

Published in final edited form as:

J Biomed Inform. 2009 June ; 42(3): 413–425. doi:10.1016/j.jbi.2009.03.003.

A Review of Auditing Methods Applied to the Content of Controlled Biomedical Terminologies

Xinxin Zhu¹, Jung-Wei Fan¹, David M. Baorto², Chunhua Weng¹, and James J. Cimino^{1,3}

¹Department of Biomedical Informatics, Columbia University, New York, NY

²New York Presbyterian Hospital, New York, NY

³Laboratory for Informatics Development, NIH Clinical Center, Bethesda, MD

Abstract

Although controlled biomedical terminologies have been with us for centuries, it is only in the last couple of decades that close attention has been paid to the quality of these terminologies. The result of this attention has been the development of auditing methods that apply formal methods to assessing whether terminologies are complete and accurate. We have performed an extensive literature review to identify published descriptions of these methods and have created a framework for characterizing them. The framework considers manual, systematic and heuristic methods that use knowledge (within or external to the terminology) to measure quality factors of different aspects of the terminology content (terms, semantic classification, and semantic relationships). The quality factors examined included concept orientation, consistency, non-redundancy, soundness and comprehensive coverage. We reviewed 130 studies that were retrieved based on keyword search on publications in PubMed, and present our assessment of how they fit into our framework. We also identify which terminologies have been audited with the methods and provide examples to illustrate each part of the framework.

Keywords

terminology; auditing method; quality factor; knowledge source; manual; automated; systematic; heuristic

1. Introduction

The quality of a controlled terminology can be characterized from any of several different perspectives. The *design* of a terminology can, from the outset, determine much about the future capabilities of the terminology. Many aspects of terminology design have been identified and characterized as desirable or undesirable [1, 2]. Standards development organizations have paid much attention to creating guidelines for quality control in terminology development. For example, the ISO/TC215 WG3 (Health Informatics - Semantic Content) has been working on such guidelines,* and the latest American National

© 2009 Elsevier Inc. All rights reserved.

Mailing Address: 622 West 168th Street, VC-5 New York, NY 10032 Katie.Zhu@dbmi.columbia.edu Fax: 212-305-3302.

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.

*See Documents under <http://www.tc215wg3.nhs.uk/pages/default.asp>

Standards Institute guidelines for designing controlled terminologies (ANSI/NISO Z39.19-2005) serves as a comprehensive reference [3]. In some cases, there is lack of consensus about desirability of particular design features (for example, some desire multiple hierarchy [1, 2] while others feel it should be avoided [4]).

The *structure* of a terminology can be studied to determine whether it supports or contradicts the stated design principles of the terminology. For example, Logical Observation Identifiers Names and Codes (LOINC) is designed to have meaningless identifiers; its use of sequential integers with check digits satisfies this requirement [5]. Similarly, the relationships in the Unified Medical Language System (UMLS) are designed to be reciprocal; the MRREL file provides a mechanism for delivering this information, as described in [6].

Finally, the *content* of a terminology can be assessed to determine if is comprehensive and accurate from lexical and semantic (as opposed to structural) standpoints. For example, the list of all laboratory tests contained in LOINC can be evaluated to identify whether it in fact contains all the terms used by hospital laboratories.

To illustrate these distinctions, consider the assessment of a terminology with respect to multiple hierarchies. A terminology can be *designed* to include multiple hierarchies, but can be found to have a *structural* characteristic that interferes with true multiple hierarchies, such as the tree addresses used in the Medical Subject Headings (MeSH), as described in [7]. Even when the terminology has a high-quality structure to support multiple hierarchies, its content might be deficient if a term that should have two parents is found to have only one.

A great deal of thoughtful planning is generally applied to terminology design, construction and maintenance. Design decisions (however controversial) are made with care, while the structural integrity can generally be guaranteed through good programming and database design. The quality of a terminology's content, on the other hand, is often not immediately obvious. However well-intentioned, authoritative, and cautious a terminology builder may be, there is always the chance for errors of omission or commission.

At the very least, good quality assurance practices dictate that assessment for errors should be a standard part of terminology management [8]. However, these practices, collectively referred to as *auditing*, can be challenging. Manual, expert review of a large terminology may provide little confidence that all errors have been detected. For example, any manual attempt to identify redundant terms in a large (>100,000 term) terminology will likely require memory that goes beyond human capacity.

To address this problem, informatics researchers and terminology developers have devised a number of methods to audit terminologies in systematic ways. Their methods often use knowledge in the terminology itself to perform the assessment and use computers to support - and in some cases entirely automate - the assessment. This paper reviews the major efforts in this area and organizes them into a framework that considers the aspects of terminology content that are audited, the methods used in the audits, and the terminology content that is employed to actually support the auditing process.

2. Quality Factors to be Audited

We first identify quality factors by which terminology content can be assessed. We consider *intrinsic* quality factors that are inherent to terminology content and that can be audited independently from external reference standards. Intrinsic factors include concept orientation, consistency, soundness, and non-redundancy. We also consider *extrinsic* quality factors that are contingent on *comprehensive coverage* of external user requirements, domain-specific contextual needs, or other external reference standards. Both types of

quality factors can be further applied to the content and knowledge structure of a terminology. We describe these factors below and summarize them in Table 1.

2.1. Concept Orientation (Intrinsic)

Concept orientation refers to the principle that the units of discourse in a controlled terminology should actually be the meanings (or concepts), rather than the human-readable labels (that is, the terms) that are enumerated in a terminology with the intention of conveying the meanings. In some of the terminological literature and documentation, the words “concept” and “term” are used inconsistently or interchangeably. In this paper, we will generally use “concepts” when we are referring to the meanings being conveyed and “terms” when we are referring to the character strings that are names for concepts. In the case of the UMLS, where strings are grouped together based on their meanings and the groupings are called “concepts” and given unique identifiers, the distinction between terms and concepts is less clear; we refer to these groupings as “UMLS concepts”.

While the precise intended meanings of terms in a terminology may be difficult to audit directly, concept orientation at a minimum requires that the items in a terminology must correspond to at least one meaning (“non-vagueness”) and to no more than one meaning (“non-ambiguity”) [1]. For example, Poison Ivy can be used to refer to a disease or a plant [9], which can cause concept ambiguity. Methods aimed at auditing concept orientation have therefore sought to identify undefined terms and terms with multiple meanings (polysemy) that are nevertheless mapped to a single term identifier (implying a single concept) [9-11].

2.2. Consistency (Intrinsic)

Consistency refers to adherence to semantic and/or linguistic rules for representing terms in a terminology. Linguistic inconsistency may occur when the same lexical modifier applied to different terms implies different relationships to the original terms. For example, there is a hierarchical relationship between congenital porphyria and porphyria, but a sibling relationship between congenital Addison’s disease and Addison’s disease (as provided by SNOMED) [12]. Inconsistency may also occur among hierarchical relationships in ways that are unrelated to lexical phenomena. Most terminologies include some kind of hierarchical relationship among their terms. These may be in the form of is-a relationships, broader-narrower relationships, part-whole relationship, or a mixture of these; in some cases, the meanings are not explicitly or implicitly determined [1]. We consider all of these representations to convey some type of organization into groups of terms with similar properties – that is, classification – including the semantic type assignments in the UMLS. Whatever the intent of such classifications within a particular terminology, the consistent application of that classification is a desirable quality. [13]. In the UMLS, for example, “adulthood” and “old-age” are assigned both semantic types “Idea or Concept” and “Age Group”, while other terms, such as “childhood”, “juvenile”, and “young adults” are assigned only to the latter [9]. The UMLS is particularly sensitive to classification inconsistencies, since it attempts to combine information from multiple terminologies that may be consistent with themselves but not with each other. For example in the Computer Retrieval of Information on Scientific Projects (CRISP) terminology, “colorectal neoplasm” is broader than “colon neoplasm”, while in MeSH the relationship is reversed; the UMLS contains both relationships, such that the two terms appear to be each other’s parent [14].

2.3. Non-redundancy (Intrinsic)

Non-redundancy refers to the absence of unwanted repetitive information. Typical redundancy errors include redundant terms, where terms with the same meaning are represented with separate, independent identifiers (rather than being linked as synonyms). For example, in the UMLS the following pairs are redundant [10]: C0000760: ABNORMAL

PAP SMEAR, C0240660: PAP SMEAR ABNORMAL, and C0002965: Angina, Unstable, C0235466: ANGINA UNSTABLE.

Redundant classifications can also occur in a terminology. Redundant is-a assignments can be found as assignments wherein a term is identified as being in two classes, with the first class being a descendant of the second (the second assignment is implied by the first assignment due to the transitivity of the is-a relationship between classes) [13, 15]. For example, in the UMLS, “year” has two semantic types: “Temporal Concept” and “Idea or Concept”, but the former is a child of the latter. Therefore, these semantic type assignments are redundant.

2.4. Soundness (Intrinsic)

Soundness refers to the accuracy of the knowledge represented in a terminology. Typical terminology features that are audited for their soundness include term classifications [16], and term names and definitions [17]. There are 9,296 UMLS concepts with the semantic type “Pharmacologic Substance” that have as children UMLS concepts of type “Clinical Drug”; however, the former refers to chemicals and the latter refers to manufactured objects consisting (in part) of chemicals [6]. This is an example of incorrect classification.

2.5. Comprehensive Coverage (Extrinsic)

Comprehensive coverage is, in some ways, the converse of soundness. While soundness refers to the accuracy of what a terminology *does* contain, comprehensive coverage refers to what a terminology *should* contain. Auditing for comprehensive coverage, therefore, must be related to the terminology’s intended domain in the world outside the terminology itself. Auditing methods assess a number of different aspects of terminology coverage, including: 1) Terms and synonyms. Some composite terms exist for which the derivative terms are missing. In the UMLS, for example, “monitor for seizure activity” and “seizure activity not present” exist, but “seizure activity” (an electroencephalographic finding) is missing [18]. 2) Defining attributes. As an example of incomplete definitions, the UMLS does not support formal definitions and includes narrative definitions for only a small portion of its concepts. 3) Hierarchical classification. In UMLS, “Inert Gas Narcosis” is an “Injury or Poisoning”, but based on common sense, it is also an “Occupational Disease”. The inference that “Injury or Poisoning” is-descendant-of “Disease or Syndrome” should be added to the semantic network. This is an example of missing Ancestor–Descendant [6] in the UMLS Semantic Network. 4) Non-hierarchical semantic relationships. For example, “cleft lip” is a “disease” but has no relation to “finding of appearance of lip” in UMLS [19].

2.6. Mixed Cases

Terminology errors may sometimes relate to multiple quality factors. For example, as mentioned previously, Gu, et al. [9] found that the UMLS concepts “adulthood” and “old-age” were classified as being of the semantic type “Idea or Concept” while other similar UMLS concepts such as “childhood”, “juvenile”, and “young adults” were not. This indicates inconsistent classification. At the same time, these UMLS concepts referred to multiple meanings, for example, the state of age and the group of people in that state, which violates non-ambiguity of concept orientation.

3. Knowledge Used to Perform Audits

Auditing a terminology can be considered a comparative process, in which the content of terminology is compared to some source of truth. There are numerous potential sources of knowledge that can be *used* to audit the knowledge a terminology *contains*, as both the literature and our own experience with Columbia University’s Medical Entities Dictionary

(MED) clearly demonstrate [20]. In a certain respect, all of these sources lie outside of the actual targeted internal representation in the terminology. However, for the purposes of this discussion, we define *intrinsic knowledge* as information derived from the classification scheme, hierarchy, semantic relationships to other terms, or attributes such as lexical information that are present within the terminology itself. We define *extrinsic knowledge* as a comparative standard deriving from an outside source, such as other terminologies, user requirements, or human expert knowledge. A recent example of automated auditing using extrinsic knowledge can be found in [21]. The National Cancer Institute Thesaurus (NCIt)'s Gene hierarchy was audited for missing "has associated process" relationships using two sources, the NCBI Entrez Gene database and the NCIt's Biologic Process hierarchy, to check hierarchical relationships. A given audit process could potentially use intrinsic knowledge, extrinsic knowledge, or a combination of both.

When a terminology seeks to maintain synchronization with some other terminology, as is the case with the UMLS and the MED, the external terminology itself becomes an extrinsic source. Maintaining synchronization with source terminologies entails a persistent ongoing audit to determine what has been added, changed or deleted. The UMLS must remain synchronized with over 100 standard source terminologies, while the MED must remain synchronized with terminologies from multiple ancillary systems (e.g., multiple laboratory, pharmacy, clinical documentation, and radiology systems) on at least a weekly basis, as well as the annually updated set of ICD9 diagnoses and procedures [22-24].

Although a source terminology may provide sufficient information for the auditing process, occasionally additional external sources are required. For example, if a new laboratory test term is to be added to a terminology such as the MED or LOINC, the terminology maintainer must consult a trusted external information source (whether published or based on his or her own knowledge) to determine whether it should be assigned to an existing class in the terminology or if a new class is needed.

External sets of clinical terms have frequently been used to assess coverage, in what are known as coding exercises, which are generally semi-automated or manual attempt to find terms. Targets for these studies have included SNOMED-CT [25], UMLS [26, 27] nursing taxonomies [28], Read Codes [29], and multiple terminologies in large studies [30, 31].

Expert user review is another external knowledge source for audits. Whether by direct perusal of a terminology [32], post hoc review to correctly classify and semantically link external data (discussed above), or to review output of audits based on intrinsic factors (discussed below), expert knowledge is the final arbiter in many evaluations.

The intrinsic knowledge that can be used in auditing processes includes hierarchical relationships, the non-hierarchical semantic relationships between terms, and the lexical knowledge about terms. We start with an example from the experience of two of the authors (DMB and JJC). Because the MED is integrated with a live clinical information system, updating content is a multistage process, in which changes are first entered into an editing environment, then a test environment, and finally to the production environment. During each editing cycle, there are more than 25 automated audits that test for primary violations of terminology rules and structure [33]. These audits are inextricably tied to the initial design. Some use hierarchical information rules ("cannot remove the last parent - all terms must have a parent", "there can be no hierarchical cycles", "a term should not have two hierarchically related parents"), some use semantic relationships rules between terms ("all semantic slots must have a reciprocal", "no redundant semantic relationships"), and some use rules combining classification and semantics ("a term cannot have two hierarchically related values in the same semantic slot; the more specialized value should be used - i.e.,

refinement is enforced”). This is a short sampling of the checks on more commonly identify errors.

A number of distributed biomedical terminologies have been assessed for similar adherence to design rules or terminological principles using intrinsic knowledge. These include assessment for adherence to basic ontological principles of SNOMED-CT [34] and the Foundational Model of Anatomy (FMA) [35], assessment for internal consistency in terminologies such as the READ Thesaurus [36, 37], and various approaches to removing cycles in the UMLS [38, 39]. Outside the domain of biomedicine, a set of formal consistency checking rules based on intrinsic hierarchical and semantic inputs has been proposed for the WordNet™ 1.5 lexical database [40] that share similarities with the routine MED audits.

Intrinsic knowledge can be used for more than highlighting violations of design principles; it has also been used in a variety of interesting ways to correct or suggest correction to content. In general the correction of content using intrinsic knowledge involves the additional input of extrinsic knowledge, often in the form of an expert to manually review items brought to light by an automated process [41].

A repeating theme in the literature is the application of semantic relationship patterns to partition a terminology into more manageable pieces for manual review. Occurrence of concepts having identical relationships (or associations to other concepts) and hierarchy are then brought to the attention of human reviewers. Using intrinsic knowledge for this kind of partitioning has been applied to the MED [42, 43], SNOMED [44-46], and the NCI [8] to reveal issues such as missing classifications and redundancy.

Another example of this approach is to use knowledge in the UMLS to derive a metaschema for the UMLS Semantic Network by grouping semantic types with identical relationship sets [47, 48]. In [13, 49], assignment of UMLS concepts to multiple UMLS semantic types, especially when those types are considered to be in different groups of semantic types of a metaschema [9, 50] or mutually exclusive [10], has been used to suggest classification errors, ambiguity and inconsistency. These methods almost always use expert manual review as a follow-up knowledge source. The MED regularly uses internal semantic relationships to determine classification; Cimino, et al. [51] described an algorithmic approach to enrich hierarchical structure in the MED and determine the correct location in the hierarchy for newly added terms.

Lexical information embedded in terminologies has also been used for auditing processes. In the simplest case, lexical information is used to fix lexical targets, such as spelling errors and uniqueness of term names, an audit performed with every MED update. Lexical information has also been applied to reveal other quality issues. For example, Campbell, et al. used term substrings in SNOMED to suggest classification omissions [52]. In other work, synonymous terms in the UMLS were used to compile a list of keyword synonyms that, in combination with semantic types, was used to detect redundancy [10].

Consistent use of linguistic phenomena, such as adjectival modifiers, has been used to assess potential inconsistency in the UMLS and SNOMED. In one study by Bodenreider, et al. [12], the intrinsic characteristics of lexical usage of adjective pairs such as “acute”/“chronic” and “primary”/“secondary” were reviewed in the context of known, extrinsic knowledge to reveal inconsistencies. In another lexical study, drug descriptions from four leading pharmacy system knowledge base vendors were compared across each field to determine lexical consistency of usage in the pharmacy domain [53]. In yet another application of intrinsic lexical knowledge, all terms, synonyms and headings contain the conjunctions “and” or “or” in SNOMED were assessed compared to editorial board policy on usage,

which specifies that “and” should imply logical AND (both must be present), “and/or” when one or both must be present and “either_or” should be used when one but not both must be present. Usage in this regard was found to be inconsistent [54].

4. Auditing Methods

The knowledge described in previous sections can be applied to the content of controlled biomedical terminologies in different ways. We summarize auditing methods into four major categories. The most straightforward but probably most labor-intensive method is *manual review* (with or without the support of a computerized user interface), by which a terminology reviewer (often a domain expert) audits the terminology with respect to certain quality factors. *Automated systematic methods* involve implementing knowledge into rule-checking programs that scan the terminology for potential problems with respect to particular quality factors, usually in a “batch” mode, to identify errors and inconsistencies. The automated systematic methods are generally reproducible and reduce the need for detailed, costly manual review. *Automated heuristic methods* involve use rules that make inferences about terminology content and then seek to identify those inferences that lead to illogical or inconsistent conclusions.

For the above three categories, we classify published reports based on the terminology attributes being audited:

- *Terms and concepts* -- we pool all quality factors related to terms and concepts under this category, including orthographic accuracy, comprehensive coverage, non-vagueness, non-ambiguity, and non-redundancy.
- *Semantic classification* -- we define semantic classification as a terminology’s taxonomic scaffold (with or without “is-a” structure), for which classes the terms of the terminology are assigned.
- *Semantic relationships* -- we define semantic relationships as relations (either hierarchical or non-hierarchical) between terms that convey meaning.

In addition to the above three major categories, we discuss separately some *high-level change management methods* that deal with logistics issues involved in auditing terminologies.

In the following subsections, we examine examples of each of these methods from the published literature and our own experience, with consideration of the terminology attributes audited and the knowledge used to support the processes.

4.1. Manual Auditing Methods

4.1.1. Terms and Concepts—Because controlled medical terminologies are typically in a constant state of development, expansion and refinement, a formal representation and taxonomy were proposed by Cimino and Clayton [22, 24, 55] to characterize changes in a terminology, based on its syntactic properties (i.e., addition, deletion, name change, and code change) in order to characterize the semantic changes they represented. For example, a name change could represent a change in meaning (major name change) or not (minor name change). Once the semantic changes were characterized, they could be dealt with formally to maintain concept orientation and concept permanence. For example, a major name change would generally require the retirement of an existing concept (corresponding to the previous version of the term) and creation of a new one (corresponding to the new version). Those changes were reconciled through manual review by domain experts.

Subsequently, Fischer [40] developed formal rules to check redundancy and consistency in a lexical database. Later, Wroe collaborated with Cimino and Rector [56] to model and integrate different drug formulation terminologies based on formal definitions that were manually created using the OpenGalen knowledge representation scheme. Terms in each terminology could be compared to corresponding ones in other terminologies based on the definitions, to identify inconsistencies between the two. Recently, Smith et al. presented a case study of applying similar formal principles to the Gene Ontology (GO) [16].

Several other manual auditing methods involved the coding of clinical records with terminologies, with their adequacy and accuracy judged by experts. For example, Chute, et al. [30] assessed various clinical terminologies for their content coverage by parsing 14,247 words into 3,061 distinct concepts. These concepts were grouped into Diagnoses, Modifiers, Findings, Treatments and Procedures, and Other. An attempt was made to manually code each concept in ICD-9-CM, ICD-10, CPT, SNOMED III, Read V2, UMLS 1.3, and NANDA, with the result scored as “no match”, “fair match”, and “complete match”. Coding consistency was assured by a secondary reviewer.

To audit attributes of controlled clinical terminologies, such as completeness, term definitions, “clarity” (represented as the inverse of the rate at which the same data might be coded in duplicate ways), consistency in clinical taxonomy, and administrative mapping, a Computer-based Patient Record Institute (CPRI) workgroup [31] assembled 1929 source records based on an initial expert evaluation and an organized consensus discussion within the workgroup. The source records were coded in each terminology scheme by an investigator and checked by the coding scheme owner. The coding was then scored by an independent panel of clinicians for acceptability on a Likert scale. The investigator for each scheme exhaustively searched a sample of coded records for duplications.

Many studies have been conducted to audit comprehensive coverage of a terminology in different settings. Humphreys, et al. [27] carried out a distributed national experiment using the Internet and the UMLS to determine the extent to which a combination of existing machine-readable health terminologies covered the terms and concepts needed for a comprehensive controlled terminology for health information systems. Several studies were performed within specific clinical information system settings: Wasserman and Wang [57] evaluated the breadth of terms and concepts for the coding of diagnosis and problem lists by clinicians within a physician order entry system; Kushniruk, et al. [32] conducted an observational study to identify concepts missing from an outpatient information system.

As Rector [58] argued, explicit information is key to most existing coding and classification systems, and different types of content must be separated based on their conceptual, linguistic, inferential and pragmatic correctness. Many research studies have evaluated comprehensive coverage of different domain-specific terminologies. Cieslowski, et al. [59] and Moss, et al. [60] worked on nursing terminology integration and evaluation. Chiang, et al. [61] extracted ophthalmology concepts from patient reports and manually reviewed their coverage in five terminologies. Warnekar and Carter [62] assessed HIV term coverage of a commercial terminology. Smith and Kumar [63, 64] audited semantic appropriateness of term names and definitions of the GO terms. While combining laboratory data sets across terminologies, Baorto, et al. [65] used the LOINC knowledge model to identify missing concepts and synonyms. Zhu, et al. [66] created a terminology model for the missing acupuncture terms in the UMLS.

4.1.2. Semantic Classification—By examining violations of ontological semantics, Kumar and Smith [64] were able to assess the appropriateness of semantic classification of GO terms. Similarly, Schulze-Kremer, et al. [67] applied five ontological principles to audit

knowledge classification in the UMLS semantic network, and Smith, et al. [16, 63] applied a set of ontological principles against GO to audit the assignment of hierarchical relations. Several years later, Mendonça, et al. [54], Spackman and Reynoso [68] used formal ontological definitions to audit misclassification in SNOMED, and Bodenreider, et al. [34] investigated subsumption in large description logic-based biomedical terminologies based on unique ontological principles.

Specifically, these ontological principles specified that all relationships of a parent class must either be inherited by each child or refined in the child, and refinement from parent to child should uniquely result in every case either from refinement of the value of a common role or introduction of a new role.

In a similar way, a qualitative, rather than a quantitative analysis of the NCI was performed by Ceusters, et al. [17] in order to assess NCI's ontological principles. Using an OWL-representation, their inspection of the system was performed breadth first, top down, entry-by-entry to detect inconsistencies with respect to the term-formation principles used, the underlying knowledge representation, and missing or inappropriately assigned textual and formal definitions.

4.1.3. Semantic Relationships—Schulz and Hahn [69] proposed a semi-automatic knowledge engineering approach for converting the human anatomy and pathology portion of the UMLS Metathesaurus into a terminological knowledge base. Their approach consisted of an integrity check of the emerging taxonomic and partonomic hierarchies, and elimination of terminological cycles and inconsistencies. They used LOOM to represent the Metathesaurus followed by a medical expert review. In detecting missing and incorrect is-a, part-of, and has-part relations, special attention was paid to the proper representation of part-whole hierarchies, while running experiments on 164,000 UMLS concepts and 76,000 relations in the terminological knowledge base. Their approach provided a formal description-logic framework to support taxonomic and partonomic reasoning. Consequently, Arts, et al. [70] used a semi-automated method to evaluate the structure of diagnoses terms used in intensive care. A description-logic reasoner was designed to find incorrect relations, along with manual review of relations performed by domain experts.

Table 2 summarizes the quality factors and knowledge sources in studies that applied manual auditing methods to various terminologies.

4.2. Automated Systematic Auditing Methods

4.2.1. Terms and Concepts—To assure unambiguous concept representation, Schulz, et al. [37] encoded simple rules into a system used to manage Read Codes, in order to control the uniqueness of concept identifiers (primary keys) and assure that each concept had a unique preferred name. Cimino [10] processed the UMLS concept strings and created an index consisting of normalized and synonymous lexical tokens to search equivalent and possibly duplicate UMLS concepts; the specificity was further improved by applying constraints based on semantic types. For example, “Angina, Unstable” and “ANGINA UNSTABLE” were found to be duplicates with different concept identifiers. Starting with a similar approach, Hole and Srinivasan [80] introduced more sophisticated algorithms for normalizing terms and enriching the lexicon of synonymous tokens so as to increase the sensitivity of their methods for identifying duplicate UMLS concepts.

Ceusters, et al. [19] identified duplicate concepts in SNOMED-CT by using a commercial medical ontology, LinKBase, and its associated search algorithm based on flexible string matching and ontology relation traversal; SNOMED-CT terms that mapped to the same LinKBase concept were considered redundancy-prone. For frame-based terminologies

(again using Read Codes as a case study), Schulz, et al. [37] searched concepts with identical definitions (attribute-values) and considered them possible duplicates.

Rogers, et al. [81] proposed a method that translated Read Codes into a description logic (DL)-based representation and applied a DL reasoning program to audit the formal definitions of the terms; problems such as missing attributes in the definitions and missing inherited attributes from parent terms to child terms were found. Similarly, Cornet and Abu-Hanna [82, 83] audited the formal definitions of the Diagnoses for Intensive Care Evaluation (DICE) terminology by translating its frame-based representation into DL and applied a DL reasoning engine to check for problems such as redundant and ambiguous terms.

To audit DL-based terminologies (using the NCI as a case study), Min, et al. [8] proposed a method to partition a terminology's classification into networks containing terms with identical DL roles (e.g., *has initiator process*), organized into single-rooted sub-networks. In particular, their method assumed that smaller subnetworks of networks with few subnetworks were highly error-prone. The underlying idea is that concepts with a rare combination of DL roles and classification are highly error-prone. Using their approach, they were able to find missing synonyms, missing concepts, and duplicate terms. For auditing more complicated DL-based terminologies such as the multi parented Specimen hierarchy of SNOMED-CT, Wang, et al. [45] advanced Min's method by refining the procedure of generating single-rooted sub-networks, explicitly differentiating whether each DL role of the sub-networks is inherited or newly introduced. They also added the assumption that sub-networks containing many DL roles but few terms (i.e., semantically specific) and the assumption that sub-networks with multiple inherited roles (i.e., semantically heterogeneous) are error-prone [46] or overlapping sub-networks [84]. They found inaccurate concept naming and incorrect synonyms, in addition to missing synonyms, missing concepts, and duplicate terms.

4.2.2. Semantic Classification—Cimino [10] checked UMLS concepts assigned to mutually exclusive semantic types, assuming those classifications are error-prone. For example, if a UMLS concept is classified both as “Animal” and as “Plant”, then one of the semantic type assignments should be wrong. Implementing the above principle from a preventive perspective (which can be considered pre-release auditing), Schulz, et al. [37] imposed restrictions in Read Codes to disallow assigning a term to semantically exclusive classes. Geller, et al. [13] and Gu, et al. [49] advanced Cimino's method with an algorithm that refined the UMLS Semantic Network into pure and intersected semantic types so that incorrect, redundant, and missing classifications were more easily exposed. Similar to Geller, Gu, et al. [9, 50] grouped the semantic types into broader meta-types [47, 48] first and then checked meta-type intersections with small number of UMLS concepts, assuming that rare combinations between the different semantic groups strongly imply erroneous classification.

Another method proposed by Cimino [6, 10] for auditing erroneous UMLS semantic type assignments was based on the principle that the hierarchical relation between two semantic types should be consistent with the parent-child relations of the UMLS concepts assigned to the types, i.e., a child UMLS concept should always be assigned a semantic type no broader than the semantic type of its parent UMLS concept. For example, the concept “Lys-Lys” should not be classified as the semantic type “Organic Chemical”, because its parent “Dipeptides” is classified as the semantic type “Amino Acid, Peptide, or Protein”, which is the child of type “Organic Chemical”. This expected consistency was used to automatically identify suspicious concepts needing manual auditing in the extent (set of concepts) assigned a given semantic type [85].

Applying a similar principle, Peng, et al. [15] audited redundant semantic type assignment for UMLS concepts that were assigned to both a particular semantic type and a parent (or ancestor) of that type, taking the parent type as redundant according to the rule of semantic type assignment specificity [86]. Fan, et al. [87] built automatic classifiers using lexical features from the UMLS concept strings and contextual features from a PubMed corpus to reclassify the UMLS concepts into broad classes. Their method found erroneous and missing classifications by checking the disagreement between their broader classification and the original semantic types.

4.2.3. Semantic Relationships—Schulz, et al. [36] audited the hierarchical relationships in the Read Codes by automating two rules: 1) the attributes of a child term should be the same as or more detailed than that of its parent term (this helps audit the correctness of the is-a relations) and 2) a term with more detailed attributes than another term should be considered a child of that term (this helps audit the completeness of the is-a relations). Campbell, et al. [52] examined the problem of missing hierarchical and non-hierarchical semantic relations in SNOMED by using lexical algorithms to suggest the existence of relationships between terms with common substrings.

Ceusters, et al. [19] used two algorithms to audit incorrect and missing relations in SNOMED-CT: 1) a DL-based classification algorithm and 2) a search algorithm that estimated semantic distance by implying correct subsumption relations.

The methods by Cimino [6, 10] mentioned above that examine consistency of semantic classification of UMLS concepts was used simultaneously to audit erroneous parent-child relations between them. Cimino [10] also suggested inferring non-hierarchical semantic relations between UMLS semantic types from the relations between the terms specified by source terminologies to improve the completeness of the semantic relationships.

Bodenreider, et al. [34] audited the hierarchical relations of SNOMED-CT by automatically checking against four ontological principles: 1) each hierarchy must have a single root, 2) each class (except for the root) must have at least one parent, 3) non-leaf classes must have at least two children, and 4) each child must differ from its parent and siblings must differ from one another.

Due to partition of a hierarchy of DL-based terminology such as NCIt and SNOMED-CT into sub-networks of identical DL-roles, [8, 45] support detecting of wrong and missing roles. Furthermore, a more refined partition of those sub-networks into single rooted sub-network helps to highlight wrong and missing hierarchical relationships.

The following table (Table 3) summarizes the quality factors and knowledge sources in studies that applied automated systematic auditing methods to various terminologies.

4.3. Automated Heuristic Auditing Methods

4.3.1. Terms and Concepts—Cimino and Barnett [94] created frame-based definitions manually, and automated a translation method for comparing a source term with all terms in each of the other target terminologies. The approach was applied to cardiac procedures in ICD9-CM, MeSH, SNOMED and CPT, with a score produced based on semantic distance. The scores were then ranked and manually reviewed.

To identify synonymy and near-synonymy, Barrows, et al. [95] explored lexical and morphologic text matching techniques to map clinically useful terms into a controlled medical terminology. Hole and Srinivasan [80] attempted to discover missed word and

phrase synonyms in a large concept-oriented metathesaurus through lexical matching, selective algorithms, and expert reviews.

Tulipano, et al. [96] used manually created knowledge-based descriptions to represent molecular imaging terms, which were then automatically mapped to GO. This process was created to address issues related to synonymy, redundancy, and ambiguity in concepts mapping. Recent work from Huang, et al. [97] introduced ways to generate new synonyms. Terms with multiple words were decomposed into single words. Synonyms for single words were identified; these provided the basis for new multiword terms to be reconstructed. In [98], techniques for limiting the combinatorial explosion caused by substituting WordNet synonyms for the single words of multiword-terms are introduced. Patel and Cimino [99] used network analysis to support decompositional terminology translation in order to identify, using a clustering coefficient, those primitive concepts that should be related to more complex concepts.

Bodenreider, et al. [12] assessed the systematic use of linguistic phenomena for both lexical and semantic features in SNOMED and the UMLS Metathesaurus. Frequently co-occurring adjectival modifiers were identified syntactically and studied in combination with the contexts of each modifier. Bodenreider, et al. [14] also evaluated the content coverage of the UMLS with an attempt to find exact matches first, followed by normalization and semantic incompatibility checking. Five broad classes of UMLS concepts were extracted using their system (LocusLink) and mapped to the UMLS Metathesaurus. The search also covered contents of gene products, phenotypes, molecular functions, biological processes, and cellular components.

4.3.2. Semantic Classification—Cimino, et al. [100] automated term subsumption, using manually created knowledge, followed by manual review of the system's suggestions for subclass partitioning, that is, the creation of new subclasses and inclusion of terms in those subclasses. The subsumption rules suggested partitioning large classes based on characteristics of subsets within the class, for example, chemical tests could be separated into subsets such as hormone tests, lipid tests, drug tests, etc. Semantic definitions for the new terms were created and added to the knowledge-base through a combination of automatic and manual means. For instance, in order to partition the large class "Chemistry Laboratory Tests", the system found that many children had a "substance measured" relation to a drug term, while many others had relations to non-drug chemical terms. The system therefore proposed a new semantic class "Drug Measurement Tests" to subsume the former, as a new subclass of "Chemistry Laboratory Tests". The same group [101] also built a rule-based terminology management system prototype in an object-oriented (OO) environment to check the inconsistency in classification.

Gu, et al. [49, 102] further experimented with OO modeling through the development of an OO database schema, using visualization techniques to identify places in the UMLS where errors and inconsistencies of semantic type assignments occurred. A different OO modeling [42, 43] was applied to the MED yielding an upper level schema, which helped highlighting ambiguity and inconsistency in the MED modeling. Moreover, to enhance comprehensibility and usability of a terminology system, Gu, et al. [103] developed an OO representation that provided an abstract or skeleton view, using a theoretical paradigm and a methodology that partitioned schemas into manageably sized fragments (they refer to this as a "forest subschema"). A set of rules was applied sequentially to simplify classification schemes such that OO classes were grouped into partitions that were relatively independent of each other and contained highly interrelated UMLS concepts. Classification errors could be detected by this step-wise heuristic model. Subclass hierarchy was refined by a medical domain expert in conjunction with a computer.

Recent work from Gu, et al. [9, 50] included a combination of an auditing technique and an expert review that determined the pure intersections of meta-semantic types of the metaschema [47], which yielded a compact abstract view of the UMLS Semantic Network. Ambiguity, incorrect classification, and inconsistencies were readily identified in the pure intersections that were examined. New conjugate and complex semantic types are suggested in [104] to better capture the semantics of concepts assigned multiple structurally viewed chemical semantic types. For such concepts the proper semantics expresses a chemical reaction or mixture rather than the typical conjunctive semantics of multiple semantic types.

4.3.3. Semantic Relationships—Zhang and Bodenreider [35] provided an operational definition of fifteen ontological principles and investigated the degree to which a large ontology of anatomy complied with them. Three rules were proposed to detect incompatible relationships. One such rule was that two terms cannot stand both in taxonomic and partitive relations, that is, for every pair of terms x and y , x and y do not have both IS-A and PART-OF relationships. For example, “Myocardium proper of right atrium” has both “REGIONAL PART OF” and “IS-A” relations with “Myocardium”, which are considered to be incompatible relationships.

Since redundant hierarchical relations are generally semantically consistent, semantic inconsistency detected in redundant hierarchical relations could be used as an indicator of potential mis-classification of one or both terms, and to trigger a review of these terms by editors of biomedical terminologies. Bodenreider [89] calculated an index of redundancy between pairs of hierarchically related terms to detect errors such as multiple pathways. Burgun and Bodenreider [105] proposed a method for assessing the semantic and hierarchical relationships between MeSH terms that co-occurred in literature citations. Zhang, et al. [93, 106, 107] expanded the UMLS Semantic Network into a multiple subsumption structure with a directed acyclic graph IS-A hierarchy, which allows a semantic type to have more than one parent. They argued that parent-child relations in the Metathesaurus implied the same in the Semantic Network, and finding the name of one type contained in the other implied the latter was a refinement of the former. New is-a relations were added to the Semantic Network, and new semantic types were created to support the multiple subsumption framework, from which new connected groups could be derived for the meta-semantic types of the metaschema for the expanded semantic network [108]. Two methodologies were applied to identify and validate new is-a relations: a lexical-based string matching process (involving names and definitions of various semantic types in the Semantic Network) and a process for converting the partition's disconnected groups of the Semantic Network [109] into connected ones.

Table 4 summarizes the quality factors and knowledge sources in studies that applied automated heuristic auditing methods to various terminologies.

4.4. High-level Change Management Methods

The auditing of contemporary biomedical terminologies often involves feedback from multiple users with different requirements. If we consider a temporal axis to auditing methods (i.e., auditing a terminology at different time points), it appears to be a dauntingly complex task that deserves robust management techniques to handle the vast amount of changes made during the auditing process. Those high level change management methods are different from the methods directly auditing terminology content. Research closely related to change management of terminologies has been found in the literature on ontology evolution and versioning. A typical method used in ontology evolution (or *versioning*) is to log the “who”, “what”, “when”, and “why” concerning the changes made. A benefit of detailed logging is the reversibility of changes if they are later found to be erroneous.

The circular evolution process by Stojanovic, et al. [117] provides a framework for considering the methods involved in change management. The six phases in their framework are: 1) Change capture (deciding what changes to be made; this corresponds to the output of the manual, automated systematic, and automated heuristic methods covered in the previous subsections), 2) Change representation (e.g., *extract_superconcept* is a change that occurs when a single concept is split into several subconcepts, with distribution of properties among them and their associated metadata; e.g., *auditor*, *timestamp*, and *reason_for_change*), 3) Semantics of change (e.g., to spot inconsistencies that could be introduced by certain change operations), 4) Change implementation (referring to both pre-implementation proofreading and the implementation itself), 5) Change propagation (applications or other terminologies that depend on the changed terminology need to be updated correspondingly), and 6) Change validation (referring to field-testing and retracting inappropriate changes based on the result of the testing).

The methods reviewed in the previous subsections are generally to be performed at certain fixed time points. However, considering Stojanovic's framework, failing to characterize potential risks at Phase 3 would unwittingly allow illegal changes to the terminology; in Phase 5, either failing to propagate correct changes or propagating incorrect changes would spoil the terminology. Additionally, if an error occurs in Phase 3 and is propagated in Phase 5, a subsequent Phase 1 can identify corrections to be made in a subsequent Phase 2.

Tools have been developed or augmented to support change management of auditing terminologies. Stojanovic and Motik [118] evaluated three generic ontology editors (Protégé, OntoEdit, and OilEd) and found that they satisfy some functions related to the temporal processes related to auditing, each with individual strength and weakness. The Karlsruhe ONtology Management Infrastructure (KAON) [117, 119] is another ontology editor that supports change management functions such as evolution strategies, consistency checking, and transparency of actions. Noy, et al. [120] developed the CHange and Annotation Ontology (CHAO) to facilitate collaborative change management, with emphasis on functions such as controlling access privileges and resolving conflicts in a multi-editor environment. CHAO was implemented as a set of plugins to Protégé and has been tested on the NCI. The usability of such collaborative editing tools can be generalized to multi-auditor tasks. Oliver and Shahar [23] proposed the CONCORDIA (CONcept and Change Operation Representation for DIAlects) model to address synchronization and logging requirements peculiar to maintaining a local terminology that diverges from an original shared terminology.

5. Discussion

Rogers has presented a framework for quality assurance methods applied to logic-based biomedical ontologies. He concludes that there are four aspects: philosophical validity, meta-ontological commitment, content correctness, and fitness for purpose [121]. Our review extends beyond ontologies to include all forms of controlled biomedical terminologies and characterizes the actual methods used to assess such properties. We find that most of the methods are focused on content correctness, in one form or another, and assume that good quality content implies fitness of purpose and vice versa.

Our framework divides terminology auditing methods into manual, systematic and heuristic methods. Some of these methods are best suited for assessing the terms and concepts in a terminology, while others are better suited for assessing semantic classification and relationships. Some methods can be used to audit multiple terminology attributes at the same time, e.g., simultaneously auditing problematic semantic type assignments and hierarchical relations between terms [10]. On the other hand, different methods can be used to audit the

same terminology attribute and cross-validate each other, e.g., identifying the mis-assignment of “Lys-Lys” to the semantic type Organic Chemical (see section 4.2.2) could be found through either the by a rule-based [10] or an automated reclassification [87] approach.

Each of these methods makes use of some knowledge. Most interesting are those that use knowledge that is intrinsic to the terminology itself, making the auditing process an exercise in introspection [7]. The disadvantage of such an approach is that it relies on “extra” knowledge that might not normally be present in a terminology. This may require terminology developers to exert effort beyond basic terminology construction (that is, enumerating and arranging the terms) to add such knowledge. The advantage of this approach is that automated auditing methods can be created that can operate independently from human expertise. Despite the added effort, many modern terminologies, such as LOINC, RxNorm, the FMA, and SNOMED, as well as the UMLS Metathesaurus, now include formal definitional information that can be exploited by auditing techniques. Thus, methods that rely on intrinsic knowledge are becoming increasingly practical.

Also interesting are the methods that are largely automated, since many terminologies are too large for complete comprehension by individual human auditors. Systematic methods (such as referential integrity) are already being incorporated directly into the terminology maintenance processes being used by standards development organizations. The methods described in this review offer additional ways to assure that terminologies are following their own rules.

Most interesting, in our opinion, is the combination of the use of intrinsic knowledge with heuristic methods. These methods can identify terminology errors that might escape deterministic automated methods and human-centered manual methods. Inherent in the nature of heuristic approaches is their imperfection, resulting in false positive findings. However, as Min, et al. [8] point out, a modest false positive rate is an acceptable trade-off when attempting to identify the parts of a terminology that should be scrutinized by human reviewers, especially when the availability of such reviewers is limited by resources or the limits of human attention.

As this review shows, there is now a rich tradition of formal auditing methods for controlled terminologies that has largely arisen in the last decade or so. Although much of the work began with theoretical approaches on limited data sets, the growing presence of rich biomedical ontologies is providing a fertile ground for further development of practical, usable auditing procedures. Indeed, much of the work presented here has resulted in feedback to UMLS, SNOMED, ICD9, MeSH and GO, leading to their improvement. Our own auditing methods are applied daily to the MED at Columbia University to support clinical, administrative and research information systems at New York Presbyterian Hospital. At the same time, the appreciation of high-quality controlled biomedical terminologies is currently on the rise, providing impetus for the actual use of such procedures to improve the terminologies that, in turn, are being increasingly relied upon to improve biomedicine.

This paper summarizes a wealth of terminology auditing methods and provides only the briefest descriptions of the work of many creative researchers. We hope that the framework we present serves to guide those who seek to develop their own auditing methods based on work that has come before. At the same time, the absence of entries in Tables 2-4 points to many heretofore unexplored opportunities for exploiting the rich knowledge in modern terminologies to support their improvement.

6. Conclusions

The last twenty years have been witness to a proliferation of terminology auditing methods that employ a variety of creative methods and exploit a variety of terminological knowledge to better evaluate and improve the terminologies that are emerging today as important components of biologic, clinical and public health systems. Much of the work has gone beyond the experimental stage to become key components of standards development and information system maintenance.

Acknowledgments

Dr. Cimino was supported in part by funds from the intramural research program at the National Institutes of Health (NIH) Clinical Center and the National Library of Medicine (NLM).

Appendix

Appendix:

Publications of Auditing Methods Applied to Controlled Biomedical Terminologies

References

- [1]. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med.* 1998; 37(4-5):394–403. [PubMed: 9865037]
- [2]. Chute CG, Elkin PL, Sherertz DD, et al. Desiderata for a clinical terminology server. *Proc AMIA Symp.* 1999:42–6. [PubMed: 10566317]
- [3]. National Information Standards Organization (NISO). Guidelines for the construction, format, and management of monolingual controlled vocabularies. 2005.
- [4]. Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol.* 2007; 25(11):1251–5. [PubMed: 17989687]
- [5]. McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clin Chem.* 2003; 49(4):624–33. [PubMed: 12651816]
- [6]. Cimino JJ, Min H, Perl Y. Consistency across the hierarchies of the UMLS Semantic Network and Metathesaurus. *J Biomed Inform.* 2003; 36(6):450. [PubMed: 14759818]
- [7]. Cimino, JJ.; Hripcsak, G.; Johnson, SB., et al. Designing an Introspective, Controlled Medical Vocabulary. In: Kingsland, LW., editor. *Proc of the Thirteenth Annual Symposium on Computer Applications in Medical Care*; Washington, D.C.. November, 1989; p. 513-518.
- [8]. Min H, Perl Y, Chen Y, et al. Auditing as part of the terminology design life cycle. *J Am Med Inform Assoc.* 2006; 13(6):676–90. [PubMed: 16929044]
- [9]. Gu H, Perl Y, Elhanan G, et al. Auditing concept categorizations in the UMLS. *Artif Intell Med.* 2004; 31(1):29–44. [PubMed: 15182845]
- [10]. Cimino JJ. Auditing the Unified Medical Language System with semantic methods. *J Am Med Inform Assoc.* 1998; 5(1):41–51. [PubMed: 9452984]
- [11]. Liu H, Johnson SB, Friedman C. Automatic Resolution of Ambiguous Terms Based on Machine Learning and Conceptual Relations in the UMLS. *J Am Med Inform Assoc.* 2002; 9(6):621–636. [PubMed: 12386113]
- [12]. Bodenreider O, Burgun A, Rindfleisch TC. Assessing the consistency of a biomedical terminology through lexical knowledge. *Int J Med Inform.* 2002; 67(1-3):85–95. [PubMed: 12460634]
- [13]. Geller J, Gu H, Perl Y, et al. Semantic refinement and error correction in large terminological knowledge bases. *Data and Knowledge Engineering.* 2003; 45(1):1–32.
- [14]. Bodenreider, O.; Mitchell, JA.; McCray, AT. Evaluation of the UMLS as a terminology and knowledge resource for biomedical informatics. *Proc AMIA Symp*; 2002. p. 61-5.

- [15]. Peng, Y.; Halper, MH.; Perl, Y., et al. Auditing the UMLS for redundant classifications. In: Kohane, IS., editor. Proc AMIA Symp; San Antonio, TX. 2002; p. 612-6.
- [16]. Smith, B.; Williams, J.; Schulze-Kremer, S. The ontology of the Gene Ontology. Proc AMIA Symp; 2003. p. 609-13.
- [17]. Ceusters W, Smith B, Goldberg L. A terminological and ontological analysis of the NCI Thesaurus. *Methods Inf Med.* 2005; 44(4):498–507. [PubMed: 16342916]
- [18]. Nadkarni P, Chen R, Brandt C. UMLS concept indexing for production databases: a feasibility study. *J Am Med Inform Assoc.* 2001; 8(1):80–91. [PubMed: 11141514]
- [19]. Ceusters W, Smith B, Kumar A, et al. Mistakes in medical ontologies: where do they come from and how can they be detected? *Stud Health Technol Inform.* 2004; 102:145–63. [PubMed: 15853269]
- [20]. Cimino JJ. From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. *Journal of the American Medical Informatics Association.* 2000; 7(3):288–297. [PubMed: 10833166]
- [21]. Cohen B, Oren M, Min H, et al. Automated comparative auditing of NCIT genomic roles using NCBI. *J Biomed Inform.* 2008
- [22]. Cimino JJ. Formal Descriptions and Adaptive Mechanisms for Changes in Controlled Medical Vocabularies. *Methods Inf Med.* 1996; 35:202–210. [PubMed: 8952304]
- [23]. Oliver DE, Shahar Y. Change management of shared and local versions of health-care terminologies. *Methods Inf Med.* 2000; 39(4-5):278–90. [PubMed: 11191695]
- [24]. Cimino, JJ.; Clayton, PD. Coping with changing controlled vocabularies. In: Ozbolt, JG., editor. Proc of the Eighteenth Annual Symposium on Computer Applications in Medical Care; Washington, DC. 1994; Philadelphia: Hanley & Belfus; p. 135-139.
- [25]. Penz JF, Brown SH, Carter JS, et al. Evaluation of SNOMED coverage of Veterans Health Administration terms. *Medinfo.* 2004:540–4.
- [26]. McDonald FS, Chute CG, Ogren PV, et al. A large-scale evaluation of terminology integration characteristics. Proc AMIA Symp. 1999:864–7. [PubMed: 10566483]
- [27]. Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. *J Am Med Inform Assoc.* 1997; 4(6):484–500. [PubMed: 9391936]
- [28]. Hardiker NR, Rector AL. Structural validation of nursing terminologies. *J Am Med Inform Assoc.* 2001; 8(3):212–21. [PubMed: 11320066]
- [29]. Brown PJ, Warmington V, Laurence M, et al. Randomized crossover trial comparing the performance of Clinical Terms Version 3 and Read Codes 5 byte set coding schemes in general practice. *British Medical Journal.* 2003; 326(7399)
- [30]. 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 Codes & Structures. *J Am Med Inform Assoc.* 1996; 3(3):224–233. [PubMed: 8723613]
- [31]. Campbell JR, Carpenter P, Sneiderman C, et al. Phase II Evaluation of Clinical Coding Schemes: Completeness, Taxonomy, Mapping, Definitions, and Clarity. *J Am Med Inform Assoc.* 1997; 4(3):238–250. [PubMed: 9147343]
- [32]. Kushniruk, AW.; Patel, VL.; Cimino, JJ., et al. Cognitive Evaluation of the User Interface and Vocabulary of an Outpatient Information System. In: Cimino, JJ., editor. Proceedings of the American Medical Informatics Association Annual Fall Symposium (formerly SCAMC); Washington, DC. October, 1996; Philadelphia: Hanley & Belfus; p. 22-26.
- [33]. Baorto DM, Li L, Cimino JJ. Practical Experience with the Maintenance and Auditing of a Large Medical Ontology. *J Biomed Inform.* 2008 submitted to. (Special Issue on Auditing of Ontologies).
- [34]. Bodenreider O, Smith B, Kumar A, et al. Investigating subsumption in SNOMED CT: an exploration into large description logic-based biomedical terminologies. *Artif Intell Med.* 2007; 39(3):183–95. [PubMed: 17241777]
- [35]. Zhang S, Bodenreider O. Law and order: assessing and enforcing compliance with ontological modeling principles in the Foundational Model of Anatomy. *Computers in Biology and Medicine.* 2006; 36:674–93. [PubMed: 16144698]

- [36]. Schulz EB, Barrett JW, Price C. Semantic quality through semantic definition: refining the Read Codes through internal consistency. *Proc AMIA Symp.* 1997:615–9.
- [37]. Schulz EB, Barrett JW, Price C. Read Code quality assurance: from simple syntax to semantic stability. *J Am Med Inform Assoc.* 1998; 5(4):337–46. [PubMed: 9670131]
- [38]. Mougin F, Bodenreider O. Approaches to eliminating cycles in the UMLS Metathesaurus: naïve vs. formal. *Proc AMIA Symp.* 2005:550–4. [PubMed: 16779100]
- [39]. Bodenreider O. Circular hierarchical relationships in the UMLS: etiology, diagnosis, treatment, complications and prevention. *Proc AMIA Symp.* 2001:57–61. [PubMed: 11825155]
- [40]. Fischer, DH. Formal redundancy and consistency checking rules for the lexical database WordNet TM 1.5. *Fifth Workshop on Very Large Corpora*; 1997.
- [41]. Gu H. Evaluation of a UMLS auditing process of semantic type assignments. *Proc AMIA Symp.* 2007:294–8. [PubMed: 18693845]
- [42]. Gu, H.; Halper, M.; Geller, J., et al. Utilizing OODB Schema Modeling for Vocabulary Management. In: Cimino, JJ., editor. *Proc AMIA Symp*; Washington, DC. 1996; p. 274–278.
- [43]. Gu H, Halper M, Geller J, et al. Benefits of an object-oriented database representation for controlled medical terminologies. *J Am Med Inform Assoc.* 1999; 6(4):283–303. [PubMed: 10428002]
- [44]. Wang AY, Sable JH, Spackman KA. The SNOMED clinical terms development process: refinement and analysis of content. *Proc AMIA Symp.* 2002:845–849. [PubMed: 12463944]
- [45]. Wang Y, Halper M, Min H, et al. Structural methodologies for auditing SNOMED. *J Biomed Inform.* 2007; 40(5):561. [PubMed: 17276736]
- [46]. Halper M, Wang Y, Min H, et al. Analysis of error concentrations in SNOMED. *Proc AMIA Symp.* 2007:314–8. [PubMed: 18693849]
- [47]. Perl Y, Chen Z, Halper M, et al. The cohesive metaschema: a higher-level abstraction of the UMLS Semantic Network. *J Biomed Inform.* 2002; 35(3):194–212. [PubMed: 12669983]
- [48]. Chen Z, Perl Y, Halper M, et al. Partitioning the UMLS semantic network. *IEEE Trans Inf Technol Biomed.* 2002; 6(2):102–8. [PubMed: 12075663]
- [49]. Gu H, Perl Y, Geller J, et al. Representing the UMLS as an Object-oriented Database: Modeling Issues and Advantages. *J Am Med Inform Assoc.* 2000; 7(1):66–80. [PubMed: 10641964]
- [50]. Gu H, Min H, Peng Y, et al. Using the metaschema to audit UMLS classification errors. *Proc AMIA Symp.* 2002:310–4. [PubMed: 12463837]
- [51]. Cimino JJ, Clayton PD, Hripesak G, et al. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *J Am Med Inform Assoc.* 1994; 1(1):35–50. [PubMed: 7719786]
- [52]. Campbell KE, Tuttle MS, Spackman KA. A “lexically-suggested logical closure” metric for medical terminology maturity. *Proc AMIA Symp.* 1998:785–9. [PubMed: 9929326]
- [53]. Cimino JJ, McNamara TJ, Meredith T, et al. Evaluation of a proposed method for representing drug terminology. *Proc AMIA Symp.* 1999:47–51. [PubMed: 10566318]
- [54]. Mendonça EA, Cimino JJ, Campbell KE, et al. Reproducibility of Interpreting “And” and “Or” in Terminology Systems. *Proc AMIA Symp.* 1998:790–794. [PubMed: 9929327]
- [55]. Cimino JJ. Controlled Medical Vocabulary Construction: Methods from the Canon Group. *J Am Med Inform Assoc.* 1994; 1(3):296–297. [PubMed: 7719811]
- [56]. Wroe CJ, Cimino JJ, Rector AL. Integrating Existing Drug Formulation Terminologies Into an HL7 Standard Classification using OpenGALEN. *Proc AMIA Symp.* 2001:766–770. [PubMed: 11825289]
- [57]. Wasserman H, Wang J. An Applied Evaluation of SNOMED CT as a Clinical Vocabulary for the Computerized Diagnosis and Problem List. *Proc AMIA Symp.* 2003:699–703. [PubMed: 14728263]
- [58]. Rector AL. Thesauri and formal classifications: terminologies for people and machines. *Methods Inf Med.* 1998; 37(4-5):501–9. [PubMed: 9865048]
- [59]. Cieslowski BJ, Wajngurt D, Cimino JJ, et al. Integration of Nursing Assessment Concepts into the Medical Entities Dictionary Using the LOINC Semantic Structure as a Terminology Model. *Proc AMIA Symp.* 2001:115–119. [PubMed: 11825165]

- [60]. Moss J, Coenen A, Mills ME. Evaluation of the draft international standard for a reference terminology model for nursing actions. *J Biomed Inform.* 2003; 36(4-5):271–8. [PubMed: 14643722]
- [61]. Chiang MF, Casper DS, Cimino JJ, et al. Representation of ophthalmology concepts by electronic systems; adequacy of controlled medical terminologies. *Ophthalmology.* 2005; 112(2): 175–183. [PubMed: 15691548]
- [62]. Warnekar P, Carter J. HIV terms coverage by a commercial nomenclature. *Proc AMIA Symp.* 2003:1046. [PubMed: 14728549]
- [63]. Smith B, Köhler J, Kumar A. One the application of formal principles to life science data: a case study in the Gene Ontology (Lecture Notes in Bioinformatics Nr. 2994). *Proc DILS.* 2004
- [64]. Kumar, A.; Smith, B. The Unified Medical Language System and the Gene Ontology: some critical reflections. In: Günter, A.; Kruse, R.; Neumann, B., editors. *Advances in Artificial Intelligence (Lecture Notes in Artificial Intelligence 2821)*. Springer; Berlin: 2003. p. 135-148.
- [65]. Baorto DM, Cimino JJ, Parvin CA, et al. Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (LOINC). *Int J Med Inform.* 1998; 51(1):29. [PubMed: 9749897]
- [66]. Zhu, X.; Lee, KP.; Cimino, JJ. Knowledge representation of traditional Chinese acupuncture points using the UMLS and a terminology model. *Proceedings of the IEEE Computer Society IDEAS Workshop on Medical Information Systems: The Digital Hospital; Beijing, China. Sept. 2004; p. 40-48.*
- [67]. Schulze-Kremer, S.; Smith, B.; Kumar, A. Revising the UMLS semantic network. In: Fieschi, M.; Coiera, E.; Li, YC., editors. *Medinfo. San Francisco, CA: 2004. p. 1700*
- [68]. Spackman, K.; Reynoso, G. Examining SNOMED from the perspective of formal ontological principles: some preliminary analysis and observations. In: H., U., editor. *Proc KR-MED. Whistler, Canada: 2004. p. 72-80.*
- [69]. Schulz S, Hahn U. Medical knowledge reengineering--converting major portions of the UMLS into a terminological knowledge base. *Int J Med Inform.* 2001; 64(2-3):207–21. [PubMed: 11734387]
- [70]. Arts DG, de Keizer N, de Jonge E, et al. Comparison of methods for evaluation of a medical terminological system. *Medinfo.* 2004:467–71.
- [71]. Bales ME, Kukafka R, Burkhardt A, et al. Qualitative assessment of the International Classification of Functioning, Disability, and Health with respect to the desiderata for controlled medical vocabularies. *Int J Med Inform.* 2006; 75(5):384–395. [PubMed: 16122973]
- [72]. Tuttle, MS.; Sherertz, DD.; Erlbaum, MS., et al. Adding Your Terms and Relationships to the UMLS Metathesaurus. In: C., PD., editor. *Proc of the Fifteenth Annual Symposium on Computer Applications in Medical Care. Washington, D.C.: Nov. 1991 p. 219-223.*
- [73]. Bodenreider O, Burgun A, Botti G, et al. Evaluation of the Unified Medical Language System as a medical knowledge source. *J Am Med Inform Assoc.* 1998; 5(1):76–87. [PubMed: 9452987]
- [74]. Arts DG, Cornet R, de Jonge E, et al. Methods for evaluation of medical terminological systems -- a literature review and a case study. *Methods Inf Med.* 2005; 44(5):616–25. [PubMed: 16400369]
- [75]. Cimino, JJ. Representation of Clinical Laboratory Terminology in the Unified Medical Language System. In: Clayton, PD., editor. *Proc of the Fifteenth Annual Symposium on Computer Applications in Medical Care; Washington, D.C.. 1991; p. 199-203.*
- [76]. Cimino JJ, Johnson SB. Use of the Unified Medical Language System in Patient Care. *Methods Inf Med.* 1995; 34(1/2):158–164. [PubMed: 9082126]
- [77]. Chute CG, Cohn SP, Campbell JR. A Framework for Comprehensive Health Terminology Systems in the United States: Development Guidelines, Criteria for Selection, and Public Policy Implications. *J Am Med Inform Assoc.* 1998; 5(6):503–510. [PubMed: 9824798]
- [78]. Chiang, MF.; Hwang, JC.; Yu, AC., et al. Reliability of SNOMED-CT Coding by Three Physicians Using Two Terminology Browsers. In: Bates, DM., editor. *Proc AMIA Symp; 2006. p. 131-135.*
- [79]. Hardiker NR. Logical ontology for mediating between nursing intervention terminology systems. *Methods Inf Med.* 2003; 42(3):265–70. [PubMed: 12874660]

- [80]. Hole WT, Srinivasan S. Discovering missed synonymy in a large concept-oriented Metathesaurus. *Proc AMIA Symp.* 2000:354–8. [PubMed: 11079904]
- [81]. Rogers JE, Price C, Rector AL, et al. Validating clinical terminology structures: integration and cross-validation of Read Thesaurus and GALEN. *Proc AMIA Symp.* 1998:845–9. [PubMed: 9929338]
- [82]. Cornet R, Abu-Hanna A. Description logic-based methods for auditing frame-based medical terminological systems. *Artif Intell Med.* 2005; 34(3):201–17. [PubMed: 15994071]
- [83]. Cornet R, Abu-Hanna A. Auditing description-logic-based medical terminological systems by detecting equivalent concept definitions. *Int J Med Inform.* 2008; 77(5):336–45. [PubMed: 17693132]
- [84]. Wang Y, Wei D, Xu J, et al. Auditing Complex Concepts in Overlapping Subsets of SNOMED. *Proc AMIA Symp.* 2008:273–7. [PubMed: 18998838]
- [85]. Chen Y, Gu HH, Perl Y, et al. Structural group auditing of a UMLS semantic type's extent. *J Biomed Inform.* 2009; 42(2009):41–52. [PubMed: 18619563]
- [86]. McCray AT, Nelson SJ. The representation of meaning in the UMLS. *Methods Inf Med.* 1995; 34(1-2):193–201. [PubMed: 9082131]
- [87]. Fan JW, Xu H, Friedman C. Using contextual and lexical features to restructure and validate the classification of biomedical concepts. *BMC Bioinformatics.* 2007; 8:264. [PubMed: 17650333]
- [88]. Cimino JJ. Letter to the Editor: An approach to coping with the annual changes in ICD9-CM. *Methods Inf Med.* 1996; 35(3):220. [PubMed: 8952307]
- [89]. Bodenreider O. Strength in numbers: exploring redundancy in hierarchical relations across biomedical terminologies. *Proc AMIA Symp.* 2003:101–5. [PubMed: 14728142]
- [90]. Dalal, M. *Proc AAAI.* Paul, Minnesota; 1988. Investigations into a theory of knowledge base revision: preliminary report; p. 475-9.
- [91]. Bodenreider, O.; Smith, B.; Kumar, A., et al. Investigating subsumption in DL-based terminologies: a case study in SNOMED CT. In: Hahn, U.; Schulz, S.; Cornet, R., editors. *Proc of the First International Workshop on Formal Biomedical Knowledge Representation (KR-MED)*; Whistler, Canada. 2004; p. 12-20.
- [92]. Ceusters W, Smith B, Kumar A, et al. Ontology-based error detection in SNOMED-CT. *Medinfo.* 2004:482–6.
- [93]. Zhang L, Halper M, Perl Y, et al. Relationship structures and semantic type assignments of the UMLS Enriched Semantic Network. *J Am Med Inform Assoc.* 2005; 12(6):657–66. [PubMed: 16049233]
- [94]. Cimino JJ, Barnett GO. Automated Translation Between Medical Terminologies Using Semantic Definitions. *MD Computing.* 1990; 7(2):104–9. [PubMed: 2186251]
- [95]. Barrows, RC., Jr.; Cimino, JJ.; Clayton, PD. Mapping clinically useful terminology to a controlled medical vocabulary. In: O., JG., editor. *Proc of the Eighteenth Annual Symposium on Computer Applications in Medical Care*; Washington, DC. 1994; Philadelphia, McGraw-Hill, New York: Hanley & Belfus; p. 211-5.
- [96]. Tulipano PK, Millar WS, Cimino JJ. Linking molecular imaging terminology to the gene ontology (GO). *Pac Symp Biocomput.* 2003:613–23. [PubMed: 12603062]
- [97]. Huang KC, Geller J, Halper M, et al. Piecewise Synonyms for Enhanced UMLS Source Terminology Integration. *Proc AMIA Symp.* 2007:339–43. [PubMed: 18693854]
- [98]. Huang KC, Geller J, Halper M, et al. Using WordNet synonym substitution to enhance UMLS source integration. *Artif Intell Med.* 2008 (doi:10.1016/j.artmed.2008.11.008).
- [99]. Patel, C.; Cimino, JJ. Decompositional Terminology Translation Using Network Analysis. *Proc AMIA Symp*; Chicago, IL. 2007; p. 588-592.
- [100]. Cimino, JJ.; Hripcsak, G.; Johnson, SB., et al. UMLS as Knowledge Base - A Rule-Based Expert System Approach to Controlled Medical Vocabulary Management. In: Miller, RA., editor. *Proc of the Fourteenth Annual Symposium on Computer Applications in Medical Care*; Washington, D.C.. 1990; p. 175-180.
- [101]. Cimino, JJ.; Hripcsak, G.; Johnson, SB., et al. Prototyping a Vocabulary Management System in an Object-Oriented Environment. In: Timmers, T.; Blum, BI., editors. *Proc of the International*

- Medical Informatics Association Working Conference on Software Engineering in Medical Informatics; 1990; Amsterdam: North-Holland Press; p. 429-439.
- [102]. Gu H, Perl Y, Geller J, et al. Modeling the UMLS using an OODB. Proc AMIA Symp. 1999:82–6. [PubMed: 10566325]
- [103]. Gu H, Perl Y, Halper M, et al. Partitioning an object-oriented terminology schema. Methods Inf Med. 2001; 40(3):204–12. [PubMed: 11501633]
- [104]. Chen L, Morrey CP, Gu H, et al. Modeling multi-typed structurally viewed chemicals with the UMLS Refined Semantic Network. J Am Med Inform Assoc. 2009; 16(1):116–31. [PubMed: 18952946]
- [105]. Burgun A, Bodenreider O. Methods for exploring the semantics of the relationships between co-occurring UMLS concepts. Medinfo. 2001:171–5.
- [106]. Zhang L, Perl Y, Halper M, et al. An Enriched UMLS Semantic Network with a Multiple Subsumption Hierarchy. J Am Med Inform Assoc. 2004; 11(3):195–206. [PubMed: 14764611]
- [107]. Zhang L, Perl Y, Halper MH, et al. Enriching the Structure of the UMLS Semantic Network. Proc AMIA Symp. 2002:939–43. [PubMed: 12463963]
- [108]. Zhang L, Perl Y, Halper M, et al. Designing metaschemas for the UMLS enriched semantic network. J Biomed Inform. 2003; 36(6):433–49. [PubMed: 14759817]
- [109]. McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. Stud Health Technol Inform. 2001; 84(Pt 1):216–20. [PubMed: 11604736]
- [110]. Cimino, JJ.; Johnson, SB.; Hripcsak, G., et al. Managing Vocabulary for a Centralized Clinical System. In: Kaihara, S.; Greenes, RA., editors. Proc of the World Congress on Medical Informatics - Medinfo '95; Vancouver, Canada. 1995; Edmonton, Alberta: Healthcare Computing and Communications Canada; p. 117-120.
- [111]. Yu, A.; Cimino, JJ. A Comparison of Two Methods for Retrieving ICD-9-CM data: The Effect of Using an Ontology-based Method for Handling Terminology Changes. Proc AMIA Symp; Chicago, IL. 2007. p. 841-845.
- [112]. Andrade RL, Pacheco E, Cancian PS, et al. Corpus-based error detection in a multilingual medical thesaurus. Medinfo. 2007:529–34.
- [113]. Cimino JJ. Distributed cognition and knowledge-based controlled medical terminologies. Artif Intell Med. 1998; 12(2):155–70.
- [114]. Bitencourt JL, Cancian PS, Pacheco EJ, et al. Thesaurus anomaly detection by user action monitoring. Medinfo. 2007:655–9.
- [115]. Caviedes JE, Cimino JJ. Towards the development of a conceptual distance metric for the UMLS. J Biomed Inform. 2004; 7(2):77–85. [PubMed: 15120654]
- [116]. Mays EK, Weida RA, Dionne RA, et al. Scalable and expressive medical terminologies. Proc AMIA Symp. 1996:259–263.
- [117]. Stojanovic, L.; Maedche, A.; Motik, B., et al. User-Driven Ontology Evolution Management. Proc of the Thirteenth International Conference on Knowledge Engineering and Knowledge Management; Ontologies and the Semantic Web; 2002. p. 285-300.
- [118]. Stojanovic, L.; Motik, B. Ontology evolution within ontology editors. Proc OntoWeb-SIG3 Workshop at International Conference on Knowledge Engineering and Knowledge Management; 2002. p. 53-62.
- [119]. Yildiz, B. Ontology evolution and versioning: the state of the art. Institute of Software Technology & Interactive Systems, Vienna University of Technology; 2006.
- [120]. Noy, NF.; Chugh, A.; Liu, W., et al. A framework for ontology evolution in collaborative environments. Proc International Semantic Web Conference; 2006. p. 544-58.
- [121]. Rogers JE. Quality assurance of medical ontologies. Methods Inf Med. 2006; 45(3):267–74. [PubMed: 16685334]
- [122]. Cimino JJ. Saying What You Mean and Meaning What You Say: Coupling Biomedical Terminology and Knowledge. Academic Medicine. 1993; 68(4):257–260. [PubMed: 8466601]
- [123]. Arts DG, Cornet R, De Jonge E, et al. Comparison of methods for evaluation of medical terminological systems. Proc AMIA Symp. 2003:779. [PubMed: 14728284]

- [124]. Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. *Yearbook of Medical Informatics*. 2006:124–35. [PubMed: 17051306]
- [125]. Cimino, JJ.; Johnson, SB.; Peng, P., et al. From ICD9-CM to MeSH Using UMLS: A How To Guide. In: Safran, C., editor. *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care*; Washington, DC: McGraw-Hill, New York; Nov. 1993 p. 730-734.
- [126]. Cimino JJ. Terminology Tools - State of the Art and Practical Lessons. *Methods Inf Med*. 2001; 40(4):298–306. [PubMed: 11552342]
- [127]. Brown SH, Elkin PL, Rosenbloom ST, et al. VA National Drug File Reference Terminology: A Cross-Institutional Concent Coverage Study. *Medinfo*. 2004:477–81.
- [128]. Campbell KE, Cohn SP, Chute CG, et al. Gálapagós: computer-based support for evolution of a convergent medical terminology. *Proc AMIA Symp*. 1996:269–273.
- [129]. Cimino JJ. Battling Scylla and Charybdis: the search for redundancy and ambiguity in the 2001 UMLS metathesaurus. *Proc AMIA Symp*. 2001:120–4. [PubMed: 11825166]
- [130]. Rocha RA, Huff SM, Haug PJ, et al. Designing a controlled medical vocabulary server: the VOSER project. *Comput Biomed Res*. 1994; 27(6):472–507. [PubMed: 7895474]

Table 1

Quality Factors and Associated Errors to be Audited.

1. Concept orientation	1.1 Undefined concepts
	1.2 Concept ambiguity
2. Consistency	2.1 Lexical inconsistency
	2.2 Inconsistent classifications
3. Non-redundancy	3.1 Redundant concepts
	3.2 Redundant classifications
4. Soundness	4.1 Incorrect classifications
	4.2 Erroneous concept definitions
5. Comprehensive coverage	5.1 Incomplete concept coverage
	5.2 Incomplete synonym coverage
	5.3 Incomplete concept definitions
	5.4 Incomplete class hierarchies
	5.5 Incomplete non-hierarchical semantic relationships

Table 2

Quality Factors and Knowledge Sources in Manual Auditing Methods

Quality Factors	Typical Errors	Knowledge Sources					
		Extrinsic	Terminology	Info	Knowledge	Intrinsic	
					Hierarchy	Relationship	Lexicon
1. Concept orientation	1.1 Undefined concepts	[22] [24]		[71]			
	1.2 Concept ambiguity	[7] [22] [24]	[7] [54] [71]				[72]
2. Consistency	2.1 Lexical inconsistency	[22] [24]		[54]			
	2.2 Inconsistent classifications	[7] [22] [24] [31] [73]	[7] [17] [32]	[38] [40] [68] [69]			[34] [38]
3. Non-redundancy	3.1 Redundant concepts	[7] [31]	[7] [72]				
	3.2 Redundant classifications		[40]				
4. Soundness	4.1 Incorrect classifications		[16] [32] [63] [67] [71] [72]		[64] [68]		[67] [74]
	4.2 Erroneous concept definitions	[22] [24] [31]	[58]	[32]		[34]	[63] [64]
5. Comprehensive coverage	5.1 Incomplete concept coverage	[7] [22] [24] [27] [29] [30] [31] [44] [56] [57] [59] [60] [61] [62] [73] [75] [76] [77]	[7] [44]	[7] [29] [32] [61] [65] [71] [75] [76] [77] [78]			[65]
		[7] [27] [30] [60] [61] [75] [76]	[7] [61] [72]				
		[66]	[17] [66]				
		[7] [44] [73]	[7] [32] [71] [75]	[79]			
		[7] [44] [73] [75] [76]	[7] [32] [71] [75] [76] [79]				
5.2 Incomplete synonym coverage	[66]	[66]					
5.3 Incomplete concept definitions	[7] [44] [73]	[7] [32] [71] [75]	[79]				
5.4 Incomplete class hierarchies	[7] [44] [73]	[7] [32] [71] [75]	[79]				
5.5 Incomplete non-hierarchical semantic relationships	[7] [44] [73] [75] [76]	[7] [32] [71] [75] [76] [79]	[79]				

Table 3
Quality Factors and Knowledge Sources in Automated Systematic Auditing Methods

Quality Factors	Typical Errors	Knowledge Sources				
		Extrinsic	Info	Knowledge	Intrinsic	
		Terminology	Hierarchy	Relationship	Lexicon	
1. Concept orientation	1.1 Undefined concepts					
	1.2 Concept ambiguity	[43]	[43]	[9] [36] [37] [48] [88]	[36] [48] [82] [83] [88]	[48] [88]
2. Consistency	2.1 Lexical inconsistency	[26]		[82]		
	2.2 Inconsistent classifications		[81]	[6] [36] [37] [38] [39] [47] [50] [91]	[8] [10] [36] [47] [51] [90]	
3. Non-redundancy	3.1 Redundant concepts			[36] [37] [45] [88]	[8] [10] [36] [45] [82] [83] [88] [92]	[80] [88]
	3.2 Redundant classifications			[15]	[13] [82]	
4. Soundness	4.1 Incorrect classifications	[20]	[20]	[20] [81]	[9] [36] [45] [50] [84] [89] [90]	[8] [10] [13] [84] [85] [87]
	4.2 Erroneous concept definitions			[84] [90]	[84] [90] [92]	[46] [84]
5. Comprehensive coverage	5.1 Incomplete concept coverage	[25] [26]		[45]	[8] [45]	
	5.2 Incomplete synonym coverage	[25]		[45]	[8] [45]	[80]
5.3 Incomplete concept definitions	5.3 Incomplete concept definitions			[92]	[13]	[52] [87]
	5.4 Incomplete class hierarchies	[21]	[84]			
5.5 Incomplete non-hierarchical semantic relationships	5.5 Incomplete non-hierarchical semantic relationships	[20]	[20]	[20] [81] [84]	[36] [45] [93]	[8] [10] [36] [45] [82] [83] [92] [93]

Table 4
Quality Factors and Knowledge Sources in Automated Heuristic Auditing Methods

Quality Factors	Typical Errors	Knowledge Sources						
		Extrinsic	Intrinsic	Info	Knowledge	Hierarchy	Relationship	Lexicon
1. Concept orientation	1.1 Undefined concepts	[101] [110]					[111]	[110] [111]
	1.2 Concept ambiguity	[96] [98]	[49] [104] [112] [113]			[42] [104] [111] [113]		[111] [113] [35] [53]
2. Consistency	2.1 Lexical inconsistency	[98]	[114] [12]					
	2.2 Inconsistent classifications	[41] [89] [94] [96] [101]	[12] [115]			[49] [89] [102] [104] [107] [113]	[35] [41] [50] [89] [101] [104] [113]	[106] [113]
3. Non-redundancy	3.1 Redundant concepts	[14] [73] [96] [98] [101] [110]				[112] [113]	[100] [111] [113]	[80] [110] [111] [113]
	3.2 Redundant classifications	[89]				[47] [89]	[89]	
4. Soundness	4.1 Incorrect classifications	[28] [41] [101] [110]	[103] [115]			[49] [102] [104] [107] [113]	[41] [42] [50] [101] [104] [113]	[110] [113]
	4.2 Erroneous concept definitions					[115]		[53]
5. Comprehensive coverage	5.1 Incomplete concept coverage	[14] [25] [94] [98] [101] [110]	[101] [114] [103]			[49]	[100] [111]	[53] [110] [111]
	5.2 Incomplete synonym coverage	[25] [95] [98] [99]					[100]	[97]
	5.3 Incomplete concept definitions	[116]						[116]
	5.4 Incomplete class hierarchies	[28] [116]	[105]			[107]		[106] [116]
	5.5 Incomplete non-hierarchical semantic relationships	[110]	[101] [105]			[113]	[100] [113]	[110] [113]

Terminologies	Publications
Clinical Terms Version 3	[29]
CPT-4 (Current Procedural Terminology)	[7] [30] [61] [94] [122]
DICE (Diagnoses for Intensive Care Evaluation)	[70] [82] [83] [123]
Proprietary Drug Terminologies	[53] [56]
FMA (Foundational Model of Anatomy)	[35] [124]
GALEN	[58] [81] [122] [124]
GO (Gene Ontology)	[16] [63] [64] [96] [124]
HHCC (Home Health Care Classification)	[59]
ICD-9-CM	[7] [22] [24] [30] [61] [88] [94] [111] [115] [122] [125]
ICF (The International Classification of Functioning, Disability, and Health)	[71]
LOINC (Logical Observation Identifiers Names and Codes)	[59] [61] [65] [124]
MED (Medical Entities Dictionary)	[22] [24] [32] [33] [42] [43] [51] [59] [61] [75] [76] [88] [95] [100] [101] [103] [110] [124] [126]
MEDCIN	[62]
MeSH	[3] [7] [94] [115] [122] [125]
MorphoSaurus (a large multilingual thesaurus for clinical medicine)	[112] [114]
NCIt (National Cancer Institute Thesaurus)	[8] [17] [21]
NDF (National Drug File)	[124] [127]
Nursing Terminologies	[28] [30] [59] [60] [79] [124]
Read	[27] [29] [30] [31] [36] [37] [81]
SNOMED	[7] [12] [19] [25] [30] [31] [44] [45] [46] [52] [54] [57] [61] [68] [78] [84] [91] [92] [94] [95] [115] [122] [124] [128]
UMLS (Unified Medical Language System)	[3] [6] [7] [9] [10] [11] [12] [13] [14] [15] [18] [26] [27] [30] [31] [38] [39] [41] [47] [48] [49] [50] [64] [66] [67] [69] [72] [73] [75] [76] [80] [85] [87] [89] [93] [97] [98] [99] [100] [101] [102] [104] [105] [106] [107] [108] [109] [115] [124] [125] [129]
VOSEER	[130]

Ambiguity	Doubtfulness or uncertainty as regards interpretation of a term. Open to or having several possible meanings. See also Polysemy and Vagueness.
Attribute	A quality or characteristic inherent in or ascribed to a term. For example, a disease term (such as "pneumonia") might have an attribute such as "site-of-disease" (which, in this case, might have the value "lung").
Audit	To examine, verify, or correct errors for good quality assurance in a terminology.
Automated heuristic auditing	Involves using rules that make inferences about terminology content and then seeking to identify those inferences that lead to illogical or inconsistent conclusions. For example, a term with many attributes might actually have multiple meanings (that is, it might be ambiguous).
Automated systematic auditing	Involves implementing knowledge into rule-checking programs that scan the terminology for potential problems with respect to particular quality factors, such as soundness or consistency. For example, if the term "lung disease" has the value "lung" for its "site-of-disease" attribute, then all terms in the class "lung disease" should also have that attribute and value (or some more specific value).
Class	A set, collection, or group containing concepts regarded as having certain traits in common; also, a term whose meaning corresponds to such a set.
Classification	The assignment of terms or concepts to defined classes (usually organized into hierarchies).
Comprehensive coverage	The extent or degree to which the terms needed to describe concepts in a particular domain of discourse contained in a terminology.
Concept	A general or abstract notion derived or inferred from specific instances or objects; concepts are associated with a corresponding representative terms in a terminology.
Consistency	Adherence to structural, semantic and/or linguistic rules for representing terms and their underlying concepts in a terminology. In a terminology with consistency, a term appears the same (and has the same attributes and children) no matter how it is arrived at through traversal of a hierarchy.
Extrinsic knowledge	Knowledge derived from an outside source, such as other terminologies, user requirements, or human expert knowledge; for example, the knowledge of a domain of discourse (which is helpful for auditing comprehensive coverage).
Hierarchy	The system of levels (usually classes) according to which a terminology is organized.
Intrinsic knowledge	Knowledge about concept derived from the classification schema, hierarchy, and semantic relationships among terms, or term attributes such as lexical information present within a terminology itself.
Lexical	A linguistic knowledge with information about the morphological variations and grammatical usage of words.
Multiple hierarchy	An arrangement of terms and their underlying concepts in which one term or concept can belong to more than one class. It is also called polyhierarchy, directed acyclic graph (DAG), lattice, or partially ordered set (poset).
Nonambiguity	A quality pertaining to a terminology wherein each distinct term in the terminology corresponds to a single concept.
Nonredundancy	A quality pertaining to a terminology wherein each distinct term in the terminology corresponds to a concept that is distinct from those concepts to which other terms in the terminology correspond.
Nonvagueness	A quality pertaining to a terminology wherein each term in the terminology must correspond to a well-formed concept.
Orthographic accuracy	Spelled correctly.
Partition	To divide the terms in a terminology into parts or sections based on some characteristics of the terms. For example, a terminology might be partitioned into groups that have common attributes.
Polysemy	A type of ambiguity wherein a term corresponds to multiple different concepts.
Redundancy	A quality pertaining to a terminology wherein multiple terms with

	synonymous meanings are included as distinct terms, implying that they refer to distinct concepts.
Relationship	A connection or association existing between terms or concepts.
Schema	A systematic orderly representation of the elements of a terminology that permits an abstract overview of the terminology; for example, the partitioning of a hierarchy into smaller subhierarchies.
Semantic	Pertaining to or arising from the meanings of terms in a terminology.
Semantic classification	A classification scheme based on the meaning of the terms (that is, their corresponding concepts) in the classification (as opposed, for example, to one based on alphabetization).
Semantic network	A graphical arrangement of nodes (in which nodes correspond to terms or classes of terms) and semantic relationships between the nodes.
Semantic relationship	A connection or association between terms that conveys knowledge about the relationships between the concepts to which the terms refer. A semantic relationship (sometimes referred to as a role) may be hierarchical or nonhierarchical.
Soundness	Accuracy of the knowledge represented in a terminology. Typical terminology features audited for soundness include term names, definitions and classifications.
Structural	As opposed to semantic and lexical, pertaining to or showing the interrelation or arrangement of terms in a terminology system, or the data structures of the terms themselves (such as their identifiers and attributes).
Synonym	A word or phrase having the same or nearly the same meaning as other words or phrases in a terminology. In some terminologies, all synonyms are considered terms related to the same concepts, possibly with one synonym considered to be the "preferred term", while in other terminologies, the preferred synonym is considered the "term" and the remaining ones are referred to simply as "synonyms". Unrecognized synonymy in a terminology leads to redundancy.
Term A	word or phrase that corresponds to a particular underlying concept.
Vagueness	A type of ambiguity wherein a term does not clearly correspond to a well-defined concept.