Cross-Mapping APACHE IV “Reasons for Intensive Care Admission” Classification to SNOMED CT

F. BAKHSHI-RAIEZ, R. CORNET, N. F. de KEIZER
Department of Medical Informatics,
Academic Medical Center, University of Amsterdam, the Netherlands

Abstract. The APACHE IV classification is used to capture diagnostic information for calculation of mortality risks in intensive care (IC). The lack of structured and formal definitions for concepts in APACHE IV classification as in any classification results in shortcomings when scaling up for re-use. The use of SNOMED CT as a reference terminology can address these shortcomings. However, all of SNOMED CT contains large amounts of information that is irrelevant for IC. By building an interface terminology (IfT) based on SNOMED CT and APACHE IV, it is possible to isolate the IC users from the complexity of SNOMED CT while enabling standardized data registration. Within this study, a mapping is realized from the APACHE IV classification to SNOMED CT. The results of the mapping will be used to identify a relevant SNOMED CT subset for the development of an IC-specific IfT. The vast majority of the diagnostic categories in APACHE IV could be mapped to one or more SNOMED CT concepts (83.8%) and for the remaining concept a partial match was identified (16%). The good mapping results will provide a SNOMED CT subset sufficient for developing an IC-specific IfT. Finally, lessons learned in this study are valuable for other researchers who intend to realize a mapping from a classification to SNOMED CT.

Keywords: SNOMED CT, APACHE IV, mapping, reference terminology, interface terminology, intensive care

Introduction

Healthcare providers documenting clinical encounters are increasingly using classifications to categorize patients with certain diseases or treatments. For instance, calculation of mortality risks in Intensive Care (IC) requires diagnostic information which is captured using the APACHE IV reasons for IC admission classification (APACHE IV classification) [1]. A classification is defined as a systematic arrangement of concepts into classes or groups based on common characteristics [2]. The lack of structure and formal semantic definitions for concepts in most classifications results in shortcomings when scaling up for re-use for multiple purposes. Accordingly, classifications such as APACHE IV generally serve a specific purpose within a certain medical domain and are mostly unsuitable for the broad registration of daily care processes.
Reference Terminologies (RT) are often used to address these scaling and re-use issues. A RT is defined as a set of concepts and relationships which provide a common reference point for comparisons and aggregation of data about the entire healthcare process and enables consistent and computer-readable coding of clinical data, which is a central feature for the use and re-use of information [3;4]. SNOMED CT is regarded as by far the most comprehensive reference terminology for coding clinical data [5]. SNOMED CT is comprehensive on its own, but also maps to other medical terminologies and classification systems already in use. This avoids duplicate data capture, while facilitating enhanced health reporting, billing and statistical analysis.

In the IC, for instance, SNOMED CT can be used to capture diagnostic information which will not only aid the calculation of mortality risks, but will also facilitate sharing and aggregation of data for other purposes. However, in case all of SNOMED CT is provided to the users, its comprehensiveness forms also its disadvantage in that it contains large amounts of information that is irrelevant for most of the specific clinical disciplines and would, therefore, be very cumbersome to use. IC users, for instance, will only be interested in a small fraction of the concepts included in SNOMED CT. Furthermore, the extensiveness of SNOMED CT for all kind of medical domains does not guarantee that it is sufficient for the collection of detailed information for daily patient care in a specific clinical field such as IC.

Therefore, the use of all of SNOMED CT as an interface terminology (IfT) in specific clinical settings is the subject of discussion in many studies [5-7]. An IfT is a terminology used for systematic collection of clinical data that supports direct entry of these data into electronic medical files. IfTs aim to facilitate the display and collection of clinical data to users in a simple way while simultaneously linking the users' own terms to structured data elements used by specific reference terminologies [4].

We aim at developing an IfT based on SNOMED CT and APACHE IV classification for IC. The local IfT to SNOMED CT will make it possible to overcome the scaling and re-use problems related to APACHE IV classification by taking full advantage of SNOMED CT’s semantic structure and description logics. On the same time it becomes possible to isolate the IC users from the complexity of SNOMED CT whose terms and language (i.e. English and Spanish) may be inappropriate and insufficient for the IC and may have an arbitrary level of details. The local IfT will initially be used by clinicians to capture diagnostic information which will be used to derive the APACHE IV diagnostic categories for calculation of mortality risks. Later on the IfT will be gradually expanded to serve multiple purposes such as registration of clinical data for medical record systems and research.

This paper describes the mapping from APACHE IV classification to SNOMED CT. Mapping is defined as “linking the content of the two systems through semantic correspondence” [2]. The results of this mapping will be used to identify the relevant subset in SNOMED CT (i.e. the group of concepts, descriptions, qualifiers and/or relationships relevant for the APACHE IV diagnostic categories) as a first step in developing the IC-specific IfT.
1. Materials and Methods

1.1. SNOMED CT

SNOMED CT is the world's largest concept-based terminological system containing 376,046 medical concepts, associated with 1,060,424 description terms for these concepts, and related to each other by a hierarchy consisting of about 1,359,435 relationships (July 2007 release) [8]. Each concept is uniquely identified and can have multiple descriptions. SNOMED CT has a dynamic organization and is compositional (i.e. it supports ‘post-coordination’). Relationships are used to define concepts and to specify how they can be refined or qualified.

1.2. APACHE IV Reasons for Admission Classification

The APACHE IV prognostic model is increasingly applied within the field of IC to adjust observed raw mortalities for case mix differences (severity of disturbance in physiological parameters and the primary reason for IC admission) in order to assess the quality of health care [1]. The relating primary reasons for IC admission are collected based on the APACHE IV classification, which contains 445 diagnostic categories belonging to 116 divisions [1]. Each reason for IC admission is first classified as non-operative or post-operative, next by body system or a transplant or trauma-related category, and then by diagnosis. A residual “other” category is used for unlisted diagnoses within each body system, transplant, and trauma category.

1.3. Mapping Procedure

Each APACHE IV diagnostic category was mapped to SNOMED CT concepts. Composite diagnostic categories (i.e. containing more than one diagnosis such as “chest/spinal trauma”) are split in atomic diagnoses which are then mapped to more than one SNOMED CT concepts. On the other hand, some composite diagnostic categories contain repeating atomic diagnoses. “Chest injury”, for instance, is part of 16 diagnostic categories in trauma group. Each such atomic diagnosis was only once mapped to SNOMED CT.

SNOMED CT was searched by the first researcher (FR) experienced in SNOMED CT and APACHE IV classification to find matches for diagnostic categories from the APACHE IV classification. The diagnostic categories which could not be mapped by FR were also searched for by a second researcher (RC), a SNOMED CT expert. In case of unclear definitions for the diagnostic categories, the researchers consulted an intensivist. The concepts in SNOMED CT were navigated using Clue browser version 2006.2.30 [9].

Mapping consisted of three consecutive activities: 1) Interpreting and analyzing the meaning of diagnostic categories. 2) Matching one APACHE IV diagnostic category to one or more SNOMED CT concept(s). The diagnostic categories were first matched to pre-coordinated concepts. In case no pre-coordinated match was available, a post-coordinated match was searched for. Concepts that did not exist in SNOMED CT were eventually matched to the appropriate superordinates. 3) Assessing each matched concept-category pair on how well they matched by marking each concept-category pair as “complete match”, “non-match” or “partial match” (i.e. matches to superordinate concepts).
Table 1: Examples of mapping types

<table>
<thead>
<tr>
<th>Examples</th>
<th>APACHE IV diagnosis category</th>
<th>SNOMED CT concept(s)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete match pre-coordinated</td>
<td>Chest/spinal trauma</td>
<td>262525000</td>
<td>chest injury</td>
</tr>
<tr>
<td>Complete match post-coordinated</td>
<td>Surgery for subdural hematoma</td>
<td>410771003</td>
<td>surgical procedure for clinical finding and/or disorder</td>
</tr>
<tr>
<td>Partial match to a superordinate concept</td>
<td>Coronary Artery Bypass Graft (CABG), redo</td>
<td>232717009</td>
<td>coronary artery bypass graft</td>
</tr>
<tr>
<td>Partial match for category “Other”</td>
<td>Other medical respiratory disorder</td>
<td>50043002</td>
<td>disorder of respiratory system</td>
</tr>
<tr>
<td>Non match</td>
<td>Other medical neuromuscular disorder</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Results

The 445 diagnostic categories in APACHE IV classification were mapped to 397 atomic SNOMED CT concepts. Table 1 provides some examples of the different mapping types. Table 2 presents the results of the mapping. SNOMED CT provided complete matches for 83.8% of the diagnostic categories and partial matches for 16.0% of the diagnostic categories. Only for one diagnostic category, “non match” could be identified.

3. Discussion

The aim of this project was to develop a mapping from APACHE IV classification to SNOMED CT. For 83.8% of the diagnostic categories in APACHE IV classification a complete match was identified in SNOMED CT. This percentage is in accordance with other studies which showed that SNOMED CT has a coverage up to 90% for various medical domains [6;10-12].

As pointed by other authors, the use of post-coordination resulted in better matching scores [13], 31% of all complete matches was realized through post-coordination. In this study post-coordination was based on the semantic possibilities of SNOMED CT representation and was not restricted to functional limitations of the Clue interface. For 12 diagnostic categories (e.g. “surgery for complications of previous spinal cord surgery”) it was semantically possible to post-coordinate them in...
SNOMED CT. However, the functionalities of Clue interface did not allow that. Since the results of our mapping aim to identify the relevant subset in SNOMED CT which is independent of the Clue interface, these concepts are marked as complete matches.

The diagnostic categories including the word “Other” (e.g. Other cardiovascular or Other Respiratory- medical) are residual categories in all body systems for conditions that can not be allocated to more specific categories. These categories form the largest part (47.8%) of the “partial match” group. The remaining diagnostic categories for which a partial match is identified are mostly named identities such as assessment scales or medical devices. We believe that these partial matches are also of great importance for the identification of the relevant SNOMED CT subset to develop a sufficient IfT for IC.

Previous studies on mapping a classification to SNOMED CT have shown that the mapping can be influenced by the structure and content characteristics of both systems [2;13]. In the APACHE IV classification some categories are provided with classification rules. The category “Respiratory Arrest” for instance is refined with “without cardiac arrest”, while the category “Cardiac arrest” is refined with “with or without respiratory arrest; for respiratory arrest only see “Respiratory arrest”. While these kind of rules are used in classification systems to make clear what should and what should not belong to a class, they are not included in terminological systems such as SNOMED CT. Since the results of our mapping will be used to identify the relevant subset in SNOMED CT for an IfT we ignored these classification rules included in APACHE IV and included every atomic concept from SNOMED CT in our mapping.

Furthermore, as in most classifications, the semantics of the diagnostic categories in APACHE IV are not formally defined, which makes mapping to other systems difficult. For instance, from the structure of the APACHE IV classification it is not clear what the exact definition of the diagnostic category “Non operative admission heart transplant” is. It can be mapped to the “Planned operative procedure for heart transplantation” which is a preoperative concept or to “cardiac transplant disorder” which is a concept meant to describe possible complications of a previous heart transplantation. Expert consultations revealed that both are possible and therefore, both mappings were included in the current mapping.

Table 2: Results of the mapping between APACHE IV classification and SNOMED CT.

<table>
<thead>
<tr>
<th>Number of categories</th>
<th>Complete match (%)</th>
<th>Partial match (%)</th>
<th>No match (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular</td>
<td>101</td>
<td>54.5</td>
<td>22.8</td>
</tr>
<tr>
<td>Gastro-intestinal</td>
<td>57</td>
<td>52.6</td>
<td>31.6</td>
</tr>
<tr>
<td>Genitourinary</td>
<td>35</td>
<td>45.7</td>
<td>14.3</td>
</tr>
<tr>
<td>Hematology</td>
<td>18</td>
<td>72.2</td>
<td>11.1</td>
</tr>
<tr>
<td>Metabolic</td>
<td>18</td>
<td>88.9</td>
<td>0.0</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>20</td>
<td>75.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Neurologic</td>
<td>54</td>
<td>66.7</td>
<td>20.4</td>
</tr>
<tr>
<td>Respiratory</td>
<td>46</td>
<td>67.4</td>
<td>21.7</td>
</tr>
<tr>
<td>Transplant</td>
<td>24</td>
<td>45.8</td>
<td>41.7</td>
</tr>
<tr>
<td>Trauma</td>
<td>72</td>
<td>47.2</td>
<td>50.0</td>
</tr>
<tr>
<td>Total</td>
<td>445</td>
<td>57.8</td>
<td>26.1</td>
</tr>
</tbody>
</table>
4. Conclusion

This study was performed as the first part of a larger study, designed to develop an IfT based on SNOMED CT and APACHE IV classification for IC. The aim of this paper was to describe the mapping from APACHE IV classification to SNOMED CT.

The vast majority of the diagnostic categories in APACHE IV could be mapped to one or more SNOMED CT concepts and for the remaining concept a partial match was identified. We believe that these good mapping results will provide us with a SNOMED CT subset sufficient for the IC-specific IfT. Furthermore, the findings of our study as described in discussion section provide valuable lessons for other researchers who intend to develop an IfT based on SNOMED CT starting with a predefined classification.

In the near future, we intend to build the IC-specific IfT based on this mapping. The IfT will initially be used to capture diagnostic information for calculation of mortality risks. After an evaluation in the IC setting, the IfT will be gradually expanded to a broad and richly-nuanced set of terms (and concepts) that accurately represent phrases and expressions occurring in the IC domain.

References


