Journal of the American Medical Informatics Association, 28(9), 2021, 2017–2026
doi: 10.1093/jamia/ocab084
Advance Access Publication Date: 21 June 2021
Review

Review

The use of SNOMED CT, 2013-2020: a literature review

Eunsuk Chang and Javed Mostafa

Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Corresponding Author: Eunsuk Chang, MD, MPH, Carolina Health Informatics Program, University of North Carolina at Chapel Hill, 335 S Columbia St., CB#7585, Chapel Hill, NC 27514, USA; eunsukc@live.unc.edu

Received 5 February 2021; Revised 30 March 2021; Editorial Decision 23 April 2021; Accepted 26 April 2021

ABSTRACT

Objective: This article reviews recent literature on the use of SNOMED CT as an extension of Lee et al's 2014 review on the same topic. The Lee et al's article covered literature published from 2001-2012, and the scope of this review was 2013-2020.

Materials and Methods: In line with Lee et al's methods, we searched the PubMed and Embase databases and identified 1002 articles for review, including studies from January 2013 to September 2020. The retrieved articles were categorized and analyzed according to SNOMED CT focus categories (ie, indeterminate, theoretical, pre-development, implementation, and evaluation/commodity), usage categories (eg, illustrate terminology systems theory, prospective content coverage, used to classify or code in a study, retrieve or analyze patient data, etc.), medical domains, and countries.

Results: After applying inclusion and exclusion criteria, 622 articles were selected for final review. Compared to the papers published between 2001 and 2012, papers published between 2013 and 2020 revealed an increase in more mature usage of SNOMED CT, and the number of papers classified in the "implementation" and "evaluation/commodity" focus categories expanded. When analyzed by decade, papers in the "pre-development," "implementation," and "evaluation/commodity" categories were much more numerous in 2011-2020 than in 2001-2010, increasing from 169 to 293, 30 to 138, and 3 to 65, respectively.

Conclusion: Published papers in more mature usage categories have substantially increased since 2012. From 2013 to present, SNOMED CT has been increasingly implemented in more practical settings. Future research should concentrate on addressing whether SNOMED CT influences improvement in patient care.

Key words: SNOMED CT, systematized nomenclature of medicine, medical ontology, Unified Medical Language System, electronic health record

INTRODUCTION

Interoperability is defined as "the ability of two or more systems to exchange information and to use the information that has been exchanged."[1] To achieve interoperability, a number of prerequisites must be met, such as mutual interest between the stakeholders involved, common architectural structures and reference models, accepted collections of coordinated reference ontologies/terminologies, and harmonized development processes.[2]

In order to get closer to the vision of comprehensive interoperability, the ontological representations used by various domain experts to represent entities in practice need to be harmonized.[3] In the biomedical domain, ontologies are getting the limelight for controlling and verifying the process interoperability and are recognized as key enablers of knowledge management, data integration, and decision support.[4]

One of the biggest, most sound and solid such ontologies is SNOMED CT. SNOMED CT is both a coding scheme to identify terms and a multi-hierarchical ontology that enables concepts to be related to each other. It is a comprehensive clinical terminology system managed by the International Health Terminology Standards Development Organization (IHTSDO), which now operates under the trade name SNOMED International and provides a standardized way of representing clinical information captured by clinicians.[5]

Although SNOMED CT is widely used in more than 50 countries,[6] there are few published reviews about its use. Cornet and de Keizer[7] analyzed over 40 years of literature (1966 -2006) on the use of SNOMED CT and its predecessors—SNOP, SNOMED, SNOMED II, SNOMED version 3.5, and SNOMED RT, classifying the use reported in 250 papers into 14 usage categories. Lee et al in 2014[8] extended the 40-year review to categorize SNOMED CT use into 15 usage categories and 5 focus categories: their review included 488 articles about SNOMED CT published between 2001 and 2012 and showed an increase in the number of papers in every focus category when compared from 2001-2006 to 2007-2012.

The purpose of this paper is to investigate the use of SNOMED CT by providing an overview of papers published between 2013 and 2020 as an extension of the review by Lee et al.[8] The current study also provides a comparative summary of research trends related to SNOMED CT use over the past 20 years, analyzed by decade on selected topics.

METHODS

Screening papers

Database searches using PubMed and Embase were performed using the same query Lee et al constructed for the 2001-2012 review, except for the "publication date" entry which was updated to retrieve articles published between January 1, 2013 and September 4, 2020. We also modified search queries for Embase since query extensions used by Lee et al, such as ".mp," were no longer available at the time of the literature search for the current review. Supplementary Appendix A describes the search queries used for the current review. A faculty librarian at the University of North Carolina at Chapel Hill confirmed the validity of our search queries. We used the resulting articles' abstracts to initially classify them and read the full article if the topic of the paper was not evident in the abstract.

Exclusion criteria

An article was excluded (1) if the article was only an abstract, (2) if the article did not mention SNOMED CT, or (3) if the article was written in a language other than English. Unlike Lee et al's review, articles that did not mention a version of SNOMED, (eg, SNOMED CT, SNOMED RT, or SNOMED III) were not necessarily excluded because the verbiage in the field of health informatics to address SNOMED CT changed to simply "SNOMED" in the 2010s.

Classifying papers

We adhered to Lee et al's criteria to classify articles by the maturity of SNOMED CT use. Lee et al defined the "SNOMED CT focus category" as the 5 stages of SNOMED use in order of maturity—"theoretical," "pre-development/design," "implementation," and "evaluation/commodity," and "indeterminate." (Table 1)

The "usage category" describes in a more detailed way how SNOMED CT is mainly used. The definition of each usage category is described in Lee et al's article.[8] Two to five usage categories are allocated to a focus category. Each usage category is assigned to a single parent focus category and does not overlap with any other usage categories. When a paper is deemed to be classifiable into more than one usage category, the most prominent usage category was selected based on the research question or the aim of the paper.

To reach consensus in the classification method between the current and the Lee et al's reviews, the corresponding author tested his classification accuracy by comparing his classification results with that of Lee et al using their article's Supplementary Appendix B.[8] The agreement rate was 64% in the first 100 randomly selected articles and improved to 78% in the next 100 random articles. To validate the consistency in the classification performed for the current review by the corresponding author, an external reviewer, a postdoctoral research fellow in health informatics, independently classified another random set of 100 articles. The agreement rate between the corresponding author and the external reviewer was 98%.

Clinical domain

The clinical domain of a study was examined to determine whether the study involved a cancerous or noncancerous disease. If it was a noncancerous disease but there were conflicting domains, the domain of the journal in which the paper was published determined the domain of the study.

Country

The country criterion refers to the country in which SNOMED CT was used. If the country could not be defined, the country where the affiliated institution of the first author was located was used. If the study looked at the use of SNOMED CT in more than one region, the paper was classified as having "multiple" countries.

RESULTS

Literature search

The literature search conducted on September 4, 2020, initially identified a total of 1002 articles after deduplication. From this result pool, 349 conference abstract-only papers were excluded. After applying the next round of exclusion criteria, an additional 31 articles were excluded, resulting in 622 papers for final review (Figure 1). Supplementary Appendix B includes a brief description of each selected article.

SNOMED CT focus category

The number of papers classified as "implementation" rose from 44 in 2001-2012 to 124 in 2013-2020, while those classified as "evaluation/commodity" increased from 8 to 60 (Figure 2). The proportions of these two focus categories were also expanded: the proportion of "implementation" articles increased from 9.0% to 19.9% and "evaluation/commodity" articles from 1.6% to 9.6%. This shows that the use of SNOMED CT in operational settings had increased in 2013-2020 compared to the 2001-2012 period.

To make a balanced comparison, we also analyzed the number of papers in each focus category by decade (Figure 3). The SNOMED CT-related research was more advanced in terms of its maturity as well as quantity in the second decade (ie, 2011–2020) compared to that of the first decade. While the number of papers in the "theoretical" category stayed almost the same over these two decades, those in the "pre-development," "implementation," and "evaluation/commodity" categories were much more numerous in 2011-2020 than in 2001-2010, increasing from 169 to 293, 30 to 138, and 3 to 65, respectively.

Table 1. Definitions of each focus category (Adopted from Lee et al[8])

Focus Category	Definition
Theoretical	Describes SNOMED CT in terms of a terminology system but not at the stage of implementation in clinical or operational environments
Pre-development/design	Refers to SNOMED CT being assessed or evaluated as to whether it can be used in full-scale implementation as a terminology standard
Implementation	Refers to SNOMED CT being used in project or operational settings
Evaluation/commodity	Evaluates SNOMED CT's effects or impacts on operational settings or demonstrates that its function has shifted from encoding data to using data recorded in SNOMED CT codes

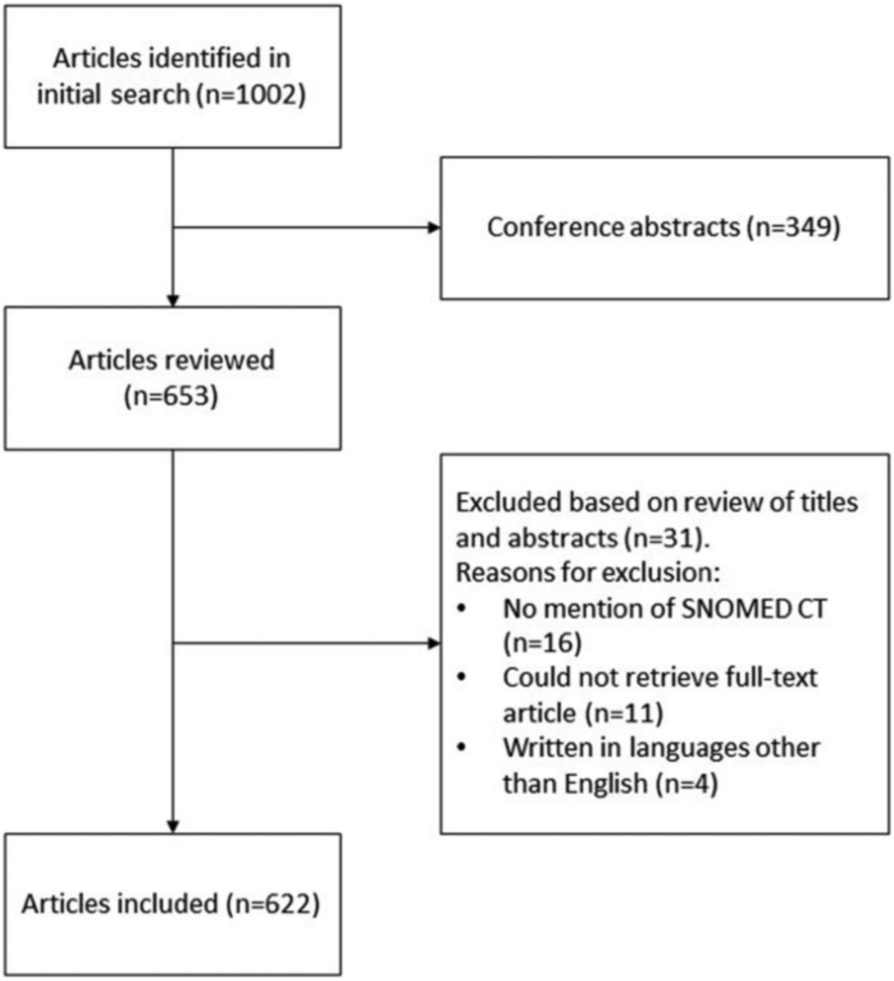

Figure 1. Flow diagram of article selection process.

Usage category

This section describes selected usage categories in each focus category except "indeterminate." Table 2 outlines the number of papers published between 2013 and 2020 in each usage category along with some of its subcategories.

Theoretical: Illustrate terminology systems theory (n = 20)

Studies in this category were generally split into two subcategories. First, frameworks and models for categorizing terminology systems were selected as one subcategory. In this category, included studies proposed ontological frameworks based on SNOMED CT-compliant coding systems for the standardization of EHR phenotypes,[9] medication data,[10] and clinical pathways.[11] The second subcategory included studies in which the semantic similarity presented by the taxonomic structure of SNOMED CT was used to classify or cluster entities like clinical models (eg, templates or archetypes),[12,13] radiology reports,[14] research articles for systematic review,[15] and clinical trials.[16]

The taxonomic structure of biomedical ontology has been shown to be useful for quantifying similarity, because it does not depend on the availability of large corpora. SNOMED CT, in combination with or without the Unified Medical Language System (UMLS), and the proliferation of textual resources in healthcare offered a rich re-

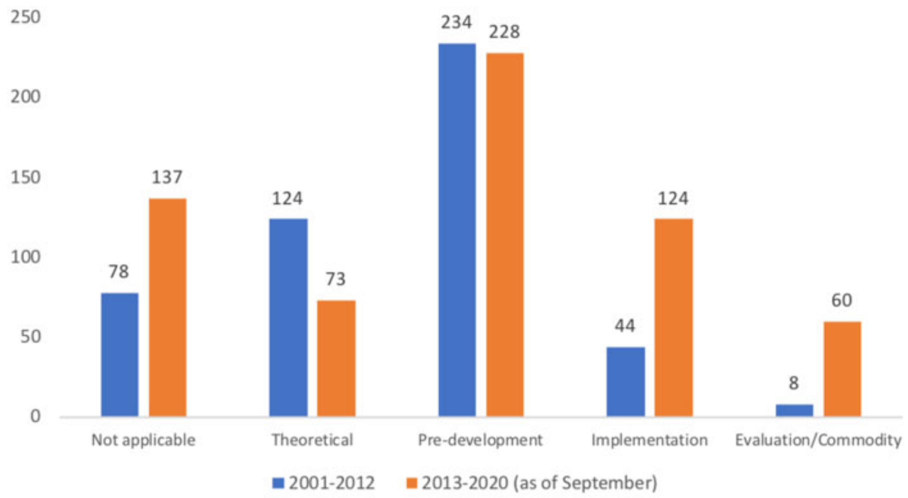

Figure 2. Number of papers by focus category, comparing those included for Lee et al's[8] (published between 2001 and 2012) with those included for current review (published between 2013 and September 2020).

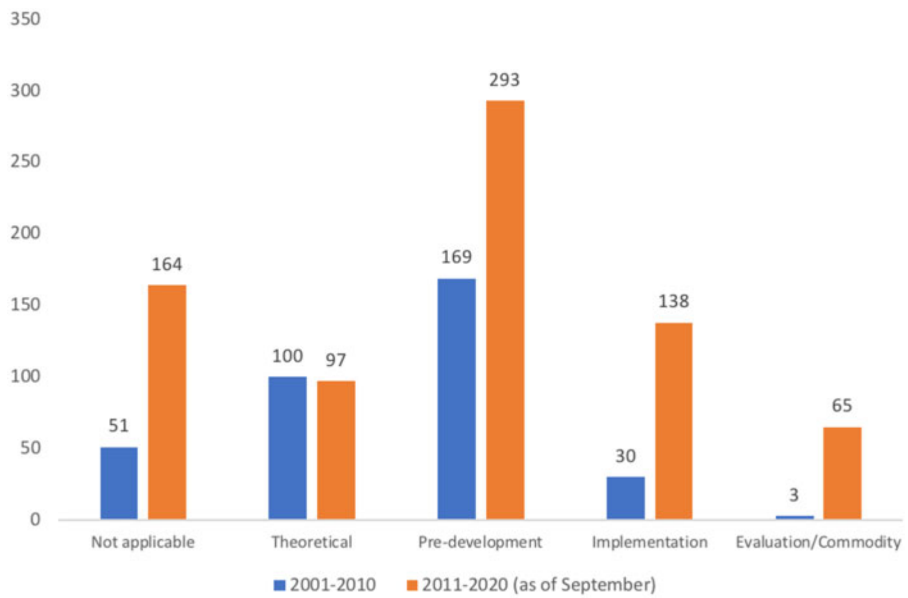

Figure 3. Number of papers by focus category and decade (2001-2010 vs 2011-2020). The numeric data for the first decade (ie, 2001-2010) and the early two years of the second decade (ie, 2011-2012) were reproduced from Supplementary Appendix B of Lee et al's review article.[8]

source for creating automated methods to measuring semantic similarity between concepts.[17] Several studies have introduced different semantic similarity metrics. Shobhana and Radhakrishnan[18] proposed a new measure of similarity that recognizes multiple inheritances in ontologies. McInnes et al[19] have introduced the *u-path* principle derived from the dense multi-hierarchical taxonomy of SNOMED CT instead of measuring a least common subsumer. Chandar et al[20] exploited a similarity-based approach to build new SNOMED CT concept from candidate n-grams. Martínez et al[21] and Sánchez et al[22] proposed a general framework to mask textually sensitive health data by exploiting semantic similarity based on the taxonomic structure of SNOMED CT.

Pre-development: Prospective content coverage (n = 67)

A total of 67 studies examined the content coverage of SNOMED CT for interface terminologies. The covered referents included: social determinants of health (n = 5), disease names/diagnoses (n = 4), eligibility criteria of clinical trials (n = 3), as well as elements of large databases and registries. The overall coverage of SNOMED was as low as 22.5% for video portal tags,[23] and as high as 99.9% for clinical charts.[24]

Content coverage of SNOMED CT was examined along with other standardized terminologies such as International Classification of Diseases (ICD)-9 and Logical Observation Identifiers, Names, and Codes (LOINC). Subset coverage was examined in three studies.[25–27] Though SNOMED CT was generally the most comprehensive terminology with great breadth, this terminology was outperformed in different specialty areas, including RadLex for radiology text,[28] NCI Thesaurus for cancer descriptions,[29] MeSH for herbal and dietary supplement terms,[30] and Omaha System for social determinants of health.[31]

Types of matching relations (eg, exact and partial matches) or post-coordination were used to describe a relationship between a

Table 2. Number of papers published between 2013 and 2020 by usage categories and subcategories

Focus category	Usage category—selected subcategories	Number
Indeter-minate (n = 137)	As an example—no subcategories (n = 64)	64
	Other—editorials (n = 24), literature reviews (n = 21), letters/correspondence (n = 10), surveys (n = 8)	73
Theoretical (n = 73)	Illustrate terminology systems theory—semantic similarity (n = 14), frameworks and models for categorizing terminology systems (n = 3), terminology theory and ontological principles (n = 3)	20
	Description of SNOMED CT and other standards—general description of SNOMED CT (n = 6), potential benefits of SNOMED CT (n = 2)	9
	Terminology auditing—structural/hierarchical (n = 11), ontological principles (n = 8), abstraction network (n = 7), lexical/linguistic (n = 7), semantic (n = 4), combination of methods (n = 7)	44
Pre-develop-ment (n = 228)	Compare to or map to other terminology systems—ICDs (n = 25), new ontology developed (n = 12), nursing terminologies (n = 6), Foundational Model of Anatomy (n = 4), MedDRA (n = 4), LOINC (n = 2)	66
	Translation—German (n = 2), Basque (n = 2)	6
	Prospective content coverage—data repositories and medical corpora (n = 26), interface terminologies (n = 6), social determinants of health (n = 5), web pages/social media (n = 5), disease names/diagnoses (n = 4), eligibility criteria of clinical trials (n = 3)	67
	Prospective inter-rater agreement—between 2 annotators (n = 1)	1
	Standard for electronic health records—electronic health records frameworks/infrastructure and integration with information models (n = 32); binding to clinical models, templates or archetypes (n = 11)	43
	Design considerations—encoding methodologies or comparison of coding techniques (n = 20); process and challenges related to the development of subsets (n = 10); search and retrieval algorithms (n = 6); general implementation challenges (n = 3); version control, management and migration (n = 3); the role and use of interface terminologies in conjunction with SNOMED CT to facilitate data capture (n = 2);	45
Implemen-tation (n = 124)	Used to classify or code in a study—free-text clinical notes/narratives (n = 29), radiology reports (n = 7), consumer-generated contents (n = 6), clinical trial-related documents (n = 2)	79
	Implementation of SNOMED CT—data repositories/registries (n = 18), query/terminology services (n = 15), clinical decision support systems (n = 8), use of data entry templates (n = 2)	45
Evaluation/Commodity (n = 60)	Prove merit—no subcategories (n = 3)	3
	Retrieve or analyze patient data—retrieval of SNOMED CT-encoded patient data (n = 53)	57

SNOMED CT concept and a covered entity. Post-coordination was employed in nine studies[24,29,32–38] to improve content coverage.

Implementation: Used to classify or code in a study (n = 79)

The studies in this category explored the semantic capabilities of SNOMED CT to normalize data components in an attempt to integrate data from heterogeneous sources. SNOMED CT was used to capture, define, and normalize heterogeneous data phenotypes from clinical free texts. The semantic interoperability features of SNOMED CT-coded entities were represented as features of classification tasks.

The coding schemes where SNOMED CT was employed were further divided into three groups. Firstly, the authors proposed a new framework to identify SNOMED CT concepts in free texts. In such cases, machine learning text mining algorithms such as support vector machines (n = 13), neural networks (n = 10), decision trees/random forest (n = 7), or ensemble/boosting algorithms (n = 3) were coupled with SNOMED CT to code or classify free texts. Secondly, open-source or commercial natural language processing (NLP) pipelines such as MetaMap (n = 9) and cTAKES (n = 6) were executed to extract or map phenotypes to SNOMED CT. Lastly, the SNOMED CT concept was extracted from free text manually by coders or physicians (n = 7).

Post-coordination was leveraged in three studies.[39–41] The addition of post-coordinated expression significantly increased the proportion of triple disorder-related identifiers that were encoded to single identifiers.[40] However, no evaluation was made as to whether the post-coordinated expression improved the performance of normalization tasks compared to only pre-coordination coding schemes.

While the majority of the published papers in this category annotated clinical records (eg, clinical notes, pathology reports, consult letters, etc.), other content such as clinical trial documents or data elements (n = 4), clinical guidelines (n = 4), and consumer-generated content (n = 6) were also encoded by SNOMED CT.

SNOMED CT concepts were further mapped to query terms to expand or support semantic searches across archetypes.[42–45] SNOMED CT concepts extracted from those kinds of free texts were then embedded into classification algorithms to classify clinical data elements by diagnosis codes or cause of death,[46–54] types of additional need attributed to patients,[55] or organ systems.[56]

Other terminology standards that were frequently used in collaboration with SNOMED CT for annotating data entities were: ICD codes (eg, ICD-9, ICD-10, or ICD-10-CM) (n = 8), Medical Subject Headings (MeSH) (n = 4), LOINC (n = 3), RxNorm (n = 3), and Medical Dictionary for Regulatory Activities (MedDRA) (n = 3). The UMLS was frequently used as a mapping layer among ontologies. In this case, the normalization process involved mapping clinical entities to the closest equivalent UMLS Concept Unique Identifier subset of SNOMED CT.

Other than free-texts in English, those in German (n = 3), French (n = 2), Swedish (n = 1), Spanish (n = 1), Czech (n = 1), and Chinese (n = 1) were encoded with SNOMED CT. Recent attempts were made to analyze Twitter mentions of disease concerns[57] and to encode COVID-19-related clinical phenotypes.[58]

Implementation: Implementation of SNOMED CT (n = 45)

In this usage category, SNOMED CT codes were electronically assigned to the data model to provide standardized terminology and were then integrated into data registries/repositories (n = 18), query/terminology services (n = 15), and knowledge-based clinical decision support systems (CDSSs) (n = 8). The clinical domains of registries/repositories included cancer (n = 5), adverse drug effects (n = 4), cardiovascular diseases (n = 2), and gene information (n = 2).

Natural language processing applications were used to extract concepts and to map them to SNOMED CT.[59,60] These applications included visualization tools to summarize or navigate terminology hierarchies[61–64] and search engines.[59] Auto-completion algorithms were employed to help providers search diseases[65,66] and to add the SNOMED concept to the severity levels of diseases.[66]

Evaluation/commodity: Prove merit (n = 3)

Three studies reported the effect or impact of SNOMED CT use on patient care in clinical or operational settings. A study on predoctoral dental students demonstrated that the use of SNOMED CT-based dental interface terminology as the sole reference in an EHR system strengthened the students' critical thinking skills.[67] Souvignet et al[68] measured the indirect utility of SNOMED CT; they showed that a forms-based web interface that represented MedDRA terms through SNOMED CT concepts and corresponding semantic relations expedited MedDRA term selection and improved search capabilities for pharmacovigilance end users. However, Dougall and colleagues found that the use of local SNOMED code was associated with the underreporting of melanoma diagnoses in hospitals.[69]

Evaluation/commodity: Retrieve or analyze patient data (n = 57)

SNOMED CT codes were used to collect patient data for retrospective observational studies. Most papers used SNOMED CT codes to identify and retrieve study subjects according to the inclusion criteria for each study. The studies conducted in the U.S. extracted SNOMED CT-coded patient data using a commercial database (ie, Explorys) (n = 16), while those conducted in Europe tended to use public national registers. The clinical domains that most frequently utilized SNOMED CT diagnoses were non-malignant gastrointestinal diseases (n = 17), followed by cancers (n = 14), pathology diagnosis (n = 7), and non-malignant skin diseases (n = 5).

Medical domain

We categorized 622 articles into 40 clinical domains. The most common domain was cancer (n = 37), followed by drugs (n = 34) and nursing (n = 26). Cancer studies aimed to develop registries or terminology services for identifying cancer patients or to retrieve patients with cancers of interest by searching databases for SNOMED CT-encoded diagnoses. Drug studies intended to map drug-related ontologies (eg, MedDRA) or to encode drug-related entities (eg, drug names or drug indications) into SNOMED CT. In the nursing domain, the authors illustrated the mapping of the elements of domain-specific concepts or the standardization of nursing terminologies using SNOMED CT.

Country

Since the publication of Lee et al's review, the number of member countries of SNOMED International has grown from 19 to 40.[70] The number of countries that produced SNOMED CT-related papers has also grown during this period. Between 2013 and 2020, papers from 43 countries were published, which was an increase from 22 over the period of 2001-2012 (ie, Lee et al's period of review). The most publishing countries in order of the number of papers were: U.S. (n = 290), France (n = 42), U.K. (n = 35), Spain (n = 33), and Australia (n = 25). About half (46.6%) of the papers came from the U.S. The studies conducted in Australia and Spain tended to be conducted in more operational settings than those in the U.K., U.S., and France (Figure 4).

DISCUSSION

Adopting the methodology from Lee et al's review, we reviewed papers on SNOMED CT published between 2013 and 2020. We could observe that the research on SNOMED CT has evolved to be implemented in more practical settings in the past seven years. The articles classified as "implementation" increased from 44 at the time of Lee et al's review (ie, 2001-2012) to 124 in 2013-2020; those classified as being focused on "evaluation/commodity" increased from 8 to 60 in the same period. On the other hand, the number of studies classified as "theoretical" has declined from 124 in 2001-2012 to 73 in 2013-2020.

Theoretical

When analyzed by decades, the number of papers in this focus category remained relatively the same. The semantic similarity was studied intensively in this category. In the general language domain, corpus-based measures of similarity revealed limitations that stem from the imbalance, sparseness, and textual ambiguity of corpora.[71,72] More recent studies in the biomedical domain[73,74] showed that ontology-based measures such as intrinsic information content (IC) outperformed the corpus-based approaches.

As the similarity measurement depends on the taxonomic structure of an ontology system, it demands extensive auditing to ensure the reliability and consistency of SNOMED CT. While the number of papers in the "theoretical" focus category declined, papers in the "terminology auditing" usage category have increased by more than 60% (from 27 to 44) from 2001-2012 to 2013-2020. We interpret this increment as a growing demand for a formal resource for various computational linguistic tasks, such as information retrieval and NLP of patient records. As the inconsistencies in SNOMED CT have a significant impact on how patient records are recorded and retrieved, it is necessary for the responsible authorities to keep good track of the quality assurance of SNOMED CT through extensive auditing.

Pre-development

In a few of the articles included in this review, SNOMED CT lacked sufficient definitional expressions needed to describe the clinical phenotypes. A possible explanation for this problem is not the lack of acceptable terms in SNOMED CT but both the type of text from which the concepts are extracted and the use of modifiers.[75,76] Another reason may be that researchers investigated SNOMED CT's coverage in areas where its use had seldom been previously explored, such as in social determinants of health[77–79] and family history in cancer guidelines.[80]

Several attempts have been made to extend the coverage of SNOMED CT; post-coordinations were used to add modifiers,[81] and subsets and reference sets were developed to cover frequently used terms.[25] These localized attempts are, however, likely to serve as a barrier for quality assurance in the ontology composed of large

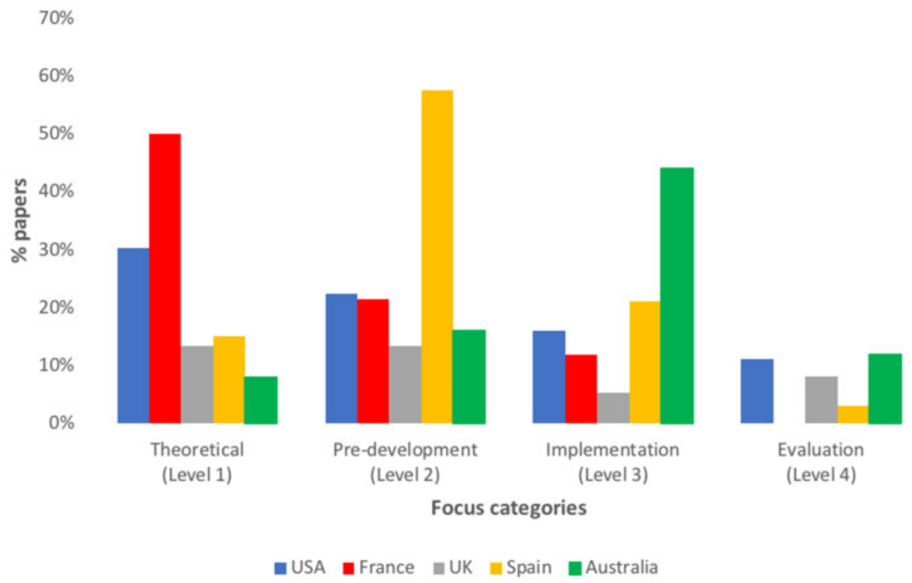

Figure 4. SNOMED CT research trends in U.S., France, U.K., Spain, and Australia, 2013-2020, by percentage of papers published. The y-axis represents the proportion of papers in each focus category. The maturity of research increases from "theoretical (level 1)" to "evaluation (level 4)" focus category.

hierarchies.[82,83] Efforts to improve content coverage should not impede the quality and scalability of the ontology.[84,85]

The coverage of content by SNOMED CT is intended to be broad and complete. Since SNOMED CT is such a large terminology system, the granularity of a SNOMED CT's concept coverage may vary by domains. For example, SNOMED CT had less granular coverage for psychological assessment instruments[86] and had fewer exact matches for cancer descriptions[29] than other terminology systems. This result is not surprising given that some terminology systems were tailored to serve specific areas unlike SNOMED CT. As Schulz et al[29] pointed out, the larger the size of the terminology, the less the inter-coder agreement is reached when coding human language with controlled vocabularies. This issue of inter-coder disagreement, which reflects an incomplete match to the original meaning of a concept, should be addressed when designing a large ontology like SNOMED CT.

Implementation

The number of papers in this focus category has substantially increased since 2012. The number of studies that used commercial or open-source NLP platforms (eg, cTAKES,[87] MedTex,[88] etc.) to automatically identify, extract, and normalize terms and provide them with SNOMED CT concepts corresponding to the constituent significantly increased. While the spectrum of NLP tasks that these tools can provide varies—from named entity recognition to the classification of documents—more attention has been paid to researchers who are looking for a convenient way to formally annotate free-text data.

Population-based registries have recently garnered the attention of researchers, clinicians, and public health investigators to serve as both a reference and a supplement to data collected from sources such as clinical trials and supporting clinical research related to disease epidemiology, interventions, and outcomes. These registries are also expected to facilitate evidence-based medicine by promoting the enrollment of fully characterized patients in clinical trials. A number of population-based registries are under development or in-service worldwide to leverage these merits.[89,90]

The capabilities of SNOMED CT to capture the meanings of data elements and to input them into the database will further demonstrate CDSSs' strengths to support better clinical decision making. The implementation of SNOMED CT in CDSSs has shown that the knowledge-based systems can augment machine learning platforms by describing results from a medical perspective.[91] To ensure seamless interoperability in large healthcare systems, further collaborations should also be facilitated between laboratory information management systems and hospital systems that may share a common EHR system.[92]

Evaluation/commodity

Attempts to evaluate the impact of SNOMED CT use on patient care were scanty; only the three studies listed in the "prove merit" usage category examined the effect or impact of SNOMED CT use. Beyond the internal validity researchers had demonstrated thus far, future research should be directed at showing the external utility of SNOMED CT, namely actual merits to patient care in clinical settings. We expect that research questions such as, "How did the use of SNOMED CT improve healthcare professionals' order skills?" and "Did the use of SNOMED CT detect the deterioration of a patient earlier?" will be addressed in the next decade.

The number of studies classified in the "retrieve or analyze patient data" usage category increased from 8 at the time of Lee et al's review to 53 in 2013-2020. As data warehouses supporting SNOMED CT codes became available and widely used in the early 2010s, researchers began to use SNOMED CT-coded patient data for retrospective epidemiological studies. Researchers can now access and collect patient data from various hospitals and healthcare systems to classify and store according to SNOMED CT, resulting in consistent querying across all sources.[93] Given SNOMED CT's ability to encode a variety of biomedical entities and its transferability among different institutions, we envision that SNOMED CT-encoded data will be the dominant form of patient data for large-scale retrospective analysis in the days to come.

Limitations

The limitations of this review include the fact that only the articles archived in PubMed and Embase were examined. It should be noted that many existing SNOMED CT implementations did not appear in the scientific literature. As Lee et al mentioned,[8] only a small portion of known implementations tends to be published in scientific publications. We examined past studies and implementations presented at SNOMED CT EXPOs (formerly SNOMED CT Implementation Showcase) to gauge the completeness of the coverage of our literature review. Only 62 out of 405 works presented at SNOMED CT EXPOs (excluding educational sessions) held between 2013 and 2020 were searchable in PubMed and Embase and these were included in this review. For a more thorough review of SNOMED CT use, we expect future research can examine those studies presented at SNOMED CT EXPOs and other proceedings but not included in this review. Adopting the same methods as Lee et al's review, however, is a key element to examine research trends in the past 20 years in a consistent way, which enabled us to appreciate how the research on SNOMED CT has evolved from 2001 to the present time.

CONCLUSION

The current review examined 622 articles published between 2013 and 2020 as an extension of Lee et al's review of those published on the same topics in 2001-2012. The papers in more mature usage categories have increased compared to those included in Lee et al's review. From the early 2010s to the present day, SNOMED CT has been increasingly implemented in more practical settings, such as building data repositories or CDSSs. However, we found only a few papers that demonstrated the impact or the merit of SNOMED CT use. Future research should concentrate on addressing whether SNOMED CT use influences improvement in patient care.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

AUTHOR CONTRIBUTIONS

JM conceived and proposed the design of the work. EC reviewed and classified papers and drafted the initial manuscript. JM critically edited the manuscript and provided insights on the review discussion. All authors participated in reviewing the comments by the reviewers and contributed to addressing the concerns raised. The final version was approved by all authors.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online

ACKNOWLEDGEMENTS

The authors acknowledge Roderick A. Gladney, Pharm.D., at the University of North Carolina at Chapel Hill School of Information and Library Science for the classification of papers.

CONFLICT OF INTEREST STATEMENT

The authors have no competing interests to declare.

DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material.

REFERENCES

1. Institute of Electrical and Electronics Engineers (IEEE). *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries*. New York, NY: The Institute of Electrical and Electronics Engineers; 1990: 114.
2. Blobel B. Architectural approach to eHealth for enabling paradigm changes in health. *Methods Inf Med* 2010; 49 (02): 123–34.
3. Oemig F, Blobel B. Natural language processing supporting interoperability in healthcare. In: Biemann C, Mehler A, eds. *Text Mining: From Ontology Learning to Automated Text Processing Applications*. Switzerland: Springer International Publishing; 2014:144.
4. Bodenreider O, Cornet R, Vreeman DJ. Recent developments in clinical terminologies - SNOMED CT, LOINC, and RxNorm. *Yearb Med Inform* 2018; 27: 129–39.
5. Komenda M, Schwarz D, Švancara J, *et al.* Practical use of medical terminology in curriculum mapping. *Comput Biol Med* 2015; 63: 74–82.
6. National Library of Medicine. Overview of SNOMED CT. https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.htmlAccessed October 14, 2020.
7. Cornet R, de Keizer N. Forty years of SNOMED: a literature review. *BMC Med Inform Decis Mak* 2008; 8 (S1): S2.
8. Lee D, de Keizer N, Lau F, *et al.* Literature review of SNOMED CT use. *J Am Med Inform Assoc* 2014; 21: e11–9.
9. Denaxas S, Gonzalez-Izquierdo A, Direk K, *et al.* UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *J Am Med Inform Assoc* 2019; 26: 1545–59.
10. Richesson RL. An informatics framework for the standardized collection and analysis of medication data in networked research. *J Biomed Inform* 2014; 52: 4–10.
11. Alahmar A, Crupi ME, Benlamri R. Ontological framework for standardizing and digitizing clinical pathways in healthcare information systems. *Comput Methods Programs Biomed* 2020; 196: 105559.
12. Gøeg KR, Cornet R, Andersen SK. Clustering clinical models from local electronic health records based on semantic similarity. *J Biomed Inform* 2015; 54: 294–304.
13. Gøeg KR, Chen R, Højen AR, *et al.* Content analysis of physical examination templates in electronic health records using SNOMED CT. *Int J Med Inform* 2014; 83: 736–49.
14. Mabotuwana T, Lee MC, Cohen-Solal EV. An ontology-based similarity measure for biomedical data-application to radiology reports. *J Biomed Inform* 2013; 46: 857–68.
15. Ji X, Ritter A, Yen P-Y. Using ontology-based semantic similarity to facilitate the article screening process for systematic reviews. *J Biomed Inform* 2017; 69: 33–42.
16. Wei DH, Fu G. Using SNOMED distance to measure semantic similarity of clinical trials. *Stud Health Technol Inform* 2017; 245: 1341.
17. Pakhomov S, McInnes B, Adam T, *et al.* Semantic similarity and relatedness between clinical terms: an experimental study. *AMIA Annu Symp Proc* 2010; 2010: 572–6.
18. Shobhana B, Radhakrishnan R. Estimation of semantic similarity between concepts and fuzzy rules optimization with modified genetic algorithm (MGA). *IIOAB J* 2016; 7: 52–60.
19. McInnes BT, Pedersen T, Liu Y, *et al.* U-path: an undirected path-based measure of semantic similarity. *AMIA Annu Symp Proc* 2014; 2014: 882–91.
20. Chandar P, Yaman A, Hoxha J, *et al.* Similarity-based recommendation of new concepts to a terminology. *AMIA Annu Symp Proc* 2015; 2015: 386–95.
21. Martínez S, Sánchez D, Valls A. A semantic framework to protect the privacy of electronic health records with non-numerical attributes. *J Biomed Inform* 2013; 46: 294–303.

22. Sánchez D, Batet M, Viejo A. Utility-preserving privacy protection of textual healthcare documents. *J Biomed Inform* 2014; 52: 189–98.
23. Konstantinidis S, Fernandez-Luque L, Bamidis P, *et al.* The role of taxonomies in social media and the semantic web for health education. A study of SNOMED CT terms in YouTube health video tags. *Methods Inf Med* 2013; 52: 168–79.
24. Oluoch T, de Keizer N, Langat P, *et al.* A structured approach to recording AIDS-defining illnesses in Kenya: a SNOMED CT based solution. *J Biomed Inform* 2015; 56: 387–94.
25. Fung KW, Xu J. An exploration of the properties of the CORE problem list subset and how it facilitates the implementation of SNOMED CT. *J Am Med Inform Assoc* 2015; 22: 649–58.
26. Taylor HL, Siddiqui Z, Frazier K, *et al.* Evaluation of a dental diagnostic terminology subset. *Stud Health Technol Inform* 2019; 264: 1602–3.
27. Chu L, Kannan V, Basit MA, *et al.* SNOMED CT concept hierarchies for computable clinical phenotypes from electronic health record data: comparison of intensional versus extensional value sets. *JMIR Med Inform* 2019; 7 (1): e11487.
28. Kahn CE. Annotation of figures from the biomedical imaging literature: a comparative analysis of RadLex and other standardized vocabularies. *Acad Radiol* 2014; 21: 384–92.
29. Schulz S, Daumke P, Romacker M, *et al.* Representing oncology in datasets: standard or custom biomedical terminology? *Inform Med Unlocked* 2019; 15: 100186.
30. Manohar N, Adam TJ, Pakhomov SV, *et al.* Evaluation of herbal and dietary supplement resource term coverage. *Stud Health Technol Inform* 2015; 216: 785–9.
31. Monsen KA, Rudenick JM, Kapinos N, *et al.* Documentation of social determinants in electronic health records with and without standardized terminologies: a comparative study. *Proc Singapore Healthc* 2018; 28: 201010581878564.
32. Campbell WS, Campbell JR, West WW, *et al.* Semantic analysis of SNOMED CT for a post-coordinated database of histopathology findings. *J Am Med Inform Assoc* 2014; 21: 885–92.
33. Ivory CH. Mapping perinatal nursing process measurement concepts to standard terminologies. *Comput Inform Nurs* 2016; 34: 312–20.
34. Le GM, Vachon R, Petit R, *et al.* SNOMED CT coding and analytics of in vitro diagnostics observations. *Stud Health Technol Inform* 2019; 264: 1460–1.
35. Matney SA, Settergren TT, Carrington JM, *et al.* Standardizing physiologic assessment data to enable big data analytics. *West J Nurs Res* 2017; 39: 63–77.
36. Sass J, Essenwanger A, Luijten S, *et al.* Standardizing Germany's electronic disease management program for bronchial asthma. *Stud Health Technol Inform* 2019; 267: 81–5.
37. Paterson GI, Christie S, Bonney W, *et al.* Synoptic operative reports for spinal cord injury patients as a tool for data quality. *Health Informatics J* 2016; 22: 984–91.
38. Højen AR, Brønnum D, Gøeg KR, *et al.* Applying the SNOMED CT concept model to represent value sets for head and neck cancer documentation. *Stud Health Technol Inform* 2016; 228: 436–40.
39. Kogan A, Tu SW, Peleg M. Goal-driven management of interacting clinical guidelines for multimorbidity patients. *AMIA Annu Symp Proc* 2018; 2018: 690–9.
40. Osborne JD, Neu MB, Danila MI, *et al.* CUILESS2016: a clinical corpus applying compositional normalization of text mentions. *J Biomed Semantics* 2018; 9: 2.
41. Peterson KJ, Liu H. Automating the transformation of free-text clinical problems into SNOMED CT expressions. *AMIA Jt Summits Transl Sci Proc* 2020; 2020: 497–506.
42. Allones JL, Taboada M, Martinez D, *et al.* SNOMED CT module-driven clinical archetype management. *J Biomed Inform* 2013; 46: 388–400.
43. Bucur A, van Leeuwen J, Chen N-Z, *et al.* Cohort selection and management application leveraging standards-based semantic interoperability and a Groovy DSL. *AMIA Jt Summits Transl Sci Proc* 2016; 2016: 25–32.
44. Safari L, Patrick JD. Restricted natural language based querying of clinical databases. *J Biomed Inform* 2014; 52: 338–53.
45. Allones JL, Martinez D, Taboada M. Automated mapping of clinical terms into SNOMED-CT. An application to codify procedures in pathology. *J Med Syst* 2014; 38: 134.
46. Butt L, Zuccon G, Nguyen A, *et al.* Classification of cancer-related death certificates using machine learning. *Australas Med J* 2013; 6: 292–9.
47. Koopman B, Zuccon G, Nguyen A, *et al.* Automatic ICD-10 classification of cancers from free-text death certificates. *Int J Med Inform* 2015; 84: 956–65.
48. Koopman B, Zuccon G, Nguyen A, *et al.* Extracting cancer mortality statistics from death certificates: a hybrid machine learning and rule-based approach for common and rare cancers. *Artif Intell Med* 2018; 89: 1–9.
49. Lin C, Hsu C-J, Lou Y-S, *et al.* Artificial intelligence learning semantics via external resources for classifying diagnosis codes in discharge notes. *J Med Internet Res* 2017; 19 (11): e380.
50. Mujtaba G, Shuib L, Raj RG, *et al.* Classification of forensic autopsy reports through conceptual graph-based document representation model. *J Biomed Inform* 2018; 82: 88–105.
51. Roldán-García MDM, García-Godoy MJ, Aldana-Montes JF. Dione: an OWL representation of ICD-10-CM for classifying patients' diseases. *J Biomed Semantics* 2016; 7: 62.
52. Ternois I, Billard-Pomares T, Carbonelle E, *et al.* Using SNOMED-CT to help the transition from microbiological data to ICD-10 sepsis codes. *Stud Health Technol Inform* 2019; 264: 1604–5.
53. Banerjee I, Ling Y, Chen MC, *et al.* Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. *Artif Intell Med* 2019; 97: 79–88.
54. Zuccon G, Wagholikar AS, Nguyen AN, *et al.* Automatic classification of free-text radiology reports to identify limb fractures using machine learning and the SNOMED CT ontology. *AMIA Jt Summits Transl Sci Proc* 2013; 2013: 300–4.
55. Hitchins ARC, Hogan SC. Outcomes of early intervention for deaf children with additional needs following an Auditory Verbal approach to communication. *Int J Pediatr Otorhinolaryngol* 2018; 115: 125–32.
56. Hier DB, Pearson J. Two algorithms for the reorganisation of the problem list by organ system. *BMJ Health Care Inform* 2019; 26 (1): e100024.
57. Barros JM, Duggan J, Rebholz-Schuhmann D. Disease mentions in airport and hospital geolocations expose dominance of news events for disease concerns. *J Biomed Semantics* 2018; 9: 18.
58. Jani BD, Pell JP, McGagh D, *et al.* Recording COVID-19 consultations: review of symptoms, risk factors, and proposed SNOMED CT terms. *Br J Gen Pract Open* 2020; 4 (4): bjgpopen20X101125.
59. Song T-M, Park H-A, Jin D-L. Development of health information search engine based on metadata and ontology. *Healthc Inform Res* 2014; 20: 88–98.
60. Wang Y, Luo J, Hao S, *et al.* NLP based congestive heart failure case finding: a prospective analysis on statewide electronic medical records. *Int J Med Inform* 2015; 84: 1039–47.
61. Perl Y, Geller J, Halper M, *et al.* Introducing the Big Knowledge to Use (BK2U) challenge. *Ann N Y Acad Sci* 2017; 1387: 12–24.
62. Randorff Højen A, Sundvall E, Rosenbeck Gøeg K. Visualizing sets of SNOMED CT concepts to support consistent terminology implementation and reuse of clinical data. *Stud Health Technol Inform* 2013; 192: 1160.
63. Silva Layes E, Bondarenco M, Machiavello D, *et al.* Implementation of a terminology server with SNOMED CT in graph databases. *Stud Health Technol Inform* 2019; 264: 1584–5.
64. Sun M, Zhu W, Tao S, *et al.* COBE: a conjunctive ontology browser and explorer for visualizing SNOMED CT fragments. *AMIA Annu Symp Proc* 2015; 2015: 2092–100.
65. Danahey K, Borden BA, Furner B, *et al.* Simplifying the use of pharmacogenomics in clinical practice: building the genomic prescribing system. *J Biomed Inform* 2017; 75: 110–21.
66. Noussa-Yao J, Boussadi A, Richard M, *et al.* Using a snowflake data model and autocompletion to support diagnostic coding in acute care hospitals. *Stud Health Technol Inform* 2015; 210: 334–8.

67. Reed SG, Adibi SS, Coover M, *et al.* Does use of an electronic health record with dental diagnostic system terminology promote dental students' critical thinking? *J Dent Educ* 2015; 79: 686–96.
68. Souvignet J, Declerck G, Trombert-Paviot B, *et al.* Semantic queries expedite MedDRA terms selection thanks to a dedicated user interface: a pilot study on five medical conditions. *Front Pharmacol* 2019; 10: 50.
69. Dougall B, Gendreau J, Das S, *et al.* Melanoma registry underreporting in the veterans health administration. *Fed Pract* 2016; 33: 55S–9S.
70. SNOMED International. Members. In: SNOMED International. https://www.snomed.org/our-customers/members Accessed September 30, 2020.
71. Budanitsky A, Hirst G. Evaluating WordNet-based measures of lexical semantic relatedness. *Comput Ling* 2006; 32 (1): 13–47.
72. Lin D. Automatic retrieval and clustering of similar words. In: Proceedings of the 36th Annual Meeting on Association for Computational Linguistics. Morristown, NJ, USA: Association for Computational Linguistics; 1998: 768–74.
73. Zare M, Pahl C, Nilashi M, *et al.* A review of semantic similarity measures in biomedical domain using SNOMED-CT. *J Soft Comput Decis Support Syst* 2015; 2 (6): 1–13.
74. Sánchez D, Batet M. Semantic similarity estimation in the biomedical domain: an ontology-based information-theoretic perspective. *J Biomed Inform* 2011; 44: 749–59.
75. Zivaljevic A, Atalag K, Warren J. Utility of SNOMED CT in automated expansion of clinical terms in discharge summaries: testing issues of coverage. *Health Inf Manag* 2020; 1833358320934528.
76. Rastegar-Mojarad M, Sohn S, Wang L, *et al.* Need of informatics in designing interoperable clinical registries. *Int J Med Inform* 2017; 108: 78–84.
77. Arons A, DeSilvey S, Fichtenberg C, *et al.* Documenting social determinants of health-related clinical activities using standardized medical vocabularies. *JAMIA Open* 2019; 2: 81–8.
78. Arons A, DeSilvey S, Fichtenberg C, *et al.* Documenting social determinants of health using standardized EHR vocabularies. *J Gen Intern Med* 2018; 33: 168.
79. Bettencourt-Silva JH, Mulligan N, Sbodio M, *et al.* Discovering new social determinants of health concepts from unstructured data: framework and evaluation. *Stud Health Technol Inform* 2020; 270: 173–7.
80. Aziz A, Pflieger L, O'Connell N, *et al.* Compatibility of family history cancer guidelines with meaningful use standards. *JCO Clin Cancer Inform* 2017; (1): 1–9.
81. Dhombres F, Winnenburg R, Case JT, *et al.* Extending the coverage of phenotypes in SNOMED CT through post-coordination. *Stud Health Technol Inform* 2015; 216: 795–9.
82. López-García P, Schulz S. Structural patterns under X-rays: is SNOMED CT growing straight? *PLoS One* 2016; 11: e0165619.
83. Rector AL, Brandt S, Schneider T. Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications. *J Am Med Inform Assoc* 2011; 18: 432–40.
84. Rosenbloom ST, Miller RA, Johnson KB, *et al.* Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *J Am Med Inform Assoc* 2006; 13: 277–88.
85. Rector AL. Thesauri and formal classifications: terminologies for people and machines. *Methods Inf Med* 1998; 37: 501–9.
86. Ranallo PA, Adam TJ, Nelson KJ, *et al.* Psychological assessment instruments: a coverage analysis using SNOMED CT, LOINC and QS terminology. *AMIA Annu Symp Proc* 2013; 2013: 1333–40.
87. Savova GK, Masanz JJ, Ogren PV, *et al.* Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010; 17: 507–13.
88. Nguyen A, Lawley M, Hansen D, *et al.* A simple pipeline application for identifying and negating SNOMED Clinical Terminology in free text. *Health Inform Conf* 2009; 188–93.
89. Siddiqui AH, Zafar SN. Global availability of cancer registry data. *J Glob Oncol* 2018; 4: 1–3.
90. World Health Organization. Existence of population-based cancer registry. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/existence-of-population-based-cancer-registryAccessed September 25, 2020.
91. Müller L, Gangadharaiah R, Klein SC, *et al.* An open access medical knowledge base for community driven diagnostic decision support system development. *BMC Med Inform Decis Mak* 2019; 19: 93.
92. Abhyankar S, Goodwin RM, Sontag M, *et al.* An update on the use of health information technology in newborn screening. *Semin Perinatol* 2015; 39: 188–93.
93. SNOMED International. 8.2 Data warehouse - data analytics with SNOMED CT. https://confluence.ihtsdotools.org/display/DOCANLYT/8.2+Data+Warehouse Accessed October 16, 2020.