Response to reviewers (24 Jan 2018)

In our responses to the reviewers below we have responded to each point separately where possible. In many cases the reviewer's suggestions have been implemented in our revised manuscript.

Ms. No.: JBI-17-720
Title: Mismatches between major subhierarchies and semantic tags in SNOMED CT
Corresponding Author: Dr. Jonathan Bona
Authors: Werner Ceusters, MD

Reviewers' comments:
We have received three reviews of the manuscript that widely differ in their evaluations. Reviewer 1 and Reviewer 2 are positive, with the exception of small numbers of concrete improvements that they are suggesting. Reviewer 3 is critical of a number of aspects of the manuscript. Some of the comments of Reviewer 2 and Reviewer 3 may be addressed by adding a section on "future work."

An effort should be made to address Comments 3 to 6 of Reviewer 3. This should be done as editorial work on the manuscript as opposed to embarking on new research. Some of it may be placed in the "future work" section suggested above. Comment 4 is particularly "sensitive" and in fact is a whole cluster of comments.

These points are addressed with edits in the resubmission and in the response to Reviewer 3.

Both Reviewer 1 and Reviewer 2 point out a few English corrections. To these may be added: Capitalize Python. The authors might consider checking and/or rephrasing/simplifying the following sentences. Page numbers here refer to "PDF pages" not manuscript pages.

The resubmission capitalizes 'Python' and implements all of the editor's others suggestions below.

Page 17: Possible codes for mismatched concepts are 'NYY' (in which case the mismatched concept has at least two ancestors one of which is mismatched and one isn't) and 'NYN' (in which case no ancestors of the mismatched concept are themselves mismatched) while for non-mismatched concepts 'YYN' and 'YYY'.

(1) whether at least one of their relationships suggest that the concept qualifies to be a disorder despite it not being classified as such.

Page 18: 'Due to (attribute)' with any destination which is either itself in the disease sub-hierarchy, or suggest explicitly a disease without being classified as one, e.g. 'Systemic disease (finding)';

Page 20: Two are present initially but are removed later. [Consider changing the tense.]

Page 21: Table 7 provides - as an example restricted to the concepts that were ever mismatched and at least in one version tagged as 'context-dependent entity' - how the sort of transitions displayed in Table 6 evolve over time.

Page 22: A third observation is that during the period from July 2007 till July 2008 mismatches are by large primarily of the NYY-type.

Page 25: Each phase correspond to a specific semantic tag (ST) /mismatch configuration (MMC) or inactive/de-activated status.

Reviewer #1: This paper presented a quality assurance method for auditing semantic tags of SNOMED CT concepts. This method was applied to SNOMED CT releases from January 2003 to January 2017. The results showed that there exist concepts whose semantic tags do not match their placements in the hierarchy over time. This authors performed root cause analysis of mismatched disorders, and suggested that the SNOMED CT authoring environment should be equipped with a formal mechanism, which is capable of suggesting stated relationships for very similar concepts, in order to keep the hierarchy in sync with semantic tags.

Overall, the manuscript is well-presented with sufficient details. The topic is suitable for the special issue on quality assurance of biomedical terminologies and ontologies. The reviewer has the following comments to further improve the paper.

The reviewer suggests the authors to move the subsection "2.4 Root-cause analysis for mismatched disorders" to the subsection "3.3 Mismatched disorders", since the method for detecting mismatches is general, while the root cause analysis was mainly for concepts with semantic tag "disorder".

We preferred to keep it in the current place since it is a description of a methodology after all, be it used only for a specific part of the analysis.

Section 1:
- "when in the user interface of, for example, an electronic healthcare record system concepts are presented without showing the entire hierarchy" => "when in the user interface of, for example, an electronic healthcare record system where concepts are presented without showing the entire hierarchy"

This has been changed in the resubmission.

- "a FSN" => "an FSN" (two places)

This has been changed in the resubmission.

- It is a bit hard to follow the sentence: "Further confusing the question of what exactly counts as a semantic tag are consistently-used strings that appear as substrings within parentheses near the end of some FSNs that make it look at first glance as if those concepts might have multiple semantic tags." Please rephrase it.

This sentence has been split in two and reorganized for clarity in the resubmission.

** Section 3.2:*
- "Table provides a condensed view of the results by eliminating the semantic tags with no mismatches at all": the table number is missing.

This has been changed to read "**Table 4** provides …"

- Table 6 caption: "they were was changed" => "they were changed"

This has been changed in the resubmission.

** Section 3.3:*
- It is a bit hard to follow the paragraph explaining Table 8. Please rephrase the paragraph.

This section has been revised for clarity.

- Table 9: It seems that the row information is missing (it is not clear what each row represents).

A paragraph has been added to explain the table more in detail and an extra column has been inserted.

** Section 4.2:*
- "as are child concepts other concepts in the figure": please fix the grammatical error.

This has been changed to read "as are child concepts **of** other concepts …"

** Section 4.3:*
- "This transition is illustrated in Figure 4, which shows the relevant concepts in 2015 and in 2015" => "This transition is illustrated in Figure 4, which shows the relevant concepts in 2015 and in 2016"

This has been changed in the resubmission.

Reviewer #2: This manuscript investigates the consistency of semantic tags in SNOMED CT fully specified terms, which are supposed to be indicative of the broad semantics of a concept, with the hierarchal structure of SNOMED CT IS-A relations. More specifically, the authors map the semantic tags to high-level SNOMED CT concepts and check whether a concept with a given semantic tag is found in the set of descendants of the corresponding high-level concept. Additionally the authors analyze change over time for this phenomenon (across subsequent versions of SNOMED CT). A limited number of discrepancies were identified (89 among the 340k concepts in the latest version of SNOMED CT) and the detailed analysis of their evolution over time is provided. The authors make recommendations to the developers of SNOMED CT for avoiding such errors and provide a broader analysis of inconsistent semantic tags among similar concepts.

This analysis of semantic consistency between semantic tags and the hierarchical organization of SNOMED CT concepts is generally interesting and squarely relevant to this special issue on quality assurance in biomedical terminologies. The manuscript is well written, not overly technical, and accessible to a large audience. Although a limited number of errors have been identified, and even though the proposed method is relatively straightforward, the detailed analysis and the recommendations provided make this paper interesting and practically useful. This reviewer only has a small number of minor suggestions detailed below. Overall, this is a solid contribution to this special issue.

Minor comments
- The semantic tags and the corresponding concepts can be seen as representing "intension", while the set of concepts having a given semantic tag and the set of descendants of the corresponding concept represent "extension". This work could have been framed as comparing 2 different extensions for the same intension. As such, it is reminiscent of earlier work comparing UMLS concepts, which were assigned a given semantic type, to the set of descendants of the Metathesaurus concept corresponding to this semantic type.
See: Bodenreider O, Burgun A. Aligning knowledge sources in the UMLS: methods, quantitative results, and applications. Stud Health Technol Inform. 2004;107(Pt 1):327-31. PubMed PMID: 15360828; PubMed Central PMCID: PMC4303371.

This interpretation of ‘extension’ in terms of descendants rests on a confusion of subtypes with instances: extension is standardly defined as the collection of instances that satisfy the intension. If an ontology would have the concept ‘Named star’, subsuming both ‘Evening star’ and ‘Morning Star’, then this paper would argue the extension of ‘Named star” to consist of 2 entities. According to the standard definition, there would be only one entity, since the extension of Evening Star and Morning Star is in both cases just Venus, thus one entity. To avoid confusions of this sort, and since we did not work with extensions in the standard sense at all, we prefer not to mention this paper or its approach at all in our revision.

- The introduction should mention how SNOMED CT hierarchies are created (i.e., automatically by a classifier, as the byproduct of description logic definitions).

We have added a sentence and reference on this to the introduction.

- The authors could have investigated another interesting hypothesis: All the descendants of a concept corresponding to a given semantic tag should exhibit this semantic tag (or one of its descendants).

Thank you for the suggestion. This question is complicated by multiple inheritance in SNOMED CT, and by the absence of an official semantic tag hierarchy. We have added section 5, Future Work, which mentions this possibility.

- In the figures (e.g., first Fig. 1), the direction of the arrows is inconsistent with the semantics of the relations -- IS-A is expected from child concept to parent concept.

We have corrected the direction of arrows in Figures 1, 3, and 4.

- It is unclear what is expected from RDFS inference applied to the triple store (since this work is already based on the inferred version of SNOMED CT), probably just creating the transitive closure of subClassOf relations. Please clarify and summarize what RDFS inference produces.

We have modified the text in section 2.2 to clarify that the precomputed subClassOf transitive closure is the main contribution of RDFS inference here.

- In section 4.3, the most obvious and directly applicable recommendation is presented last ("second shortcoming"). This reviewer would recommend reversing the order. Moreover, while interesting, it is not entirely clear how the first part relates to the analysis of semantic tags.

We have reversed the order of these two points in that section (4.3), and edited to clarify how the absence of a mechanism to suggest states relationships for similar concepts relates to our analysis of semantic tags.

Grammar and typos
- "Some of these concepts are second order concepts" -> second-order (missing hyphen)
- "semantic tag are consistently-used strings" -> consistently used (extraneous hyphen)
- The second figure is labeled "Figure 1"
- "the semantic tag turned out to be identical to the name of the corresponding concept modulo capitalization" -- "modulo" -> "ignoring"
- "Table provides a condensed view" -> Table 4
- "shows the relevant concepts in 2015 and in 2015" -> in 2015 and in 2016

The second of reviewer #2's Grammar and typographical error points was obviated by other edits made to accommodate reviewer #1's suggestions. The other suggestions here have been implemented verbatim.

Editor: An effort should be made to address Comments 3 to 6 of Reviewer 3. This should be done as editorial work on the manuscript as opposed to embarking on new research. Some of it may be placed in the "future work" section suggested above. Comment 4 is particularly "sensitive" and in fact is a whole cluster of comments.

Reviewer #3: The authors indicate a problem within the SNOMED CT management and use of semantic tags within the Fully Specified Names (FSN) of several SNOMED CT concepts. The authors highlight a phenomenon within the SNOMED CT releases in which the semantic tag can be inconsistent with the root node of the subhierarchy concept that first employs the semantic tag. The authors describe a method in which the FSN's for all SNOMED CT concepts across all international releases are analyzed and whether the semantic tag used in each FSN is consistent with the subhierarchy's root node tag.

The authors describe the computational methods employed to analyze the semantic tag question rapidly. The methods are quite involved and impressive from a computational vantage point, and the volume of analysis is extensive. A total of 466 concepts were identified over the course of all SNOMED CT releases between the years of 2003 and 2017 that had a semantic tag mismatch. The conclusion of the research is that SNOMED CT tooling is deficient because it apparently does not have the capacity to check semantic tag hierarchical and subhierarchical correctness.

While the computational methods and analyses employed by the authors are impressive, I find several issues with this research.

Comment 1: The findings are trivial in both scope and content. Given the number of concepts and relationships (roughly 300,000 concepts and 1,000,000 inferred relationships) within a single SNOMED CT release, much less over 15 releases, a cumulative error of 466 concepts is insignificant. Moreover, semantic tags are employed as a user aid for a human to disambiguate like terms (i.e., natural language descriptions) based on their top-level hierarchy. They are not used in the concept model to categorize or classify concepts. Furthermore, mismatches of semantic tags within the same top level hierarchy (i.e., one of the 19 concept hierarchies of the SNOMED CT concept model) do not represent a problem with significant consequences in terms of concept use or logical inference. If the semantic tag associated with the SNOMED CT concept is NOT aligned with the top-level concept hierarchy tag, then a problem may be present.
The example presented in Figure 3 represents a child concept and its semantic tag that is mismatched AND the tag of the child concept is associated with a different top level hierarchy than its parent concept. This is an error worth noting, but upon further review, the concept in question, Sickle cell test kit, is primitive with no defining attributes. Therefore, this was not a modeling issue, but merely a concept naming issue with minimal, if any, consequence in terms of classification of SNOMED CT content. (Note: it is a leaf node concept)
If there was substantial prevalence of semantic mismatches, this research could be of use. In particular, if the mismatches were between one of the 19 top level hierarchies, potential misuse of SNOMED CT concepts could be propagated. The issue is less significant for subhierarchy mismatches and devolves to an inconvenience for users (e.g., Finding vs. Disorder) vs. a classification and description logic issue.

We have added text to section 4.1 clarifying the significance of semantic tag errors for users of SNOMED CT.

Reviewer 3 is invited to reconsider the stance that detecting and preventing errors that would otherwise be implemented in health record systems responsible for keeping people alive and well is “trivial”. We provide arguments in the paper to the contrary.

Comment 2: The authors do not articulate why this is important research. Given the small number of semantic tag mismatches and the minimal degree of impact these mismatches have on the overall use and classification of the concept model, why was this work done? I have to believe that the findings presented were a by-product of another research effort and not an end in itself. Perhaps providing context to this research would help a reader understand its potential importance?

Section 4.4 “Related work” in the original submission mentions the context in which this work arose:

> “*It is a continuation of earlier efforts in which we examined patterns of semantic tag changes between releases of SNOMED CT and observed that certain change patterns occur frequently among certain subsets of the total set of semantic tags [7]*”

This text is retained in the revised manuscript.

We have also added a section on Future Work that gives more context on the larger body of work that this is a part of, and we have added text to the Conclusion and Section 4.1 that clarifies why it is important to detect and eliminate these errors.

Comment 3: The authors attempt to analyze SNOMED CT in light of a different concept model, OGMS, in section 3.3. This section is orthogonal to the research presented in this manuscript. If the authors would like to compare and contrast differing ontologies/terminologies, that can be done in a separate manuscript. The comments and conclusion seem out of place and disjointed from the balance of the contents.

This work is not a comparison between SNOMED CT and OGMS. OGMS is a realist ontology that makes a clear distinction between diseases, disease courses, and disorders. The distinctions are used in 3.3 to illustrate the point that tag mismatches appear to be related to confusion between these distinct categories of entities. We have added text to this section to clarify this point also in section 2.4 where we provided a motivation for this analysis.

Comment 4: The overarching conclusion is that SNOMED CT tooling is inadequate because the FSN and semantic tag can be mismatched with parent concepts. True, a feature in the tooling could be added to check for these anomalies, but given their low number and limited impact on the overall concept model, such efforts would be a misdirected use of resources unless all other SNOMED CT quality issues of consequence are corrected.

The authors have implemented computational tools that efficiently detect these errors. They are logically and computationally simple, and will be easy to reproduce in any reasonable tooling environment. We will gladly offer to assist SNOMED CT in implementing this tooling feature in order to minimize the already low amount of resources required to correct this issue. If the Reviewer 3 is aware of more consequential systemic quality issues in SNOMED CT that may be computationally detected and fixed, Reviever 3 is encouraged to share those suggestions with the authors.

Furthermore, the authors claim to have completed a root cause analysis. If so, what were their steps? How was the analysis performed?

This now explained in section 2.4

Did they discuss the situation with the editorial committee or staff at SNOMED International?

We didn't. We preferred to perform an objective independent analysis of the problems by only using the tools that non-privileged researchers have at their disposal.

Did they have access to SNOMED CT tooling? Are the authors aware that tooling has changed at SNOMED International within the past two years?

We do not have access to SNOMED CT tooling. These tooling changes have apparently not resulted in a reduction of the sort of errors we detect and describe in this manuscript. Automated quality control methods to detect errors are even more desirable when tooling changes are occurring than otherwise.

In their analysis, did the authors take into consideration the concept model changes and development efforts ongoing throughout the evolution of SNOMED CT?

We did. That has now been clarified in the paper and examples provided.

Did the authors take into account editorial policy and practice changes at SNOMED International?
We did. Already in the original submission, we had made several explicit references to the editorial policies document issued by IHTSDO.

A true root cause analysis would include investigating and identifying these details and more.

Information about the internal operations of SNOMED CT that is not publicly available was not available to the authors at the time of this work, or subsequently. This work is primarily an analysis of the published ontology,

not of SNOMED CT business practices, tooling, policies, etc, although those surely are related to the contents of the ontology. This comment contains good suggestions for future projects. The authors will be in touch with the SNOMED CT team with recommendations arising from the work described in this manuscript. We are happy to accept from the IHTSDO, Reviewer 3, or any other interested parties a grant enabling us to work further on the types of analyses suggested by Reviewer 3 here.

We have provided more explanations about our “root cause analysis”.

Comment 5: Tables are difficult to interpret and require a legend of meaning for the notation used throughout the manuscript. The authors' classification system of concept mismatches is unique and not readily retained while digesting the manuscript. A reader would benefit from some form of legend.

We have added more explicit legends to the tables where needed

Comment 6: While the authors provide many tables and figures, they are short on examples. Given the limited number of mismatched concepts, comprehensive list could be included in an addendum or supplemental materials section. This would enable the reader to determine if these mismatches are truly impactful and noteworthy.

We have added Appendix A, a complete listing of concepts that are mismatched in the January 2017 release of SNOMED CT.

Response to reviewers (Feb 13, 2018)

In our responses to the reviewers below we have responded to each point separately where possible. In many cases the reviewer's suggestions have been implemented in our revised manuscript.

Ms. No.: JBI-17-720
Title: Mismatches between major subhierarchies and semantic tags in SNOMED CT
Corresponding Author: Dr. Jonathan Bona
Authors: Werner Ceusters, MD

Reviewers' comments:

The authors have done a good job in reacting to the reviewers' comments.
Minimal work remains.

Comments on Annotated Version:

PDF page 58, line 19: which make it look -> which makes it look

> This has been corrected in the resubmission.

page 75, line 16: The use of the word "than" indicates an intended comparison, but the text surrounding it does not show such a comparison.

> Line 16 on page 75 is a figure. The only text in that figure is concept names. The word 'than' appears in a sentence on line 16 of page 76, and on line 16 in no other pages. The authors assume that page 76 is the instance of 'than' in question, and have edited that sentence for clarity.

Appendix A:

Please see the web page:

https://ihtsdo.freshdesk.com/support/solutions/articles/4000095393-why-did-you-change-your-name-from-ihtsdo-to-snomed-international-