Reviewer's report

Title: A SNOMED CT encoding method for palliative care

Version: 1 Date: 23 May 2010

Reviewer: Rachel Richesson

Reviewer's report:

This is a well-written and comprehensive description of a method for using SNOMED CT to code various types of data. The suggested methodology is very sound and logical, and undoubtedly would be a good resource for implementers using SCT in any number of domains. However, I think the paper needs some restructuring and focus to be readable as a research paper for the BMC Informatics audience. It currently reads more like a how-to guide or a text book. The detailed methods (which are all good methods) are described in the first 17 pages of the article, but as a reader I could not get a sense of where these methods came from - are they "best practice" (and whose), recommendations from IHTSDO, borrowed from some other implementation, generic methods that would be applicable to other terminologies outside of SCT, or... do these methods emerge from this pilot project planning/discussion/experience? If the latter is the case, then the authors have not really provided much detail on that project - including scope, purpose, quantity, and funding, etc. - until the results section.

I think this paper just needs to be reorganized a bit, which would allow some sections to be streamlined, and it would be much more readable and more valuable. I have the following suggestions:

Major Compulsory Revisions

*Provide a clear and stated purpose for the use of SNOMED CT in this dataset. Because of the early discussion of data elements, coded values & free text values, I get the sense that the goal is to use SNOMED as completely and thoroughly as can - in a sense it seems like you wanted to use it for any and all and at whatever level of abstraction. The mention of SCT inappropriate for dates and audit concepts seems a given - not even worth mentioning really. If the purpose were defined up front (e.g., "use SCT to represent coded values for dx problems" or "use SCT to represent xx data elements" or "...xxx parts of a patient record") this would allow you to cut much of your text, and would allow the reader to see what you were doing and why (i.e., the VALUE and SIGNIFICANCE of what you are doing.)

*Move the details about the Palliative Care data set, along with the motivation and objectives of the pilot to the introduction. This is the context of your work and the reader needs that context right up from, I believe. The methods could also be considered background (if they are documented elsewhere) and your methods section could talk about what pieces of those methods that you applied where
and how on your database. The reason I say this is because the current description of methods - which is comprehensive, logical and well-written - seems abstract and removed from your data set. There is no reference to specifics of your data on the methods. The generic method that you describe could also be shortened considerably. This is a very long manuscript with many figures.

*Keep in the results section the characterization of what you found. But the numbers about how many of each type of term you looked at should be in background - that in essence is part of your starting "problem".

*Consider clarifying the term "interface terminology". I see it used a lot in methods, but only on p. 23 do you state that the interface terminology could be implemented to standardize initial coding. Was that a goal of this project? Is that an option to implement in the system? If that is a goal, then that should be specified early in the paper. I had found that term confusing especially in the figures.

[I made a similar note in Fig. 3. I loved the figure, but then the box on interface terminology confused me, because I thought - was that the purpose of the project - a data-driven approach to defining an interface terminology? On a same note, I love Figure 2, but would remove the "Interface Terminology Set" completely.]

*The term "candidate terms" was also confusing. As you describe, the "candidate terms" include data elements, coded values, and free text. I think deciding what to code would be determined as a pre-requestite of the project, not generated by the nature of the data. Regardless, for section 2.1 you might re-title it to "Indentifying Scope of Project/Target Components/Data Entities for coding".. something like that... The use of 'terms' confused me some because you are talking about raw or source data, and terms are in SNOMED, and it just adds to confusion in my mind.

*Similarly "Extracting Coded Values" (p.5) could really be called 'mapping'. (On that note, you describe a lot fo NLP but never refer to it by name. That might be helpful, and pointing to other works on NLP could allow you to reduce the length of your description of some well-established methods.

[Another confusing term was "combining candidate terms" (section 2.1.2.4). Itruly love your mention of prioritizing the workload by frequency of use, but I do not understand what you are combining. I think you are just counting, right?

With all of this said, I think all of the content is in this paper and the methods, discussion, and conclusion are right on the mark! I strongly believe that this is a valuable contribution and would recommend it for publication if it could be reorganized. The idea of a general method is great, but I think it has to first be presented in the context of the work, then describe your experience as the results, and then generalize it in the discussion - or point to why it is generalizable. But you have to ground the method in some context first.

I think all the content is there, and just needs to be reorganized. I consider this
major, so I had to select that radio button option regarding acceptance. But with some reworking of context and focus, I think this would be a valuable and solid paper.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.