Comparing Heterogeneous SNOMED CT Coding of Clinical Research Concepts by Examining Normalized Expressions

James E. Andrews[*],
School of Library and Information Science, University of South Florida, Tampa, FL

Timothy B. Patrick,
Health Care Administration and Informatics Program, College of Health Sciences, University of Wisconsin-Milwaukee

Rachel L. Richesson,
Pediatrics Epidemiology Center, University of South Florida, Tampa, FL

Hana Brown, and
School of Library and Information Science, University of South Florida, Tampa, FL

Jeffrey P. Krischer
Pediatrics Epidemiology Center, University of South Florida, Tampa, FL

Abstract

Objective—A continual problem confronting the implementation of standardized vocabularies such as SNOMED CT is that their expressive flexibility and power provide more than one way to represent a given concept. The goal of this study was to investigate how the CliniClue™ Expression Transformer tool could be used to help in discerning similarities and differences among three separate sets of clinical research concepts coded in SNOMED CT by three different paid expert coding companies.

Methods—Initial editing of the companies' coded datasets was required to enable accurate input into CliniClue Version: 2006.2.0030 Expression Transformer tool. The normal forms of the company codings for the 319 clinical research question/answer sets were compared to determine whether they were equivalent or otherwise related (e.g., if one was subsumed by the other). Basic frequencies were computed for (957) pairwise comparisons of each of 319 concepts each coded by the three expert coders, and the implications of the results discussed.

Results—The primary finding from this study was that, for each of the paired comparisons, approximately half of the time the companies' codings could be related, primarily via subsumption. The greatest percentage of equivalent concepts between any two companies was 33%. These same two companies also agreed most often on the core clinical concept measure from an earlier study by the authors.

Conclusion—Heterogeneity among coders using the same controlled terminology appears inescapable despite the extensive efforts of terminological standards developers and implementers. In our study, the computable determination of equivalence of discordantly coded concepts still failed

*Corresponding Author: James E. Andrews, Ph.D. School of Library and Information Science, University of South Florida, 4202 E. Fowler Ave., CIS 1040 Tampa, FL 33620, jandrews@cas.usf.edu, 813.974.2108, Fax: 813.974.6840.

to yield acceptably comparable data. A clearer articulation, and perhaps a simplification, of rules for the consistent use for terminologies such as SNOMED CT is needed.

Keywords

SNOMED CT (Systemized Nomenclature of Medicine); CliniClue Browser; Controlled Terminologies; Coding Consistency; Semantic Variation

---

INTRODUCTION

Controlled terminologies are intended to be standards for insuring consistent and accurate communication across a community of users. However, in practice discordant coding is not uncommon, and is becoming an important research area for determining if the implementation of controlled terminologies actually produces consistently coded data. This is particularly true regarding relatively complex terminologies such as SNOMED CT (SCT). While clinically-rich and near-comprehensive, the expressive power of SCT can be limiting since it provides more than one way to represent a given concept. This was the key conclusion in a previous study by the authors (JA, RR, JK), where a sample of clinical research concepts coded by three separate professional coding companies showed high levels of variance [1].

At some level, coding differences are to be expected and may be unavoidable. Language is inherently complex even when attempts are made at control or standardization. Individuals with different interests or training may use the same terminology to code the same concepts differently; alternatively, they may use the same terminology to code different facets of the same phenomena; finally, they may use different terminologies to code the same or different aspects of the phenomena. It has long been recognized that a fundamental problem in information retrieval is the reconciliation of such differences [2]. In this paper we are specifically concerned with the case where different coders use the same terminology but code differently. We are not concerned with the problem of reducing coding variation *per se*, but rather are concerned with the larger issue of whether the sameness or difference in meaning underlying different codings can be detected and perhaps reconciled algorithmically. Under controlled conditions, in which all coders utilize the same explicit coding instructions, there may be less coding variation than in more natural settings in which the different coders are located in different organizational contexts, are driven by different interests and goals, and are not bound by the same explicit coding instructions. Detecting and reconciling differences in meaning across data collected in diverse natural settings will be an important challenge for systems interoperability and data sharing, and the degree to which such detection is possible may depend on a number of factors. Nevertheless, for any integration of data collected in multiple distinct institutions, methods for the detection and resolution of semantic variation are critical.

Novice and expert users of SCT have a perception that SCT relationships already facilitate this kind of comparison. Although there are more than 350,000 concepts in SCT, many are not fully defined -- meaning that the collections of explicitly asserted relationships are necessary and sufficient to unambiguously define concepts. It would seem that the dominance of primitive (i.e. not fully-defined) concepts in SCT would limit the determination of concept equivalence of various post-coordinated expressions. Under the new International Health Terminology Standards Development Organization (IHTSDO), SCT is freely available to users from the 9 member countries, including the U.S. [3,4]. Along with the SCT data, a free-ware called CliniClue™ [5] also is freely available, and likely the most commonly used SCT browser.

The goal of this study was to investigate how the CliniClue Expression Transformer tool could be used to help in discerning similarities and differences among three separate sets of clinical

research concepts coded using SCT by three independent coding companies. This tool facilitated the transformation of the codes into normal expressions, and distinguished subsumption from equivalence. Though outside the scope of this preliminary study, our ultimate goal is the identification of specific methods for reconciling discordant coding.

BACKGROUND

The development and use of standards in health care is guided by the need to consistently and accurately represent and communicate information, often across heterogeneous contexts, to ensure effective storage and retrieval for various purposes [4]. Standardization of language, in particular, remains a major challenge in medical informatics due to the inherent complexities of language and the information involved. While major efforts have been underway for decades in controlled terminology development, consistent adaptation is by no means ubiquitous. In part, this is due to the fact that standards such as SCT, designed to be expressive and comprehensive, are complex to implement and use. Moreover, independent healthcare institutions may find it more expedient to use locally-developed terminologies, or those provided by technology vendors, for structured data entry and thus generally rely on proprietary information and communication methods and tools unless otherwise compelled (e.g. by federal regulations).

Still, SCT has emerged as the best candidate for representing a variety of clinical data, and is also likely to be a major standard in clinical research. In 2004, a collaboration of federal agencies seeking agreement on health care data collection, known as the Consolidated Health Informatics (CHI) initiative, named SCT as the standard to use for diagnoses, problem lists, anatomy, and procedures [4]. These various agencies have worked together in recommending the use of the best data standards in a variety of areas, and all federally-funded clinical research is obligated to follow the CHI-recommended data standards, including SCT [4]. However, there is still no consensus of use of SCT in local applications, and the underlying conceptual and logical models may not be sufficiently developed to determine equivalence across semantic variations in coding strategy [6].

There is a natural tension between language control and expressiveness; that is, the very features meant to enable complete and flexible expressions also lead to divergent representations from different users, even within the same or similar contexts. Arguably, an initial measure of the potential usefulness of a controlled terminology should be how well users are able to consistently express meaning. The surface level expression will likely vary from one user to the next depending on how they utilize or understand a terminology and its structure and rules, but the essential meaning (despite syntactical variation) should be demonstrably equivalent by applications utilizing the terminology. While consistent coding (syntactic and semantic) is the ideal, coding which can be deemed *computationally equivalent* should be a reasonable surrogate. In light of current findings and trends in development and implementation of controlled terminologies, an emerging focus seems to be to allow varying local coding styles and interfaces that nevertheless afford some level of interoperability.

An advantage of SCT is the numerous, clinically-rich atomic concepts and pre-coordinated concept expressions, and that it offers even more expressiveness through post-coordinating more complex expressions [7]. In the context of clinical coding, pre-coordination refers to those terms which are fully formed and assigned a code within the terminology (even if expressing a complex or compound concept); for instance (in SCT), *190419001|diabetes mellitus, adult onset, with other specified manifestation*. Post-coordination in terminologies such as SCT allows more complex concept expressions and relevant qualifiers to be expressed where pre-coordinated or single concepts are not available or sufficient. For example, in order to note the laterality of a hip replacement, the following coding would be needed: *52734007|*

total replacement of hip| 272741003|laterality|=7771000|left| [8, p. 60]. There are tradeoffs between these two approaches: pre-coordination is often easier for coders, and can ensure consistency of coding, and possibly of meaning. On the other hand, one can create quite elaborate post-coordinated expressions (allowing users the representation of detailed concept expressions while simplifying terminological maintenance) but risk data integrity if there is no demonstrable equivalence of meaning. Similar challenges have been noted in standard texts on thesaurus construction [2,9–10]. In particular, when the post-coordination power of a terminology increases to a great scale, as is the case apparently with SCT, then certainly we expect increase in variation when the coders are located in different organizational contexts and are driven by different interests and goals. It is therefore the case that (as logically developed, for instance, by Cimino [11]) post-coordination implies a pressing need for determining concept equivalence.

Determining Equivalence among SNOMED CT Expressions

If heterogeneous coding is more or less inevitable, an important activity is to determine if multiple representations of medical concepts (i.e., variant SCT concept expressions) are otherwise comparable. A coder's choice of how to construct a representation of a concept may be driven by various contextual and coder-related factors. For instance, a person engaged in structured data entry for an electronic medical record may seek a certain level of expressiveness that, while relatively rich in the sense of capturing clinical content, may not require the granularity of some other context, such as in a clinical research trial.

Other critical factors complicating heterogeneous coding come from the rules or guidelines for coding in a controlled terminology such as SCT. While it is necessary that users have the ability to post-coordinate concepts to form more detailed or explicit expressions, this is a process that is exacerbated by difficult to use or under-specified compositional rules within the vocabulary. Variation among coders is therefore to be expected (this is evident even among professional coders, as shown in Andrews et al. [1]) particularly where complex/compound concepts or qualifying values need to be coded. At the time of this study, a consumer-ready version of the rules for post-coordination in SCT have not been fully articulated. Most of this information is available in the SCT documentation, but it is not yet in a form that is readily consumable by those who may not have a detailed knowledge of the representational rules and models of SCT, but who nevertheless are charged with coding and/or structured data entry.

As noted, not all concepts need to be coded in exactly the same format to enable comparisons, including equivalence, and thereby support effective retrieval. Transforming SCT concepts into *normal forms* is one approach that has the potential to accommodate this [8,12,]. The idea of normalization is one found in a variety of contexts. It is based on the general idea that stripping statements of merely extraneous information can allow for like comparisons. For instance, in predicate logic differently stated assertions can be compared by reducing statements to a common form. Normalization is also necessary in such contexts as processing text strings (e.g. using word-stemming algorithms) and, of course, in relational database modeling. In the context of SCT, the documentation notes that, "The purpose of generating normal forms is to facilitate complete and accurate retrieval of pre and post-coordinated SNOMED CT expressions from clinical records or other resources" [8 p.5, *Transforming Expressions to Normal Forms*]. The goal is primarily to allow *close-to-user* expressions to be logically transformed into a *normal* expression (or a common, most proximate primitive form) so that more than one concept representation can be compared [8, p.5]. Specific details on the algorithms for transforming expressions into normal forms are outlined in the SCT technical documentation [8].

One way of utilizing this approach to facilitate more effective information retrieval was outlined by Dolin, Spackman, and Markwell [13]. Although an information model (HL7 RIM)

was used to further structure SCT codes, the results nevertheless seem promising for allowing what they have termed "selective information retrieval,' which requires a way to navigate broader or narrower concepts to include in retrieval sets depending on the query. Elkin et al. [7] extend such approaches to develop a formalism that is meant to enable distinguishing comparability among compositional expressions being communicated across systems. A key theme in these projects is that it is important to have the ability to algorithmically explore semantic relationships when comparing seemingly discordant expressions.

These and other studies suggest a key restriction to effective use of SCT and similarly complex controlled terminologies. That is, system designers and other terminology implementers must understand and be able to adopt and implement the underlying logical models within a complex vocabulary, and adjust for discordance by algorithmically determining equivalence or subsumption so as to reconcile variant expressions and generally assist data managers and other users concerned with information storage, retrieval, and management. Moreover, this should be something that crosses health contexts; for instance, by better enabling the translation of research data into a clinical care context.

DATA SOURCE AND METHODS

The goal of this study was to determine if utilizing existing SCT tools could help in understanding whether seemingly discordant coding of the same concepts can be semantically related. The central question was: once normalized, are these expressions equivalent, subtypes or supertypes of one another, or not formally related at all within the SCT structure?

The dataset for this study was derived from the aforementioned research project that sought to determine coding variance among three professional coding companies [1]. In that study, each company was sent a set of 319 randomly selected clinical research question/answer sets culled from Case Report Forms (CRFs) from observational and interventional research protocols in the Rare Disease Clinical Research Network (RDCRN)[14]. The companies were asked to code these clinical concepts using SNOMED CT (implemented using whatever tool or application they chose). A set of instructions were included, which: 1) outlined the purpose of the project, 2) sought to reduce unnecessary variability by requesting they code the concepts using a specified SCT "preferred hierarchy" (e.g., we would specify whether the question-answer pair "…" should be coded as a clinical finding or an observable entity), 3) requested comments by the coders in cases where they might find it difficult or impossible to sufficiently code certain items, and, 4) requested they use the SCT post-coordination syntax found in SCT documentation [8].

A key finding from that study was that, even when comparing only the "core clinical concept" (i.e. not including qualifiers or other attributes, since there was so much variation in how each company added such attributes) used by each company for each clinical research concept, there was a large degree of variation among the professional coders, which was the impetus for the study presented here. Despite variation in syntax (which we could transform to a common syntax) of post-coordinated and pre-coordinated codes, could we find a way to compare them in a normalized form?

Data Preparation

Initially, we found that the codings we examined, in the forms in which each company submitted them, were not comparable since the companies appeared to use different approaches. For instance, one company did not appear to intend to write SCT expressions, *per se* (in the sense of utilizing the compositional grammar), while another company appeared to write definitions, and the third appeared, in some cases but not all, to intend to write SCT

compositional grammar expressions. So, for example, for the question/answer pair, "Abdomen; Abnormal":

Company A wrote:

116680003|Is a (attribute)|=119415007|General finding of abdomen (finding),

363713009|Has interpretation (attribute)|=263654008|Abnormal (qualifier value)

Company B wrote:

119415007 General finding of abdomen (finding),

and, **Company C** wrote:

119415007 General finding of abdomen (finding)

263654008 Abnormal (qualifier value).

Company A's coding, though syntactically incorrect with respect to the SCT compositional grammar rules, is clearly intended to constitute a definiens. It is not clear whether Company B or Company C intended a SCT expression. For instance, it is possible that Company B (and the other companies, for other codes) may have felt that *General finding of the abdomen (finding)* sufficiently covered such explicit modifiers as "Abnormal." As another example, consider the question/answer pair, "Abdomen-Hepatosplenomegaly. No."

Company A wrote:

116680003|Is a (attribute)|=36760000|Hepatosplenomegaly (disorder),

408729009|Finding context (attribute)|=410516002|Known absent (qualifier value),

Company B wrote:

243796009|situation with explicit context|:

{246090004|associated finding|=36760000|Hepatosplenomegaly (disorder)|

,408729009|finding context|=410516002|known absent|

,408731000|temporal context|=410512000|current or specified|

,408732007|subject relationship context|=410604004|subject of record|},

and, **Company C** wrote:

36760000 Hepatosplenomegaly (disorder).

Again, Company A's coding, though syntactically incorrect, is clearly intended to constitute a definiens. Company B's coding is clearly intended to be a SCT expression -- and in fact is a legal SCT expression -- while Company C's coding is not a post-coordinated concept expression but rather a pre-coordinated concept.

Minor modifications were therefore necessary, meaning that some assumptions had to be made regarding the intent of each company. Essentially, we felt that only a minimum level of modifications to the codings should be done, based on consensus of the authors, but which we felt were necessary to allow for logical transformations and comparisons. The goal was not to change the coding, but to enable comparisons using CliniClue tools.

In light of these issues, and to allow comparison of the codings produced by the three companies, we used the following heuristic.

1. a coding in the form of a definition or definiens was converted to a legal SCT expression by:

    a. removing the initial "116680003|Is a (attribute)|=" for company A

    b. adding missing "|" characters around concept names,

    c. replacing the initial "," with ":".

The first above example is used for illustration:

116680003|Is a (attribute)|=119415007|General finding of abdomen

(finding),363713009|Has interpretation (attribute)|=263654008|Abnormal (qualifier value)

was transformed to:

119415007|General finding of abdomen (finding)|:363713009|Has interpretation

(attribute)|=263654008|Abnormal (qualifier value)|

The assumption was that this transformation was justified in that treating the attribute-value pair (or pairs) as a refinement of the initial concept would be consistent with the coder's original intention. For example, treating

363713009|Has interpretation (attribute)|=263654008|Abnormal (qualifier value)

as a refinement of the initial concept

119415007|General finding of abdomen (finding)

would be consistent with the coder's original intention. To be clear, however, qualifying values (such as "Abnormal") were not added to codes; rather, where such terms were present we simply made appropriate syntactical adjustments to allow for comparability.

1. A coding of a single concept ID and term such as:

119415007 General finding of abdomen (finding)

was converted to a legal SCT expression of the form ("*ws*" = whitespace)

ws conceptId ws ["|" ws term ws "|" ws] [3, p.58]

by adding the required "|" characters, as in

119415007|General finding of abdomen (finding)|

1. A coding consisting of a pair of a single non-attribute concept ID and term followed by a single attribute value concept ID and term, such as

119415007 General finding of abdomen (finding)

263654008 Abnormal (qualifier value)

was converted to a legal SNOMEDCT expression of the form

concept: attributeName ws "=" attributeValue [3, p.58]

by supplying an attribute name concept that appears consistent with the intent of the coding, as in

119415007|General finding of abdomen (finding)|:363713009|Has interpretation

(attribute)|=263654008|Abnormal (qualifier value)|

Methods

The company codings for all question/answer pairs were transformed utilizing the above heuristic. Our goal was to compare normal forms of each code to determine whether they were equivalent or otherwise related (e.g. if one was subsumed by the other). Thus, the resulting SCT expressions were compared pair-wise using the CliniClue Version: 2006.2.0030 Expression Transformer, with the following selections: *Component = Expression, Transform = Normal form - with context*, and *Display = Compositional (all)*. Basic frequencies and statistics were calculated using SPSS v16.0.

RESULTS

The data were formatted so that simple frequencies could be computed. Table 1 shows the percentages for each category of comparison for each of the pair wise comparisons (percentages are rounded to nearest whole number). Table 2 shows these same comparisons, modified to reflect only the cases where the companies had previously agreed (in the aforementioned study) on the "core clinical concept" (i.e. without regard to syntax or qualifiers or other additional SCT codes). As a measure of goodness-of-fit, Chi-square statistics were calculated for each of the paired comparisons. Each of these was shown to be statistically significant (all $p < .001$).

It should be noted that although the frequencies reported between Companies B and C were highest in terms of percent equivalent (in both Table 1 and Table 2), this is somewhat misleading. Of the cases that were equivalent (n=108), 86% were cases where both companies chose the same expression, and only 14% were classified as *computationally equivalent* by the CliniClue tool but initially expressed differently. The percentages in both tables for the "Number Equivalent" between "Co_B with Co_C" were based on the same 108 equivalent terms; yet, the percentage is higher in Table 2 given the smaller overall number of cases.

Some further examination of the data was useful for qualifying whether comparing normal forms was useful. In many cases, this exercise was indeed helpful in making comparisons. For instance, for the question/answer pair, "Chest and Lungs Specifications/No retractions," one company coded this as (after heuristic above applied):

> 67909005|Chest wall retraction (finding)|: 408729009|Finding context
>
> (attribute)|=410516002|Known absent (qualifier value)|

A second company coded this:

> 243796009|situation with explicit context|:
>
> {246090004|associated finding|=67909005|Chest wall retraction (finding)|
>
> ,408729009|finding context|=410516002|known absent|
>
> ,408731000|temporal context|=410512000|current or specified|
>
> ,408732007|subject relationship context|=410604004|subject of record|}

When compared using the CliniClue expression transformer tool, with "<no transform>" selected (i.e. no normalization of the terms for comparison), these two expressions were deemed unrelated. However, selecting "Normal form – with context" (per our method), they were shown to be equivalent. Examples such as this were common throughout the set.

On the other hand, regarding the higher percentage of equivalent matches between Companies B and C, it was more difficult to determine the usefulness of our approach. That is, as noted earlier, a relatively small percentage of these were concepts expressed differently by the companies yet determined to be equivalent when compared in normal forms. A closer look at

these revealed that nearly every one of these cases involved refinements involving body structure or finding site related to the core concept. For example, for the question/answer pair, "Mouth hemorrhage specify/Tongue,"

Company B coded:

> 22490002 | Bleeding from mouth (disorder)

and Company C:

> 22490002|Bleeding from mouth (disorder)|21974007|Tongue structure (body structure)|

The additional context qualifier of "Tongue Structure" added as a finding site to the primary concept did not change the relationship noted in CliniClue between the two expressions, or any of the others with similar structures. This may be because such qualifiers, in these instances, are subsumed by the primary expression, or perhaps due to syntactical errors that precluded comparison.

Discussion

The primary finding from this study was that, for each of the paired comparisons, approximately half of the time the companies' codings could be related, primarily via subsumption. This could imply that nearly half of the information coded by different experts could conceivably be lost (i.e. would not be retrievable).

Worse is the very low percentage of actual "equivalent" comparisons, particularly between Companies A and B, and A and C. Absent of any in-depth personal interviews of the coders, or other evidence of intent, one might still be able to speculate on a few reasons for this discordance. First, it could be argued that the data is not representative of the kind of information for which SCT was intended. Our dataset came from assessment items on Case Report Forms (CRFs) from actual clinical research studies, and is not from a clinical care context, *per se*. However, we did limit the culling of questions/answer pairs to those from forms used to record physical examination and assessments, which presumably could be used in clinical care. Still, our source data may reflect specific questions from clinical research and rare diseases, and so are not necessarily commonly coded concepts. Regardless of the generalizability of the results, our sample clearly represents concepts from a domain (human research in rare diseases) with a need for consistent and accurately coded data.

The greatest percentage of equivalent concepts appeared to be between Companies B and C, which are also the two companies that agreed most often on the core clinical concept measure from the earlier study. In fact, as noted, it turned out that of these a high percentage were exact matches. That is more encouraging to some degree since it shows that at least two of the companies reflected similar approaches during the times that they were in agreement.

A more obvious reason for the overall differences (i.e. lack of comparability) and lack of relatedness among the common forms of these expressions probably came from each company's general approach to coding with SCT, as well as some anomalies in the data. For instance, there were cases where one company chose a concept from a SCT hierarchy different than what was suggested in the original dataset (e.g. they chose an expression from the Observable Entity axis instead of Finding). Such cases led the normalized expressions to be reported as unrelated. This particular issue may illustrate a fundamental problem in polyhierarchical terminologies. Generally, our examination of the data revealed how easily, even at the initial level of classifying a concept, one person's perspective could be quite different from another's, perhaps due to stylistic or context-specific issues, or simply natural interpretive variances.

This study was motivated by a use case to integrate data from three distinct institutions for research purposes. Any future data integration endeavors would require detection of similarity and difference in meaning, as well as the resolution of difference. Specifically, our goal was to investigate how and if the manipulation of description logics that represent coded concepts could assist us in determining sameness or difference in the meaning of underlying SCT codings. To this end, because coding variation seems inescapable, the use of tools such as CliniClue will be most valuable in the secondary use of data and any data sharing project.

The existence of competing information models, such as HL7 RIM, CDISC and BRIDG, has already been noted as a major obstacle in achieving interoperability [4], and adding terminological variation simply worsens the problem. Information models – especially granular ones - might limit the semantics that are placed in the terminology by restricting which, if any, post-coordination would be used, which could be beneficial in terms of simplification of coding. Abstract models, such as HL7 RIM, leave ambiguity as to which pieces of complex terminologies such as SNOMED CT could or should be used. This research should underscore the importance of including terminology guidelines in discussions and modeling of information model standards.

CONCLUSIONS

Heterogeneity among coders using the same controlled terminology appears inescapable despite the extensive efforts of terminological standards developers and implementers. All areas of medicine are information and communication intensive, but effective care and research are dependent upon the data quality required for effective retrieval. Problems of coding heterogeneity due to complexities inherent in language usage and control are far from being solved. Yet, even a basic recognition of the natural tendencies of people to perceive and express complex concepts differently (even when supposedly following the same rules) is a start toward building tools that better accommodate such heterogeneity. This study is one example illustrating this pressing need.

Clearly, a new model for articulating rules of use for terminologies such as SCT is needed. Guiding users who are not experts in terminology design and general ontological principles is one necessary step toward enabling wider adoption and, ultimately, more consistent results. Preferably, such a model would be designed around various use cases from real world scenarios; ones which might provide a quicker reference for users after some initial training.

Computational approaches can never address all issues stemming from the ambiguities of language, but they do hold promise to go a long way toward restricting semantic variance, enforcing some rules, and allowing for post-coding comparisons. Some approaches may include an enhanced description logic that formalizes SCT's semantic structure in a manner more reflective of tested use cases, or coding tools that limit the variation of post-coordination expression. One approach toward this end might emphasize structured data capture in more explicit clinical or research settings to ensure greater consistency. A variation of this approach could include the combination of more structured standards with SCT that allow for more directed post-coordination of complex concepts. For instance, LOINC is a standard that might help reduce such heterogeneity given its more faceted nature. LOINC's 6 information components [component/analyte; property measured; timing; type of sample; type of scale; method] can be used to codify attributes of interest in questions such as those examined in this study – e.g., *timing* (e.g., point in time or time interval), *scale type* (e.g., nominal, ordinal, etc.). This basic LOINC model has been tested by a few groups to represent questions item content in standardized questionnaires. [15–17] Lastly, future research should formally test the efficacy of the SNOMED CT description logic to determine semantic equivalence between pre-coordinated and post-coordinated concept expressions, as differences in meaning among such

expressions are both predictable and unavoidable. An ultimate goal of any of these approaches is not to eliminate heterogeneity of expression, but rather is to minimize uncertainty regarding the sameness or difference in meaning.

Acknowledgements

References

1. Andrews JE, Richesson R, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. J Am Med Inform Assoc 2007;21(2):105–14.

2. Lancaster, FW. Vocabulary control for information retrieval. 2. Arlington, VA: Information Resources Press; 1986.

3. International Health Terminology Standards Development Organization. [Organization website]. Copenhagen; Denmark: [updated November 23, 2007' cited November 27, 2007]. Available from: http://www.ihtsdo.org/

4. Richesson RL, Krischer J. Data standards in clinical research: gaps, overlaps, and challenges and future directions. J Am Med Inform Assoc 2007 Nov–Dec;14(6):687–696. [PubMed: 17712081]

5. CliniClue Browser. Version 2006.2.0030. United Kingdom: The Clinical Information Consultancy; 2006.

6. Richesson RL, Andrews JE, Krischer JP. Use of SNOMED CT to represent clinical research data: a semantic characterization of data items on case report forms in vasculitis research. J Am Med Inform Assoc 2006;13:536–546. [PubMed: 16799121]

7. Elkin PL, Brown SH, Lincoln MJ, Hoqarth M, Rector A. A formal representation for messages containing compositional expressions. Int J Med Inform 2003 Sep;71(2–3):89–102. [PubMed: 14519402]

8. College of American Pathologists. SNOMED Clinical Terms [SNOMED CT] Technical Reference Guide. January 2007 Release.

9. Aitchison, J.; Gilchrist, A.; Bawden, D. Thesaurus construction and use: a practical manual. Chicago: Fitzroy Dearborn; 2000.

10. Soergel, D. Indexing languages and thesauri: construction and maintenance. Los Angeles: Melville Publishing Co; 1974.

11. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998 Nov;37(4–5):394–403. [PubMed: 9865037]

12. Spackman KA. Normal forms for description logic expressions of clinical concepts in SNOMED RT. Proc AMIA Symp 2001:627–631. [PubMed: 11825261]

13. Dolin RH, Spackman KA, Markwell D. Selective retrieval of pre- and post-coordinated SNOMED concepts. Proc AMIA Symp 2002:210–214. [PubMed: 12463817]

14. Hampton T. Rare disease research gets boost. JAMA 295(24):2836–2838. [PubMed: 16804140]

15. Matney S, Bakken S, Huff SM. Representing nursing assessments in clinical information systems using the logical observation identifiers, names, and codes database. Journal of Biomedical Informatics 2003;36(4–5):287–293. [PubMed: 14643724]

16. White TM, Hauan MJ. Extending the LOINC Conceptual Schema to Support Standardized Assessment Instruments. Journal of the American Medical Informatics Association 2002;9(6):586–599. [PubMed: 12386110]

17. Bakken S, Cimino JJ, Haskell R, et al. Evaluation of the Clinical LOINC (Logical Observation Identifiers, Names, and Codes) Semantic Structure as a Terminology Model for Standardized Assessment Measures. Journal of the American Medical Informatics Association 2000;7(6):529–538. [PubMed: 11062226]

Table 1

Relationships among Companies (n=319)

Pair wise Comparisons	Number Equivalent	Number Super/Subtype	Number Unrelated	Missing Data[*]
Co_A with Co_B	9%	41%	38%	11%
Co_A with Co_C	4%	37%	54%	4%
Co_B with Co_C	34%	18%	38%	10%

[*] The percentage of cases wherein one of the companies in the pair chose not to code the question/answer pair.

Table 2

Relationships when Companies Agreed on Core Concept

Pair wise Comparisons	Number Equivalent	Number Super/Subtype	Number Unrelated	Missing data[*]
Co_A with Co_B (n=160)	10%	38%	40%	12%
Co_A with Co_C (n=148)	5%	39%	53%	3%
Co_B with Co_C (n=151)	72%	9%	17%	3%

[*] The percentage of cases wherein one of the companies chose not to code the question/answer pair.