Using SNOMED-CT as a Reference Terminology: Mapping VA Disability Terminology to ICD-9-CM

Peter L Elkin, Steven H Brown, Casey Husser, Dietlind Wahner-Roedler, Sandra Bailey, Linda Nugent, Karla Porter Brent, A Bauer

Abstract:

We hypothesized that SNOMED-CT could be used to assist direct International Classification of Diseases 9 – Clinically Modified [1]. (ICD-9-CM) mappings in creating a valid crosswalk to the Veterans Benefits Administration (VBA) code set. This was accomplished by creating an ICD-9-CM terminology server [2] and directly mapping that to the textual descriptions of the VBA codes. ICD-9-CM and the VBA codes were both mapped to SNOMED-CT. The SNOMED-CT mappings were matched and terms from related concepts were displayed for an expert coder’s review. We report the rate of direct ICD-9 mapping (26%), the success of the crosswalk (95%) and the rate at which the reviewer had to add codes to complete the mapping (99%). The method using the SNOMED-CT crosswalk provided significantly better coverage than the ICD-9-CM direct mapping alone (Pearson Chi Square test; p<0.001). We conclude that SNOMED-CT when used as part of the method described below can be a useful adjunct to direct mapping between administrative classifications.

1. Introduction

Precoordinated terminologies used for local administrative purposes not infrequently need to be mapped to national and international administrative code sets such as ICD-9 or ICD-10 [3]. This presents a problem as many of these precoordinations do not have exact morph-syntactic equivalence in the corresponding terminology. We suggest that one way to perform this crosswalk is using an intermediate reference terminology. Here we used SNOMED-CT v1.0 to map the Veterans Administration’s (VA) VBA codes to ICD-9-CM. The VBA codes are a set of administrative codes used at the VA for assigning coverage of a veteran’s benefits as well as other uses. For example, a veteran may be service connected for “Knee Pain” due to an injury jumping off a structure during a battle.

Comparable data is the basis for the practice of evidence-based medicine. Many clinical questions will never be addressed in a randomized, blinded, controlled clinical trial. The best hope for providing clinicians answers to these questions lies with a deeper understanding of the clinical record. Recording information at the granularity with which we practice medicine holds the promise to provide the data needed to gain an improved understanding of what constitutes the “best practice” of medicine. In order to accomplish this goal we need controlled health vocabularies to insure comparable data. Compositional vocabularies are one potential answer to the problem of providing enough content completeness to be clinically useful.
The degree to which a compositional mechanism can provide coverage for the concepts used by clinicians at the point of care is not currently known. The data from the LSVT trial [4] provides a benchmark for the coverage of these concepts by a large set of atomic and pre-coordinated concepts. By making use of standard vocabularies and adding a mechanism for creating composite terms from individual terms, we can test the value of these constructions.

A previous article that reviewed SNOMED-CT’s ability to represent more complex statements was entitled “Coverage of Oncology Drug Indication Concepts and Compositional Semantics by SNOMED-CT” [5]. Here we reviewed 115 oncology drug indications containing 1527 unique concepts. We found that SNOMED-CT was only capable of representing 86.3% of the concepts completely and 60.1% of the linking semantics. Of the errors, 50.5% were drug names that were unknown to the terminology. The next most common type of concepts, which were missing from SNOMED-CT, involved specific treatments at 40.5%. Therefore it was unclear whether or not SNOMED-CT could serve effectively as a reference terminology for the purpose of performing a crosswalk between two administrative pre-coordinated classifications.

2. Methods

The crosswalk was accomplished by first creating an automatic mapping of the terms from the ICD-9-CM and VBA classifications to SNOMED-CT. In many cases, this produced compositional expressions using SNOMED-CT codes that were representative of the original pre-coordinated term from the classification. We categorized the SNOMED-CT concepts as either Kernel concepts, Modifiers, Qualifiers, Severity, or Laterality [6]. These distinctions were hand assigned by hierarchy to the SNOMED-CT reference terminology (ple). Kernel concepts are defined as a main point of the concept, modifiers change the meaning of a term in a clinical sense (e.g., Stage, Grade), qualifiers change the meaning of a term in a temporal or administrative sense (e.g. History of, Recurrent), Severities and Laterality follow the SNOMED-CT hierarchies for these concepts. Using SNOMED-CT definitions augmented by these distinctions, the investigators studied how well each of these coding systems could be linked together.

The investigators used the following general rules in the creation of the automated crosswalk:

1) First the investigators matched the Kernel concepts (at least one has to match, if one exists), then the modifiers, then qualifiers, then severity, then laterality in that order. This continues until either all concepts are matched or zero matches are returned. When zero matches are returned, the list of matches in the set before it will be returned in a list to be adjudicated by an expert reviewer.

2) The output of the pair-wise crosswalk between these three classifications (three output files were generated) were then reviewed by two expert reviewers (clinicians), and when they disagree a third reviewer was employed to resolve the issue. When no exact match was available, the reviewers made their determination using a browser capable of displaying each of the target terminologies, whether:

a) There was no appropriate crosswalk between the two terminologies (hence SNOMED-CT legitimately could not and should not have connected the two terminologies or

b) When there was a legitimate match between the two classifications and SNOMED-CT did not provide the appropriate mapping: In the case when SNOMED-CT failed to map two synonymous concepts, a failure analysis was
employed that noted what was missing in SNOMED-CT that would have facilitated the mapping and as well as which classification contained the unrecognized content and/or semantics.

c) These failures were double checked by the reviewers using a SNOMED-CT terminology browser to avoid errors in the mapping software leading to an undeserved check mark against SNOMED-CT.

The crosswalk strategy includes categorizing each of the SNOMED-CT concepts as Kernel Concepts, Modifiers, Qualifiers, Laterality, and Severity. After these types have been established, we give preference to matches of Kernel concepts then Modifiers then Qualifiers then Severities and lastly to Lateralities. The two concepts must have at least one Kernel Concept in common and after that, the more prioritized commonality the better. The concepts with the best matches are returned. Human review can then determine whether the appropriate concept was identified. Of course, if there is a direct match (all words or their synonyms are represented in a concept from each terminology), then that takes precedence over the crosswalk.

Examples of such mappings are:

**Case 1: Disseminated intravascular coagulation with renal cortical necrosis**

Which maps to:

- [WITH]
  - Disseminated intravascular coagulation (disorder) [67406007] [K]
  - Renal infarction (disorder) [45456005] [K]

where [with] is an operator which applies a Boolean “AND” to the concepts identified by the text / semantic type / concept ID and concept class.

This maps to two ICD-9-CM concepts:

- **ACUTE RENAL FAILURE WITH LESION OF RENAL CORTICAL NECROSIS** (disease) [584.6] [K]

AND

- **DEFIBRINATION SYNDROME** (disease) [286.6] [K]

**Case 2: Dementia of unknown etiology**

Which maps to:

- - [AND]
  - Dementia (disorder) [52448006] [K]
  - [UNKNOWN]
    - Etiology (attribute) [134198009] [M]

where AND is a Boolean operator and [unknown] is an uncertainty operator.

This term has four possible matches in ICD-9-CM:

- **PRESENIILE DEMENTIA** (disease) [290.1] [K]
- **DEMENTIA IN CONDITIONS CLASSIFIED ELSEWHERE** (disease) [294.1] [K]
- **PRESENIILE DEMENTIA, UNCOMPLICATED** (disease) [290.10] [K]
- **SENILE DEMENTIA, UNCOMPLICATED** (disease) [290.0] [K]

The correct match would need to be chosen by an experienced coder in this situation.
Case 3: Paralysis of musculocutaneous nerve

Which maps to:

- Paralysis (finding) [44695005] [K]
- [has Finding Site]
  - Entire musculocutaneous nerve (body structure) [181019000] [M]

where [has finding site] is a SNOMED relationship.

ICD-9-CM has the single concept:

INJURY TO MUSCULOCUTANEOUS NERVE (disease) [955.4] [K]

which only approximates the meaning of the original input string.

Case 4: Thrombo-angiitis obliterans (Buerger’s disease)

Which maps simply to:

Thromboangiitis obliterans (disorder) [52403007] [K]

ICD-9-CM has the exact concept match:

THROMBOANGIITIS OBLITERANS (BUERGER’s DISEASE) (disease) [443.1] [K]

A web-based system (four-tier architecture with a relational database back end) was designed and implemented to facilitate the review and grading of the acceptability of the crosswalk. The system showed the reviewer the original terms; their matches and the compositional expressions to which they mapped in SNOMED-CT. The interface was available to the reviewers to view and grade the output of the systems matches to the VA disability terminology and to ICD-9-CM. We also provided a set of Web based services and interfaces to allow coders to review the crosswalk between the VA disability terminology and to ICD-9-CM using SNOMED-CT, and this interface allowed the coders to pick the correct mapping or approve the automated mapping as warranted by the review.

The results are reported from a SNOMED-CT perspective. True positives are concepts from the classifications that were found to have at least one valid match between the two classifications using SNOMED-CT. The mappings were reviewed by an expert ICD-9-CM coder who had access to the VBA staff in the case that questions arose regarding a VBA code’s meaning. A correct mapping rate of ≥ 93% was arbitrarily judged to be acceptable accuracy for clinical use. This rate is similar to or better than many of the sensitivities for common confirmatory diagnostic tests [7].
3. Results

Table 1: SNOMED-CT Mapping to a VBA Code (VBA Code|VBA Text|SNOMED-CT)

<table>
<thead>
<tr>
<th>VBA Code</th>
<th>VBA Text</th>
<th>SNOMED-CT</th>
</tr>
</thead>
</table>
| 8310     | Vagus nerve Neuritis. | Neuritis (disorder) [84299009] [K]  
|          |          | [has Finding Site]  
|          |          | . Entire vagus nerve (body structure)  
|          |          | [362466001] [M] |

Table 2: VBA Code, then SNOMED Mapping, then ICD-9 Direct Map, and ICD-9 Map. In the first row the direct map was an exact match while in the second row there was a SNOMED-CT Match and a Crosswalk match but no Direct mapping to ICD-9.

<table>
<thead>
<tr>
<th>VBA Code</th>
<th>VBA Text</th>
<th>SNOMED-CT</th>
<th>ICD-9 Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>8020</td>
<td>Abscess brain</td>
<td>Abscess brain</td>
<td>INTRACRANIAL ABSCESS (disease) [324.0] [K]</td>
</tr>
<tr>
<td>6030</td>
<td>Accommodation, paralysis of (eye)</td>
<td>Accommodation, paralysis of (eye)</td>
<td>PARESIS OF ACCOMMODATION (disease) [367.51] [K]</td>
</tr>
</tbody>
</table>

Overall, the ICD-9-CM direct map only correctly mapped 26% of the VBA codes. The addition of the crosswalk on top of the direct map provided some suggested code in 98% of the cases and at least one code was considered an appropriate match in 95% of the cases. As the goal of this project was to provide a mapping from the VBA codes to all possibly appropriate ICD-9-CM codes, the expert coder added additional codes to almost all of the cases (99%). Many inappropriate codes were removed as we had expected given that the goal of the project was to maximize the sensitivity of the process. When we compared the rate of direct ICD-9-CM mapping to that of the ICD-9-CM direct map plus the crosswalk the later provided significantly more coverage (Pearson Chi Square test; p<0.001).

4. Conclusions

SNOMED-CT was useful in providing suggested mappings for a crosswalk between the VBA administrative terminology and ICD-9-CM. However, almost all of the mappings required adjustment using expert human review. We believe that this is due to the fact that these coding systems have considerable meaning that is not expressed in the core classification but is defined in the user’s manual. When these rules are incorporated into the system the authors believe the system will yield greatly improved results. We did not do a comparative study regarding how long it would have taken to perform the same task without the crosswalk but we believe that this method linked with a good user interface significantly shortened the review time. Further, we believe that the suggestions helped the reviewer to be more complete as it found matches across hierarchies which may have been missed if not cued to the reviewer by the crosswalk system. The crosswalk provided significantly more correct mappings than a direct ICD-9 match alone.

SNOMED-CT should be considered as a useful reference terminology for providing a set of suggested matches using the crosswalk method outlined in this manuscript to map between and among administrative classifications.
5. Acknowledgements

This work was funded in part by a grant from the National Library of Medicine (LM06918). The authors would like to thank Raju Kalidindi for his technical assistance with this project.

References: