: PHOSPHATE.INORGANIC versus PHOSPHATE; (b) Sample type: SERUM/PLASMA versus PLASMA (for analytes where serum and plasma would be the same) or URINE versus URINE SEDIMENT; and (c) Property: NUMBER FRACTION versus NUMBER CONCENTRATION or NUMBER FRACTION versus VOLUME FRACTION. An example of LOINC matching failure due to 'other laboratory factors' is that the laboratory information system at one hospital reports ABO and RH antigen blood type together as one result, while the other reports them as two separate test results.

4. Discussion

We found that, even for common tests, there can be major differences in how individual hospitals code laboratory tests into LOINC. Coding error in other arenas, such as hospital discharge abstracts using ICD-9 has been studied by many investigators [6–

11]. Incorrect principal diagnosis coding errors between 18.5 and 42.8% have been observed [12]. For DRG encoding, error rates of 14.7–20.8% have been reported [12]. Such large error rates are not surprising because of the large degree of subjective interpretation and domain-specific knowledge that is required for encoding clinical diagnoses. The amount of disagreement found in the present study was similar to that found in the pilot study and surprising because encoding laboratory test names initially appeared to involve far less ambiguity than would encoding clinical diagnoses or DRGs. In practice this may not be true.

Disagreements among local experts cause differences in LOINC coding even for common identical tests. The power of LOINC to be highly specific also makes it complex and correct translation to LOINC requires a high resolution of knowledge of laboratory testing, precisely what properties are being measured and on what entity and by what method for each test. We found that in practice, physicians with good understanding of ordering and interpreting laboratory tests at their local institution, frequently don't have the resolution of knowledge to successfully translate all tests to LOINC. In our study, we found the 'kind of property' field created significant disagreements, even with the LOINC database itself. For example, the standard automated hematocrit, while frequently done by automated cell counting is a calculated value that represents a 'volume fraction' (VFRC). However, in the LOINC database, this hematocrit was called a 'number fraction' (NFRC), probably because an initial cell count is done prior to the calculation. Mismatches between our hospitals for tests such as 'erythrocyte mean corpuscular volume' (MCV) were caused by one hospital assigning 'entity volume' (ENTVOL) to the kind of property field (the entity being 'RBC'

(red blood cell) in the specimen field), while another hospital assigned simply 'blood' (BLD) to the specimen field. The coding for this test was different from the pilot study, but still resulted in a mismatch between the hospitals. The LOINC database assigned ENTVOL to the kind of property field and RBC to the sample type field in this case. However, for a conceptually similar laboratory test, the 'platelet mean volume', the LOINC database assigned ENTVOL to the kind of property field and BLOOD to the sample type field instead of the entity, PLATELET, which would have been consistent with their choice for the MCV. While such differences may seem trivial, they could, in practice, prevent common lab tests, which are actually identical between institutions, from being considered the same.

The 'sample type' field frequently caused mismatches because of the issue of serum versus plasma. Most analytes yield very similar results on serum or plasma and should be coded as 'SER/PLAS' even if a laboratory does primarily one or the other. Certain analytes, however, yield different results on serum or plasma (e.g. total protein and phosphorus) and < sample type > should be encoded as 'SER' or 'PLAS'. All hospitals in this study coded < sample type > for the most part based on what samples are handled by their laboratory and not based on analyte properties. This is the primary reason why reducing matching stringency from two LOINC fields (< analyte > : < sample type >) (Fig. 1) to 1 LOINC field (< analyte >) in this study greatly increased the number of matches. The serum versus plasma issue continues to remain an issue in the LOINC database. One of the hospitals in the present study chose to allow LOINC codes to be assigned to their tests if the only discrepancy with the LOINC database involved 'serum' in place of 'plasma' or 'plasma' in

place of 'serum'. That approach works for a laboratory utilization study, but may become problematic in a study comparing results for tests where plasma and serum yield divergent results.

Similarly, tests that differ only by laboratory methodology would need to be matched for a utilization study. For such applications where matching by LOINC code would be too specific, relaxing conditions required for a match by selecting subgroups of test name fields may be preferred. Attempting to reduce the matching stringency in our study did increase number of true matches attained (Fig. 1), but minimally so. One of the advantages of the LOINC approach is that coverage in the LOINC database is not always necessary, because tests can be given a fully-specified name using LOINC naming conventions. String matching individual fields of the LOINC test name can then be used to align tests from different sites. In the pilot study, this approach was advantageous, but after recoding for this study, matching by LOINC code was about as good as matching by LOINC test name fields.

Occasionally, a facility will not have distinct internal test codes for clinically significant specimen type differences. The lack of distinction of arterial from venous blood for blood gas testing for one of the hospitals was just one example of this. In general, such tests cannot be assigned a single LOINC code because they would lack the granularity of LOINC and encompass several distinct LOINC codes. Receiving systems must have an approach for handling incoming fully-specified LOINC names that are more general than those in the LOINC database or in their own system. Such situations would need to be addressed on a case-by-case basis and the solution would be dependent on the reason for the data-exchange. The inverse of this situation where preexisting codes in the LOINC database were not granular enough to explicitly define a test also occurred for several tests in the top 100 lists.

The knowledge-based approach of LOINC allows laboratory tests and more recently clinical results, to be explicitly defined by their code and/or name. Theoretically, this allows unambiguous pooling of data from diverse sites without the requirement for post-coding human inspection. This follow-up study confirms that this goal can potentially be reached only if a careful, standard LOINC coding procedure is used at all sites, carried out by individuals with significant domain-specific expertise and well-educated in the LOINC system. Even following recoding to a more recent version of LOINC, automated matching of the top 100 tests done on CHF patients between two academic hospitals results in a number of LOINC matching failures primarily due to differences in local coding choices. Institution to institution variability in LOINC coding could be partly alleviated if laboratory instrument manufacturers were to determine LOINC codes for assays carried out on their instruments.

Acknowledgements

References

[1] D.M. Baorto, J.J. Cimino, C.A. Parvin, M.G. Kahn, Using logical observation identifier names and codes (LOINC) to exchange laboratory data among three academic hospitals, Proc. Am. Med. Inform. Assoc. Annu. Fall Symp., 1997, pp. 96–100.

[2] Institute of Medicine, in: R.S. Dick, E.B. Steen (Eds.), Committee on Improving the Patient Record, The computer-based patient record: An essential technology for health care, National Academy Press, Washington DC, 1991.

[3] A.W. Forrey, C.J. McDonald, G. DeMoor, S.M. Huff, D. Leavelle, D. Leland, T. Fiers, L. Charles, B. Griffin, F. Stalling, A. Tullis, K. Hutchins, J. Baenziger, 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 (1) (1996) 81–90.

[4] LOINC user's guide. Available from http://www.mcis.duke.edu.standards/termcode/loinc.htm.

[5] M.G. Kahn, J.J. Cimino, D.M. Baorto, Utilization of the hospital laboratory in congestive heart failure: Geographic variability. Manuscript in preparation.

[6] L.K. Demlo, P.M. Campbell, S.S. Brown, Reliability of information abstracted from patient's medical records, Med. Care 16 (1978) 995–1005.

[7] R.F. Corn, The sensitivity of prospective hospital reimbursement to errors in patient data, Inquiry 18 (1981) 351–360.

[8] E. Cohen, D. Bernier, S. Tam, D. Schimel, A.H. Postel, S. Scheidt, J.B. Stamm, Data quality and DRGs: An assessment of the reliability of federal beneficiary discharge data in selected Manhattan hospitals, J. Commun. Health 10 (1985) 238–246.

[9] L.I. Iezzoni, S. Burnside, L. Sickles, M.A. Moskowitz, E. Sawitz, P.A. Levine, Coding of acute myocardial infarction: Clinical and policy implications, Ann. Intern. Med. 109 (1988) 745–751.

[10] D.C. Hsia, W.M. Krushat, A.B. Fagan, J.A. Tebbutt, R.P. Kusserow, Accuracy of diagnostic coding for Medicare patients under the prospect payment system, New Engl. J. Med. 318 (1988) 352–355.

[11] D.C. Hsia, C.A. Ahern, P.B. Ritchie, L.M. Moscoe, W.M. Krushat, Medicare reimbursement accuracy under the prospective payment system, 1985 to 1988, J. Am. Med. Assoc. 268 (1992) 896–899.

[12] J. Green, N. Wintfeld, How accurate are hospital discharge data for evaluating effectiveness of care?, Med. Care 31 (1993) 719–731.