



Published in final edited form as:

*J Biomed Inform.* 2012 February ; 45(1): 1–14. doi:10.1016/j.jbi.2011.08.016.

## Auditing Complex Concepts of SNOMED using a Refined Hierarchical Abstraction Network

Yue Wang, MS<sup>1</sup>, Michael Halper, PhD<sup>2</sup>, Duo Wei, BS<sup>1</sup>, Huanying Gu, PhD<sup>3</sup>, Yehoshua Peri, PhD<sup>1</sup>, Junchuan Xu, MD<sup>1</sup>, Gai Elhanan, MD<sup>1,4</sup>, Yan Chen, PhD<sup>5</sup>, Kent A. Spackman, MD, PhD<sup>6</sup>, James T. Case, DVM, PhD<sup>7</sup>, and George Hripcsak, MD<sup>8</sup>

<sup>1</sup>Computer Science Dept., New Jersey Institute of Technology, Newark, NJ 07102 USA

<sup>2</sup>Information Technology Dept., New Jersey Institute of Technology, Newark, NJ 07102 USA

<sup>3</sup>Computer Science Dept., New York Institute of Technology, New York, NY 10023 USA

<sup>4</sup>Halfpenny Technologies, Inc., Blue Bell, PA 19422 USA

<sup>5</sup>Computer Information Systems Dept., BMCC, CUNY, New York, NY 10007 USA

<sup>6</sup>IHTSDO, 2300 Copenhagen S, Denmark

<sup>7</sup>NLM/NIH, Bethesda, MD 20817 USA

<sup>8</sup>Dept. of Biomedical Informatics, Columbia University, New York, NY 10032 USA

### Abstract

Auditors of a large terminology, such as SNOMED CT, face a daunting challenge. To aid them in their efforts, it is essential to devise techniques that can automatically identify concepts warranting special attention. “Complex” concepts, which by their very nature are more difficult to model, fall neatly into this category. A special kind of grouping, called a *partial-area*, is utilized in the characterization of complex concepts. In particular, the complex concepts that are the focus of this work are those appearing in intersections of multiple partial-areas and are thus referred to as *overlapping concepts*. In a companion paper, an automatic methodology for identifying and partitioning the entire collection of overlapping concepts into disjoint, singly-rooted groups, that are more manageable to work with and comprehend, has been presented. The partitioning methodology formed the foundation for the development of an abstraction network for the overlapping concepts called a *disjoint partial-area taxonomy*. This new disjoint partial-area taxonomy offers a collection of semantically uniform partial-areas and is exploited herein as the basis for a novel auditing methodology. The review of the overlapping concepts is done in a top-down order within semantically uniform groups. These groups are themselves reviewed in a top-down order, which proceeds from the less complex to the more complex overlapping concepts. The results of applying the methodology to SNOMED’s Specimen hierarchy are presented. Hypotheses regarding error ratios for overlapping concepts and between different kinds of overlapping concepts are formulated. Two phases of auditing the Specimen hierarchy for two releases of SNOMED are reported on. With the use of the double bootstrap and Fisher’s exact test

---

© 2011 Elsevier Inc. All rights reserved.

All communications to: Michael Halper, Information Technology Dept., New Jersey Institute of Technology, Newark, NJ 07102-1982 USA, Tel: (973) 596-5752, Fax: (973) 596-5324, mikhahalper@yahoo.com.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

(two-tailed), the auditing of concepts and especially roots of overlapping partial-areas is shown to yield a statistically significant higher proportion of errors.

## Keywords

SNOMED; Terminology; Auditing; Quality Assurance; Partitioning; Abstraction Network; Taxonomy; Complex Concept; Neighborhood Auditing; Group Auditing

---

## 1 Introduction

SNOMED CT [1] is one of the leading biomedical terminologies in use today. This is evidenced, for example, by the fact that it is slated to become an integral component of standardization in health information technology [2]. In one particular application, the encoding of patients' problems in Electronic Health Records (EHRs) by concepts derived from SNOMED has been proposed as part of the requirements for "meaningful use" of such systems [2]. Due to this, quality assurance is a critical task facing SNOMED's maintenance personnel. Given SNOMED's expanding content and attendant complexity, its quality assurance is a non-trivial matter. There is assuredly a need to provide automated and semi-automated methodologies for aiding editors in this endeavor. For example, methodologies that can automatically identify concepts likely to exhibit higher rates of errors and thus warranting special attention fit the bill.

One of our driving research themes has been that "complex" concepts, as defined by various criteria, are worth concentrating on in auditing efforts. By their very nature, such concepts are more difficult to model and should therefore be scrutinized more closely by auditors. (Various kinds of complex concepts targeted for auditing are discussed further in Section 2.3.) In [3], we identified a category of concepts that can be deemed complex based on our previously introduced abstraction network called the *partial-area taxonomy*. We presented a methodology for hierarchically clustering such concepts—called *overlapping concepts*—and automatically constructing a novel abstraction network for their presentation. A portion of the new network, the *disjoint partial-area taxonomy*, is a directed acyclic graph of nodes representing groups of overlapping concepts where increased conceptual complexity is encountered as one navigates downward in the terminological hierarchy.

In this paper, which is a companion to [3], we again follow the theme of focusing auditing on complex concepts and introduce a methodology for auditing the overlapping concepts based on the disjoint partial-area taxonomy first presented in [3]. Our methodology constitutes a systematic review of the overlapping concepts as determined by their hierarchical ordering within the disjoint partial-area taxonomy. The methodology is applied to the July 2009 release of SNOMED's Specimen hierarchy. The results are compared to those obtained from an audit carried out on the July 2007 release and based on a preliminary methodology that also focused on overlapping concepts [4]. That methodology just reviewed all overlapping concepts without utilizing any grouping structures or ordering.

Because this paper is meant to be a companion to [3], we do not present all aspects of the disjoint partial-area taxonomy herein. However, an overview is given in Section 2.2.

## 2 Background

### 2.1 SNOMED CT Concepts

SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) [1] uses description logic [5] to model a wide variety of biomedical concepts, arranged in a

collection of 19 IS-A (subsumption) hierarchies. The foundational construct used in the definition of concepts is the *attribute relationship* (or simply *relationship*) that serves to connect one concept to another. For example, the definition of the concept *Ear problem* includes the relationship *finding site1* directed to the concept *Ear structure*. Each concept in SNOMED also has a unique identifier and a set of naming terms, among which we find the fully specified name, the preferred term, and synonyms.

## 2.2 Taxonomy Paradigm

The notion of a *disjoint partial-area taxonomy*, an abstraction network affording a high-level view of the content of a SNOMED hierarchy, underlies the auditing methodology presented in this paper. The reader is invited to see the companion paper [3] for all details of this kind of network. In the following, we present a brief overview, including important definitions.

The disjoint partial-area taxonomy has its basis in two other abstraction networks, the area taxonomy and the partial-area taxonomy [6]. Foundational to all three is the notion of *area*, which is defined to be a set comprising all concepts within a hierarchy that have the exact same set of relationships, regardless of the targets of those relationships. A given area is denoted by its complement of defining relationships. For example, with respect to the Specimen hierarchy, we find areas named {*morphology*}, {*identity*}, and {*procedure, morphology*}. The entire collection of areas serves to partition a hierarchy as each concept will belong to one and only one area according to its relationship structure. The areas are abstracted into nodes to form the area taxonomy (network) whose links are called *child-of* relationships [3].

The partial-area taxonomy refines the area taxonomy with a collection of embedded nodes inside the main area nodes. These additional embedded nodes stand for concept groupings called *partial-areas*, based on the notion of *root* of an area. Specifically, a root is a concept having none of parents in its area. A partial-area is a set comprising one root and all its descendants (within the area). An area may have more than one root, so it may have more than one partial-area. The partial-areas serve to provide singly rooted and hierarchically cohesive divisions of an area. Each is named after its unique root. They are also linked together using *child-of*'s [3].

A given concept may reside in more than one partial-area, a situation that occurs when the concept is a descendant of two or more roots. Such a concept is called an *overlapping concept*. An example of this can be seen in Figure 1, where the concept at the bottom *Dialysis fluid specimen* belongs to the partial-areas rooted at *Fluid sample* and *Drug specimen*. The presence of overlapping concepts somewhat degrades the categorization power of partial-areas. When looking at a specific partial-area, one can encounter concepts belonging solely to that partial-area and therefore elaborating the semantics of its root only. But other concepts—the overlapping concepts—would belong to additional partial-areas at the same time and elaborate the semantics of multiple roots. The concept *Dialysis fluid specimen* from Figure 1 is both a fluid sample and a drug specimen, unlike its parent *Dialysate sample* which is only a kind of fluid sample. Moreover, overlapping concepts constitute knowledge convergence points within the hierarchy. As such, they warrant the designation “complex” and thus should be separated out from other concepts for the sake of auditing review.

---

<sup>1</sup>Concept names are written in italics with the first letter capitalized; relationships are in italics.

In order to address these issues, we have developed—in the companion paper [3]—an additional abstraction network, the disjoint partial-area taxonomy, to properly model and highlight the overlapping portions of partial-areas as nodes in their own right. Our aim in formulating the disjoint partial-area taxonomy was to partition the overlapping concepts to obtain a collection of concept groups satisfying single-rootedness. For details, see [3]. The basis for the partitioning is the notion of *overlapping root*. Basically, such a concept is one that sits at the top of overlapping concepts, with none of its parents themselves being overlapping concepts. In a recursive fashion, additional overlapping roots are identified below the top overlapping roots. As an illustration, the 15 overlapping roots of the area {*substance*} (SNOMED 2009) are shown as multi-colored boxes in Figure 2. The multi-coloring is used to indicate which area roots—appearing singly-colored at the top—the overlapping roots are descended from. (The other roots of area {*substance*} are not shown.) For example, the tricolored overlapping root *Acellular blood (serum or plasma) specimen* is a descendant of *Blood specimen*, *Body substance sample*, and *Fluid sample*. The uncolored concepts are non-root overlapping concepts.

Each overlapping root will be the root of its own newly formed concept group called a *disjoint partial-area* (or *d-partial-area*, for short). (Again, see [3] for details.) The disjoint partial-area taxonomy is constructed from the d-partial-areas, which become nodes. The portion of the disjoint partial-area taxonomy for the area {*substance*} in Figure 3 is derived from the excerpt in Figure 2.

### 2.3 Terminology Auditing

Auditing (quality assurance) is an important aspect of any terminology's life cycle [7], particularly one as large and comprehensive as SNOMED. A variety of systematic auditing techniques have been proposed and applied to SNOMED.

SNOMED's conceptual coverage and its completeness have been assessed using comparative approaches involving external sets of clinical terms [8, 9, 10]. An evaluation of the semantic completeness of SNOMED's content has also been done using a formal concept analysis (FCA)-based model [11]. Following that work, a highly-scalable approach was utilized to determine how well SNOMED conformed to a lattice structure and to suggest possible content extensions [12].

Lexical information (specifically, term substrings) was used to detect potential classification omissions [13]. In other work, lexical analysis of SNOMED concepts' textual descriptions has yielded a large collection of underspecified concepts and possibilities for refining SNOMED's content [14]. Our own lexical approach has identified a variety of inconsistencies between SNOMED terms and the underlying logical modeling of seemingly similar concepts [15]. Inconsistent usage of the words “and” and “or” in SNOMED terms has been studied [16].

Ontological and linguistic techniques were utilized to identify duplicates and redundancy [17, 18]. SNOMED has been analyzed to determine how well its hierarchical relations adhere to four basic ontological principles [19, 20]. Since SNOMED is based on a description-logic (DL) formalism, it is amenable to algorithms developed in the context of DL representations for the detection of terminological inconsistencies [21] and synonymy [22]. The impact of SNOMED revisions was assessed by investigating the manual mappings between a proprietary interface terminology to two versions of SNOMED [23]. In [24], we find a comprehensive review of auditing methodologies used for SNOMED along with a useful general glossary pertaining to auditing.

In general, the typically limited availability of auditing resources makes it imperative to develop systematic techniques that focus efforts on concepts or groups of concepts that are likely to have higher rates of errors. In this way, a better return, measured in the number of errors found, can be expected for a given amount of auditing work. We have proposed and implemented SNOMED auditing regimens that make use of the two programmatically derived taxonomies introduced above [6]. We have shown that the taxonomies are helpful in promoting more efficient and effective auditing. Examples of groups of concepts having an increased likelihood of being in error based on taxonomies include small partial-areas [25], areas with but a few small partial-areas (for NCIT) [7], and regions that have so-called strict inheritance obtainment patterns [6, 25]. Different kinds of concept errors have been found to manifest themselves as anomalies at the taxonomy level, allowing for efficient discovery. In [26], the taxonomy framework was extended to hierarchies having no outgoing relationships by utilizing implicit converse relationships. The connection between auditing and complexity measures expressed with the taxonomy framework was explored in [27]. The auditing methodology presented in this paper is based on additional refinements to the partial-area taxonomy.

Complex concepts, the focus of the present work, can be characterized in a variety of ways. As one might expect, complex concepts of different kinds show higher likelihood of error and have been the focus of certain auditing methodologies. For example, in the context of the partial-area taxonomy of SNOMED, we have shown that *strict-inheritance regions*, groupings based on more “tangled” inheritance patterns and thus naturally containing more complex concepts, tended to experience larger percentages of errors [25]. With respect to the UMLS, we have shown that concepts assigned multiple semantic types are of higher complexity expressed as compound semantics [28, 29]. Small groups of concepts sharing the same multiple typing were found to have a higher likelihood of error, due mostly likely to the uncommon compound semantics that they elaborated [30, 31]. Concepts appearing in the extents of *intersection classes*, multiply inheriting classes within the Medical Entities Dictionary (MED) [32] schema, also were inherently more complex and showed increased error rates [33]. These examples illustrate the theme that complex concepts, due to their multiple categorizations, offer fertile ground for gleaning errors. We have exploited this theme in the context of a variety of auditing methodologies for a number of different terminologies and terminological systems. The current paper further shows the benefit of this kind of approach in a DL-based terminology.

Many important terminologies and terminological systems, aside from SNOMED and the others mentioned above, have been the focus of systematic auditing regimens. In fact, a special issue of JBI (June 2009 [34]) has been devoted exclusively to terminology auditing methodologies. In [24] in that issue, a framework was introduced to help classify the large body of disparate techniques based on various criteria. For example, distinctions were made based on the kind of terminology attribute that was the focus of the audit, e.g., terms and concepts vs. semantic classification. Moreover, the methodologies were categorized according to their uses of various knowledge and their levels of automation in the identification of problems. According to the classification, the methodology presented herein can be described as “automated systematic.”

Some of the methodologies surveyed in [24] that were designated automated systematic involved some kind of rule specification. For example, the work in [35] used rules to assess certain uniqueness constraints in Read Codes. In the context of the UMLS, a search for concept redundancy was aided by constraints on semantic types [36]. Our own algorithm [37] for finding all redundant semantic-type assignments is based on a rule for the UMLS Semantic Network [38]. Concept redundancy was also addressed in SNOMED with the use of rules based on a mapping to LinkBase, a medical ontology [18]. A number of automated

systematic methods have exploited DL representations of terminologies. The methodology of [22] is such an example. Our methodology does not utilize any DL classifier functions or any features of SNOMED's underlying DL framework, except for its systematic definition of relationships and their inheritance via the IS-A hierarchy. Our taxonomy-based auditing has been used in complement to an approach using a DL classifier in [39]. The current taxonomy-based auditing approach does not employ rules in determining whether a potential error condition exists. Instead, a classification of a collection of complex concepts is made and an abstraction network is defined on top of that collection to guide the auditing efforts. Any rules that might be employed by the auditor are extraneous to the basic methodology. In general, we view our taxonomy-based auditing methodology as complementary to other auditing approaches. Since different auditing techniques typically expose some kinds of errors while missing others, there is a need for a suite comprising a variety of techniques to provide quality-assurance support for terminologies.

### 3 Methods

Different auditing methodologies are applied in the first phase and the second phase of our study. The former is with respect to the July 2007 release of SNOMED; the latter, with respect to the July 2009 release.

#### 3.1 Phase 1: Unordered Auditing

As we have discussed, the overlapping concepts are complex concepts due to their multiple classification with respect to the partial-area taxonomy and are thus targeted for auditing. For Phase 1, we call upon two of our domain-expert authors (GE and JX), each of whom has training in medicine as well as training and experience in medical terminologies. The overlapping concepts of the July 2007 Specimen hierarchy are reviewed individually by each of the two auditors. The concepts are presented to the auditors with the following data for each: concept ID, preferred term, area, and d-partial-area. The auditor is given a standardized form containing two fields for completion. The first field is used to indicate the error type (if any). The choice is to be made from a menu of seven types of errors: incorrect parent, missing parent, incorrect child, missing child, incorrect relationship type, missing relationship, and incorrect relationship target. The second field is used by the auditors to suggest a correction for the error discovered.

The auditors' review in this phase involves the examination of all overlapping concepts without regard to any specific order [4]. After that, the two auditors together review concepts for which their individual reports differ, and analyze the discrepancies until a consensus is reached. A consensus report is then given to another author (KAS)—who is currently the Chief Terminologist of IHTSDO [40]—for further review. Only his accepted results are reported for Phase 1.

#### 3.2 Phase 2: Topologically Ordered Auditing

We have seen in [3] that some overlapping concepts are more complex than others as we move down through the hierarchy. With this idea in mind, we propose the following auditing regimen that utilizes the paradigm of “group-based” auditing [6]. In the group-based approach applied to overlapping concepts, the concepts are reviewed in groups exhibiting semantic uniformity, that is, all the overlapping concepts of a d-partial-area are reviewed together with an eye toward the overlapping root which expresses the overarching semantics of the group. Furthermore, the concepts in the immediate neighborhoods of the overlapping concepts (consisting of parents, children, siblings, and targets of relationships) are audited. This “neighborhood auditing” may help to uncover propagated errors, which might otherwise be missed if the review were limited to the overlapping concepts alone.

Since SNOMED is description-logic based [5], relationships are inherited by a child concept from its parent(s) along the IS-A hierarchy. Thus, an error such as an incorrect relationship will be inherited, too. Furthermore, even an error such as an omitted relationship may be “inherited” in the sense that if it is missing from the parent, it will probably be missing from the child (unless it is explicitly defined at the child).

As a consequence, it is preferred in an audit of a group of hierarchically related concepts that the review follow a top-down order. Following such an order may help in detecting more errors as well as in accelerating the review process. In particular, when a child is scrutinized, the auditor is already aware of any errors with the parents and is alert to their potential propagation. The topological sort [41] of a directed acyclic graph (DAG)—the structure exhibited by a SNOMED hierarchy—offers a traversal of concepts in a manner where each is processed only after all its parents have been processed. Because the d-partial-areas and their *child-of* relationships also constitute a DAG [3], the disjoint partial-area taxonomy enables the utilization of the topological sort order at two different levels: the d-partial-area level and the concept level, with the latter nested in the former.

The following describes the auditing methodology for overlapping concepts based on the disjoint partial-area taxonomy. It should be noted that overlapping roots come in two varieties: *base* and *derived*. The details can be found in [3]. The important distinction between the two in this context is that the base overlapping roots occur toward the top of the concept hierarchy and are above all the derived overlapping roots. Also note that some d-partial-areas do *not* have any overlapping concepts at all. They are the ones at the very top of the disjoint partial-area taxonomy that were residually left over after the lower-level d-partial-areas—containing overlapping concepts—were removed from their original partial-areas. For example, the top d-partial-area *Drug specimen* (1), comprising a single, non-overlapping concept, was left over as a result of extracting the d-partial-areas *Intravenous infusion fluid sample* (2) and *Dialysis fluid specimen* (1) (see Figure 3) from the original partial-area also named “*Drug specimen*” that contained a total of four concepts. Those upper-level d-partial-areas are not considered in our auditing methodology.

1. **Taxonomy level:** The d-partial-areas are processed in topological sort order starting with those having base overlapping roots. The processing proceeds through their children, grandchildren, etc., down to the very bottom of the disjoint partial-area taxonomy. As discussed in [3], the lower d-partial-areas are rooted at more complex overlapping concepts.
2. **Concept level:** On arrival at a particular d-partial-area in (1), all its constituent concepts are reviewed in a topological sort order starting with its unique root and progressing downwards. The concepts are presented to the auditor in an indented hierarchical (textual) format for inspection. The indented display neatly supports the top-down processing where each concept is reviewed only after all its respective parents are reviewed.

We note that the topological sort order leaves degrees of freedom with regards to the order with which the nodes of the graph are visited—and reviewed. For example, in a level-by-level traversal, all nodes on a given level are processed before any node on the next level. Another choice is a “preorder traversal,” where the processing proceeds from a parent node to its children and even its grandchildren, assuming all their parents were already processed at that point. For the effectiveness of the auditing regimen, we recommend the preorder traversal. In this way, the scrutiny of a child follows that of the parent as quickly as possible, allowing an auditor to more readily retain knowledge of errors discovered at the parent and potentially propagating to the child.

To illustrate the Taxonomy level, the review will begin with the bicolored d-partial-areas in Figure 3, including *Exhaled air specimen*, *Inhaled air specimen*, etc. Once the review reaches *Body fluid sample*, the only bicolored d-partial-area with children, it proceeds to the bottom level containing eight tricolored d-partial-areas, i.e., *Acellular blood (serum or plasma) specimen*, *Peripheral blood specimen*, and so on. When all child d-partial-areas of *Body fluid sample* have been audited, the processing continues with the rest of the bicolored d-partial-areas, e.g., *Dialysis fluid specimen*. Again, the d-partial-areas of one color in Figure 3 do not have overlapping concepts and are therefore not part of the auditing regimen.

Within the d-partial-area *Body fluid sample*, the Concept level processing would begin with the root *Body fluid sample* and then proceed to its 22 children, including *Exudate sample* and *Discharge specimen* (Figure 2). When a concept with children is encountered, the children are processed immediately after the parent to support the auditor in detecting error propagation from parent to child. For example, *Amniotic fluid specimen* is followed by its child *Cytologic fluid specimen obtained from amniotic fluid*. An example of a propagation of an error that is easily detectable when reviewing a d-partial-area can be seen with the concept *Synovial fluid specimen* in the d-partial-area *Body fluid sample* (Figure 2). A missing *topography* relationship is detected with the target *Articular space* in the Body Structure hierarchy. The same missing relationship is detected for its three children: *Multiple joint synovial fluid*, *Cytologic material obtained from synovial fluid*, and *Synovial fluid joint NOS*. Arriving later at the d-partial-area *Acellular blood (serum or plasma) specimen*, the root would be examined first. Note that the root's overlapping parent *Body fluid sample* would already have been examined according to the Taxonomy level ordering. The review of its child *Serum specimen* and its four children would follow. Only after that would the review of the sibling *Plasma Specimen* and its three descendants occur (see Figure 2).

For further illustrative purposes, Figure 4 shows an excerpt of four d-partial-areas, *Body fluid sample*, *Acellular blood (serum or plasma) specimen*, *Venous blood specimen*, and *Peripheral blood specimen*, of the area {*substance*}, where both the d-partial areas, drawn as boxes, and the concepts, listed inside the boxes, are displayed in an indented format to illustrate the topological-sort-order processing. The auditing proceeds left-to-right and downward, following the indentation. Only a sample of the concepts are shown for the d-partial-area *Body fluid sample*.

For this phase, the auditing is performed by three of our domain-expert authors (GE, JX, and YC), each of whom has training in medicine as well as training and experience in medical terminologies. All the overlapping concepts of SNOMED's Specimen hierarchy (July 2009), within all its areas, are audited. The data presented to them for each concept are exactly the same in this phase as they are in Phase 1. Additionally, the same error-reporting form is used. In Section 4, a sample of the various types of errors is listed.

In the Phase 2 review, we seek to achieve a better agreement regarding the combined reported results. Thus, the auditors' findings are anonymized and summarized. The three experts are then requested to review the summary report and mark whether they agree or disagree with the errors listed. One expert might overlook an error discovered by another, and may eventually agree with it once the potential error is reported. All errors asserted by at least one auditor are reviewed by another author (JTC) who is in charge of the SNOMED United States National Release Center (NRC). Only errors confirmed by him are considered in the results. Let us note that any changes approved by him for inclusion in the US extension of SNOMED are eventually transferred to the IHTSDO for review and potential inclusion in SNOMED's international release.



### 3.3 Hypotheses and Control Sample

There are two hypotheses that we wish to investigate in regard to this study. The first distinguishes between overlapping concepts and non-overlapping concepts. The second distinguishes between overlapping roots of d-partial-areas and other overlapping concepts.

*Hypothesis 1:* Concepts residing in d-partial-areas having overlapping roots (i.e., overlapping concepts) are more likely to have errors than concepts residing in d-partial-areas containing no overlapping concepts.

*Hypothesis 2:* Overlapping roots of d-partial-areas are more likely to have errors than non-root overlapping concepts.

The first hypothesis asserts that these more complex concepts indeed exhibit a higher number of errors. The second hypothesis refers to the more significant overlapping concepts as the overlapping roots, where the convergence of multiple inheritance paths occurs and where we expect higher concentrations of errors.

As a basis for comparison, we also audit a control sample comprising concepts gleaned from partial-areas having no overlaps whatsoever. Both kinds of concepts are audited by the same auditors. Figure 5 presents a flow diagram that summarizes our study.

To compare overlapping concepts with those in the control sample, we look at the proportion of erroneous concepts. We use the d-partial-area as the unit of analysis, and we aggregate across levels (because of the small number of concepts at Level 2). Both hypotheses are tested for Phases 1 and 2 of the auditing on the two releases of SNOMED, two years apart. We employ the double bootstrap [42] and Fisher's exact test two-tailed [43] to calculate the statistical significance of the difference of the proportions, for Hypothesis 1 and 2, respectively.

## 4 Results

The results are reported for Phase 1 in Section 4.1 and for Phase 2 in Section 4.2. The results pertaining to the hypotheses (see Section 3.3) are distributed in these sections according to the respective phase.

### 4.1 Phase 1: Auditing of July 2007 SNOMED

The July 2007 release of the Specimen hierarchy consists of 1,056 active concepts, of which 162 are overlapping. For its partial-area taxonomy, see Figure 2 in [3]. Most of the overlapping concepts reside in Level 1 areas, i.e., those having one relationship. In fact, roughly one third (155 out of 468) of the Level 1 concepts are overlapping. And these are found primarily in the areas  $\{substance\}$  and  $\{topography\}$ . A portion of the disjoint partial-area taxonomy of  $\{substance\}$  can be seen in Figure 6, which should be compared with the 2009 version appearing in Figure 3. The d-partial-area of  $\{topography\}$  can be found in Figure 10 in [3]. Overlapping concepts also appear in the partial-areas of areas with two relationships but in far fewer numbers. In fact, there are only seven of them. Six are in  $\{topography, procedure\}$ , and the other is in  $\{topography, morphology\}$ .

Table 1 presents the results of auditing the 35 overlapping concepts (see Figure 8 in [3]) distributed across nine d-partial-areas in the area  $\{substance\}$  (Figure 6). For each d-partial-area, the following are listed: number of overlapping concepts  $V$ , number of erroneous overlapping concepts  $V_{err}$ , the number of errors  $E_{root}$  exhibited by the overlapping root, and the total number of errors  $E$  for all overlapping concepts.<sup>2</sup> For example, the largest d-partial-area *Blood specimen* has 13 concepts, of which five were found to be in error. The

root *Blood specimen* had two errors, and overall the d-partial-area's concepts had seven. For this d-partial-area, 50% (six out of 12) of the non-root overlapping concepts are erroneous, while the root itself exhibits two errors. This result, for one example of a d-partial-area, gives support to Hypothesis 2.

The auditing results for all overlapping concepts are listed by area in Table 2. For each area, we show its total number of concepts  $C$ , number of overlapping concepts  $V$ , number of overlapping roots  $D$ , number of erroneous overlapping concepts  $V_{err}$ , total number of errors  $E$  for the overlapping concepts, number of erroneous overlapping roots  $D_{err}$ , number of errors  $E_{root}$  exhibited by the set of overlapping roots, and a number of relevant ratios. For example, {*substance*} has 81 concepts, of which 35 are overlapping. Eleven (31%) of the latter were found to have a total of 31 errors or an average of 2.8 per erroneous concept, as detailed in Table 2. The ratio of the total number of errors at the overlapping concepts to the number of overlapping concepts is 0.89. Of the nine overlapping roots, five (56%) were found to be in error—with a combined 24 errors among them (or 4.8 errors per erroneous root). But only 23% ( $= (11 - 5)/(35 - 9)$ ) of the non-root overlapping concepts had errors. Let us note that for some areas (e.g., {*procedure*}), the ratio in the last column is not applicable (undefined) since singletons (i.e., d-partial-areas containing just one concept) have no non-root overlapping concepts. Other ratios may not be applicable due to a lack of errors. Nevertheless, the total ratios at the bottom of the table are defined across all the areas with overlapping concepts.

Most overlapping concepts in {*topography*} are found in intersections of partial-areas involving *Tissue specimen* containing 126 concepts. We have tabulated these results separately in Table 3. For example, the partial-area *Specimen from eye* has 18 concepts. Its intersection with *Tissue specimen* has 12 of them. Eight of those are in error.

The control sample was gleaned from partial-areas that had no intersections whatsoever with other partial-areas and from d-partial-areas having no overlapping concepts (i.e., those left over after the removal of the d-partial-areas *with* overlapping concepts from a partial-area; see, e.g., the six d-partial-areas at Level 1 of Figure 3). Furthermore, we used only partial-areas that contained more than one concept. The reason for the last requirement is that, as we alluded to, partial-areas of one concept are already known to be error-prone [7, 25]. Thus, they do not make for a proper control sample.

We used a control sample of 78 concepts from Level 1, half of its overlapping concepts (155). From Level 2, we gathered seven concepts for the control sample, an equal number to the overlapping concepts. Hence, there are  $155 + 7 = 162$  overlapping concepts, and the control sample has  $78 + 7 = 85$  concepts. Since our purpose was to audit overlapping concepts, we used a smaller control sample that was large enough to support statistical significance for the result presented below.

Table 4 gives the results of the auditing carried out on these two groups of concepts.  $C$  denotes the number of concepts,  $E$  (Column 3) denotes the total number of errors, and  $C_{err}$  is the number of erroneous concepts (Column 5)—with a given concept potentially having more than one error. The average erroneous-concept rate among the overlapping concepts was 55%, and among the control sample it was 29% (Column 6). The difference was significant (using the double bootstrap [42]) at the 0.05 level, supporting Hypothesis 1. Let us point out that there was nearly one error (0.98) on average per overlapping concept as compared to 0.36 on average within the control sample (Column 4). Moreover, erroneous

<sup>2</sup>Again, an overlapping concept may have more than one error.

concepts in the overlapping group had 1.8 errors on average (last column) versus 1.2 errors on average for the control sample, showing further difference between the two.

In examining the auditing results, we found that overlapping roots are more error-prone than other overlapping concepts. For example, in  $\{procedure\}$  and  $\{topography, procedure\}$ , all errors are found in overlapping roots. As shown in Table 2, in the area  $\{substance\}$ , five out of nine roots (55%) versus six ( $= 11 - 5$ ) out of 26 ( $= 35 - 9$ ) non-root overlapping concepts (23%) were found to be erroneous. To assess Hypothesis 2, we use the data from Table 2 for the entire collection of overlapping concepts. The percentage of erroneous concepts for overlapping roots is 69% ( $= 50/72$ ). The percentage of erroneous concepts in the set of non-root overlapping concepts is 43% ( $= (89 - 50)/(162 - 72)$ ). The difference in the percentages of erroneous concepts between the overlapping roots (69%) and the non-root overlapping concepts (43%) is statistically significant (Fisher's exact test two-tailed [43],  $p$ -value = 0.0014), supporting Hypothesis 2.

#### 4.2 Phase 2: Auditing of July 2009 SNOMED

The results of Phase 1 were submitted to CAP [44] for consideration and incorporation into the Specimen hierarchy. As a result, there were many changes in the overlapping concepts of this hierarchy as reflected in SNOMED's July 2009 release. The area taxonomy and the partial-area taxonomy for the July 2009 release appear in Figures 7 and 8, respectively. A comparison of the area taxonomies of 2007 (Figure 1 in [3]) and 2009 (Figure 7) exposes many differences in the Specimen hierarchy. For example, the total number of concepts with one relationship—which is equal to the sum of the sizes of the (green) areas on Level 1—went down from 468 to 420. At the same time, the area  $\{substance\}$  grew from 81 to 107 concepts. The number of areas with three relationships went down from seven to five with the loss of the two areas  $\{morphology, procedure, substance\}$  and  $\{topography, identity, procedure\}$ . On the other hand, the area  $\{procedure, topography, substance\}$  grew from 26 concepts in 2007 to 288 concepts in 2009.

Similarly, comparing the partial-area taxonomies for 2007 and 2009 reveals many differences. For example, the area  $\{substance\}$  changed from having ten to 11 partial-areas. But that small numerical change is misleading, as one can guess, considering the 32% increase in the size of the area. Only six partial-areas did not change. A new partial-area is *Blood specimen* with 25 concepts. Note that there was a d-partial-area with that name consisting of 13 concepts in 2007 (Figure 6). At the same time, *Drug specimen* shrank from 23 to four concepts, mainly due to the removal of blood specimen concepts. *Body substance sample* expanded from 47 to 67 concepts, while *Fluid sample* grew from 44 to 55 concepts. Such large changes on the partial-area level seem to indicate an increase in the overlap size when compared to the overall increase of 26 concepts observed on the area level. As another example, the area  $\{morphology, topography, substance\}$  went from having three partial-areas to 12. The area  $\{morphology, topography, procedure, substance\}$  grew from one to ten.

The number of overlapping concepts increased by 48 from 162 to 210 (30%). Clearly, the landscape of the overlapping portions of partial-areas changed meaningfully from the time of the July 2007 release. For example, as was predicted above, in the area  $\{substance\}$ , there were 35 overlapping concepts in nine d-partial-areas in 2007 (Figure 9 in [3]), but 48 overlapping concepts in 15 d-partial-areas in 2009 (Figure 3).

These changes motivated the application of our new methodology based on the disjoint partial-area taxonomy in this phase to the July 2009 release's overlapping concepts. Our expectation was also that our new methodology employing a detailed order of review would expose errors missed during Phase 1.

A sample of different types of errors agreed upon by all three auditors and confirmed after a review (by author JTC) is listed in Table 5. For example, it was agreed that *Serum specimen from blood product* is missing a parent *Blood specimen from blood product* that should be added. Table 6 summarizes the number of occurrences for each type of error found in the overlapping concepts of the July 2009 release. Missing parents, for example, were found for 23 concepts.

The auditing results for Phase 2 are listed by area in Table 7, in the same format used in Table 2 for Phase 1. In this case, for example, {*topography*} has 249 concepts, with 110 of them being overlapping. Fifty-two out of the 110 (47%) were found to have a total of 57 errors or an average of 1.10 per erroneous concept. The ratio of the total number of errors to the number of overlapping concepts is 0.52. Twenty of the 37 overlapping roots (54%) were found to be in error—with a combined 22 errors among them (or 1.10 errors per root). Finally, 44% ( $= (52 - 20)/(110 - 37)$ ) of the non-root overlapping concepts had errors.

For the entire set of overlapping concepts summarized in the bottom row of Table 7, 127 out of 210 (60%) were found to be erroneous. This result is applicable in assessing Hypothesis 1 (as shown in Table 8).

The control sample for Phase 2 was taken strictly from partial-areas and d-partial-areas that had no intersections whatsoever. As with Phase 1, we used only partial-areas that contained more than one concept. The sample consisted of 111 concepts from the same areas as the overlapping concepts. And as in Phase 1, the number of sample concepts taken from areas with small numbers (i.e., 2–16) of overlapping concepts was about the same as the number of overlapping concepts taken from those areas. The sample concepts numbered about half the overlapping concepts for areas with larger numbers of overlapping concepts. As with Phase 1, our purpose was to audit overlapping concepts, and we used a smaller control sample that was nevertheless big enough to support statistical significance of the result.

Like Table 4, Table 8 juxtaposes the results of auditing the overlapping concepts and those in the control sample. The average erroneous-concept rate among the overlapping concepts was 60%, versus 13% for the control sample (Column 6). The difference was significant (using the double bootstrap [42]) at the 0.05 level, supporting Hypothesis 1. Let us note that there were 0.77 errors on average per overlapping concept as compared to 0.13 on average within the control sample (Column 4). Erroneous concepts in the overlapping group had 1.27 errors on average (last column) versus 1.00 errors on average for the control sample, showing further difference between the two samples.

For the assessment of Hypothesis 2, we used the results obtained for all overlapping concepts, reflected in the bottom row of Table 7. Among the 87 overlapping roots, 61 (70%) were erroneous, while for the 123 ( $= 210 - 87$ ) non-root overlapping concepts, 66 ( $= 210 - 87$  or 54%) were found to be in error. The difference in the percentages of erroneous concepts between the overlapping roots (70%) and the non-root overlapping concepts (54%) is statistically significant (Fisher's exact test two-tailed,  $p$ -value = 0.0217).

## 5 Discussion

### 5.1 Auditing Theme: Complex Concepts

This study is motivated by a general theme that more “complex” concepts tend to have more errors than simpler concepts. The theme of being more complex may manifest itself in a variety of ways. One manifestation of this theme for partial-areas was the group of concepts residing in “strict inheritance” partial-areas [25]. In the context of the present work, this theme appears twice: the first time in identifying overlapping concepts as more complex

than non-overlapping concepts due to their elaborating the multiple semantics of the multiple partial-areas they belong to; the second in the distinction between overlapping roots and non-root overlapping concepts. The reason for the higher complexity of overlapping roots stems from their being at the junction points where multiple hierarchical paths from ancestors converge. Each such path contributes a portion of a diverse collection of inherited knowledge at the overlapping root. Hypothesis 1 addresses the first appearance. Hypothesis 2 pertains to the second.

As was also shown in [25] with regards to strict inheritance partial-areas, the results of our study confirm the auditing theme that complex concepts have relatively more errors. In view of the fact that modeling complex concepts is more challenging than modeling simpler concepts, it is not really surprising to find more errors in the former. The research challenge is to discover various characterizations of “complex” concepts. In particular, it is fruitful to identify structural characterizations that can be computed automatically, as in the current study and in [25]. The higher error rate shown here and in [25] will help achieve higher productivity from quality-assurance personnel in their review of such concepts. It is suggested that the design of partial-area taxonomies and the auditing of the complex concepts discussed here and in [25] should become integral parts of the design cycle for terminologies such as SNOMED and the NCIt [7]. Such techniques will also help interface terminologies such as Kaiser-Permanente’s CMT [45] or the VA’s ERT [46], which were derived initially from SNOMED and were enhanced with local vocabulary as well as integrated parts of other terminologies. It is a research challenge to identify more manifestation of complex concepts using taxonomies or other structural techniques for SNOMED and similar terminologies.

One may wonder why there are more errors in overlapping roots than there are in other overlapping concepts (as stated in Hypothesis 2), in spite of the expectation that our methodology will expose error propagation from parents to children, which implies that errors at an overlapping root would be “inherited” by the other concepts in its d-partial-area. One should realize that indeed missing or incorrect relationship errors are “inherited,” but that is not true of other errors, e.g., an incorrect parent. Furthermore, many d-partial- areas have just a single concept (which serves as the respective root), with no children below to inherit the errors. Hence, our methodology is designed to expose the cross-generational error propagation to the extent that it exists.

## 5.2 Repeated Application of an Auditing Methodology

In previous papers [6, 25], we presented various methodologies for auditing a SNOMED hierarchy. A question to consider is whether there is a reason to reapply the same auditing technique to the hierarchy obtained following corrections derived from the earlier auditing phase that used the same technique. Should we assume that not all errors were found and corrected? In the context of this paper, the question was: should we audit the overlapping concepts again following the first phase reported in [4]? Furthermore, how many times should the same technique be applied? Another way to phrase this last question is: how do we identify the convergence of the auditing process?

We had several reasons to re-audit the overlapping concepts. First, in Phase 1, we just audited the set of all overlapping concepts without utilizing any structure among them. In this paper, we introduced the new “group auditing” methodology of overlapping concepts where d-partial-areas were utilized as the grouping unit following the new framework described in [3]. Furthermore, the new methodology employs a top-down ordering within each d-partial-area and among various d-partial-areas.

Another reason for repeating the auditing on the overlapping concepts is the large increase in their numbers and the number of d-partial-areas. For example, see Figure 3 for the d-partial-areas in the area {*substance*} in comparison to the corresponding Figure 9 that appeared in [3]. In Figure 9 of [3], we see only four d-partial-areas without overlapping concepts at the first level and nine d-partial-areas comprising overlapping concepts. In Figure 3, showing the overlapping concepts of {*substance*} in 2009, there are six top d-partial-areas without overlapping concepts and 15 d-partial areas with overlapping concepts. Moreover, when one reviews the details of the two figures, many internal changes can be seen. For example, the d-partial-area *Body fluid sample* had 11 concepts in 2007 and 23 in 2009. *Blood specimen* had 13 overlapping concepts in Level 3 originally, and in 2009 it is a top d-partial-area of one concept only. It has eight child d-partial-areas containing 18 overlapping concepts on Level 3, which are shared jointly by the parent d-partial-area *Body fluid sample* (see Figure 3). The latter was a parent of *Blood specimen* in Figure 9 of [3]. Obviously, such changes reflect an entire remodeling of many overlapping concepts.

When realizing the extent of the changes, it was possible that new errors were introduced and that the new disjoint partial-area taxonomy would lead to exposure of errors not reported in the review of the 2007 release. The results shown in Table 7 justify the decision for the second auditing phase. While we expected a meaningful amount of errors to be found in Phase 2, we were surprised by their magnitude. Both the percentages of erroneous concepts among overlapping concepts (60% vs. 55%) and among overlapping roots (70% vs. 69%) were little changed in spite of this being a second round of auditing. Part of the explanation may be the improved methodology employed in this study. Another reason may be the large increase in the number of overlapping concepts (from 162 to 210). A further factor might be that in practice the proper modeling of these complex concepts demands more than one iteration.

On the other hand, the ratio of errors per erroneous concept was reduced (0.93 to 0.77) for all overlapping concepts, as was the ratio for erroneous overlapping roots (2.1 to 1.3). Hence, while the percentage of erroneous concepts persisted, the average number of errors fell. That is, we found less concepts with multiple errors. This last observation seems in line with the speculation above that multiple iterations are required for the proper modeling of complex concepts.

One could certainly question the expectation of the need for an additional phase of auditing after all corrections from the overlapping-concept regimen have been implemented. That is particularly true when the corrections have made their way into SNOMED's international release following the report of one of the authors (JTC) at the NRC to IHTSDO. To better understand the phenomenon of finding more errors in a subsequent phase of auditing overlapping concepts mentioned above, one needs to keep in mind the restructuring undergone by d-partial-areas due to the discovered errors. For example, in the description of the methodology in Section 3, we mentioned a concept *Synovial fluid specimen* in the d-partial-area *Body fluid sample*, which together with its children is missing the relationship *topography* to *Articular space*. But reviewing the complete audit report for the overlapping concepts in {*substance*}, one may realize that the same concept was found to have an incorrect parent, *Body fluid sample*, which was replaced by *Joint fluid specimen*. This latter concept was independently found to be missing the same *topography* relationship, as was its child *Cytologic material obtained from joint fluid*. Furthermore, another concept *Synovial fluid cells* in the area {*topography*} was also made a child of *Synovial fluid specimen* instead of *Synovial sample*. What we see is a movement of many concepts into the d-partial-area rooted at *Joint fluid specimen*, which before had only one child. Moreover, this d-partial-area would move from the area {*substance*} to the area {*substance, topography*} due to the additional *topography* relationship. When all these corrections are incorporated into a future

release of SNOMED, the disjoint partial-area taxonomy will convey the refined modeling of all joint fluid specimen concepts, contributing to better overall comprehension. However, this new modeling may expose errors not yet detected and deserves the analysis provided by the disjoint partial-area taxonomy.

If the new disjoint partial-area taxonomy for the Specimen hierarchy obtained as a result of the Phase 2 audit, and possibly reflecting a future release of SNOMED, were to differ meaningfully from the disjoint partial-area taxonomy of the 2009 release of SNOMED, then it may be advisable to reapply the auditing regimen utilizing this new view.

### 5.3 Error Rates and the Complexity of the Disjoint Partial-area Taxonomy

In Phase 1 of the auditing, the bulk of the erroneous overlapping concepts and the overlapping concept errors occurs for the areas *{substance}* and *{topography}*. It is interesting to compare the various ratios of errors for these two areas. The percentage of erroneous overlapping concepts in *{topography}* (61%) is about double that in *{substance}* (31%). However, when measuring the ratios of errors to overlapping concepts, the values for the two areas, 0.95 and 0.89, respectively, are close. This is a result of a much higher ratio of errors to erroneous concepts for *{substance}* (2.8) than for *{topography}* (1.6). This observation indicates a correlation between the ratio of the number of errors to the number of erroneous concepts and the level of complexity of overlapping concepts, as expressed in the structure of the disjoint partial-area taxonomy. As was discussed and shown in Figures 9 and 10 in [3], the nature of the overlap is much more complex for *{substance}* with several levels in its disjoint partial-area taxonomy, while it is simpler and relatively flat for *{topography}*.

### 5.4 An Audit Report from Several Auditors

The auditing in Phase 1 was performed by two of the authors (GE, JX), and their error report was obtained by a consensus from their individual findings. Anecdotal evidence from the auditors was that the face-to-face consensus process seemed to follow more of a social give-and-take rather than a deep investigation about the concepts. Similar anecdotal evidence was obtained for a study of auditor performance regarding a consensus-building stage [31].

As a result, we decided in the Phase 2 auditing to avoid the discussion-based, consensus-building effort. Instead, we circulated a combined report derived from the three auditors' Phase 2 reports. This report was anonymized and contained listings of the number of auditors for each identified error. In this second stage, each auditor was asked to indicate their agreement with each of the errors. Errors that had the support of at least one auditor were passed on for further review. It seems that a second review of others' audit reports carried out by each auditor individually without the pressure of direct social interaction is functioning well in achieving an agreement level. Not only was a better level of agreement reached, but we also witnessed auditors backing off from certain errors, when noticing that the other auditors did not mark them.

### 5.5 Limitations and Future Work

As we can see from Tables 4 and 8, according to all reported measures, there is a significantly higher return for the auditing effort obtained for the overlapping concepts compared to concepts in partial-areas without overlaps. Such higher return seems to justify concentrating auditing efforts on the more complex overlapping concepts. The results confirm Hypothesis 1. More experiments with different and larger hierarchies of SNOMED and similar terminologies, e.g., NCI [7], are needed to further confirm our finding. One idea expressed in [3] that was not confirmed by our study was that "derived" overlapping roots (of d-partial-areas) would be more error-prone than "base" overlapping roots due to their

higher complexity. Our results did not support such a phenomenon. Future studies should look again at whether this extra inherent complexity manifests itself in higher error rates in other SNOMED hierarchies.

Our interest in this paper was not in studying the auditing process *per se*, but in the distribution of the unquestionable errors resulting from it. We may investigate auditor performance and the impact of various protocols in achieving better agreement among a group of auditors in the future.

## 6 Conclusion

We proceeded from the assumption that “complex” concepts warrant particular attention in quality-assurance activities pertaining to SNOMED. Toward that end, we presented an auditing methodology based on a refined abstraction network for a SNOMED hierarchy, called the disjoint partial-area taxonomy, formulated in a companion paper [3]. The complex concepts in this study were taken to be those residing in elements of the disjoint partial-area taxonomy that represented certain overlapping subsets of portions of a SNOMED hierarchy. These so-called overlapping concepts in the Specimen hierarchy (in two different releases of SNOMED) were identified programmatically and then put through rigorous audits. Comparing these auditing results with those from control sets, we found a statistically significant higher error rate among the overlapping concepts. Furthermore, among the overlapping concepts, roots have a statistically significantly higher error rate than do non-roots. Thus, our auditing methodology based on the disjoint partial-area taxonomy and its overlapping concepts can be seen as an important addition to the existing suite of SNOMED and SNOMED-related terminology auditing regimens.

- Kinds of complex concepts (overlapping concepts) warranting auditing attention are characterized with an abstraction network.
- A methodology (from a companion paper) partitions the collection of overlapping concepts into disjoint, singly-rooted groups.
- An abstraction network ("disjoint partial-area taxonomy") for the overlapping concepts is derived from the partition.
- The abstraction network is used as the basis for auditing that involves top-down hierarchical review of overlapping concepts.

Statistical analysis shows that overlapping concepts (and their "roots") exhibit significantly higher proportions of errors.

## Acknowledgments

This work was partially supported by the NLM under grant R-01-LM008912-01A1.

## References

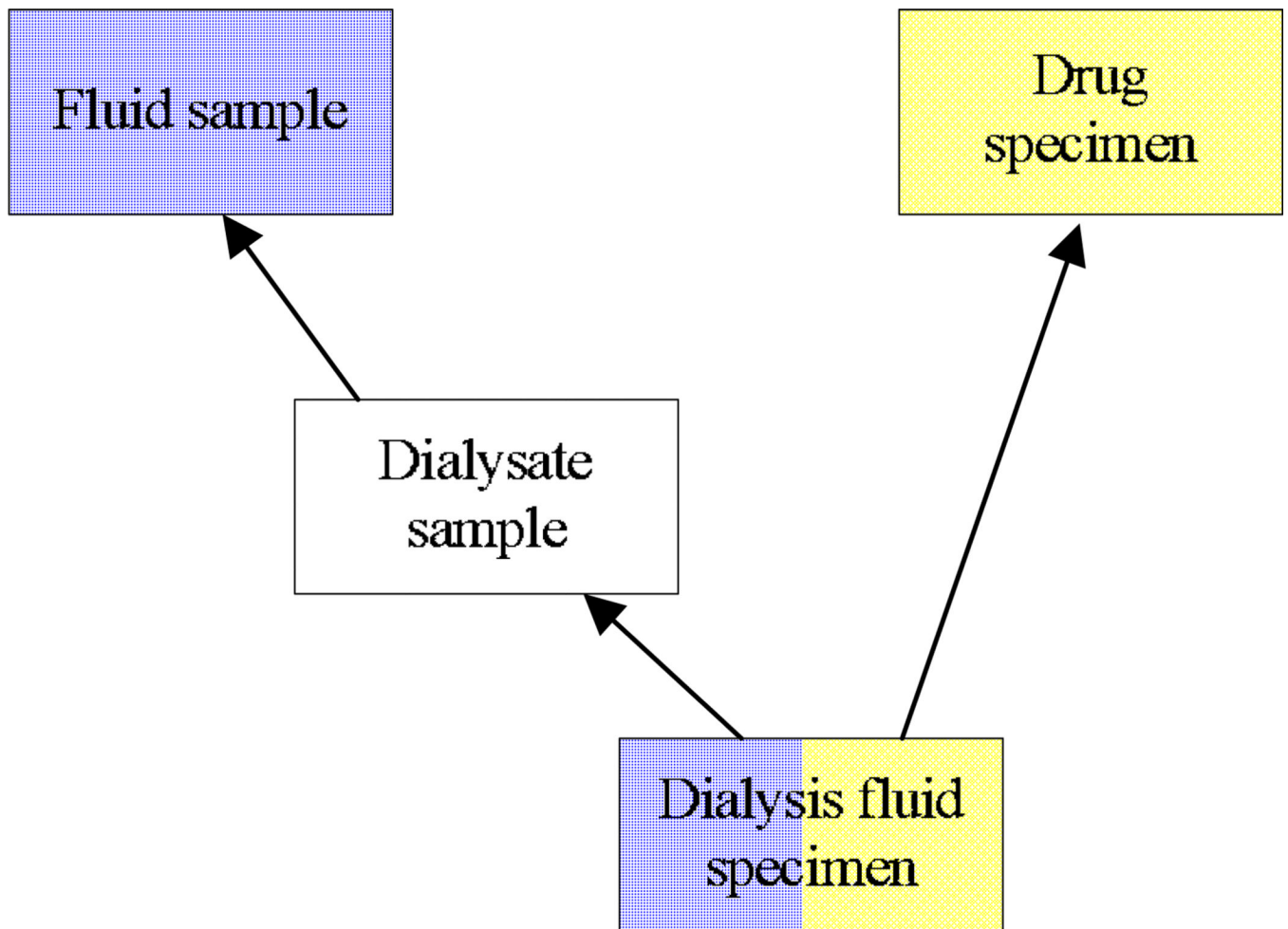
1. IHTSDO: SNOMED CT. available at <http://www.ihtsdo.org/snomed-ct>.
2. Department of Health and Human Services, Health Information Technology. Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology; Final Rule, 45 CFR Part 170. 2010 July 28.
3. Wang Y, Halper M, Wei D, Perl Y, Geller J. Abstraction of complex concepts with a refined partial-area taxonomy of SNOMED, Submitted in parallel for publication in JBI as a companion paper.



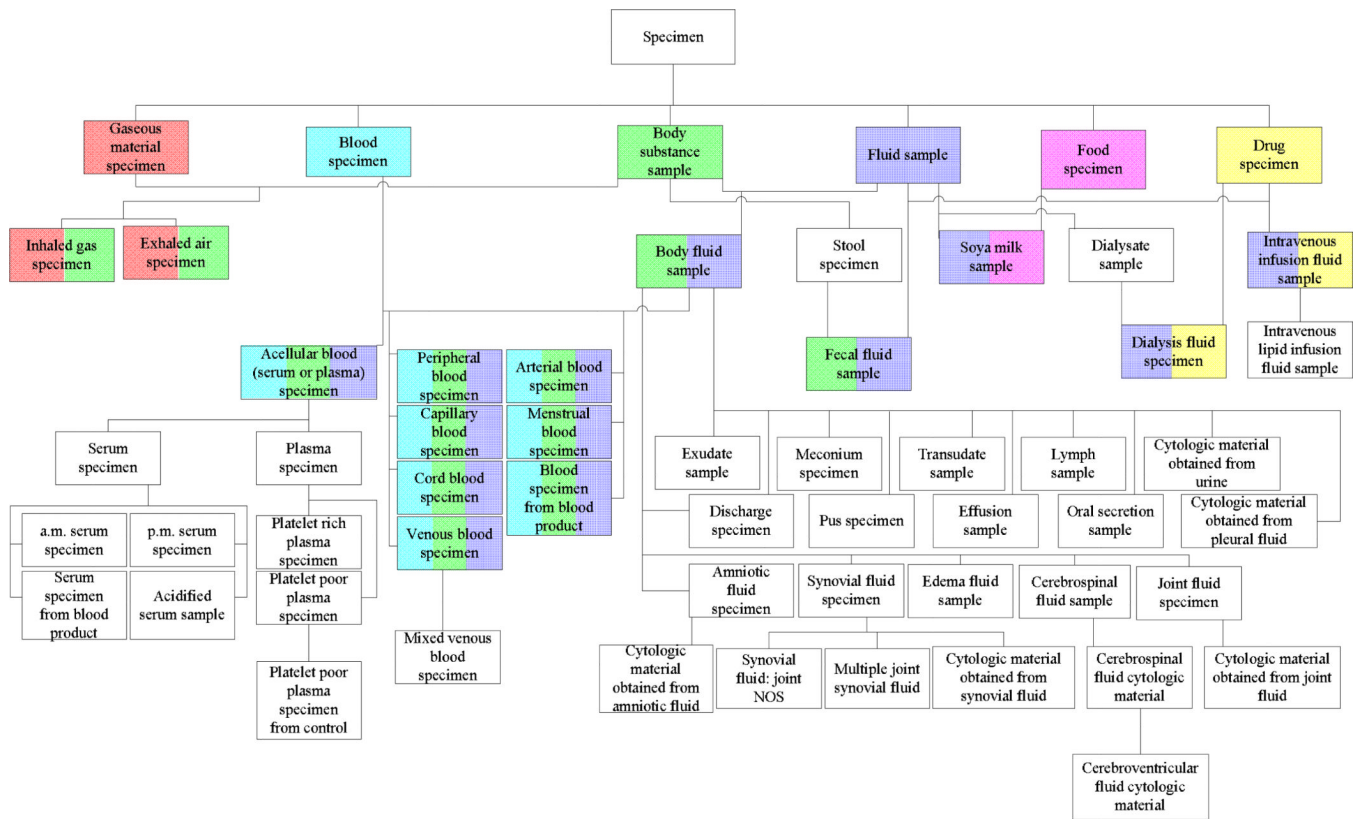
4. Wang, Y.; Wei, D.; Xu, J.; Elhanan, G.; Perl, Y.; Halper, M.; Chen, Y.; Spackman, KA.; Hripcsak, G. Auditing complex concepts in overlapping subsets of SNOMED. In: Suermondt, J.; Evans, RS.; Ohno-Machado, L., editors. Proc. 2008 AMIA Annual Symposium; Washington, DC. 2008. p. 273-277.
5. Nardi, D.; Brachman, RJ. An introduction to description logics. In: Baader, F.; Calvanese, D.; McGuinness, DL.; Nardi, D.; Patel-Schneider, PF., editors. The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge, UK: Cambridge University Press; 2003. p. 1-40.
6. Wang Y, Halper M, Min H, Perl Y, Chen Y, Spackman KA. Structural methodologies for auditing SNOMED. *Journal of Biomedical Informatics*. 2007; 40(5):561–581. [PubMed: 17276736]
7. Min H, Perl Y, Chen Y, Halper M, Geller J, Wang Y. Auditing as part of the terminology design life cycle. *JAMIA*. 2006; 13(6):676–690. [PubMed: 16929044]
8. Penz, JF.; Brown, SH.; Carter, JS., et al. Evaluation of SNOMED coverage of veterans health administration terms. In: Fieschi, M.; Coiera, E.; Li, Y-C., editors. Proc. Medinfo 2004. San Francisco, CA: 2004. p. 540-544.
9. Chute CG, Cohn SP, Campbell KE, et al. The content coverage of clinical classifications. for the computer-based patient record institute's work group on code & structures. *JAMIA*. 1996; 3(3): 224–233. [PubMed: 8723613]
10. Campbell JR, Carpenter P, Sneiderman C, et al. Phase ii evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. *JAMIA*. 1997; 4(3):238–250. [PubMed: 9147343]
11. Jiang G, Chute CG, et al. Auditing the semantic completeness of SNOMED CT using formal concept analysis. *JAMIA*. 2009; 16(1):89–102. [PubMed: 18952949]
12. Zhang, G-Q.; Bodenreider, O., et al. Large-scale, exhaustive lattice-based structural auditing of SNOMED CT. Proc. 2010 AMIA Annual Symposium; Washington, DC. 2010. p. 922-926.
13. Campbell, KE.; Tuttle, MS.; Spackman, KA., et al. A “lexically-suggested logical closure” metric for medical terminology maturity. In: Chute, CG., editor. Proc. 1998 AMIA Annual Fall Symposium; Orlando, FL. 1998. p. 785-789.
14. Pacheco, E.; Stenzhorn, H.; Nohama, P.; Paetzold, J.; Schulz, S. Detecting underspecification in SNOMED CT concept definitions through natural language processing. Proc. 2009 AMIA Annual Symposium; San Francisco, CA. 2009. p. 492-496.
15. Agrawal, A.; Elhanan, G.; Halper, M. Dissimilarities in the logical modeling of apparently similar concepts in SNOMED CT. Proc. 2010 AMIA Annual Symposium; Washington, DC. 2010. p. 212-216.
16. Mendonca, EA.; Cimino, JJ.; Campbell, KE., et al. Reproducibility of interpreting “and” and “or” in terminology systems. In: Chute, CG., editor. Proc. 1998 AMIA Annual Fall Symposium; Orlando, FL. 1998. p. 790-794.
17. Ceusters, W.; Smith, B.; Kumar, A.; Dhaen, C. Ontology-based error detection in SNOMED-CT. In: Fieschi, M.; Coiera, E.; Li, Y-C., editors. Proc. Medinfo 2004. San Francisco, CA: 2004. p. 482-486.
18. Ceusters, W.; Smith, B.; Kumar, A.; Dhaen, C. Mistakes in medical ontologies: Where do they come from and how can they be detected?. In: Pisanelli, DM., editor. *Ontologies in Medicine: Proc. Workshop on Medical Ontologies*. Rome: 2003. p. 145-164.
19. Bodenreider, O.; Smith, B.; Kumar, A.; Burgun, A. Investigating subsumption in DL-based terminologies: A case study in SNOMED CT. In: Hahn, U.; Schulz, S.; Cornet, R., editors. Proc. First Int'l Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004). Whistler, Canada: 2004. p. 12-20.
20. Bodenreider O, Smith B, Kumar A, et al. Investigating subsumption in SNOMED CT: an exploration into large description logic-based biomedical terminologies. *Artificial Intelligence in Medicine*. 2007; 39(3):183–195. [PubMed: 17241777]
21. Schlobach S, Huang Z, Cornet R, Van Harmelen F. Debugging incoherent terminologies. *Journal of Automated Reasoning*. 2007; 39:317–349.
22. Cornet R, Abu-Hanna A. Auditing description-logic-based medical terminological systems by detecting equivalent concept definitions. *int'l Journal of Medical Informatics*. 2008

23. Wade G, Rosenbloom T. The impact of SNOMED CT revisions on a mapped interface terminology: terminology development and implementation issues. *Journal of Biomedical Informatics*. 2009; 42(3):490–493. [PubMed: 19285570]
24. Zhu X, Fan J-W, Baorto DM, Weng C, Cimino JJ. A review of auditing methods applied to the content of controlled biomedical terminologies. *Journal of Biomedical Informatics*. 2009; 42(3): 413–425. [PubMed: 19285571]
25. Halper, M.; Wang, Y.; Min, H.; Chen, Y.; Hripcsak, G.; Perl, Y.; Spackman, KA. Analysis of error concentrations in SNOMED. In: Teich, JM.; Suermondt, J.; Hripcsak, G., editors. *Proc. 2007 AMIA Annual Symposium*; Chicago, IL. 2007. p. 314-318.
26. Wei, D.; Halper, M.; Elhanan, G.; Chen, Y.; Perl, Y.; Geller, J.; Spackman, KA. Auditing SNOMED relationships using a converse abstraction network. *Proc. 2009 AMIA Annual Symposium*; San Francisco, CA. 2009. p. 685-689.
27. Wei, D.; Wang, Y.; Perl, Y.; Xu, J.; Halper, M.; Spackman, KA. Complexity measures to track the evolution of a SNOMED hierarchy. In: Suermondt, J.; Evans, RS.; Ohno-Machado, L., editors. *Proc. 2008 AMIA Annual Symposium*; Washington, DC. 2008. p. 778-782.
28. Geller J, Gu H, Perl Y, Halper M. Semantic refinement and error correction in large terminological knowledge bases. *Data & Knowledge Engineering*. 2003; 45(1):1–32.
29. Representing the UMLS as an OODB: Modeling issues and advantages, *JAMIA* 7 (1) (2000) 60–80, selected for reprint. In: Gu, H.; Perl, Y.; Geller, J.; Halper, M.; Liu, L.; Cimino, JJ., editors; Haux, R.; Kulikowski, C., editors. *Yearbook of Medical Informatics: Digital Libraries and Medicine* (International Medical Informatics Association). Stuttgart, Germany: Schattauer; 2001. p. 271-285.
30. Gu H, Perl Y, Elhanan G, Min H, Zhang L, Peng Y. Auditing concept categorizations in the UMLS. *Artificial Intelligence in Medicine*. 2004; 31(1):29–44. [PubMed: 15182845]
31. Gu, H.; Hripcsak, G.; Chen, Y.; Morrey, CP.; Elhanan, G.; Cimino, JJ.; Geller, J.; Perl, Y. Evaluation of a UMLS auditing process of semantic type assignments. In: Teich, JM.; Suermondt, J.; Hripcsak, G., editors. *Proc. 2007 AMIA Annual Symposium*; Chicago, IL. 2007. p. 294-298.
32. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *JAMIA*. 1994; 1(1):35–50. [PubMed: 7719786]
33. Gu H, Halper M, Geller J, Perl Y. Benefits of an object-oriented database representation for controlled medical terminologies. *JAMIA*. 1999; 6(4):283–303. [PubMed: 10428002]
34. Geller J, Perl Y, Halper M, Cornet R. Guest editorial: Special issue on auditing of terminologies. *Journal of Biomedical Informatics*. 2009; 42(3):407–411. [PubMed: 19465342]
35. Schulz EB, Barrett JW, Price C. Read Code quality assurance: From simple syntax to semantic stability. *JAMIA*. 1998; 5(4):337–346. [PubMed: 9670131]
36. Cimino JJ. Auditing the Unified Medical Language System with semantic methods. *JAMIA*. 1998; 5(1):41–51. [PubMed: 9452984]
37. Peng, Y.; Halper, M.; Perl, Y.; Geller, J. Auditing the UMLS for redundant classifications. In: Kohane, IS., editor. *Proc. 2002 AMIA Annual Symposium*; San Antonio, TX. 2002. p. 612-616.
38. McCray AT, Nelson SJ. The representation of meaning in the UMLS. *Methods of Information in Medicine*. 1995; 34:193–201. [PubMed: 9082131]
39. Wei, D.; Bodenreider, O. *Proc. Medinfo 2010*. Cape Town, South Africa: 2010. Using the abstraction network in complement to description logics for quality assurance in biomedical terminologies - a case study in SNOMED-CT; p. 1070-1074.
40. IHTSDO. International Health Terminology Standards Development Organisation. available at <http://www.ihtsdo.org>.
41. Even, S. *Graph Algorithms*. Potomac, MD: Computer Science Press; 1979.
42. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press; 1993.
43. Good, P. *Permutation, Parametric, and Bootstrap Tests of Hypotheses: A Practical Guide to Resampling*. 3rd Edition. New York, NY: Springer; 2005.
44. College of American Pathologists – CAP Home. available at <http://www.cap.org>.

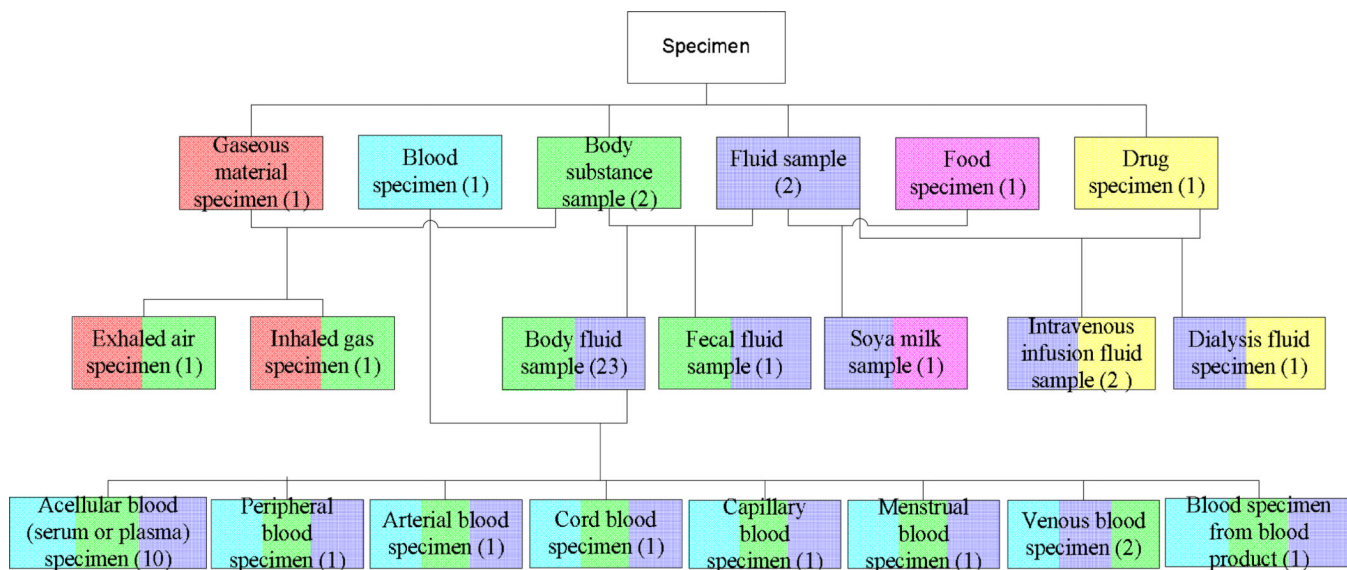
45. Dolin, RH.; Mattison, JE.; Cohn, S.; Campbell, KE.; Wiesenthal, AM.; Hochhalter, B., et al. Kaiser Permanente's Convergent Medical Terminology. In: Fieschi, M.; Coiera, E.; Li, Y-C., editors. Proc. Medinfo 2004. San Francisco, CA: 2004. p. 346-350.
46. Lincoln, MJ.; Brown, SH.; Nguyen, V.; Cromwell, T., et al. U.s. department of veterans affairs enterprise reference terminology strategic overview. In: Fieschi, M., editor. Proc. Medinfo2004. San Francisco, CA: 2004. p. 391-395.



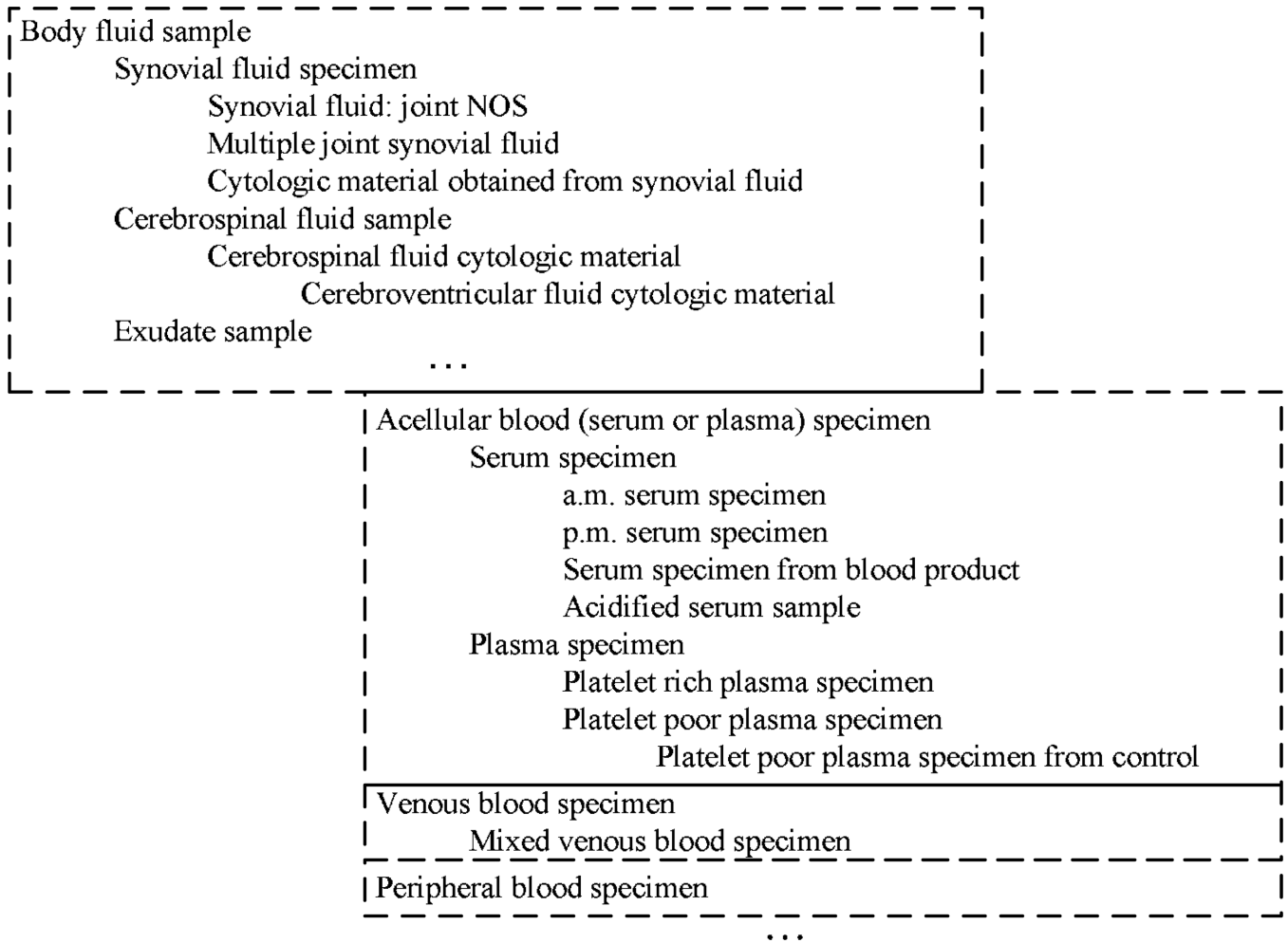
**Figure 1.** The concept *Dialysis fluid specimen* belongs to the partial-areas rooted at *Fluid sample* and *Drug specimen*, as indicated by the color coding. It is therefore an example of an overlapping concept



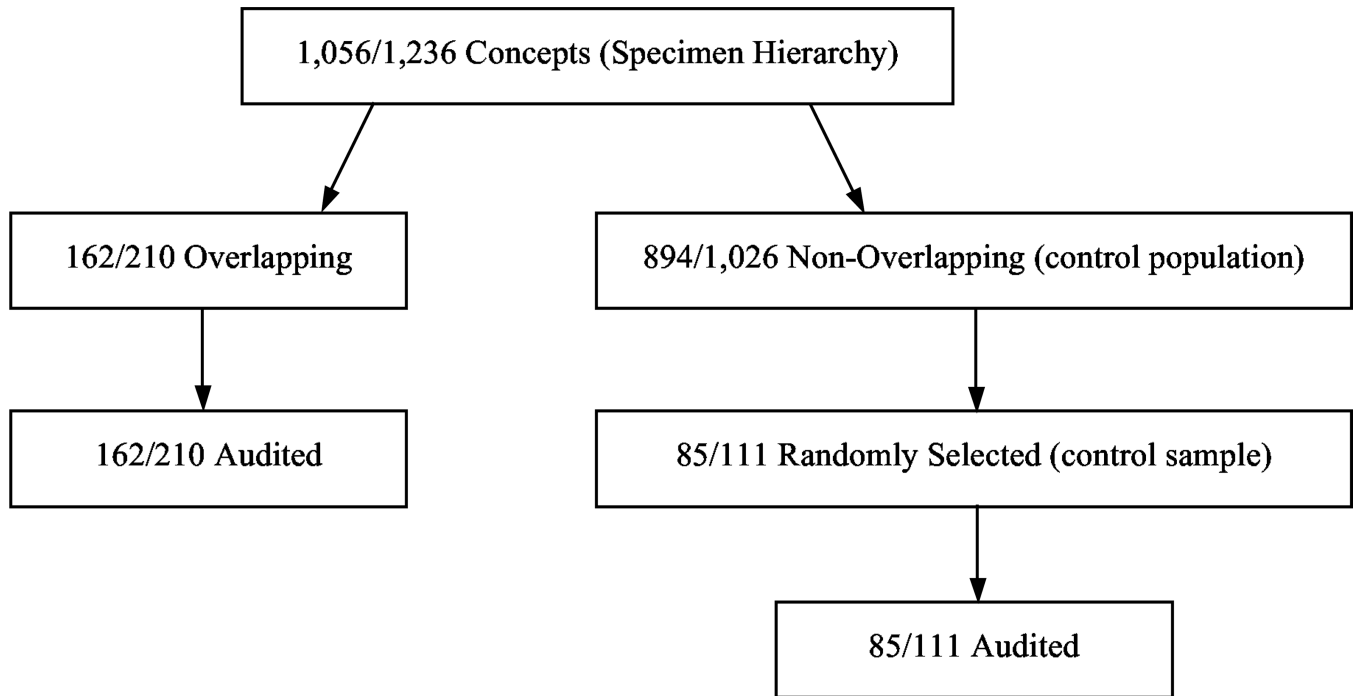
**Figure 2.** The 15 overlapping roots from the area {*substance*} of the Specimen hierarchy (July 2009) are shown as multi-colored boxes among other concepts. The coloring indicates their ancestry



**Figure 3.** The portion of the disjoint partial-area taxonomy for the area *{substance}* corresponding to the concept network in Figure 2 (SNOMED 2009)

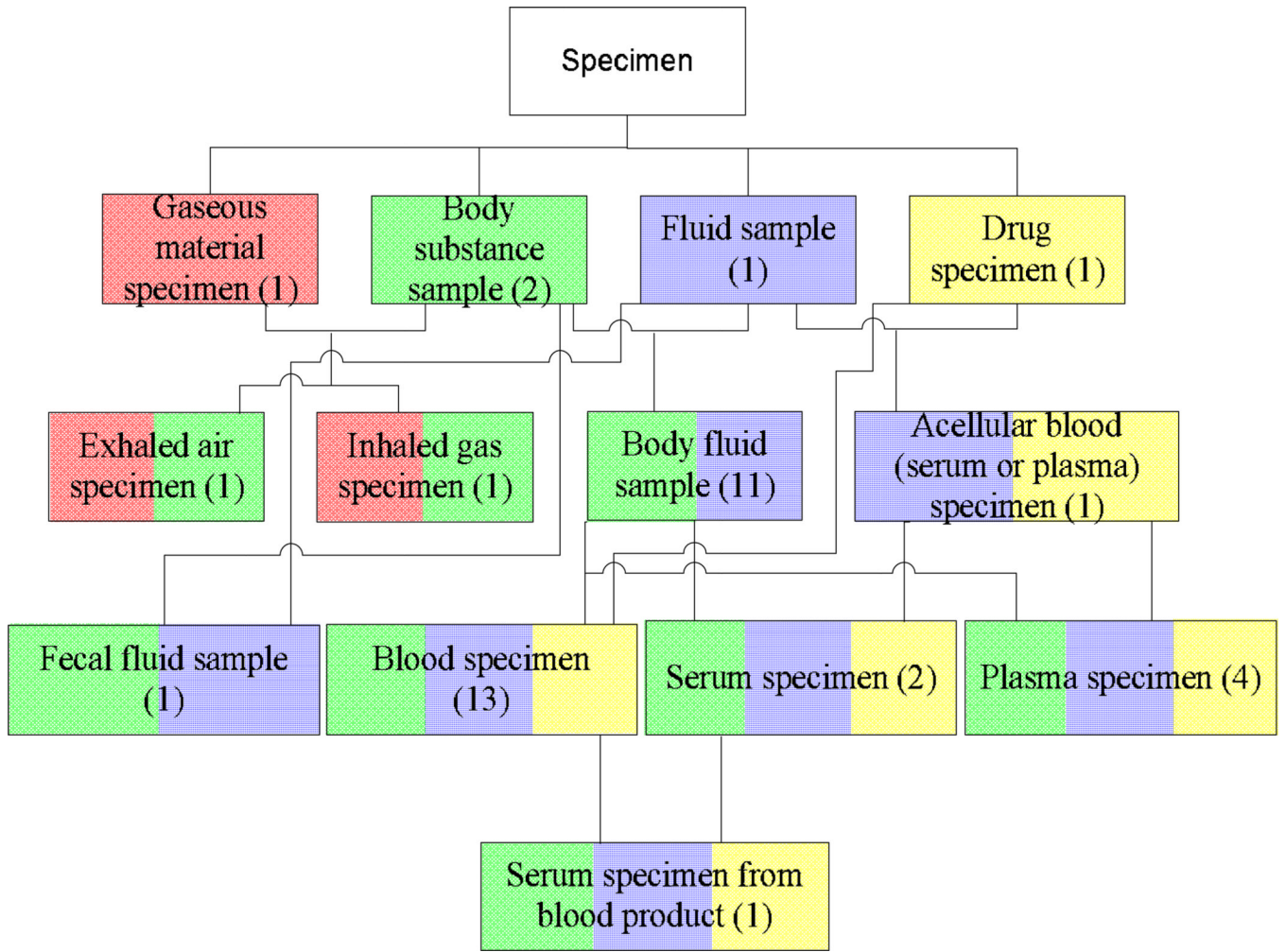


**Figure 4.**  
An indented display of four d-partial-areas and their constituent concepts illustrating the topological-sort-order processing

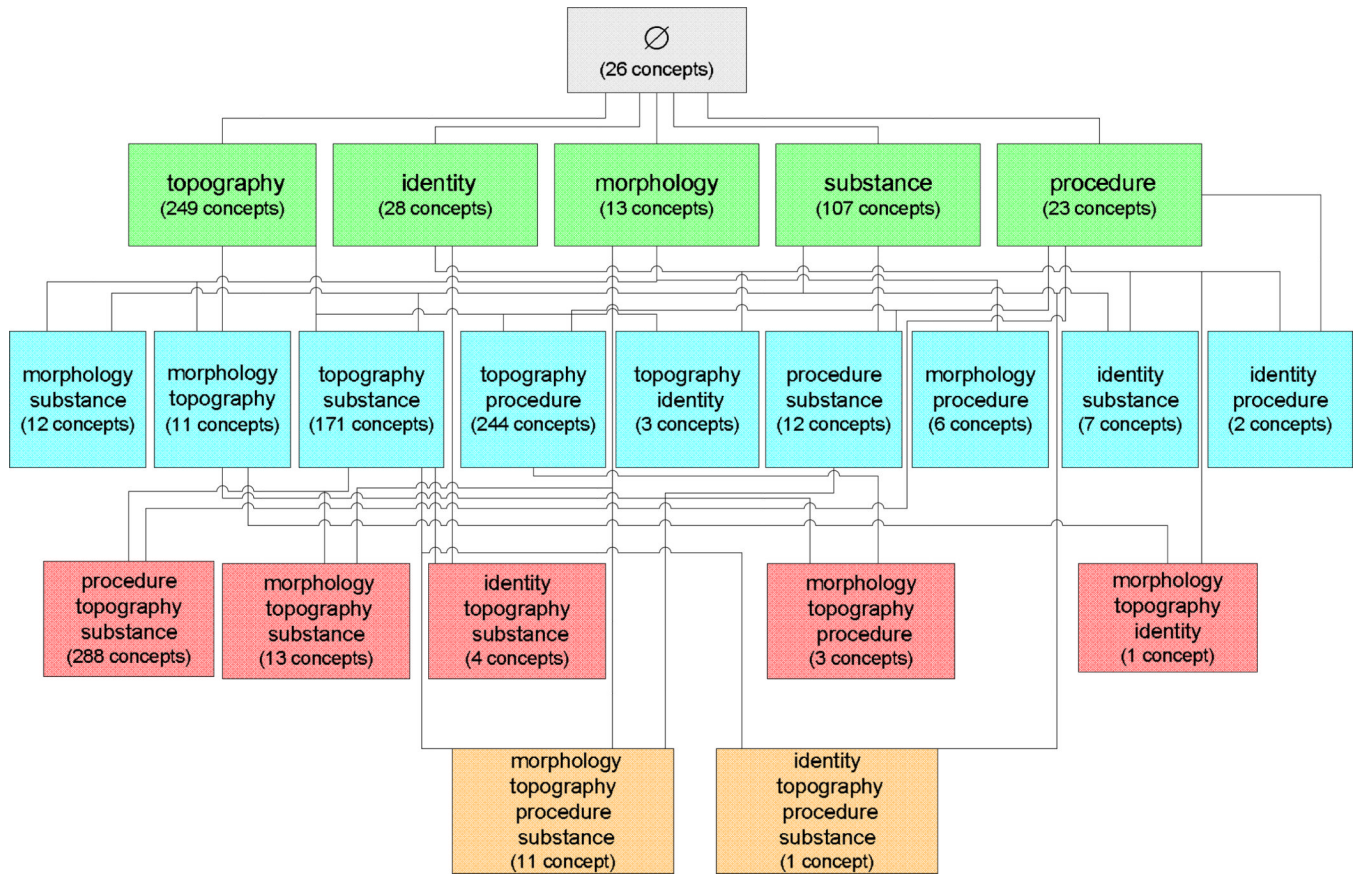


**Figure 5.** Flow diagram summarizing the audits of SNOMED 2007 and 2009. The numbers in each box represent the respective numbers from the 2007 and 2009 versions of SNOMED. For example, “1,056/1,236” in the top box indicates that there are 1,056 concepts in the Specimen hierarchy in SNOMED 2007 and 1,236 in 2009

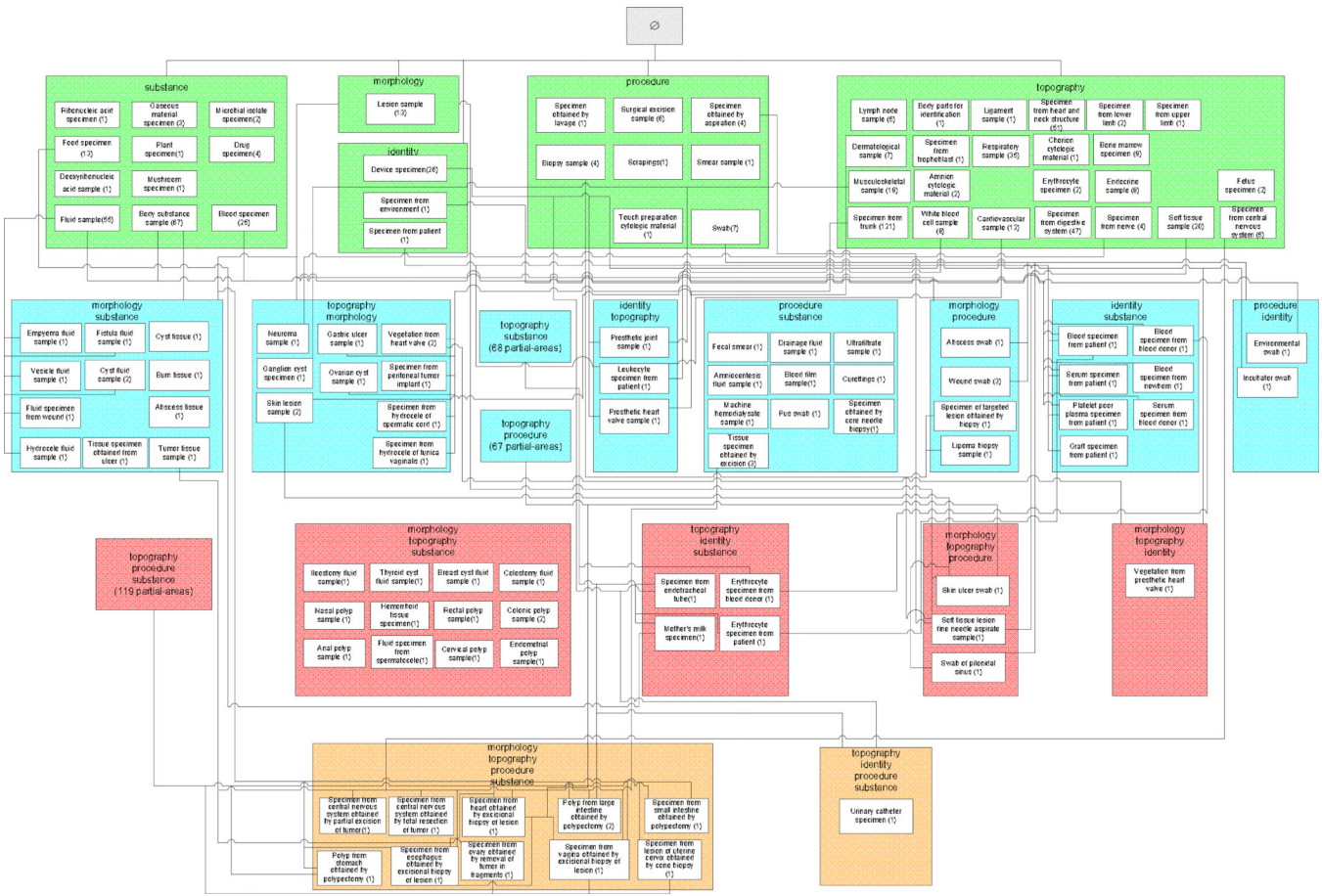




**Figure 6.** A portion of the disjoint partial-area taxonomy for the area {*substance*} (July 2007). The multi-colored boxes are the d-partial-areas containing overlapping concepts



**Figure 7.** Area taxonomy for SNOMED’s Specimen hierarchy (July 2009). The boxes are the areas, and the lines are the *child-of* relationships



**Figure 8.** Partial-area taxonomy for the Specimen hierarchy (July 2009). The main boxes are the areas, and the lines are the *child-of* relationships. The embedded boxes are the partial-areas

**Table 1**Auditing results for overlapping concepts of {*substance*} arranged by d-partial-area

<b>D-partial-area</b>	<b><i>V</i></b>	<b><i>V<sub>err</sub></i></b>	<b><i>E<sub>root</sub></i></b>	<b><i>E</i></b>
Exhaled air specimen	1	0	0	0
Inhaled gas specimen	1	0	0	0
Fecal fluid sample	1	0	0	0
Acellular blood (serum or plasma) specimen	1	1	1	1
Serum specimen from blood product	1	1	3	3
Serum specimen	2	0	0	0
Plasma specimen	4	1	1	1
Body fluid sample	11	3	17	19
Blood specimen	13	5	2	7
<b>Total:</b>	<b>35</b>	<b>11</b>	<b>24</b>	<b>31</b>

*V* = # overlapping concepts; *V<sub>err</sub>* = # erroneous overlapping concepts; *E<sub>root</sub>* = # errors at the overlapping root; *E* = total # errors at overlapping concepts

Table 2

Auditing results for overlapping concepts by area

Area	C	V	D	V <sub>err</sub>	E	E/V <sub>err</sub>	E/V	D <sub>err</sub>	E <sub>root</sub>	E <sub>root</sub> /D <sub>err</sub>	D <sub>err</sub> /D	(V <sub>err</sub> - D <sub>err</sub> )/(V - D)
substance	81	35	9	11	31	2.8	0.89	5	24	4.8	56%	23%
topography	333	116	52	71	110	1.6	0.95	39	62	1.59	75%	50%
procedure	20	3	3	3	9	3.0	3.0	2	9	4.5	66%	N/A
identify	20	1	1	0	0	N/A	0	0	0	N/A	0%	N/A
topog., proc.	380	6	6	4	9	2.3	1.5	4	9	2.3	66%	N/A
topog., morph.	18	1	1	0	0	N/A	0	0	0	N/A	0%	N/A
<b>Total:</b>	<b>852</b>	<b>162</b>	<b>72</b>	<b>89</b>	<b>159</b>	<b>1.8</b>	<b>0.93</b>	<b>50</b>	<b>104</b>	<b>2.1</b>	<b>69%</b>	<b>43%</b>

C = # concepts; V = # overlapping concepts; D = # overlapping roots; V<sub>err</sub> = # erroneous overlapping roots; E = total # errors at overlapping concepts; D<sub>err</sub> = # erroneous overlapping roots; E<sub>root</sub> = # errors at the overlapping roots; N/A = Not applicable

**Table 3**Results of auditing intersections involving partial-area *Tissue specimen*

<b>Second Partial-Area</b>	<b>C</b>	<b>V</b>	<b>V<sub>err</sub></b>	<b>V<sub>err</sub> / V (%)</b>
<i>Specimen from eye</i>	18	12	8	67
<i>Ear sample</i>	2	1	0	0
<i>Specimen from breast</i>	8	4	2	50
<i>Cardiovascular sample</i>	13	3	1	33
<i>Products of conception tissue sample</i>	12	1	1	100
<i>Genitourinary sample</i>	73	20	17	85
<i>Dermatological sample</i>	6	2	0	0
<i>Specimen from digestive system</i>	74	29	18	62
<i>Musculoskeletal sample</i>	35	22	15	68
<i>Respiratory sample</i>	41	6	5	83
<i>Endocrine sample</i>	12	3	0	0
<i>Specimen from central nervous system</i>	4	1	0	0
<i>Spec. from thymus gland</i>	2	1	0	0
<i>Specimen from trophoblast</i>	2	1	0	0

C = # concepts; V = # overlapping concepts; V<sub>err</sub> = # erroneous overlapping concepts;

**Table 4**

Auditing results for overlapping concepts vs. control sample (Phase 1)

	$C$	$E$	$E/C$	$C_{err}$	$C_{err}/C$ (%)	$E/C_{err}$
Overlapping	162	158	0.98	89	55	1.8
Control Sample	85	31	0.36	25	29	1.2

 $C$  = # concepts;  $E$  = # errors;  $C_{err}$  = # erroneous concepts;

**Table 5**

Sample of error types of overlapping concepts for July 2009 release

Concept	Partial-areas	Error Type(s)	Correction(s)
<i>Serum specimen from blood product</i>	<i>Blood specimen / Fluid sample / Body substance sample</i>	Missing parent	Add parent: Blood specimen from blood product
<i>Dentin specimen</i>	<i>Specimen from digestive system / Specimen from head and neck structure</i>	Incorrect Parent: Oral cavity sample	Correct parent: Specimen from tooth
<i>a.m serum specimen</i>	<i>Blood specimen / Fluid sample (specimen) / Body substance sample</i>	Missing relationship	Add relationship: TIME_ASPECT with the value of - am - ante meridiem
<i>Specimen from tooth</i>	<i>Specimen from digestive system / Specimen from head and neck structure</i>	Incorrect relationship target: Oral cavity structure	Refine with : Tooth structure
<i>Specimen obtained by fine needle aspiration procedure</i>	<i>Specimen obtained by aspiration / Biopsy sample</i>	Missing child	Add children: <ul style="list-style-type: none"> <li>* Breast fine needle aspirate sample;</li> <li>* Soft tissue lesion fine needle aspirate sample;</li> <li>* Specimen from heart obtained by fine needle aspiration procedure;</li> <li>* Specimen from kidney obtained by fine needle aspiration procedure</li> <li>* Specimen from thymus gland obtained by fine needle aspiration biopsy</li> </ul>
<i>Tissue specimen from placenta</i>	<i>Tissue specimen from genital system / Products of conception tissue sample</i>	Other error type: missing ancestor "Soft tissue sample"	Create a proper concept to parent it in the "Soft tissue sample" tree.



**Table 6**

Distribution of types of errors in the second phase of auditing overlapping concepts

Error Type	# Concepts
Missing parent	23
Incorrect parent	22
Missing child	6
Incorrect child	2
Missing relationship	55
Incorrect relationship target	2
Other error type	6

Table 7

Phase 2 auditing results for overlapping concepts by area

Area	C	V	D	V <sub>err</sub>	E	E/V <sub>err</sub>	E/V	D <sub>err</sub>	E <sub>root</sub>	E <sub>root</sub> /D <sub>err</sub>	D <sub>err</sub> /D	(V <sub>err</sub> - D <sub>err</sub> )/(V - D)
substance	107	48	15	28	36	1.29	0.75	8	11	1.38	53%	61%
topography	249	110	37	52	57	1.10	0.52	20	22	1.10	54%	44%
procedure	23	2	1	1	1	1.00	0.50	1	1	1.00	100%	0%
topog., proc.	244	29	16	28	38	1.36	1.31	15	19	1.27	94%	100%
topog., subst.	171	5	4	3	4	1.33	0.80	3	4	1.33	75%	0%
subst., topog., proc.	288	16	14	15	25	1.67	1.56	14	23	1.64	100%	50%
<b>Total:</b>	<b>1,082</b>	<b>210</b>	<b>87</b>	<b>127</b>	<b>161</b>	<b>1.27</b>	<b>0.77</b>	<b>61</b>	<b>80</b>	<b>1.30</b>	<b>70%</b>	<b>54%</b>

C = # concepts; V = # overlapping concepts; D = # overlapping roots; V<sub>err</sub> = # erroneous overlapping roots; E = total # errors; D<sub>err</sub> = # erroneous overlapping roots; E<sub>root</sub> = # errors at the roots;

**Table 8**

Auditing results for overlapping concepts vs. control sample (Phase 2)

	$C$	$E$	$E/C$	$C_{err}$	$C_{err}/C$ (%)	$E/C_{err}$
Overlapping	210	161	0.77	127	60	1.27
Control Sample	111	14	0.13	14	13	1.00

 $C$  = # concepts;  $E$  = # errors;  $C_{err}$  = # erroneous concepts;