

# SNOMED CT Concept Hierarchies for Sharing Definitions of Clinical Conditions Using Electronic Health Record Data

Duwayne L. Willett<sup>1,2</sup> Vaishnavi Kannan<sup>2</sup> Ling Chu<sup>1,2</sup> Joel R. Buchanan<sup>3</sup> Ferdinand T. Velasco<sup>4,5</sup>  
 John D. Clark<sup>1</sup> Jason S. Fish<sup>1,5</sup> Adolfo R. Ortuzar<sup>2</sup> Josh E. Youngblood<sup>2</sup> Deepa G. Bhat<sup>5</sup>  
 Mujeeb A. Basit<sup>1,2</sup>

<sup>1</sup>Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, United States

<sup>2</sup>Health System Information Resources Department, University of Texas Southwestern Medical Center, Dallas, Texas, United States

<sup>3</sup>Department of Medicine, University of Wisconsin School of Medicine and Public Health, University of Wisconsin–Madison, Madison, Wisconsin, United States

<sup>4</sup>Texas Health Resources, Arlington, Texas, United States

<sup>5</sup>Southwestern Health Resources, Dallas, Texas, United States

**Address for correspondence** Duwayne L. Willett, MD, MS, University of Texas Southwestern Health System, 5323 Harry Hines Boulevard, Dallas, TX 75390-9047, United States  
 (e-mail: Duwayne.Willett@UTSouthwestern.edu).

Appl Clin Inform 2018;9:667–682.

## Abstract

**Background** Defining clinical conditions from electronic health record (EHR) data underpins population health activities, clinical decision support, and analytics. In an EHR, defining a condition commonly employs a diagnosis value set or “grouper.” For constructing value sets, Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) offers high clinical fidelity, a hierarchical ontology, and wide implementation in EHRs as the standard interoperability vocabulary for problems.

**Objective** This article demonstrates a practical approach to defining conditions with combinations of SNOMED CT concept hierarchies, and evaluates sharing of definitions for clinical and analytic uses.

**Methods** We constructed diagnosis value sets for EHR patient registries using SNOMED CT concept hierarchies combined with Boolean logic, and shared them for clinical decision support, reporting, and analytic purposes.

**Results** A total of 125 condition-defining “standard” SNOMED CT diagnosis value sets were created within our EHR. The median number of SNOMED CT concept hierarchies needed was only 2 (25th–75th percentiles: 1–5). Each value set, when compiled as an EHR diagnosis grouper, was associated with a median of 22 International Classification of Diseases (ICD)-9 and ICD-10 codes (25th–75th percentiles: 8–85) and yielded a median of 155 clinical terms available for selection by clinicians in the EHR (25th–75th percentiles: 63–976). Sharing of standard groupers for population health, clinical decision support, and analytic uses was high, including 57 patient registries (with 362 uses of standard groupers), 132 clinical decision support records, 190 rules, 124 EHR reports, 125 diagnosis dimension slicers for self-service analytics, and 111 clinical quality measure calculations. Identical SNOMED CT definitions were created in an EHR-agnostic tool enabling application across disparate organizations and EHRs.

## Keywords

- ▶ SNOMED CT
- ▶ ICD-10
- ▶ electronic health records and systems
- ▶ disease management
- ▶ health information exchanges
- ▶ registries
- ▶ translational research

received  
 February 12, 2018  
 accepted after revision  
 June 29, 2018

DOI <https://doi.org/10.1055/s-0038-1668090>.  
 ISSN 1869-0327.

© 2018 Georg Thieme Verlag KG  
 Stuttgart · New York

License terms



**Conclusion** SNOMED CT-based diagnosis value sets are simple to develop, concise, understandable to clinicians, useful in the EHR and for analytics, and shareable. Developing curated SNOMED CT hierarchy-based condition definitions for public use could accelerate cross-organizational population health efforts, “smarter” EHR feature configuration, and clinical–translational research employing EHR-derived data.

## Background and Significance

How should we define various “populations” for population health? Identifying patients who share a common clinical condition (phenotype) proves valuable for clinical care, for population health management, and for clinical–translational research (→Table 1; see also →Table 2 for definitions used in this article). So how can we define a clinical condition most practically and effectively in the era of electronic health records (EHRs)? And how can we work with partners in population health initiatives and clinical research to define any condition similarly?

Previously, defining clinical conditions frequently employed the best and only data available digitally: claims data, tagged with International Classification of Diseases (ICD) diagnosis codes.<sup>1</sup> Thus, shared condition definitions traditionally have been based on published lists of ICD codes (“value sets”), either nationally or locally defined. But in the EHR era, richer sources of digital data abound at the level of clinical events. In many EHRs, clinicians select diagnoses associated with these events using clinical terms rather than ICD codes. These clinical terms, often sourced from a vendor such as Intelligent Medical Objects (IMO) or Health Language, map both to ICD codes (for billing and coding purposes) and to Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) concepts (formerly SNOMED Clinical Terms), as the required nomenclature for EHR interoperability in the United States via health information exchanges (HIEs). Being more numerous and granular, clinical terms and SNOMED CT concepts often enable higher

fidelity to clinical diagnostic thinking than ICD codes (→Fig. 1).

## International Classification of Diseases Value Sets to Define Conditions

### International Classification of Diseases Coding and Classification

The ICD coding system traces its roots to initial efforts at statistically analyzing causes of death, beginning with John Graunt’s London Bills of Mortality in the 17th century.<sup>2–4</sup> Although rooted firmly in international epidemiology, additional uses emerged, and ICD-7 was adopted in the United States for hospital coding. ICD-9, released in 1979, remained in use in the United States as ICD, Ninth Revision, Clinical Modification (ICD-9-CM) until 2015, when it was replaced by the then 25-year-old ICD-10 (with clinical modifications as ICD-10-CM).

ICD’s subdivisions are largely body-system based, and employ a strict, disjoint classification scheme. That is, any ICD code has only a single path up to the root (top) of the hierarchy. For statistical reporting, a single path up avoids double-counting of epidemiologic events in summarized data. In the United States, the Centers for Medicare and Medicaid Services produces lists of ICD codes (value sets) for use in calculating performance measures based on ICD-coded claims data.<sup>5</sup> These value sets are publicly available via the National Library of Medicine’s Value Set Authority Center (VSAC), used consistently throughout the country, and

**Table 1** Example use cases for defining patient clinical conditions

Clinical care	Population health	Clinical–translational research
Deliver real-time clinical decision support: <ul style="list-style-type: none"> <li>• Best practice guidelines and pathways for a condition</li> <li>• Drug-disease interactions</li> </ul> Configure “smarter” EHRs, automatically tailored to the patient’s condition(s): <ul style="list-style-type: none"> <li>• Documentation templates</li> <li>• Order sets</li> </ul> Construct registries: <ul style="list-style-type: none"> <li>• Lists of “my patients” or “our patients” with a given condition, for operational or care coordination reasons</li> </ul>	For patients with a given condition: <ul style="list-style-type: none"> <li>• Design care pathways</li> <li>• Close care gaps</li> <li>• Conduct targeted outreach and communications</li> <li>• Calculate and report on process measures and outcome measures</li> <li>• Combine clinical data from disparate EHRs and other sources for a 360° view of the patient</li> <li>• Perform risk stratification and other predictive analytics</li> </ul>	<ul style="list-style-type: none"> <li>• Find patients with a given clinical condition to invite to participate in clinical/translational research (potentially across disparate EHRs)</li> <li>• Define a clinical condition as a “phenotype” of interest to study</li> <li>• Define conditions as comorbidities or covariates</li> </ul>

Abbreviation: EHR, electronic health record.

**Table 2** Words or phrases with their associated meaning as used in this article

Word or phrase used in this article	Meaning in the context of this article	Alternative phrases
Value set	A uniquely identifiable set of valid codes, concepts, or clinical terms from a terminology, used together to represent a useful clinical grouping	Grouper
[Clinical] condition	A diagnosis, illness, or disease. May be general (as in “Arrhythmia (atrial or ventricular)”) or more specific (as in “AV nodal reentrant tachycardia”). Patient registries typically consist of patients with either a shared clinical condition, or a shared exposure (e.g., to a procedure, drug, or environmental agent). <sup>26</sup> In an EHR, entries on the patient’s Problem List are usually clinical conditions	Medical condition; clinical phenotype; illness; disease
[EHR] grouper	A set of codes, concepts, or clinical terms from a terminology as implemented in an EHR, used together to represent a useful clinical grouping. EHR diagnosis groupers are commonly used to define a clinical condition within the EHR	Diagnosis grouper; value set
SNOMED CT top-level hierarchy	The current 19 top-level hierarchies, under which all other SNOMED CT concepts are included as more specific subtypes (see “Subtype relationship” below). Examples of top-level hierarchies include “Clinical finding,” “Procedure,” and “Body structure”	SNOMED CT hierarchy
SNOMED CT concept hierarchy	Any given SNOMED CT concept along with all its descendants (subordinate concepts) as defined by subtype relationships <sup>56</sup>	SNOMED CT concept including all descendants
Subtype relationship	A relationship between two SNOMED CT concepts where one concept is a more specific subtype of another, more general concept. The most widely used type of relationship in SNOMED CT, also known as an “is a” relationship	“Is a” relationship; parent–child relationship
Subpopulation	The subset of persons/patients resulting from some segmentation algorithm. Used here primarily for patient registries which identify patients with a shared condition or a shared exposure	Registry population; population

Abbreviations: EHR, electronic health record; SNOMED CT, Systematized Nomenclature of Medicine–Clinical Terms.

employed for both financial payment programs and public-reporting of quality measures. Additionally, ICD code lists have been commonly employed in epidemiologic studies, and have a long history of use in scientific publications.

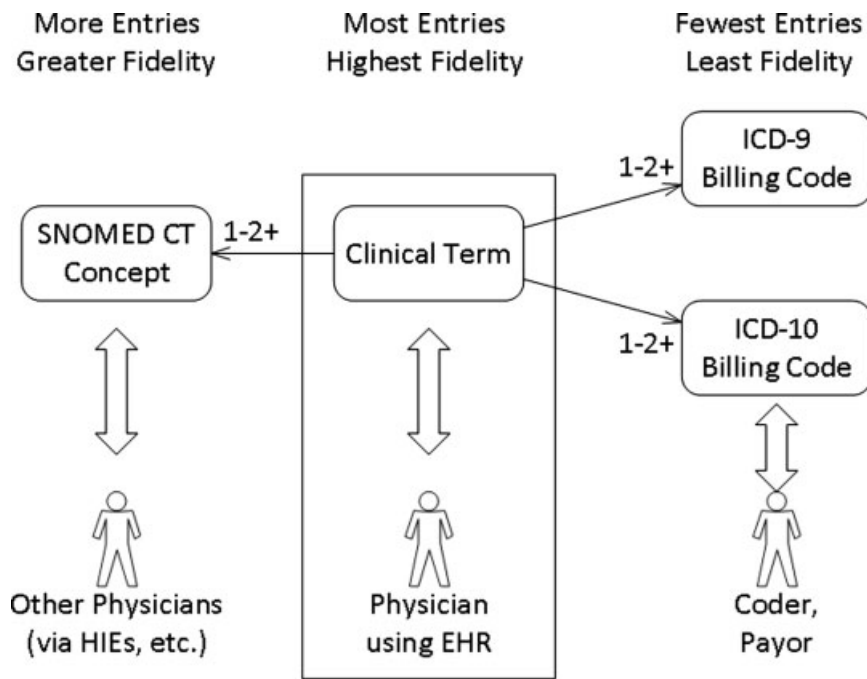
#### Why not International Classification of Diseases Value Sets for Defining Conditions?

Given these well-established uses of ICD code value sets, why even look at an alternative? The original intent of ICD as a high-level classification for epidemiological understanding of broad causes of mortality and morbidity makes it less useful for most accurately capturing clinical information at the point of care. For clinical input into electronic systems, a clinical terminology proves better suited.<sup>6,7</sup>

In the EHR era, content-rich clinical data are now available electronically: patient-level conditions (problem lists), orders, encounters, test results, procedures, and timed events such as sequential process steps or displays of clinical decision support (CDS) advisories with corresponding clinician responses. Diagnoses at these patient or event levels are typically entered in the EHR as clinical terms (from a vendor’s

clinical terminology or using SNOMED CT itself). Since these clinical terminologies are more granular than ICD,<sup>6,8,9</sup> employing ICD value sets alone risks loss of clinical fidelity to the patient’s specific condition.

Accountable Care Organizations (ACOs) and other population health initiatives increasingly need to identify subgroups of patients with a given condition, and to combine clinical data from disparate EHRs for near real-time CDS and care coordination for that condition. Claims data alone can be inadequate for this purpose, both from a timing and an information content standpoint. For instance, claims data are necessarily delayed by the claims submission and adjudication process; consequently, diagnoses from claims are not immediately available for real-time CDS upon entry of the diagnosis in the EHR. Claim diagnoses typically also lag behind transmission of the Continuity of Care Document (CCD) SNOMED CT-encoded diagnosis data via an HIE, which occurs upon completion of an encounter. Increasingly, clinical data derived from EHRs and rooted in clinical terminologies will be a cornerstone for ACO data repositories, analytics, and interventions.



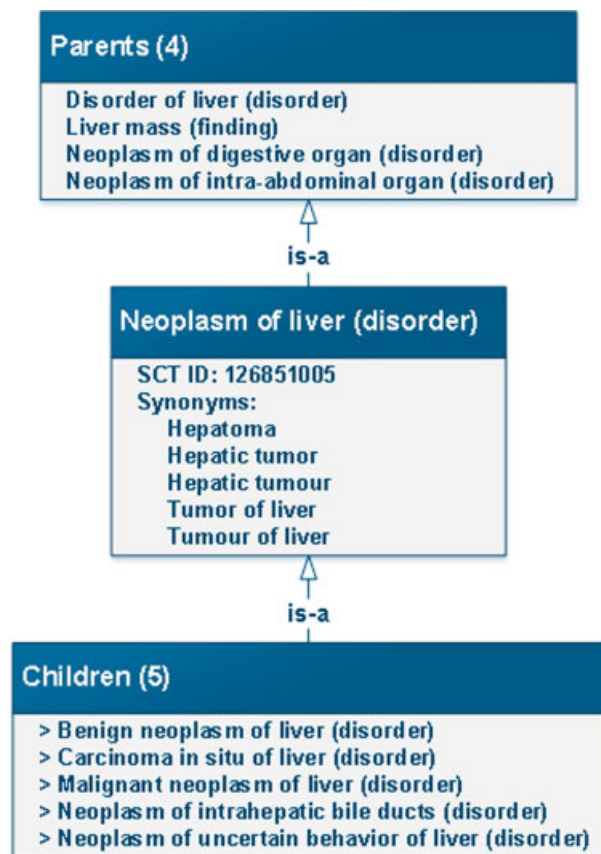
**Fig. 1** Relation of clinical terms used by physicians within the electronic health record (EHR) to Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) concepts for interoperability, and to International Classification of Diseases (ICD) codes for billing.

### SNOMED CT Concept Hierarchies to Define Conditions

#### SNOMED CT and Concept Hierarchies

SNOMED CT serves as a publicly available, international clinical terminology for use in electronic health care applications. As the most comprehensive clinical terminology, SNOMED CT aids in inputting and retrieving coded clinical information, and in interoperability between clinical systems.<sup>3</sup> Originated in 1965 as the Systematic Nomenclature of Pathology by the College of American Pathologists, SNOMED later merged with the “Read Codes” developed by clinicians in England’s National Health Service to form SNOMED CT, released in 2002.<sup>10</sup> Iterative development of SNOMED CT occurs through a governing body, SNOMED International.<sup>11</sup>

In contrast to ICD, SNOMED CT is an ontology supporting multiple types of relationships between clinical concepts.<sup>3,12,13</sup> The subtype relationship (also known as an “is a” relationship) defines one concept as a subtype of another, more general, “parent” concept. This enables efficient classification of a clinical condition by including references to a SNOMED CT concept with all its hierarchical “children” and further descendant subtype concepts.<sup>14</sup> Additionally, SNOMED CT supports polyhierarchies. For instance, a “Neoplasm of liver” is both a subtype of “Disorder of liver” and a subtype of “Neoplasm.” In SNOMED CT, one can find “Neoplasm of liver” (and from there any specific type of liver cancer) traversing down either path (→ Fig. 2). In ICD-10, one would have to choose whether to classify “Liver cancer” under Chapter 11 “Diseases of the digestive system” or Chapter 2 “Neoplasms.” In addition to the concept hierarchy-defining subtype relationship, over 50 attribute relationships can be used to connect concepts among different SNOMED CT top-level hierarchies.<sup>15</sup>



**Fig. 2** Example of a Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) polyhierarchy.<sup>24</sup> A “Neoplasm of liver (disorder)” has 4 parents, including both “Disorder of liver (disorder)” and “Neoplasm of digestive organ (disorder).” Also shown are 5 “children” such as the concept “Malignant neoplasm of liver (disorder),” which in turn has more specific descendants.

SNOMED CT International Edition is released in January and July of each year and the U.S. version is released in March and September. The September 2017 SNOMED CT (U.S. edition) release includes 560,985 concepts,<sup>11</sup> while the August 2017 ICD-10-CM release (for U.S. fiscal year 2018) includes only 71,704 codes.<sup>16</sup> As a result, SNOMED CT concepts match many conditions more specifically than does ICD-10. For instance, SNOMED CT concepts, but not ICD-10 codes, distinguish among different types of kidney cancer and of acidosis, for which clinical management varies substantially (→Table 3).

### SNOMED to Define Patient Conditions in EHRs

Many EHRs support the creation of *EHR diagnosis groupers* to define and reuse a group of clinical conditions. These diagnosis groupers (a type of “content reference set”<sup>3</sup>) can be created using SNOMED CT concept hierarchies, ICD codes, or lists of individual clinical terms. SNOMED CT-based diagnosis groupers offer the potential to be:

1. Simple to define, because of the logical supertype–subtype nature of SNOMED CT parent–child relationships. Specifying one or a few SNOMED CT supertype (parent) concepts defines a grouper containing multiple more granular subtype (“descendant”) disorders.
2. A natural high-fidelity match for the clinical vocabulary of clinical experts who know which subconditions should or should not be included in a subpopulation of clinical interest (→Fig. 3).
3. More resilient to future changes within the coding system, such as addition of new concepts or deprecation of old ones, without requiring grouper redefinition.

Within the EHR, potential reuses of a standard, vetted diagnosis grouper for the same condition include CDS, dynamic (rule-based) appearance of tailored documentation tools and order sets, and population of patient registries.<sup>17,18</sup>

### SNOMED to Define Conditions for ACOs

ACOs frequently employ an HIE strategy for combining data from disparate EHRs. HIEs support the federally defined CCD standard, which uses SNOMED CT for exchanging diagnosis information.<sup>19</sup> SNOMED CT value sets thus provide a straightforward way to define conditions of interest from HIE or other CCD-derived data. SNOMED CT condition definitions can then be shared for a variety of population health purposes, such as care coordination, targeted outreach, clinical quality measure (CQM) calculations, and as variables in risk stratification and predictive analytic algorithms.

### SNOMED to Define Conditions for Clinical and Translational Research

Clinical and translational scientists leverage the richer clinical data now being stored in EHRs, to conduct analyses not otherwise possible using claims and other administrative data alone.<sup>20–24</sup> Such research also benefits from the higher fidelity of SNOMED CT-encoded diagnoses for clinically important distinctions among subtypes of conditions. Even if a given research project requires an idiosyncratic definition of the primary study population of interest, covariate conditions can still share existing definitions, rather than having to redundantly create new ones for each project.

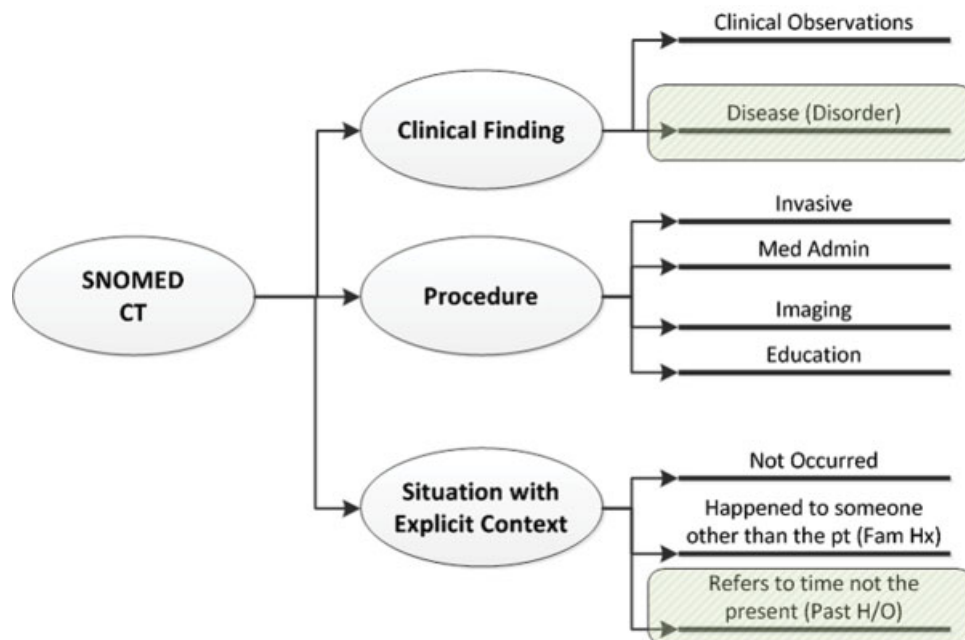
In summary, significant benefit can be derived from defining each clinical condition (such as “renal cell carcinoma”) once, with a single vetted SNOMED CT value set

**Table 3** Clinical precision differences between ICD and SNOMED CT

Clinical condition	ICD-9-CM	ICD-10-CM	SNOMED CT
Renal cell carcinoma	189.0 Malignant neoplasm of kidney, except pelvis	C64.9 Malignant neoplasm of unspecified kidney, except renal pelvis	702391001 Renal cell carcinoma
Transitional cell carcinoma of kidney	189.0 Malignant neoplasm of kidney, except pelvis	C64.9 Malignant neoplasm of unspecified kidney, except renal pelvis	408642003 Transitional cell carcinoma of kidney
Nephroblastoma	189.0 Malignant neoplasm of kidney, except pelvis	C64.9 Malignant neoplasm of unspecified kidney, except renal pelvis	302849000 Nephroblastoma
Metabolic acidosis	276.2 Acidosis	E87.2 Acidosis	59455009 Metabolic acidosis
Respiratory acidosis	276.2 Acidosis	E87.2 Acidosis	12326000 Respiratory acidosis
Lactic acidosis	276.2 Acidosis	E87.2 Acidosis	91273001 Lactic acidosis
Neurosarcoidosis	135 Sarcoidosis	D86.89 Sarcoidosis of other sites	231093008 Neurosarcoidosis (and descendants)

Abbreviations: ICD-9 (10)-CM, International Classification of Diseases, Ninth (Tenth) Revision, Clinical Modification; SNOMED CT, Systematized Nomenclature of Medicine–Clinical Terms.

Note: Three examples are shown: SNOMED CT concepts distinguish among different types of kidney cancer (3 types shown), and also differentiate different types of acidosis. ICD-10 codes do not distinguish among these clinically relevant subtypes.<sup>55</sup> Neurosarcoidosis is a relatively rare condition: it can be searched for by its SNOMED CT concept hierarchy, but not by ICD-9 or ICD-10 code.



**Fig. 3** Most commonly used Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) branches for constructing hierarchical subsets defining a diagnostic condition. Primary defining conditions most often are found in the Disease (Disorder) section of the “Clinical Finding” branch. History of a condition, if desired, is within the “Situation with Explicit Context” branch. Occasionally, entries in the Procedure branch prove relevant.

which then can be shared for clinical, population health, and clinical–translational research purposes.<sup>25</sup>

### The Present Project

In 2015, we undertook an Ambulatory Quality Outcomes project to develop at least one EHR-based specialty-specific registry for each of 30 specialties at the University of Texas (UT) Southwestern.<sup>18</sup> Registries included combinations of:

- Clinician documentation tools in the EHR.
- CDS tools.
- Patient questionnaires for patient-reported outcomes (for some registries).
- Patient registry list(s) viewable within the EHR.
- Data warehouse–derived clinical quality performance measures, fed back into the EHR.

We relied on SNOMED CT–based diagnosis groupers to define conditions of interest, either as primary conditions for registries or as covariate conditions. These SNOMED CT groupers were designated as health system “standard” groupers, and reused in CDS tools, rules, reports, performance measure calculations, and tailoring of relevant content to patients on our patient portal. Sharing SNOMED CT definitions of conditions aided in rapid-cycle development of multiple specialty patient registries, accelerating implementation of our population health initiatives.<sup>18</sup> As of January 2018, over 80,000 distinct patients were actively managed on one or more of 57 registries, with 10,875 patient-reported outcome questionnaires completed as part of registry-related data collection.

In this report we (1) describe creation of SNOMED CT groupers to define multiple specialty conditions for our

registry project, (2) evaluate the relative complexity involved to construct and maintain them, and (3) assess the shared reuse of these groupers for a variety of clinical and analytic purposes.

### Objective

This article demonstrates a practical approach to defining clinical conditions with combinations of SNOMED CT concept hierarchies, and evaluates the potential for sharing definitions for a broad range of clinical, population health, and analytic uses.

### Materials and Methods

#### Software

#### EHRs and Health Information Exchanges

Our organizations each operate a separate instance of EHR software from Epic (Verona, Wisconsin, United States). SNOMED CT–encoded diagnoses are exchanged between these Epic instances via a standard CCD format using the included HIE capability (Care Everywhere). Our organizations also participate in one or more national HIEs (eHealth Exchange, CareEquality), enabling CCD exchange with EHRs from any participating EHR vendor and with the Veterans Administration EHR.

The UT Southwestern clinically affiliated network of physicians comprises practices using a variety of individual EHRs. For population health purposes, we are linking these practices to each other and to Epic via an internal HIE (dbMotion, from Allscripts, Chicago, Illinois, United States).

To date, the following EHRs have been linked to our dbMotion HIE and can exchange SNOMED CT-encoded diagnosis data: Allscripts Sunrise, eClinicalWorks (Westborough, Massachusetts, United States), Epic, and NextGen (Horsham, Pennsylvania, United States), with other EHRs continuing to be added.

### Clinical Terminology Vocabulary

The clinical terminology vocabulary employed in UT Southwestern's Epic instance during this study was IMO's proprietary Problem (IT) Terminology, version 2018 R1, corresponding to the SNOMED CT International Edition July 2017 release and the SNOMED CT U.S. Edition September 2017 release.

### Clinical Terminology Mapping

Our dbMotion HIE employs clinical terminology mapping software (Symedical, from Clinical Architecture, Carmel, Indiana, United States) for mapping EHR-specific content identifiers to reference clinical terminologies and ontologies. Additional content subset modeling capabilities of Symedical were employed for defining EHR-agnostic and consistent condition definitions using SNOMED CT concept hierarchies.<sup>3</sup>

### Standard Diagnosis Groupers

For specialty patient registries based on a shared condition,<sup>26</sup> we chose to use a SNOMED CT concept hierarchy-based definition for the primary condition, as well as for any comorbid conditions. That is, SNOMED CT-defined conditions were employed both to select the list of patients included in a registry (i.e., displayable on rows in a registry report), and typically also to determine one or more registry metrics (each displayable as a column of information about a registry patient). Specialty registries were developed on a staggered basis during a series of 2-week development iterations (mean, 4 iterations per registry);<sup>18</sup> any new standard diagnosis grouper construction required by a registry took place during one or more of those iterations. An initial set of 125 diagnosis groupers requested as part of this specialty registry development were selected for this study, without restriction on the specialties or conditions involved.

## Approach

### Defining and Constructing SNOMED CT Concept Groupers

Given a request to define a condition within the EHR, a clinical informaticist initially searched for the condition with a SNOMED CT hierarchy browser—either one within the EHR or SNOMED International's Web-based SNOMED CT Browser.<sup>27</sup> Searching on common clinical synonyms for the condition invariably identified one or more matching SNOMED CT concepts, most commonly within the "Clinical finding" top-level hierarchy (which includes "Disorders" or diagnoses). Each initial concept located by searching will be referred to as an "index" concept below.

Once found, the index SNOMED CT concept choice was refined with a "drill-up, drill-down" approach. First, a "drill-up" examination of each parent concept of the index concept

was done (by selecting a parent within the SNOMED CT browser software). This helped gauge (1) if the parent itself more accurately included all the intended condition, and thus should be used instead of the index concept, or (2) if some of the parent's other "child" concepts (i.e., "siblings" of the index concept) were also relevant and should be included in addition to the index concept. For instance, in constructing a grouper for the condition "Coronary artery disease (CAD)," drilling up from the SNOMED CT concept "Coronary atherosclerosis (disorder)" yields the parent "Disorder of coronary artery (disorder)" which proves too broad. Examination of the siblings of "Coronary atherosclerosis (disorder)" reveals some siblings should be included as indicating the presence of CAD, such as "Mechanical complications of coronary bypass (disorder)," while other siblings should be excluded, such as "Congenital anomaly of coronary artery (disorder)." Next, for each identified index concept, a "drill-down" examination of the concept's "descendants" was done, to see if any should be excluded. For instance, for a candidate concept of "Malignant neoplasm of breast (disorder)," the child "Malignant melanoma of the breast (disorder)" could be selectively excluded.

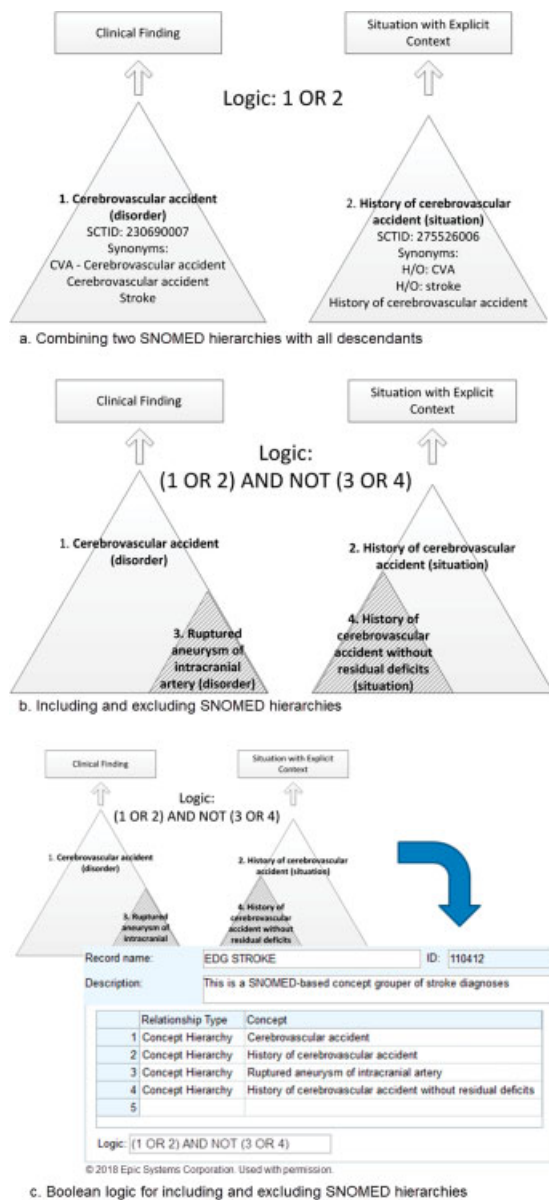
The above sequence (search, drill-up, drill-down) was then repeated as needed to look for other condition-defining concepts. Typical additional searches would be for (1) a "history of" concept (within the "Situation with explicit concept" top-level hierarchy), (2) complications of the condition implying its presence (e.g., "diabetic ketoacidosis" for diabetes), or (3) condition-defining procedures, e.g., "Coronary bypass grafting" implying presence of CAD, within the "Procedures" top-level hierarchy (→Fig. 3).

Then, each included and excluded concept was numbered from 1 to  $n$ , and Boolean logic was written to combine them. By convention, included concept hierarchies were listed first, followed by the excluded concepts. As an example, for 2 included concepts and 2 excluded ones, the Boolean logic would be "(1 OR 2) AND NOT (3 OR 4)"—see →Fig. 4. By using concept hierarchies, this method leverages the subtype relationships within the SNOMED CT ontology; attribute relationships were not employed.

Finally, from the grouper Boolean logic definition, a full list of grouper contents was generated. In the EHR browser, this was a list of included clinical terms from the clinical terminology vocabulary (→Fig. 5). In the Symedical clinical terminology and mapping software, this was a list of SNOMED CT concepts, including descendants (→Fig. 6). Each list was then reviewed for any terms or concepts that should have been excluded or were notably missing. If so, the process was repeated to further refine the SNOMED CT concept selection and Boolean logic.

### Vetting SNOMED CT Grouper Design with Clinicians

The more closely a SNOMED CT grouper validly represents a unique, clinically important condition, the greater its value for shared use. Accordingly, subject matter expert clinician vetting proves advantageous. In vetting a given grouper with specialist experts, we sought their clinical judgment about which real-world diagnosis subtypes to include in, or exclude



**Fig. 4** Constructing Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) hierarchy-based condition definitions. (A) In this hypothetical example meant to include any patients who have ever had a stroke, concept hierarchies from the Disorder branch and the Situation with Explicit Context branch are combined with Boolean logic. (B) Some descendants of a hierarchy can be excluded if desired. In this example, “Ruptured aneurysm” is excluded, perhaps to be included in a separate “intracranial bleeding” condition definition, and “History of cerebral vascular accident (CVA) without residual deficits” is excluded for being potentially unverified (again, all hypothetical for illustration purposes only). Boolean logic handles both the inclusion and exclusion criteria. (C) Within the electronic health record (EHR), construction of the condition-defining diagnosis grouper is straightforward, using Boolean logic.

from, the target subpopulation. We avoided making them sift through either a long list of ICD-10 codes or the even longer list of clinical terms within the EHR. Rather, we posed questions based on the much smaller set of relevant SNOMED

CT concepts and descendants, as in **Fig. 6**. (Example questions: should “gestational diabetes” and/or “preexisting diabetes mellitus in pregnancy” be included in a definition of “diabetes mellitus”?; should “stunned myocardium” and/or “hibernating myocardium” be included in a definition of “ischemic cardiomyopathy”?) Once vetted, the grouper’s name in the EHR was appended with a specific suffix “(Standard),” to streamline recognition and promote reuse.

### Employing SNOMED CT Groupers in the EHR

Standard groupers were reused in rules, decision support advisory records, and reports. Rules can be evaluated at multiple points throughout the EHR, for instance, dynamically presenting condition-specific documentation templates or banners to clinicians. Whenever a rule needed to check for the presence or absence of a diagnostic condition, we encouraged searching for and reusing a “(Standard)” SNOMED CT grouper, rather than creation of a duplicative one-off grouper for isolated use by the rule. Similarly, CDS advisory records frequently evaluate one or more conditions as criteria. Reports within the EHR—both patient-specific detail reports as well as lists of patients—often include conditions as report parameters, i.e., for column display or for patient inclusion in a list. Use of standard SNOMED CT groupers was encouraged during design reviews of CDS advisories and reports.

### Employing SNOMED CT Groupers in Clinical Quality Measure Calculations

Beyond their use in the EHR, the same SNOMED CT groupers were used to analyze extracted EHR data within the enterprise data warehouse (EDW) for:

- Population definition,
- Comorbid condition definition, and
- electronic CQM (eCQM) calculations (ones developed locally to support quality improvement initiatives).

Formulae to calculate the denominator, numerator, and exclusions for a given eCQM often involve checking whether each patient has one or more conditions. We defined these conditions using the same standard SNOMED CT groupers employed within the EHR, employing a table-driven approach in the calculation engine that referred to standard groupers.<sup>18</sup>

### Potential for Sharing SNOMED CT Grouper Definitions across Organizations and EHRs

To demonstrate the potential for sharing SNOMED CT grouper definitions more broadly, we set out to construct exactly equivalent SNOMED CT groupers (content subsets) in our HIE’s associated clinical terminology management system (Symedical) as in the Epic EHR. These SNOMED CT content subsets are EHR-agnostic, applicable to CCDs received from any EHR within the clinically affiliated network of practices participating in our HIE.

### Evaluation and Measurement Methods

We assessed three aspects of using SNOMED CT concept hierarchy-based groupers: (1) simplicity of grouper



ID	Name	Mapped ICD-9-CM Codes	Mapped ICD-10-CM Codes
63 242498	Acute ischemic right ACA stroke	434.91	I63.521
64 426477	Acute ischemic right anterior cerebral artery (ACA) stroke	434.91	I63.521
65 242501	Acute ischemic right ICA stroke	434.91	I63.231
66 426478	Acute ischemic right internal carotid artery (ICA) stroke	434.91	I63.231
67 242505	Acute ischemic right MCA stroke	434.91	I63.511
68 426479	Acute ischemic right middle cerebral artery (MCA) stroke	434.91	I63.511
69 242507	Acute ischemic right PCA stroke	434.91	I63.531
70 426480	Acute ischemic right posterior cerebral artery (PCA) stroke	434.91	I63.531
71 242834	Acute ischemic stroke	434.91	I63.50
72 243897	Acute ischemic VBA brainstem stroke	434.91	I63.219, I63.22
73 1031997	Acute ischemic VBA brainstem stroke, left	434.91	I63.212
74 1040073	Acute ischemic VBA brainstem stroke, right	434.91	I63.211
75 1033199	Acute ischemic VBA brainstem stroke, unspecified laterality	434.91	I63.219
76 243899	Acute ischemic VBA thalamic stroke	434.91	I63.219, I63.22
77 1033680	Acute ischemic VBA thalamic stroke, left	434.91	I63.212
78 1033612	Acute ischemic VBA thalamic stroke, right	434.91	I63.211
79 1034678	Acute ischemic VBA thalamic stroke, unspecified laterality	434.91	I63.219

Loaded 1089 of 1089 records  
© 2018 Epic Systems Corporation. Used with permission.

**Fig. 5** Sample of 1,089 clinical terms (from Intelligent Medical Objects [IMO]) within a Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) hierarchy-based condition definition. Alternate names for the same diagnosis are outlined in red. Two different clinical concepts with the same International Classification of Diseases (ICD)-9 and -10 codes are outlined in blue.

construction, (2) shared use within the EHR for clinical, population health, and analytic purposes, and (3) whether SNOMED CT hierarchy-based groupers could be implemented across organizations on disparate EHRs.

**Evaluation of Simplicity of Grouper Construction**

To evaluate the simplicity of grouper construction for the set of vetted “standard” groupers, we calculated the median, minimum, 25th percentile, 75th percentile, maximum value, and mean of several grouper characteristics. We planned a priori to use the median as the primary measure of central tendency, due to expected skew.

From each SNOMED CT grouper definition, we assessed:

- Number of defining SNOMED CT concepts needed in the Boolean logic expression for the grouper.
- Number of total SNOMED CT concepts contained within the grouper, including all descendants of the defining concepts.

From the resulting (compiled) diagnosis groupers within the EHR, we assessed each of the following:

- Number of distinct ICD-10 codes included (mapped to the included IMO clinical terms).

Term	Code	Retired
Anterior circulation stroke of uncertain pathology	230714009	No
Benedikt's syndrome	41713005	No
Brainstem stroke syndrome	195212005	No
Cardioembolic stroke	413758000	No
Cerebellar stroke	16371781000119100	No
Cerebellar stroke syndrome	195213000	No
Cerebral ischemic stroke due to extracranial large artery atherosclerosis	724426006	No
Cerebral ischemic stroke due to intracranial large artery atherosclerosis	724425005	No
Cerebral ischemic stroke due to occlusion of extracranial large artery	724993002	No
Cerebral ischemic stroke due to small artery occlusion	724424009	No
Cerebral ischemic stroke due to stenosis of extracranial large artery	724994008	No
Cerebrovascular accident	230690007	No
Cerebrovascular accident due to left carotid artery stenosis	16026991000119107	No
Cerebrovascular accident due to occlusion of left anterior cerebral artery	16000471000119107	No
Cerebrovascular accident due to occlusion of left posterior communicating artery	16024191000119100	No

90 items

**Fig. 6** Individual Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) concepts within a SNOMED hierarchy-based condition definition. Although the Boolean logic definition only uses 4 concepts (see → Fig. 4), a total of 90 SNOMED CT concepts are included as relevant descendants. Additions or deletions to descendants within the Boolean logic-defined hierarchy are automatically incorporated with future updates. (Screen capture from Symedical® clinical terminology management software, (c) 2018 Clinical Architecture. Confidential.)

- Number of distinct ICD-9 codes included.
- Number of total distinct ICD codes (9 and 10) included.
- Number of IMO clinical terms included.

We also counted the number of physician subject matter experts engaged in diagnosis grouper content discussions, in addition to physician informaticist review.

#### Evaluation of EHR and Analytic Shared Use

To evaluate shared use of standard groupers within the EHR, we counted the following EHR record types incorporating use of a SNOMED CT standard diagnosis grouper:

- Registry inclusion criteria and metric definitions.
- CDS records, such as best practice advisories and health maintenance reminders.
- Rules for evaluating EHR data that drive other EHR behavior (such as dynamic appearance of a documentation tool or order set).
- Real-time reports of various types within the EHR.

To evaluate shared use of standard groupers for analytics, we counted the number of:

- Conditions available to clinicians for self-service analytics within the EHR, and
- eCQM numerator, denominator, and/or exclusion calculations in the EDW which employed one or more standard diagnosis groupers.

#### Qualitative Evaluation of Cross-Organization and Cross-EHR Sharing Potential

To assess the feasibility of employing shared condition definitions for clinical data received from disparate EHRs via CCDs, we constructed SNOMED CT hierarchy-based groupers in our HIE's associated clinical terminology management software (Symedical). We evaluated the feasibility of constructing EHR-agnostic groupers in Symedical to

exactly match the SNOMED CT concepts and Boolean logic in the Epic-based groupers.

## Results

### Simplicity of Grouper Construction

In our set of 125 standard groupers, the median number of SNOMED CT concept hierarchies needed for grouper definition was only 2 (range of 1–30, 25th percentile = 1 and 75th percentile = 5). Thirty-five of 125 groupers (28%) were defined with a single SNOMED CT concept hierarchy; the remaining majority (90 groupers, 72%) employed Boolean combinations of concept hierarchies. Once defined, SNOMED CT hierarchy-based diagnosis groupers generally took 5 to 15 minutes each to create in the EHR development environment (bulk creation via import file is also possible). The number of subject matter expert physicians engaged in discussion on grouper contents (in addition to review of each grouper by one or more physician informaticists) was 43.

Among them, the 125 groupers included a total of 525 references to SNOMED CT concept hierarchies: 413 of 525 concepts (79%) were within the “Clinical finding” top-level hierarchy, 80 (15%) within “Situation with explicit context,” and 20 (4%) within “Procedure,” accounting for 98% of all concept references (→Fig. 3). The remaining 12 concepts (2%) were distributed among the “Social context,” “Body structure,” and “Observable entity” top-level hierarchies. Among the 413 Clinical finding concepts, 351 (85%) had a semantic tag of “disorder” and 62 (15%) of “finding.”

Following grouper implementation within our Epic EHR, we assessed the resulting number of distinct SNOMED CT concepts, ICD codes, and clinical terms represented in the compiled grouper contents (→Table 4). Three of these diagnosis grouper definitions are shown in →Table 5 as examples, and all 125 are available in the →Supplementary Material (available in the online version).

**Table 4** Distribution of number of SNOMED CT codes needed to define this set of 125 groupers, and numbers of individual SNOMED CT concepts, clinical terms, and ICD codes included per grouper

	Minimum	25th percentile	50th percentile (median)	75th percentile	Maximum	Mean
Grouper definition:						
SNOMED CT concepts used to define (no. of hierarchies used in Boolean logic)	1	1	2	5	30	4.2
Grouper contents:						
SNOMED CT concepts included	1	11	32	97	4,658	157
Clinical terms included	1	63	155	976	116,043	2,217
ICD-10 codes associated with clinical terms	1	5	13	53	10,528	164
ICD-9 codes associated	1	3	9	27	851	41
Total ICD codes associated	2	8	22	85	11,234	204

Abbreviations: ICD, International Classification of Diseases; SNOMED CT, Systematized Nomenclature of Medicine–Clinical Terms.

**Table 5** Examples of condition definitions with SNOMED CT concept hierarchies and Boolean logic

Condition	Boolean logic	SNOMED CT concepts (with Symedical version of Boolean logic)	SNOMED CT concepts to define	SNOMED CT concepts included	EHR clinical terms included
Arterial thromboembolism	(1 OR 2 OR 3) AND NOT (4 OR 5)	(Arterial thrombosis (disorder) [65198009] including descendants OR Arterial embolism (disorder) [54687002] including descendants OR History of artery embolism (situation) [10824251000119108] including descendants) AND NOT Pulmonary embolism (disorder) [59282003] including descendants AND NOT H/O: pulmonary embolus (situation) [161512007] including descendants	5	178	976
Chronic lymphocytic leukemia (CLL)	1 OR 2	Chronic lymphoid leukemia, disease (disorder) [92814006] including descendants OR History of chronic lymphocytic leukemia (situation) [63581000119104] including descendants	2	36	138
Osteoporosis	1	Osteoporosis (disorder) [64859006], including descendants	1	43	2,287

Abbreviations: EHR, electronic health record; SNOMED CT, Systematized Nomenclature of Medicine–Clinical Terms.

To match the contents of the succinct SNOMED CT concept hierarchy grouper definitions, other list-based approaches to create diagnosis groupers would involve a markedly larger median quantity of items. Compared with the median 2 SNOMED CT hierarchies needed to define a diagnosis grouper, the resulting groupers included a median of 32 SNOMED CT individual concepts, 155 clinical terms selectable in the EHR by clinicians, and 22 different ICD codes (ICD-9 and -10) associated with those clinical terms.

### EHR and Analytic Shared Use of Standard SNOMED CT Groupers

To date, the set of 125 standard groupers have seen shared use by 57 patient registries (which include 362 separate references to standard groupers for registry-defining or comorbid conditions), 132 CDS items (alerts, health maintenance reminders), 190 EHR rules for non-CDS purposes, and 124 report definitions (→ [Table 6](#)).

Our EHR offers ad hoc or “self-service” analytics to clinicians, for instance, to find numbers of patients with certain conditions, optionally sliced further by other criteria (medications, procedural history, etc.). By making standard SNOMED CT diagnosis groupers visible in the self-service analytic tool, 125 vetted condition definitions have been made available to physicians for self-service analytics. In formal performance measurement reporting, 111 eCQM calculations (of a numerator, denominator, or exclusion) performed in UT Southwestern’s EDW use one or more of the standard SNOMED CT groupers.

### Cross-Organization and Cross-EHR Application

To demonstrate feasibility of employing a shared SNOMED CT grouper definition for evaluation of clinical data from disparate EHRs, we replicated construction of our standard Epic SNOMED CT hierarchy-based groupers in an HIE-associated clinical terminology management system (Symedical). All

125 were readily constructed in Symedical to match precisely the SNOMED CT concepts and Boolean logic of the Epic groupers, and thus applicable to SNOMED CT-encoded concepts sent by any certified EHR contributing to the HIE.

## Discussion

### Main Findings

In moving toward more personalized medicine and to value-based reimbursement, defining patient conditions becomes crucial—so that optimal care for each condition can be better

**Table 6** Reuse of standard SNOMED CT diagnosis groupers for EHR tools and for analytics

Category	Item type	Total no. of times a standard grouper used
EHR	Specialty patient registries ( $n = 57$ )	362
EHR	Clinical decision support records	132
EHR	Rules (used by the EHR’s rules engine; other than CDS)	190
EHR	Report definitions (within EHR)	124
Analytics	Clinician self-service analytics diagnosis “slicers”	125
Analytics	eCQM calculations of numerator, denominator, or exclusion	111

Abbreviations: CDS, clinical decision support; eCQM, electronic clinical quality measure; EHR, electronic health record; SNOMED CT, Systematized Nomenclature of Medicine–Clinical Terms.

specified, delivered, and measured. In the EHR era, clinical events are now being captured in richer detail than available previously with billing data alone. EHR events associated with diagnoses now commonly employ clinical terms, which are mapped to SNOMED CT for diagnosis interoperability among EHRs. Population health efforts require a facile way to define diagnostic conditions similarly across EHRs, preserving high clinical fidelity and leveraging this expanded EHR content.

In our study:

1. Starting from clinical intent for a given target population, using SNOMED CT concept hierarchies and Boolean logic proved to be a simple and concise way to define clinical conditions as diagnosis groupers, compared with listing individually all relevant SNOMED CT concepts, ICD codes, or clinical terminology terms. SNOMED CT concept hierarchy-based definitions of even highly specific conditions proved practical to construct.
2. Shared use of standard definitions of conditions via SNOMED CT has been high. Once constructed, SNOMED CT condition definitions found extensive shared use for EHR registries, rules, CDS, eCQM performance measure calculations, and self-service analytics. Benefits of standard grouper reuse include:
  - Avoiding rework costs of duplicative grouper construction.
  - Avoiding inconsistent grouper definitions, preventing later avoidable reconciliation work and “archeology” to track down discrepancies between definitions within the EHR and EDW.
  - Streamlining future rapid, iterative development of new CDS tools and reports within the EHR, and new eQMs in the EDW.
3. Sharing SNOMED CT concept-based groupers also proved simple to accomplish and can lead to straightforward definition of identical subpopulations across organizations and EHRs. SNOMED CT-based condition definitions proved feasible to construct in an EHR-agnostic clinical terminology tool, enabling application to CCD data from diverse organizations and EHRs within a clinically integrated network. Using SNOMED CT to link disparate sources together optimally leverages the vocabulary standardization requirements of the Health Information Technology for Economic and Clinical Health Act<sup>24</sup> for population health purposes.

### **Clinical Guidelines and eQMs Should Preferentially Define Conditions Using SNOMED CT Hierarchy Groupers and Boolean Logic**

Clinical guidelines often change physician practice only slowly and incompletely.<sup>28–30</sup> Data collection for eQMs can involve burdensome box-clicking by physicians—sometimes to idiosyncratically double-document an exclusionary condition already in the EHR. If guideline and eCQM authors were enabled to readily capture with SNOMED CT hierarchies their clinical intent for condition types and subtypes being

focused on, then those identical definitions could be made readily available to physicians practicing with certified EHRs. Duplicative efforts across the country to recreate guideline clinical intent with a locally developed diagnosis grouper could be eliminated. Creating diagnosis groupers to help drive associated CDS tool(s) to promote following a clinical guideline would become immediately more practical to implement, whether within the EHR or in a shareable CDS form invoked as an online service.<sup>31,32</sup> Diagnoses recorded during normal clinical care would be leveraged, avoiding redocumentation to meet a CQM. In sum, by employing SNOMED CT hierarchies to streamline practical implementation of guideline promotion within the EHR, updated guidelines could more quickly translate to a positive effect on patient care.

### **Using Standard Condition Definitions Based on SNOMED CT for Improving the Clinician EHR Experience**

Although the percentage of U.S. hospitals and physician practices on an electronic record increased dramatically with the federal EHR Incentive Program,<sup>33–36</sup> physician dissatisfaction with EHRs remains high.<sup>37,38</sup> Alert fatigue, “documentation fatigue” (click counts), and difficulty finding relevant information in the chart detract from potential EHR benefits for many physicians.<sup>39,40</sup>

While enhancing physician experience with the EHR is a far larger topic, a library of consistent, refined SNOMED CT-based condition definitions within the EHR can potentially help, by spurring:

- Smarter CDS: Condition-targeted CDS—based on real-time data within the EHR—can appear more selectively and appropriately to clinicians.
- More focused data capture: Rules can present condition-specific documentation templates only when relevant to a particular patient.
- Better signal-to-noise in information displays: Problem-oriented views can automatically collate and present the most relevant clinical data for the patient’s conditions,<sup>41</sup> potentially reducing clinicians’ cognitive burden.<sup>42</sup>

### **Sharing Standard Condition Definitions for Clinical–Translational Research and Advanced Analytics**

A library of SNOMED CT-defined conditions would also benefit clinical/translational research and analytics seeking to derive new knowledge from the growing expanse of EHR data, in the first limb of the “practice-to-knowledge, knowledge-to-practice” Learning Health System cycle.<sup>25,30,43</sup> Pragmatic studies making use of EHR data frequently need to consider one or more disorders as covariates or comorbid conditions: reusing existing SNOMED CT definitions avoids redundant work and enhances consistency. Predictive and prescriptive algorithms have potential to become more robust as their input conditions include more sophisticated EHR event data, and with clinical phenotypes consistently defined across any source EHR. Rare and/or specialized conditions can be more

easily focused on using the finer granularity of SNOMED CT concept hierarchies.

## Limitations

### Use of SNOMED CT with Billing Claims Data

One potential limitation to adopting SNOMED CT condition definitions is that billing claims data are still required for a complete understanding of a patient's interactions with the health care system—and these data will come tagged with ICD diagnoses alone. To handle this efficiently requires a reliable ICD-to-SNOMED CT map, including both ICD-9 and ICD-10 for coverage of historical data. Such maps exist, with varying coverage of ICD codes.<sup>24,44</sup> In one study of administrative claims data for 1.5 million persons from 2003 to 2007, over 99% of the ICD-9-CM diagnosis codes used could be mapped to SNOMED CT using the Observational Medical Outcomes Partnership (OMOP) common data model.<sup>45</sup> Similarly, over 99% of primary ICD-10-CM diagnosis codes on the 16.2 million claims submitted by UT Southwestern's multi-specialty physician practice during the calendar year 2016 mapped successfully to a primary SNOMED CT concept using the OMOP common data model, enabling effective use of SNOMED CT concept groupers with real-world billing data. In the future, valuable harmonization work underway will bring the next version of ICD (ICD-11) and SNOMED CT in much closer alignment.<sup>46</sup>

### Maintenance of SNOMED CT Hierarchy Groupers

SNOMED CT is updated twice yearly. While SNOMED CT hierarchy-based groupers (condition definitions) offer the major maintenance advantage of automatically including any newly added subtype descendants, they still require periodic review. SNOMED CT updates can include newly added concepts, deprecated concepts, or changes in a concept's hierarchical position (through modifications of its subtype-supertype relationships). Additions of new concepts as subtypes of an existing grouper concept are handled gracefully, as are deprecated concepts. Addition of a new concept as a sibling of a currently included concept hierarchy requires clinical review to decide whether the new entry should also be included. An algorithm to autodetect groupers with newly added sibling concepts after a SNOMED CT update would facilitate this targeted review. Newly added descendants could also be automatically detected and reviewed if desired, even though likely to remain included.

Changes in hierarchy positions conceptually could cause problems. In practice, many such migrations address a previous quality issue in the SNOMED CT tree, by improving consistency of subtype-supertype relationships. Hierarchy-based groupers generally are stable across such migrations. As an example from our list of 125 conditions, we had to reference 5 SNOMED CT concept hierarchies to define “Tinnitus,” because 4 of the many variants of tinnitus in this version of SNOMED CT lack “Tinnitus” as a supertype—instead being linked to the broader supertype “Disorder of ear.” One can envision as part of ongoing SNOMED CT quality improvement efforts that these 4 tinnitus variants will

ultimately have “Tinnitus (finding)” added as a supertype “parent”—like the several other variants of tinnitus already do. Should such a migration happen, our Tinnitus grouper will not “break.” The grouper's definition could then be simplified to consist of just “Tinnitus (finding), including descendants”—but the grouper would work identically, before or after such simplification.

Since we do not yet automatically flag groupers potentially affected by SNOMED CT updates, instead we employ periodic reviews. Any enumerated list-based ICD, SNOMED CT, or IMO term groupers (extensional value sets) are put on an annual review cycle as they are more likely to become stale with updates, while hierarchy-based diagnosis groupers (intensional value sets) are put on a 3-year review cycle. In practice, we have not encountered undesirable grouper behavior stemming from SNOMED CT semiannual updates when using this review frequency.

### Use of SNOMED CT's Subtype Relationship Only

Our use of SNOMED CT concept hierarchies (combined with Boolean logic) inherently employs the subtype relationship between concepts.<sup>13</sup> However, other connections between concepts—known as attribute relationships—are possible. Attribute relationships enable additional ways to refine a SNOMED CT content subset by adding further constraints. Sixteen attribute relationships can be used to further elaborate Clinical Finding concepts, such as “Associated morphology,” “Finding site,” “Causative agent,” and “After” (for temporal relationships).<sup>15</sup> Attribute relationships can be included in the “precoordinated” definition of a more-specific individual SNOMED CT concept: For example, “Fracture of neck of femur (disorder) [SCTID: 5913000] has two precoordinated relationships: (1) an “Associated morphology” of “Fracture (morphologic abnormality) [SCTID: 72704001], and (2) a “Finding site” of “Structure of neck of femur” (body structure) [SCTID: 29627003]. Alternatively, attribute relationships can be specified at query time, using “postcoordination.” A standard grammar exists for expressing such relationships: SNOMED CT Expression Constraint Language.<sup>47</sup> SNOMED CT Expression Constraint Language also can specify shareable combinations of concept hierarchies exactly equivalent to the Boolean logic-based approach taken in this study.

As a practical matter, when constructing diagnosis groupers within the EHR, the concept hierarchies frequently contain precoordinated concepts (such as “Fracture of neck of femur”) which make use of attribute relationships. But currently, our EHR groupers do not directly specify postcoordinated SNOMED CT attribute relationships. The Symedical content subset designer does enable such postcoordinated queries. Leveraging attribute relationships could further refine condition and patient subpopulation definitions.

### Defining Conditions from EHR Data Using Diagnoses Only

One might also question why limit the domain of condition-defining inputs to SNOMED CT diagnosis concepts only? For example, why not also use laboratory test results, current medications, and even unstructured data to define a clinical phenotype such as “Diabetes Mellitus”<sup>48</sup>? And what will

happen as more and more “conditions” are defined based on genetic data<sup>49</sup>? Because of the multiple corollary benefits to safe clinical care and to analytics from each patient having a single accurate master list of their active health conditions (their Problem List), we favor separating out those concerns. That is, we encourage using other nondiagnosis domain data (laboratory test results, medications, etc.) to enhance the quality of the Problem List through active additions and refinements of a patient’s conditions.<sup>50–52</sup> But we strongly advocate the patient’s Active Problem List serve as the single, unified focal point for clinical situational awareness, clinical communication, and analytic understanding of “all the patient’s problems.”<sup>53</sup>

## Conclusion

SNOMED CT hierarchy-based diagnosis groupers are simple to develop and maintain, understandable to clinicians, useful in both the EHR and EDW, and readily shareable. Developing curated SNOMED CT hierarchy-based condition definitions (“intensional value sets”) and disseminating them publicly (e.g., via the VSAC) could help accelerate cross-organizational population health efforts, “smarter” EHR feature configuration, and clinical–translational research.<sup>54</sup> SNOMED CT hierarchies can define clinical conditions more precisely than achievable with ICD, and closely match how clinicians think about disease subtypes. And by directly employing the terminology standard now native to EHRs, they prove highly practical to implement across multiple health care delivery organizations. Guideline-writing groups and eCQM authors who define conditions using SNOMED CT hierarchies thus could more quickly see uptake of their work efforts into EHR-based CDS and patient registries, providing clinicians and patients practical tools for improving care delivery and patient outcomes.

## Clinical Relevance Statement

With increasing focus on population health, identifying patients who share a clinical condition helps promote best practice clinical care within an electronic health record (EHR), and across clinically integrated networks. EHRs now exchange diagnoses using a standard terminology, SNOMED CT. Defining clinical conditions with SNOMED CT concept hierarchies is far simpler than alternatives, and such definitions can be readily shared for multiple clinical and analytic purposes.

## Multiple Choice Questions

1. Compared with lists of ICD codes, EHR diagnosis groupers constructed with SNOMED CT concept hierarchies:
  - a. Require more frequent updating when new codes or concepts are added or deprecated.
  - b. More closely match clinical thinking about disease subtypes to include or exclude.
  - c. Use the coding system mandated for professional billing in the United States.
  - d. Are more complex to construct and maintain.

**Correct Answer:** The correct answer is option b. The hierarchical subtype (“is a”) relationships between SNOMED CT concepts express type–subtype relationships that closely match how clinicians think about clinical disorders and their subtypes. This streamlines clinical vetting of groupers to achieve faithful representation of the intended clinical condition.

Groupers designed with SNOMED CT hierarchies are more resilient to additions or deprecations of individual codes/concepts than list-based approaches such as lists of ICD codes: addition of new “descendants” within an existing SNOMED CT hierarchy does not require a change to the grouper definition in the EHR. SNOMED CT is the terminology mandated for health information exchange of diagnoses between EHRs; professional billing employs ICD-10-CM codes in the United States, not SNOMED CT concepts. (“Maps” translating ICD-10-CM codes to SNOMED CT concepts enable groupers defined with SNOMED CT hierarchies to handle both clinical EHR data and billing claims data.) SNOMED CT hierarchy groupers require far fewer concepts/codes to define (median of 2 concepts in this report) than groupers using lists of ICD codes.

2. Diagnosis “groupers,” or value sets:
  - a. Can only be used in electronic health records, not in analytic applications such as enterprise data warehouses.
  - b. Are best constructed individually each time a condition definition is needed (one use = one grouper).
  - c. Are best constructed for each distinct real-world condition, and shared for multiple uses (one real-world condition = one grouper).
  - d. Are available exclusively as lists of codes or concepts.

**Correct Answer:** The correct answer is option c. Constructing one grouper per distinct real-world condition promotes higher quality through more clinical vetting per grouper, higher consistency, and lower total cost in time/effort via shared reuse for multiple clinical and analytic purposes. Accordingly, this is preferred over constructing multiple groupers for the same clinical condition each time a new use arises (e.g., CDS tool vs. eCQM).

Diagnosis value sets can be shared for use in analytic applications, such as data warehouses as well within EHRs. While diagnosis value sets can be list-based (“extensional,” constructed by listing out individual clinical terms, individual ICD codes, or individual SNOMED CT concepts), they also can be hierarchy-based (“intensional”)—constructed by referring to combinations of SNOMED CT hierarchies that match desired clinical inclusion/exclusion criteria.

### Protection of Human and Animal Subjects

Creation of specialty registries at UT Southwestern was performed for quality improvement purposes and deemed exempt from institutional review board (IRB) review by UT Southwestern’s IRB. Analysis of reuse and complexity of diagnosis groupers in the EHR did not involve human or animal subjects, and did not require IRB review.

### Funding

Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award Number UL1TR001105. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

### Conflict of Interest

None.

### References

- 1 Motheral BR, Fairman KA. The use of claims databases for outcomes research: rationale, challenges, and strategies. *Clin Ther* 1997;19(02):346–366
- 2 World Health Organization. WHO | ICD-11 Revision – International Classification of Diseases; 2017 [6/4/2017]. Available at: <http://www.who.int/classifications/icd/revision/en/>. Accessed July 18, 2018
- 3 Benson T, Grieve G. Principles of Health Interoperability: SNOMED CT, HL7, and FHIR. 3rd ed. London: Springer; 2016. xxix, 451 p
- 4 Moriyama IM, Loy RM, Robb-Smith AHT, Rosenberg HM, Hoyert DL, National Center for Health Statistics (U.S.). History of the Statistical Classification of Diseases and Causes of Death. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2011. Available at: [http://www.cdc.gov/nchs/data/misc/classification\\_diseases2011.pdf](http://www.cdc.gov/nchs/data/misc/classification_diseases2011.pdf). Accessed July 18, 2018
- 5 National Library of Medicine (NLM) NLoM. Value Set Authority Center (VSAC): National Library of Medicine (NLM); [updated August 02, 2017]. Available at: <https://vsac.nlm.nih.gov/>. Accessed July 18, 2018
- 6 Bowman S. Coordinating SNOMED-CT and ICD-10. *J AHIMA* 2005; 76(07):60–61
- 7 HL7. HL7 Meaningful Use NPRM (Notice of Proposed Rule Making) comments – Final [Letter]; 2010 [updated 3/15/2010]. Available at: [https://www.hl7.org/documentcenter/public\\_temp\\_85634A36-1C23-BA17-0C77886C56EADBE4/newsroom/HL7%20Meaningful%20Use%20NPRM%20comments%20FINAL.pdf](https://www.hl7.org/documentcenter/public_temp_85634A36-1C23-BA17-0C77886C56EADBE4/newsroom/HL7%20Meaningful%20Use%20NPRM%20comments%20FINAL.pdf). Accessed July 18, 2018
- 8 Winnenburg R, Bodenreider O. Metrics for assessing the quality of value sets in clinical quality measures. *AMIA Annu Symp Proc* 2013;2013:1497–1505
- 9 Morley KI, Wallace J, Denaxas SC, et al. Defining disease phenotypes using national linked electronic health records: a case study of atrial fibrillation. *PLoS One* 2014;9(11):e110900
- 10 Cornet R, de Keizer N. Forty years of SNOMED: a literature review. *BMC Med Inform Decis Mak* 2008;8(Suppl 1):S2
- 11 SNOMED International. SNOMED CT United States Edition. Bethesda, MD: U.S. National Library of Medicine; 2017. Available at: [https://www.nlm.nih.gov/healthit/snomedct/us\\_edition.html](https://www.nlm.nih.gov/healthit/snomedct/us_edition.html). Accessed July 18, 2018
- 12 Højen AR, Sundvall E, Gøeg KR. Methods and applications for visualization of SNOMED CT concept sets. *Appl Clin Inform* 2014;5(01):127–152
- 13 4. SNOMED CT Basics: SNOMED International; 2017. Available at: <https://confluence.ihtsdotools.org/display/DOCSTART/4.+SNOMED+CT+Basics>. Accessed July 18, 2018
- 14 Randorff Højen A, Rosenbeck Gøeg K. SNOMED CT implementation. Mapping guidelines facilitating reuse of data. *Methods Inf Med* 2012;51(06):529–538
- 15 6. SNOMED CT Concept Model: SNOMED International; 2017. Available at: <https://confluence.ihtsdotools.org/display/DOCSTART/6.+SNOMED+CT+Concept+Model>. Accessed July 18, 2018
- 16 CMS.gov. 2018 ICD-10-CM. Centers for Medicare and Medicaid Services; 2017 [updated 08/11/2017]. Available at: <https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs.html>. Accessed July 18, 2018
- 17 Mohd Sulaiman I, Karlsson D, Koch S. Mapping acute coronary syndrome registries to SNOMED CT: a comparative study between Malaysia and Sweden. *Methods Inf Med* 2017;56(04):330–338
- 18 Kannan V, Fish JS, Mutz JM, et al. Rapid development of specialty population registries and quality measures from electronic health record data: an agile framework. *Methods Inf Med* 2017;56(99):e74–e83
- 19 D'Amore JD, Sittig DF, Wright A, Iyengar MS, Ness RB. The promise of the CCD: challenges and opportunity for quality improvement and population health. *AMIA Annu Symp Proc* 2011;2011:285–294
- 20 Richesson RL, Green BB, Laws R, et al. Pragmatic (trial) informatics: a perspective from the NIH Health Care Systems Research Collaboratory. *J Am Med Inform Assoc* 2017;24(05):996–1001
- 21 Gottesman O, Kuivaniemi H, Tromp G, et al; eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med* 2013;15(10):761–771
- 22 Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc* 2014;21(04):578–582
- 23 Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010;17(02):124–130
- 24 FitzHenry F, Resnic FS, Robbins SL, et al. Creating a common data model for comparative effectiveness with the Observational Medical Outcomes Partnership. *Appl Clin Inform* 2015;6(03):536–547
- 25 Richesson RL, Smerek MM, Blake Cameron C. A framework to support the sharing and reuse of computable phenotype definitions across health care delivery and clinical research applications. *EGEMS (Wash DC)* 2016;4(03):1232
- 26 Leavy MB. Registries for Evaluating Patient Outcomes: A User's Guide. In: Gliklich RE, Dreyer NA, Leavy MB, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*. Rockville, MD: Agency for Healthcare Research and Quality (U.S.); 2014
- 27 SNOMED International SNOMED CT Browser. SNOMED International; 2017. Available at: <http://browser.ihtsdotools.org/>. Accessed July 18, 2018
- 28 Lenfant C. Shattuck lecture—clinical research to clinical practice—lost in translation? *N Engl J Med* 2003;349(09):868–874
- 29 Avorn J. Transforming trial results into practice change: the final translational hurdle: comment on “Impact of the ALLHAT/JNC7 Dissemination Project on thiazide-type diuretic use”. *Arch Intern Med* 2010;170(10):858–860
- 30 Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med* 2010;2(57):57cm29
- 31 Boxwala AA, Rocha BH, Maviglia S, et al. A multi-layered framework for disseminating knowledge for computer-based decision support. *J Am Med Inform Assoc* 2011;18(Suppl 1):i132–i139
- 32 Hooks CDS. HL7 and Boston Children's Hospital; 2018. Available at: <http://cbs-hooks.org/>. Accessed July 18, 2018
- 33 Blumenthal D, Tavenner M. The “meaningful use” regulation for electronic health records. *N Engl J Med* 2010;363(06):501–504
- 34 Menemeyer ST, Menachemi N, Rahrurkar S, Ford EW. Impact of the HITECH Act on physicians' adoption of electronic health records. *J Am Med Inform Assoc* 2016;23(02):375–379
- 35 Kruse CS, Kothman K, Anerobi K, Abanaka L. Adoption factors of the electronic health record: a systematic review. *JMIR Med Inform* 2016;4(02):e19. Doi: 10.2196/medinform.5525
- 36 Hsiao CJ, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2013. *NCHS Data Brief* 2014;(143):1–8
- 37 Meigs SL, Solomon M. Electronic health record use a bitter pill for many physicians. *Perspect Health Inf Manag* 2016;13:1d
- 38 Abramson EL, Patel V, Pfoh ER, Kaushal R. How physician perspectives on e-prescribing evolve over time. A case study

- following the transition between EHRs in an outpatient clinic. *Appl Clin Inform* 2016;7(04):994–1006
- 39 Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R; with the HITEC Investigators. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC Med Inform Decis Mak* 2017;17(01):36
  - 40 McCoy AB, Waitman LR, Lewis JB, et al. A framework for evaluating the appropriateness of clinical decision support alerts and responses. *J Am Med Inform Assoc* 2012;19(03):346–352
  - 41 Buchanan J. Accelerating the benefits of the problem oriented medical record. *Appl Clin Inform* 2017;8(01):180–190
  - 42 Harry E, Pierce RG, Kneeland P, Huang G, Stein J, Weller J. Cognitive load and its implications for health care [Web page]. *NEJM Catalyst: NEJM*; 2018 [updated March 14, 2018]. Available at: <https://catalyst.nejm.org/cognitive-load-theory-implications-health-care/>. Accessed July 18, 2018
  - 43 Maddox TM, Albert NM, Borden WB, et al; American Heart Association Council on Quality of Care and Outcomes Research; Council on Cardiovascular Disease in the Young; Council on Clinical Cardiology; Council on Functional Genomics and Translational Biology; and Stroke Council. The learning healthcare system and cardiovascular care: a scientific statement from the American Heart Association. *Circulation* 2017;135(14):e826–e857
  - 44 Adamusiak T, Shimoyama N, Shimoyama M. Next generation phenotyping using the unified medical language system. *JMIR Med Inform* 2014;2(01):e5
  - 45 Reich C, Ryan PB, Stang PE, Rocca M. Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases. *J Biomed Inform* 2012;45(04):689–696
  - 46 Rodrigues J-M, Robinson D, Della Mead V, et al, eds. *Semantic Alignment between ICD-11 and SNOMED CT*. MEDINFO 2015: eHealth-Enabled Health; 2015
  7. SNOMED CT Expressions SNOMED International; 2017. Available at: <https://confluence.ihtsdotools.org/display/DOCSTART/7.+SNOMED+CT+Expressions>. Accessed July 18, 2018
  - 48 Halpern Y, Horng S, Choi Y, Sontag D. Electronic medical record phenotyping using the anchor and learn framework. *J Am Med Inform Assoc* 2016;23(04):731–740
  - 49 Hoffman JM, Dunnenberger HM, Kevin Hicks J, et al. Developing knowledge resources to support precision medicine: principles from the Clinical Pharmacogenetics Implementation Consortium (CPIC). *J Am Med Inform Assoc* 2016;23(04):796–801
  - 50 Wright A, Pang J, Feblowitz JC, et al. Improving completeness of electronic problem lists through clinical decision support: a randomized, controlled trial. *J Am Med Inform Assoc* 2012;19(04):555–561
  - 51 McEvoy D, Gandhi TK, Turchin A, Wright A. Enhancing problem list documentation in electronic health records using two methods: the example of prior splenectomy. *BMJ Qual Saf* 2018;27(01):40–47
  - 52 Falck S, Adimadhyam S, Meltzer DO, Walton SM, Galanter WL. A trial of indication based prescribing of antihypertensive medications during computerized order entry to improve problem list documentation. *Int J Med Inform* 2013;82(10):996–1003
  - 53 Weed LL. Medical records that guide and teach. *N Engl J Med* 1968;278(11):593–600
  - 54 Khatipov E, Madden M, Chiang P, et al. Creating, maintaining and publishing value sets in the VSAC. *AMIA Annu Symp Proc* 2014;20;14:1459
  - 55 Fung KW. How SNOMED CT can help in the ICD-10-CM transition [PowerPoint]. NLM; 2012. Available at: [https://www.nlm.nih.gov/research/umls/mapping\\_projects/SNOMED\\_ICD10\\_KWFung.pptx](https://www.nlm.nih.gov/research/umls/mapping_projects/SNOMED_ICD10_KWFung.pptx). Accessed July 18, 2018
  - 56 5. SNOMED CT Logical Model: SNOMED International; 2017. Available at: <https://confluence.ihtsdotools.org/display/DOCSTART/5.+SNOMED+CT+Logical+Model>. Accessed July 18, 2018