Evaluation of SNOMED Coverage of Veterans Health Administration Terms

Janet FE Penz a, Steven H Brown b, John S Carter c, Peter L Elkin d, Viet N Nguyen a, Shannon A Sims a, Michael J Lincoln a

a Veterans Administration Hospital and University of Utah Department of Medical Informatics, Salt Lake City, Utah
b Veterans Administration Hospital, Nashville, Tennessee
c Apelon, Inc., Ridgefield, CT d Mayo Clinic, Laboratory of Biomedical Informatics, Rochester, MN

Abstract

Veterans Health Administration (VHA) is now evaluating use of SNOMED-CT. This paper reports the first phase of this evaluation, which examines the coverage of SNOMED-CT for problem list entries. Clinician expressions in VA problem lists are quite diverse compared to the content of the current VA terminology Lexicon. We selected a random set of 5054 narratives that were previously “unresolved” against the Lexicon. These narratives were mapped to SNOMED-CT using two automated tools. Experts reviewed a subset of the tools’ matched, partly matched, and un-matched narratives. The automated tools produced exact or partial matches for over 90% of the 5054 unresolved narratives. SNOMED-CT has promise as a coding system for clinical problems. In subsequent studies, VA will examine the coverage of SNOMED for other clinical domains, such as drugs, allergies, and physician orders.

Keywords

Veterans Health Administration, Terminology, SNOMED, Controlled Vocabulary, Medical Records Systems, Computerized

Introduction

The Veterans Health Administration (VHA) is the largest integrated health system in the United States and has a vast repository of clinical data that is currently stored as either coded or narrative text. Veterans Health Information Systems and Technology Architecture (VISTA) [1] forms the framework for VHA’s Computerized Patient Record System (CPRS). VHA is now preparing to implement a national (and locally replicated) Health Data Repository (HDR) as well as a variety of standard coding systems such as MeSH, ICD-9, and CPT. However, clinical problem expressions have proved to be considerably more diverse than Lexicon can accommodate. When clinicians fail to find a suitable code for a problem narrative, they can optionally choose their narrative as a free text problem expression. As a result, VA has been accumulating hundreds of thousands of “unresolved problem narratives” each year. These unresolved narratives are often more granular or more pre-coordinated (e.g., “chest pain radiating to the left arm”) than the existing Lexicon content.

The U.S. Federal government has recently negotiated a license for the Systematic Nomenclature of MEDicine-Clinical Terms (SNOMED-CT). This formal reference terminology is compatible with the VA’s Enterprise Reference Terminology and may offer advantages over the existing VHA terminology sources such as the Lexicon. The VA is now evaluating how to use SNOMED-CT. This paper reports upon that evaluation’s first phase, examining the coverage of SNOMED-CT for VA problem list terms and phrases. SNOMED-CT contains approximately 344,000 concepts, 913,000 English language descriptions and more than one million semantic relationships [2]. It is a multi-axial, hierarchical coding scheme that has previously shown to have broad coverage of health terminologies [3,4].

Methods

Unresolved problem narratives arise at all VA sites. A random sample of these narratives (URN’s) comprising 5,054 terms was selected for this study. These narratives have a broad scope and un-matched narratives. SNOMED-CT has promise as a coding system for clinical problems. In subsequent studies, VA will examine the coverage of SNOMED for other clinical domains, such as drugs, allergies, and physician orders.

VHA has now decided to standardize its clinical terminologies across the enterprise in order to accommodate an HDR of national scope. VA clinicians across the nation now enter practically all notes, orders, problems, etc using VA’s Computerized Patient Record System (CPRS). As these individual CPRS systems are re-engineered to feed the HDR, data standardization becomes a primary concern. For example, VISTA problem list entries are currently resolved against the VA Lexicon Utility. This utility, in existence for 10 years, contains VA specific terminology as well as a variety of standard coding systems such as MeSH, ICD-9, and CPT. However, clinical problem expressions have proved to be considerably more diverse than Lexicon can accommodate. When clinicians fail to find a suitable code for a problem narrative, they can optionally choose their narrative as a free text problem expression. As a result, VA has been accumulating hundreds of thousands of “unresolved problem narratives” each year. These unresolved narratives are often more granular or more pre-coordinated (e.g., “chest pain radiating to the left arm”) than the existing Lexicon content.

The U.S. Federal government has recently negotiated a license for the Systematic Nomenclature of MEDicine-Clinical Terms (SNOMED-CT). This formal reference terminology is compatible with the VA’s Enterprise Reference Terminology and may offer advantages over the existing VHA terminology sources such as the Lexicon. The VA is now evaluating how to use SNOMED-CT. This paper reports upon that evaluation’s first phase, examining the coverage of SNOMED-CT for VA problem list terms and phrases. SNOMED-CT contains approximately 344,000 concepts, 913,000 English language descriptions and more than one million semantic relationships [2]. It is a multi-axial, hierarchical coding scheme that has previously shown to have broad coverage of health terminologies [3,4].

Methods

Unresolved problem narratives arise at all VA sites. A random sample of these narratives (URN’s) comprising 5,054 terms was selected for this study. These narratives have a broad scope and include abbreviations, diagnoses, procedures, medical risk factors, information about patient involvement in research trials, symptoms, laboratory findings, etc. Some narrative phrases in-
volve complex, pre-coordinated concepts, for example, “radiation proctitis with lower GI bleed and severe anemia”. The narratives also include abbreviations, misspellings and non-standard phrases or words such as “transaminitis”. Therefore, not all unresolved narratives are necessarily codable.

We selected a random set of 5,054 unresolved narratives and submitted each to two automated SNOMED mapping tools. One tool was provided by Apelon, Inc. as part of its TermWorks product and this searched SNOMED for concept names matching the unresolved narratives using a multi-step algorithm. The algorithm uses word-based tokenization, word normalization and stemming to find matching concepts. From the set of results returned, we used the “best match” as determined by highest number of tokens matched as an entry point into SNOMED-CT for expert reviewers to evaluate whether or not an appropriate concept describing the health factor was found in SNOMED-CT.

The narratives were also submitted to the SmartAccess Vocabulary Server (SAVS; pronounced “Saves”), which has been developed by Dr. Peter Elkin’s laboratory at the Mayo Clinic, Rochester, MN. The SAVS system is an indexing tool with source materials using a Concept-Based Indexing Schema. The ontology of medical concepts available in the SNOMED-CT terminology provides the underlying indexing schema. The software has been extensively tested in Mayo’s Usability Laboratory. SAVS provides capability for spelling correction, synonym and coding schemes and data storage.

For each term we compared the output of the SNOMED mapping for both tools. The automated results were classified as “complete” matches when each tool produced the same SNOMED code(s). The results were classed as “partial” matches when the two tools produced non-congruent SNOMED code(s). If neither tool generated a SNOMED match, then the result was classified as “failed”. Thus the automated matching results on each narrative fell into one of these three groups. A clinician examined partial or failed matches and corrected any spelling errors and expanded acronyms or abbreviations. Subsequently several clinicians with coding experience (JFEP, MJL, SS, VN) examined samples which totaled 200 of the 603 “complete” matches, 400 of the 4148 “partial” matches, and 303 of the 5054 “failed” matches. Because of the relatively small numbers in this group the sample was made up of all 303.

Clue-5 Browser (http://www.clinical-info.co.uk/), a look-up engine for SNOMED CT terms, was used to confirm accuracy of the tool generated SNOMED concept codes and to assess for the existence of codes in the situations in which there was either incorrect code generation or where no code at all was generated.

Results

The initial comparison of the SNOMED concept ID codes generated by each tool is summarized in Table 1. Table 2 shows the results of the evaluation and expert review of these samples of URN’s.

When the sample of 200 “complete” matches were manually compared to the corresponding SNOMED concept obtained using the Clue Browser the majority did demonstrate accurate assignment of codes by the automatic tools. Determinations were made by expert reviewers regarding the existence of appropriate SNOMED concepts for any elements that were not captured by the automatic tools to assess the overall adequacy of SNOMED to cover the URN terms. As expected there was a high degree of agreement in this group having URN terms for which both tools provided identical concept codes. 83% of “complete” matches were determined to be accurate in the coverage of the URN and with manual coding a total of 93.5% of the terms were found to have acceptable and completely encompassing SNOMED concepts.

Table 1: Comparison of the degree of agreement in SNOMED concept ID codes generated for URN’s by both TermWorks and SmartAccess tools.

<table>
<thead>
<tr>
<th>Tool output</th>
<th>“Complete” agreement</th>
<th>“Partial” agreement</th>
<th>“Failed” with no output</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. (%)</td>
<td>604 (11.9)</td>
<td>4148 (82.1)</td>
<td>303 (6.0)</td>
<td>5054 (100.0)</td>
</tr>
</tbody>
</table>

“Partial” tool matches formed the largest category comprising 4148 of 5055 URN’s. The random sample of 400 were treated similarly and yielded 46.5% accuracy by the automatic tools matching and with an additional 34% being partially accurate. Common reasons for incomplete match or partially incorrect match were abbreviations and spelling errors in a portion of a composite term that the tools were unable match. Expert review was able to capture almost all remaining terms. Thus a total of 91.3% of this sample showed accurate coverage by SNOMED concepts.

The third category consisting of URN’s for which neither tool could provide a SNOMED code (“failed matches”) included 303 URN’s. Initial correction of spelling and abbreviation errors were necessary during expert review and manual assessment of the existence of appropriate SNOMED concepts to cover the URN terms. Although the reviewers could not resolve 104 terms, 197 of 199 terms (99%) could be accurately manually matched to SNOMED codes. Two remaining terms were found to have inadequate SNOMED coverage.

Table 3 demonstrates example which illustrate the limitations associated with the use of automatic coding tools for uncorrected text. Spelling errors such as “fatigue” not surprisingly result in inaccurate coding. Similarly many common colloquialisms such as “labs” instead of “laboratory tests” are not currently included in SNOMED CT and cannot be accurately coded. Unfortunately, a significant number of URN terms could simply not be interpreted at all as a result of the use of non-standard abbreviations, spelling errors, non-standard terms or a combination of these factors. For example, given the clinical source of the URN terms, “Tuber” most likely represents tuberculosis rather than a fleshy root or bulb. More interesting, though, are the terms which simply do not exist in SNOMED or, because of ambiguity in the language of the URN, are inaccurately coded. The example of “tetanus” is representative of the latter situation. In this case “tetanus”, which appeared numerous times in the list of 5054 URN’s, is almost certainly being entered as a short form
Table 2: Assessment of SNOMED codes generated by automatic tools and expert review with manual look up. The first rows demonstrate the accuracy of the tool outputs when reviewed by clinical experts. The unknown group included all URN’s which remained ambiguous or uninterpretable after spelling correction and expansion of abbreviations.

<table>
<thead>
<tr>
<th></th>
<th>Perfect agreement in tool output codes (No./%)</th>
<th>Incomplete agreement in tool output codes (No./%)</th>
<th>No codes generated by tools (No./%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tool output accurate</td>
<td>166 (83.0)</td>
<td>200 (50.0)</td>
<td>-</td>
</tr>
<tr>
<td>Tool output with partial agreement (broad)</td>
<td>24 (12.0)</td>
<td>75 (18.8)</td>
<td>-</td>
</tr>
<tr>
<td>Tool output with partial agreement (narrow)</td>
<td>0 (0.0)</td>
<td>27 (6.8)</td>
<td>-</td>
</tr>
<tr>
<td>Tool output incorrect</td>
<td>2 (1.0)</td>
<td>41 (10.1)</td>
<td>-</td>
</tr>
<tr>
<td>Tool output partially incorrect</td>
<td>0 (0.0)</td>
<td>17 (4.2)</td>
<td>-</td>
</tr>
<tr>
<td>Unknown</td>
<td>8 (4.0)</td>
<td>39 (9.8)</td>
<td>104 (34.3)</td>
</tr>
<tr>
<td>URN’s unmatched to SNOMED by tools</td>
<td>-</td>
<td>-</td>
<td>303 (100.0)</td>
</tr>
<tr>
<td>Corrected and understandable URN’s incompletely matchable to SNOMED</td>
<td>5 (2.5)</td>
<td>11 (3.1)</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td>Total understandable URN’s matchable to SNOMED</td>
<td>187 (97.4)</td>
<td>346 (95.8)</td>
<td>197 (99.0)</td>
</tr>
<tr>
<td>Sample total</td>
<td>290 (100.0)</td>
<td>400 (100.0)</td>
<td>303 (100.0)</td>
</tr>
</tbody>
</table>

Note: The oftetanus toxoid (since frequency of occurrence of this URN is greatly disproportionate to the relative incidence of the disease in the US) but is apparently reasonably coded as the disease of tetanus. A total of 108 of the sampled URN’s (11.9%) remained unresolved even with expert manual review and coding, largely due to due inability to determine the original meaning of the URN. This was generally related to the entry of ambiguous abbreviations, severe spelling errors (for example, “vertiligo” probably representing vertigo vs. vitiligo), or purely numeric data, which likely reflected ICD-9 or CPT codes.

Post-coordination was required for many of the URN’s for which manual review and coding resulted in accurate matches of the main atomic concepts. A much smaller number of URN’s could be represented by both a single pre-coordinated term and by multiple atomic elements requiring post-coordination. We assessed for the presence of necessary linking semantic concepts in 117 URN’s requiring post-coordination taken from the manually reviewed, randomized sample of 400 URN’s. In this group 89% of appropriate semantic links were found. If this is extrapolated to the remainder URN’s, it is estimated that, while 95.8 – 99% of major conceptual atoms can be found, only 85-89% of these have linking concepts to fully describe the entire, often complex, URN concept. For example, “lower extremity weakness” can be matched to “entire lower limb” (182281004) and “asthenia” (13791008) but a full association between these two atoms requires the addition of a linking concept such as “finding site”.

Table 3. Examples of manually assessed accuracy for URN terms with identical SNOMED code output from both tools.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>URN</th>
<th>SNOMED concept ID</th>
<th>SNOMED descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial (broad)</td>
<td>Cecal diverticulitis</td>
<td>307496006</td>
<td>Diverticulitis</td>
</tr>
<tr>
<td></td>
<td>Fasting labs</td>
<td>16985007</td>
<td>Fasting</td>
</tr>
<tr>
<td></td>
<td>SP BTL</td>
<td>237679004</td>
<td>Status post</td>
</tr>
<tr>
<td>Incorrect</td>
<td>Fatigue</td>
<td>256674009</td>
<td>Fat</td>
</tr>
<tr>
<td></td>
<td>Tetanus (vaccine)</td>
<td>76902006</td>
<td>Tetanus (disease)</td>
</tr>
<tr>
<td>Unknown</td>
<td>Hem</td>
<td>57813001</td>
<td>Heme</td>
</tr>
<tr>
<td></td>
<td>EP</td>
<td>50762005</td>
<td>IL-1</td>
</tr>
<tr>
<td></td>
<td>Tuber</td>
<td>75505008</td>
<td>Tuber (root)</td>
</tr>
</tbody>
</table>

Discussion

It has been reported previously that common reasons for failure to match VA URN’s with coded concepts in the VA Lexicon included use of numeric data or modifiers [7]. Similar types of problems also limited matching SNOMED codes in this study. It is also notable that a significant proportion of URN’s remained unmatched even after expert review. Obviously, the ability to appropriately codify these types of data depends upon entry of accurate and unambiguous terms into the database, in addition to the existence of matching coded concepts.
Redundancy in SNOMED concepts was an interesting finding and was illustrated by the tool outputs. The TermWorks tool provided a single best pre-coordinated code while Opticode parsed each URN into more atomic elements. There were 104 out of 303 URN’s (greater than 30% of the “failed” match category) that remained unresolved even after expert review. Clearly the quality of the data is an important factor explaining match failures in this category. There were several instances where each tool provided different codes but taken together the codes gave conceptually identical results. For example, colon cancer (SNOMED-CT code 363406005) would be considered equivalent to adenocarcinoma (35917007) of the colon (302508007). In general, more complex the term, the more theoretical synonymous combinations will be possible. In addition the original language of the term will likely affect automated coding and possibly manual coding as well. A Billroth II operation could be considered verbal shorthand for an arguably more precise or specific antrectomy (or distal gastrectomy or partial gastrectomy) with retrocolic (or antecolic) gastojejunostomy. While an abdominal surgeon may understand the nuances of the elements contained in the latter term a clinician in a different specialty might not. As a result, each clinician may code such terms slightly differently, relative to their particular understanding of the term. The use of pre-coordinated rather than more atomic concepts leads to issues surrounding the benefits and hazards of combinatorial vocabularies [13]. Currently SNOMED allows both strategies and this may create additional complexity in extracting concepts for use in population based research, decision support and other clinical purposes. Additionally, the use of post-coordination requires the presence of appropriate semantic linking concepts and our results in a small sample show that 89% of these necessary concepts do exist. However, this also implies that in 11% of cases the entire meaning of the URN cannot be fully described by the main atomic elements without a linking concept that does not appear to exist in SNOMED at this time. Further study is necessary to determine if these rates are representative of other narrative sources and clinical domains. There are several limitations of this study. The existence of linking concepts was only evaluated in a small sample of URN’s (117 terms requiring post-coordination) and further study will be needed to validate the results based on these data. Additionally, the high frequency of spelling errors and jargon in the unresolved narratives reduced the effectiveness of the automated matching tools and decreased our ability to assess their overall value. However, in data subsets having standard clinical language these tools appear promising.

Conclusion

This study demonstrates that greater than 85% of VHA’s unresolved narratives can, in fact, be coded using SNOMED-CT. It is likely that more structured sources such as orders and operative or progress notes would have similar or perhaps even better coverage by SNOMED. This suggests that the addition of SNOMED-CT coding and semantic relationships to the VHA Health Data Repository will allow considerably greater utility than currently exists.

Acknowledgements

Drs. Lincoln, Penz, Nguyen and Sims gratefully acknowledge support from the Department of Veteran Affairs and the VHA Salt Lake City Office Field Office.

The Clue-5 Browser was obtained by download from (http://www.clinical-info.co.uk/), The Clinical Information Consultancy, Reading Berkshire, UK.

References


Address for Correspondence
Janet FE Penz, MD,
Department of Surgery,
Veterans Administration Medical Center, 500 Foothill Drive,
Salt Lake City, UT 84148