Issues in Creating and Maintaining Value Sets for Clinical Quality Measures

Rainer Winnenburg, PhD, Olivier Bodenreider, MD, PhD

National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

`{winnenburgr|obodenreider}@mail.nih.gov`

Abstract

Objective: *To develop methods for assessing the validity, consistency and currency of value sets for clinical quality measures, in order to support the developers of quality measures in which such value sets are used.* **Methods:** *We assessed the well-formedness of the codes (in a given code system), the existence and currency of the codes in the corresponding code system, using the UMLS and RxNorm terminology services. We also investigated the overlap among value sets using the Jaccard similarity measure.* **Results:** *We extracted 163,788 codes (76,062 unique codes) from 1463 unique value sets in the 113 quality measures published by the National Quality Forum (NQF) in December 2011. Overall, 5% of the codes are invalid (4% of the unique codes). We also found 67 duplicate value sets and 10 pairs of value sets exhibiting a high degree of similarity (Jaccard > .9).* **Conclusion:** *Invalid codes affect a large proportion of the value sets (19%). 79% of the quality Measures have at least one value set exhibiting errors. However, 50% of the quality measures exhibit errors in less than 10 % of their value sets. The existence of duplicate and highly-similar value sets suggests the need for an authoritative repository of value sets and related tooling in order to support the development of quality measures.*

Introduction

In recent years, there has been an effort to establish quality measures for health care providers, in the objective of improving the quality of health care and comparing performance across institutions. The National Quality Forum (NQF) [1] defines quality criteria for health care and proposes uniform standards and measures. As a result, over the past 10 years standards and measures for patients care have continuously been refined and endorsed. In order to establish compatibility with electronic health record (EHR) systems and other clinical IT systems, quality measures comprise several hundred value sets, which are based on standard biomedical vocabularies and clinical terminologies including SNOMED CT (Medicine-Clinical Terms), CPT (Procedures & Supplies), ICD10 (Diseases and Health Problems), LOINC® (Diagnosis) and RxNorm (Drugs). In order to foster re-use, value sets are registered with unique Object Identifiers (OIDs) in registries such as the HL7 Object Identifier Registry.

Quality measure data are manually entered and maintained in a collection of spreadsheets for different domains of clinical care. Errors commonly deriving from manual data curation are to be expected. Furthermore, as value sets evolve over time, the underlying vocabularies also undergo regular updates. Such changes have to be propagated to the value sets in order to avoid an increasing number of stale references.

The lack of standard approaches for controlling the quality of value sets in quality measures motivates our work. The objectives of this study are to develop innovative methods for assessing the validity, consistency and currency of value sets for clinical quality measures. More specifically, we assess the validity of value sets and their individual values with regard to the latest changes in the underlying source vocabularies and identifying overlapping value sets by pairwise comparison of their code lists. We apply our techniques on the quality measures defined by the National Quality Forum (NQF). Based on our observations, we make recommendations for addressing staleness, incompleteness, and duplication of quality measure value sets.

Background

Quality measure value sets form the basis for guidelines and standards for measuring and reporting on performance regarding preferred practices or measurement frameworks. The NQF incorporates purchasers, physicians, nurses, hospitals, certification bodies to develop EHR technology and health IT standards. The Quality Data Model (QDM) is an information model established by the NQF, which defines concepts used in quality measures and clinical care and is intended to enable automation of electronic health record (EHR) use. Each QDM element is composed of a *category (e.g., Medication)*, the *state* in which that category is expected to be used and a

value set of codes in a defined *code system* (vocabulary). In a recently started effort, quality measures are converted into standardized electronic measures (eMeasures), which allow interaction with EHR systems and other clinical IT systems.

Value set identification. All the codes in an NQF value set come from the same code system, with the exception of those value sets containing other value sets (called Groupings, see also Figure 1). Each value set is identified by an OID. OIDs are global ISO (International Organization for Standardization) [2] identifiers registered in systems such as the Health Level Seven International (HL7) Object Identifier Registry [3].

NLM Terminology Services are used in this study for the quality assessment of value set codes. The Unified Medical Language System® (UMLS) terminology services (UTS) provide exact and normalized match search functions, which identify medical concepts for a given search string. The normalization process is linguistically motivated and involves stripping genitive marks, transforming plural forms into singular, replacing punctuation (including dashes) with spaces, removing stop words, lower-casing each word, breaking a string into its constituent words, and sorting the words in alphabetic order [4]. In this study, we use the UTS Java API for look up, as well as exact and normalized name mapping for code lists values. RxNorm is a standardized nomenclature for medications maintained by the U.S. National Library of Medicine (NLM) in cooperation with proprietary vendors [5]. RxNorm concepts are linked by NLM to multiple drug identifiers for each of the commercially available drug databases within the Unified Medical Language System® (UMLS®) Metathesaurus® (including NDDF Plus). In addition to integrating names from existing drug vocabularies, RxNorm creates standard names for clinical drugs. Although RxNorm is covered by the UMLS, in this study we access RxNorm data through its own API, since RxNorm is updated more frequently than the UMLS. For this study we used the February 2012 release of the RxNorm dataset.
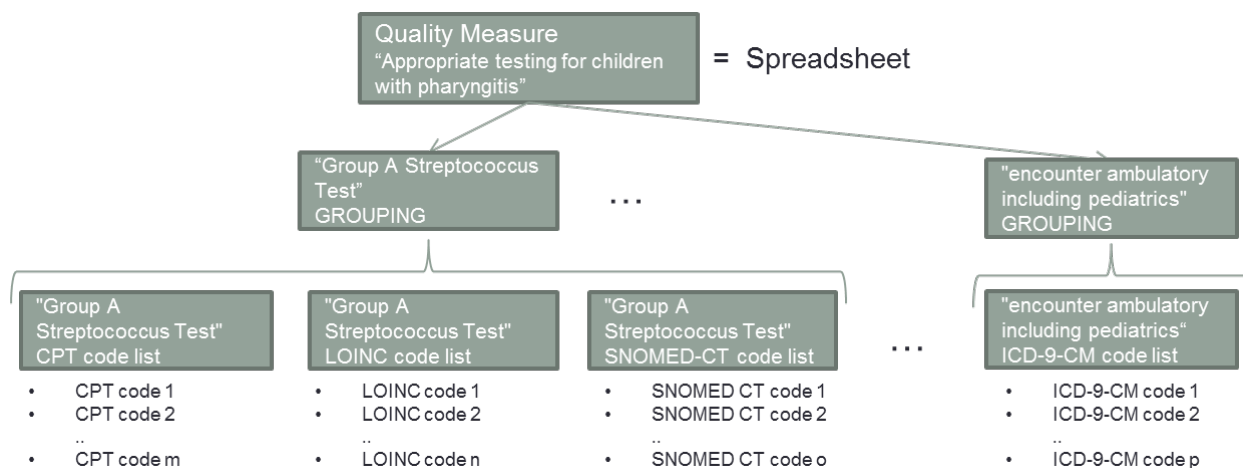

Figure 1 Organization of Quality measures. Each spreadsheet contains data for a specific part of clinical care (e.g. *Appropriate testing for children with pharyngitis*). For each aspect (e.g., *Group A Streptococcus Test*, etc.), a GROUPING complies one or several code lists, each specific to one code system (e.g., *CPT, LOINC, SNOMED CT, ICD9*, etc.).

Materials

Our study is based on the quality measures published by the NQF in December 2011, which were provided in 113 Excel spreadsheets. Each spreadsheet covers one specific Clinical Quality Measure, such as *Appropriate testing for children with pharyngitis* (see Figure 1). In total these value sets cover 164,659 (distinct 76,859) codes originating from various biomedical vocabularies. Each spreadsheet lists codes together with their name (descriptor), the name and the version of the code system they are originating from, as well as the OID and name of the value set and the QDM category they are associated with.

For example, the clinical drug with the **name** *Ampicillin 167 MG/ML / Floxacillin 167 MG/ML Injectable Solution* has the **code** *105134* in the *RxNorm* (**code system**) as of *2011* (**code system version**), and is part of the *Aminopenicillins* **value set** with the **OID** *2.16.840.1.113883.3.464.0001.157* under the **QDM category** *Medication* in the spreadsheet that describes performance measures for *Appropriate testing for children with pharyngitis*.

Methods

Overview. Our method for assessing the quality of the value sets found in the quality measures can be summarized as follows. We extract the value sets from the Excel spreadsheets and consolidate them into one file for convenience. We then assess the validity of the codes through a series of checks against terminologies. We assess the consistency of the value sets, making sure the code system from which the codes originate matches what is expected from the value set description. Finally, we assess the similarity of the value sets through pairwise comparisons.

The methods developed for the analysis of the quality measure data were implemented in Java 6. For acquiring the value set data from the Excel spreadsheets we used the Java Excel API [6]. The verification, mapping, and comparison of value set codes was realized using the UTS API version 2011AB [7] and the RxNorm API [8] as of February, 2012.

Acquiring data. The data from all *Value Set Descriptors* worksheets from the 113 Excel spreadsheets are retrieved, analyzed, accounting for differences in structure of the spreadsheets (different order of worksheets in spreadsheets and columns in worksheets), and consolidated into one file. We exclude Groupings from the value sets, since they have no direct association to the values. We keep provenance information (reference to original quality measures) for all extracted values.

Assessing the validity of codes. The purpose of this initial validity check is to identify codes that cannot be linked to valid entities in their corresponding code system. Reasons include wrong associations between a code and a code system, misspelled identifiers or names, or the use of obsolete or retired codes. Our approach goes through the steps depicted in Figure 2.
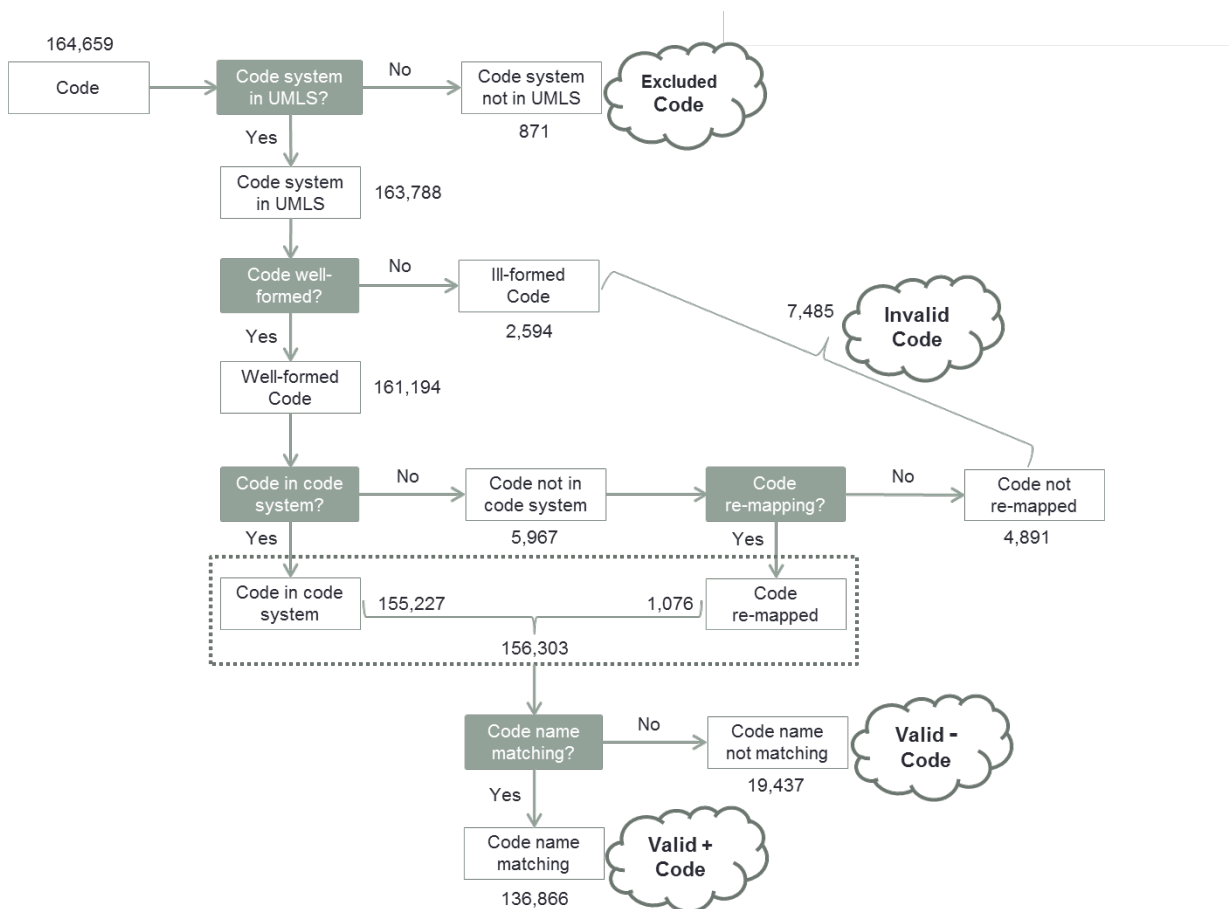

Figure 2 Assessing the validity of code.

Code system in UMLS? We check if the code system of a given value is present in the UMLS and can thus be processed by the UTS API, by mapping the code system entry in the spreadsheet to the code system abbreviation used by the UMLS, e.g., *SNOMED-CT* and *SNOMED → SNOMEDCT*. Entries of unknown code systems are put

aside as *Excluded Codes*. Only codes of known code systems are checked for *Code validity*.

Code well-formed? Codes are *well-formed* if they match the formal requirements of the corresponding vocabularies. The format of a code is checked against regular expressions specific to each code system, accompanied by a control digit check for codes from those systems that provide such a control mechanism (LOINC, SNOMED CT). Entries with codes that do not meet the designated code system's requirements are marked as *Invalid*.

Code in code system? The minimum requirement for a code to be considered valid is that it be *registered in UMLS* under the designated code system. RxNorm codes are checked against the latest version of *RxNorm* through the RxNorm API. Registered codes are considered for an additional *Code name matching* step. Codes not found in the code system are assessed in a *Code re-mapping* step.

Code re-mapping? Some codes may have become obsolete in a more recent version of a code system. If a value has been assigned to a new code, this code might be retrieved via the built-in remapping functions from the API (currently only available for RxNorm codes through the *getRxcuiStatus* method). The obsolete code will be replaced by the new code for further *Code name matching*. Entries with codes that cannot be re-mapped are marked as *Invalid*.

Code name matching? Depending on the level of accuracy required (strict/relaxed), it is additionally required that the name (descriptor) provided for this code in the spreadsheet match the label or one of its synonyms in the UMLS Metathesaurus. Compliance of code and name are determined using the exact and normalized string matching functions of the UTS API. Entries are marked as *VALID+* if a match is found; otherwise codes are marked as *VALID-*. We do not resolve such contradictions in an automated way, as it cannot be determined, whether the code or the name is correct. We still tend to consider the codes valid, because in most cases, what prevents the match is a minor phenomenon, such as truncation of the name.

Assessing the consistency of value sets. The purpose of this validity check is to identify errors regarding to the composition of value sets. Values of a particular code system are connected to a code list by the assignment of an OID and a value set name. Errors in OIDs, such as typographical errors, can result in incorrect assignment to a different code list. The validity of an OID assignment can be checked by comparing the value set names given for members of a particular value set.

Assessing the similarity of value sets. The validity assessment of values and value sets described above are pre-processing steps for the final evaluation of pairwise similarity of the value sets. We perform the evaluation with the relaxed value inclusion criteria, including both *VALID+* and *VALID-* codes (see Figure 2).

Source Code level: Giving the benefit of the doubt to those codes for which there is discrepancy between the name in the value set and the name in the code system, we calculate the similarity of two value sets from the same code system by comparing their values. Two value sets share the same value if they both contain the same valid code. We use the Jaccard index J to measure the similarity between two Value Sets A and B as follows:

$$J(A,B) = |A \cap B| / |A \cup B|.$$

UMLS CUI level: For the comparison of code lists from different code systems we introduce a preceding mapping of the individual codes to their corresponding Concept Unique Identifiers in the UMLS (UMLS CUIs). Such a mapping is possible because the UMLS Metathesaurus links terms from different source vocabularies that are synonyms of the same concept in the UMLS with one common UMLS CUI. Two value sets share the same value if they both contain a valid code mapped to the same UMLS CUI, regardless of the underlying Source CUIs.

The results can be compared easily since, at both levels, the same metric (Jaccard index) is being used.

Results

Assessing the validity of codes. We assessed the validity of the value set codes following the steps of our validation pipeline described in Figure 2.

Code system in UMLS? Of the 164,659 codes that participate in value sets in NQF quality measures, 163,788 (99%) originate from vocabularies covered by the UMLS.

Table 1 shows the number of entries for each of the source vocabularies. The value sets contained 871 codes from vocabularies not covered by the UMLS, namely the Medicare Severity – Diagnosis Related Groups (MS-DRGs), Place of Service Codes (POS), and U.S. Licensed Vaccine (CVX) Codes. These values are excluded from consideration in this study.

Table 1 Value Set Codes

	ICD10CM	SNOMED CT	ICD9	RxNorm	ICD10PCS	CPT	HL7	LOINC	HCPCS	Included	Excluded	Total
# entries	65553	31622	23571	21572	11751	7964	842	785	128	**163788**	**871**	**164659**
# entries (distinct)	33012	19706	7696	8695	3843	2298	156	577	79	**76062**	**797**	**76859**

Code well-formed? In a next step, the codes that are not well-formed with respect to the corresponding code system are analyzed. 2,594 codes were classified as ill-formed, leaving 161,194 codes for next validation step. The majority of discarded codes were ill-formed due to the following errors:

- *Excel-specific errors:* Due to Excel auto-format option there are missing decimal points and leading zeros in the spreadsheet data.
  Example: *Direct infection of unspecified elbow in infectious and parasitic disease classified elsewhere* with ICD10CM code M01.X29 has become M01x29, missing the decimal point.
  Example: *Bypass Left Subclavian Artery to Left Upper Arm Artery with Autologous Venous Tissue, Open Approach* with ICD10PCS code 0314091 has become _314091, missing the leading 0.

- *Typographical errors in codes:* Due to the manual data entry process, typographical errors can be easily introduced to names and codes.
  Example: *Diabetic cataract associated with type I diabetes mellitus* with SNOMEDCT code 42192002 was recorded as 421920<u>02</u>, containing an extra 0.

- *Wrong assignment:* The code system associated with a code in the value set might be wrong. As a consequence, well-formedness of the code cannot be assessed.
  Example: *Neomycin 25 MG/ML Oral Solution* with RXNORM code 311932 was associated with <u>SNOMEDCT</u>.

Code in code system? Of the 161,194 well-formed codes, 5,967 codes were not identified as valid in their code system when querying the UMLS and the RxNorm repositories. Since these codes match the formal requirements of the corresponding vocabularies, we expect to find many stale codes or codes with single typographical errors among them. Some stale codes of RxNorm can be re-mapped to valid codes using the RxNorm re-mapping function.

Code re-mapping? Code re-mapping was successful for 1,076 codes of code system RxNorm. The original codes were replaced with the re-mapped codes for further processing. Including the re-mapped RxNorm codes, 156,303 (95%) of 163,788 codes with code systems in the UMLS were validated.

Code name matching? For the 156,303 value set codes that could be identified in the UMLS or RxNorm we compared the names given in the value descriptions in the spreadsheets to the corresponding names in the code systems. 136,866 (83%) code names were found in the UMLS for the corresponding code, including 16,722 codes from RxNorm. These values were labeled *VALID+*, indicating that they fulfill the strict criteria for value set inclusion. For 19,437 values the corresponding code name was not found in the UMLS for this code. These values were labeled *VALID-*.

Table 2 shows the number of valid and invalid codes for all code systems, with (strict) and without (relaxed) name checking. Several reasons prevented names from being found in the UMLS for the corresponding code:

- Values originating from Current Procedural Terminology (CPT®) had no descriptor assigned, so no code/name comparison could be performed. Values originating from Logical Observation Identifiers Names and Codes (LOINC®) often had component names assigned, although the code referred to a test, e.g.

LNC|35565-1|HIV 1 p40 Ab, where the correct name for code 35565-1 would be *HIV 1 p40 Ab [Presence] in Serum by Immunoblot (IB).*

- Healthcare Common Procedure Coding System (HCPCS) descriptors are lengthy and thus error prone to typographical errors, abbreviations, truncations, etc., which makes it more difficult to find them in the UMLS. For example, the long descriptor for G8471 found in UMLS *Calculated BMI above the upper parameter and a follow-up plan was documented in the medical record* was shortened to *Bmi >= 30 was calculated and a follow-up plan was documented in the medical record* in the spreadsheet.

Table 2 Validity of Values

	ICD10CM		SNOMED CT		ICD9		RxNorm		ICD10PCS	
	Valid	Invalid	Valid	Invalid	Valid	Invalid	Valid	Invalid	Valid	Invalid
Strict (Valid+)	61608 (93%)	3945 (7%)	30812 (97%)	810 (3%)	16237 (68%)	7334 (32%)	16722 (77%)	4850 (23%)	11294 (96%)	457 (4%)
Relaxed (Valid-)	63517 (96%)	2036 (4%)	31035 (98%)	587 (2%)	23415 (99%)	156 (1%)	17641 (81%)	3931 (19%)	11390 (96%)	361 (4%)
	CPT		HL7		LOINC		HCPCS		Total	
	Valid	Invalid	Valid	Invalid	Valid	Invalid	Valid	Invalid	**Valid**	**Invalid**
Strict (Valid+)	0 (0%)	7964 (100%)	1 (1%)	841 (99%)	139 (17%)	646 (83%)	53 (41%)	75 (59%)	**136866 (83%)**	**26922 (17%)**
Relaxed (Valid-)	7834 (98%)	130 (2%)	568 (67%)	274 (33%)	781 (99%)	4 (1%)	122 (95%)	6 (5%)	**156303 (95%)**	**7485 (5%)**

Although, it could be argued that the overall number of invalid codes is relatively small, they actually affect a quite significant number of value sets.

Assessing the consistency of value sets.

Errors in value set identifiers (OIDs), such as typographical errors, can result in the incorrect association of a code with another value set. For example, if $code_1$ is associated with OID_2 instead of OID_1, $code_1$ will be not only missing from the codes in the value set identified by OID_1, but will also be incorrectly included in the value set identified by OID_2. What is true for the association between code and OID is also true of the association between value set name and OID. In fact, we identified cases where different value set names were associated with the same OID (see Figure 3 for typical examples). In many cases, the cause of the error is a minor variant in the value set name (first three cases in Figure 3). However, we have also encountered cases of errors in the OID, as in the fourth example in Figure 3, where two totally different names share the same OID. In this case, we traced the origin of the erroneous OIDs to misuse of the mouse-based copy function in Excel, which increments numeric values – including OIDs – when the values are "dragged" with the mouse. This issue can have dramatic consequences when an OID is copied (with incrementation) over a large number of cells.

- Multiple names for 2.16.840.1.113883.3.666.05.747
  - HOSPITAL MEASURES-ENDOTRACHEAL INTUBATION
  - HOSPITAL MEASURES-ENDOTRACHAEL INTUBATION
- Multiple names for 2.16.840.1.113883.3.464.0001.395
  - RUBELLA VACCINATION
  - RUBELLA VACCINE
- Multiple names for 2.16.840.1.113883.3.666.05.761
  - HOSPITAL MEASURES-IV OR PO QUINOLONES
  - HOSPITAL MEASURES-IV OR PO ANTIPNEUMOCOCCAL QUINOLONES
- Multiple names for 2.16.840.1.113883.3.464.0001.399
  - PREGNANCY
  - SEXUALLY ACTIVE WOMEN

Figure 3 Typical examples for code lists with multiple names

For example, terms for the ICD9 value set with the name *pregnancy* and OID 2.16.840.1.113883.3.464.0001.104 is also associated with OIDs ranging from 2.16.840.1.113883.3.464.0001.79 to 2.16.840.1.113883.3.464.0001.218, not

only resulting in missing terms for the original pregnancy code list, but also introducing irrelevant pregnancy terms to other existing code lists in the same range.

Assessing the similarity of value sets.

We conducted a pairwise comparison of all value sets that contained at least one valid code after assessing the validity of codes and the consistency of value sets.

Value Set comparison (Source Codes). We compared 1361 code lists from 9 different vocabularies. Table 3 shows the number of considered value sets for each code system. We calculated the pairwise comparison between all code lists within a given code system.

Table 3 Code system specific code lists considered for pairwise comparison

SNOMEDCT	RXNORM	ICD9CM	ICD10CM	CPT	LNC	HL7V3.0	HCPCS	ICD10PCS	Total
430	252	230	198	149	54	20	17	11	**1361**

Value Set comparison (UMLS CUIs). In contrast to the value set comparison described above, we calculated the similarity of all 1361 code lists against each other. We mapped the code of each value set to their equivalent UMLS CUI using the *findCUIBySabCode* function of the UTS API. Although this approach broadens the search space, it will also detect similar value sets across code systems for the same quality measure aspect.

Table 4 Results of pairwise comparison of code lists

Jaccard index J	1	>.9	>.8	>.7	>.6	>.5
# pairs (Source Codes)	67	10	15	16	25	31
# pairs (UMLS CUIs)	90	11	18	11	34	33

Simply considering the source codes, there are 67 duplicate value sets and 10 pairs of value sets exhibiting a high degree of similarity (Jaccard > .9). Value sets with a similarity of J > .9 share a large number of codes and are thus good candidates for potential redundancy. Redundancy can result from independent registration of a value set with different OIDs by different authority bodies. Thus, it should be expected to retrieve similar value sets that differ significantly in their assigned OIDs.

Discussion

Significance of findings. Although the overall proportion of invalid codes is generally modest (4-5%, when name validation is not required), these codes have important repercussions on the validity of value sets and, in turn, on the applicability of the quality measures themselves. Figure 4 shows the proportion of value sets (top left) and quality measures (top right) affected by the presence of invalid codes. A total of 291 value sets (21%) have at least one invalid code. Of the 113 quality measures, 89 (79%) have at least one value set affected by the presence of at least one invalid code. However, 50% of the quality measures exhibit errors in less than 10 % of their value sets. Additional analysis is required to assess the degree to which quality measures will be affected in practice, which is important for the practical applicability of these quality measures in the context of the "meaningful use" [9, 10].
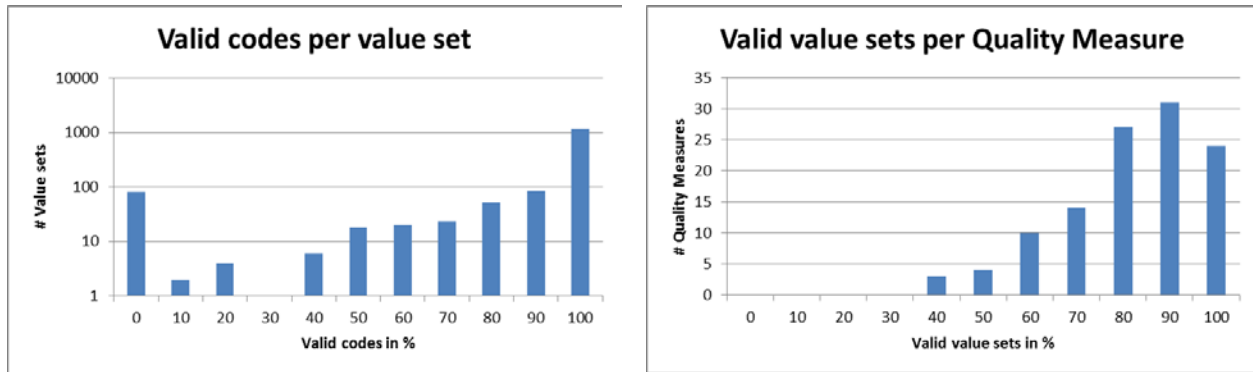
Figure 4 Effect of invalid codes on value sets and Quality Measures

Recommendations. Based on our observations and analyses, we suggest the following.

- In order to encourage reuse over reinvention, there needs to be an authoritative source of value sets, preferably a repository from which value sets could be downloaded. Metadata about the value sets should make it easy to discover existing value sets and include information about provenance and currency.

- A set of tools is needed to support the development of value sets. Such tools should support search by similarity within and across code systems, maintenance when code systems evolve, and intensional definitions (e.g., "diabetes mellitus" and all its descendants in SNOMED CT), as well as extensional definitions (lists of codes).

- Authoring tools for the clinical quality measures should respect the formatting of the codes (decimal points, left padding with 0s, etc.)

Limitations and future work. We have demonstrated the applicability of our value set assessment approach on quality measures defined by the NQF. However, our methodology is applicable to value sets from other organizations, such as the PHIN Vocabulary Access and Distribution System (VADS) [11] and the Standard and Interoperability Framework. We plan to assess these value sets as well.

In this study, we purposely ignored those value sets composed of other value sets (groupings). We intend to resolve them into their values and use them in our comparisons. More generally, we would like to extend our methods to value sets defined in intension, rather than in extension.

Finally, our assessment of overlap was originally limited to shared source codes (within code sets). We extended it to UMLS CUIs, which allowed us to compare value sets across code sets. However, our comparison is still limited to shared codes. By using semantic similarity, we would expand the comparison to closely related codes across value sets, within or across code systems.

Conclusions

We developed a method for assessing the validity, consistency and currency of value sets that works for all code systems that are covered by the UMLS. In our approach we identified errors in values related to (typographical errors, copy/paste) and stale values due to missing update mechanisms. Regarding the problem of stale codes, we could demonstrate the importance of re-mapping functions. As a consequence, the RxNorm API was extended with a status function, which provides the current status of a code (active, retired, and unknown) including re-mapping information. Assessing the validity, consistency and similarity of value sets revealed dependencies and redundancies between value sets, suggesting the need for an authoritative repository and related tools. Finally, although the overall proportion of invalid codes is generally modest, these codes have important repercussions on the validity of value sets and, in turn, on the applicability of the quality measures themselves.

References

1. *National Quality Forum (NQF)*. Available from: http://www.qualityforum.org.
2. *International Organization for Standardization (ISO)*. Available from: http://www.iso.org/iso/en/ISOOnline.frontpage.
3. *Health Level Seven International (HL7) Object Identifier Registry*. Available from: http://www.hl7.org/oid/.
4. McCray, A.T., S. Srinivasan, and A.C. Browne, *Lexical methods for managing variation in biomedical terminologies.* Proc Annu Symp Comput Appl Med Care, 1994: p. 235-9.
5. Nelson, S.J., et al., *Normalized names for clinical drugs: RxNorm at 6 years.* Journal of the American Medical Informatics Association : JAMIA, 2011. **18**(4): p. 441-8.
6. *Java Excel API*. Available from: http://jexcelapi.sourceforge.net.
7. *UMLS Terminology Service Metathesaurus Browser*. Available from: http://uts.nlm.nih.gov/metathesaurus.html.
8. *RxNorm API*. Available from: http://rxnav.nlm.nih.gov/RxNormAPI.html.
9. Blumenthal, D. and M. Tavenner, *The "meaningful use" regulation for electronic health records.* N Engl J Med, 2010. **363**(6): p. 501-4.
10. Health and Human Services Department, *Health Information Technology: Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology, 2014 Edition; Revisions to the Permanent Certification Program for Health Information Technology: A Proposed Rule by the Health and Human Services Department on 03/07/2012.* Federal Register, 2012: p. 13832-13885.
11. *PHIN Vocabulary Access and Distribution System (VADS)*. Available from: http://phinvads.cdc.gov.