



# From Concepts and Codes to Healthcare Quality Measurement: Understanding Variations in Value Set Vocabularies for a Statin Therapy Clinical Quality Measure

Raja A. Cholan, BS; Nicole G. Weiskopf, PhD; Doug Rhoton; Bhavaya Sachdeva, MPH; Nicholas V. Colin, MA; Shelby J. Martin, MS, RD; David A. Dorr, MD, MS<sup>1</sup>

## ABSTRACT

**Objective:** To understand the impact of distinct concept to value set mapping on the measurement of quality of care.

**Background:** Clinical quality measures (CQMs) intend to measure the quality of healthcare services provided, and to help promote evidence-based therapies. Most CQMs consist of grouped codes from vocabularies - or 'value sets' - that represent the unique identifiers (i.e., object identifiers), concepts (i.e., value set names), and concept definitions (i.e., code groups) that define a measure's specifications. In the development of a statin therapy CQM, two unique value sets were created by independent measure developers for the same global concepts.

**Methods:** We first identified differences between the two value set specifications of the same CQM. We then implemented the various versions in a quality measure calculation registry to understand how the differences affected calculated prevalence of risk and measure performance.

**Results:** Global performance rates only differed by 0.8%, but there were up to 2.3 times as many patients included with key conditions, and differing performance rates of 7.5% for patients with 'myocardial infarction' and 3.5% for those with 'ischemic vascular disease'.

**Conclusion:** The decisions CQM developers make about which concepts and code groups to include or exclude in value set vocabularies can lead to inaccuracies in the measurement of quality of care. One solution is that developers could provide rationale for these decisions. Endorsements are needed to encourage system vendors, payers, informaticians, and clinicians to collaborate in the creation of more integrated terminology sets.

## Introduction

Several years after the Affordable Care Act became law, there have been a number of innovative efforts that aim to improve the health care system in the United States with the intention to transform payment and improve patient outcomes by enhancing patient-centered care and the quality of care.<sup>1</sup> Health care costs continue to rise, with Americans spending almost twice as much as individuals in other developed nations, while experiencing worse health outcomes.<sup>2,3,4</sup> Although quality-based payment programs have been around for decades, recently, the Centers for Medicare and Medicaid Services (CMS) has been charged with changing 80 percent of these payments from fee-for-service to value-based, where “value” can be thought of as benefit over cost, or, for this example, the quality of care over its cost. Additionally, CMS recently announced the Merit-based Incentive Payment System (MIPS) under the Medicare Access and CHIP Reauthorization Act (MACRA), which will start paying clinicians based on quality of care beginning in 2019.<sup>5,6</sup>

One effort that intends to improve value is the U.S. Department of Health and Human Services’ Million Hearts initiative. This national effort aims to prevent one million heart attacks and strokes, in part by focusing on clinical quality measures (CQMs) that represent key steps to reduce heart attacks and strokes; providing these simpler, lower cost (and higher value) services has been shown to reduce deadly and complicated heart attacks and strokes.<sup>7</sup> Known as the “ABCS”, the measures are: Aspirin when appropriate (CMS164v5; NQF0068), Blood pressure control (CMS165v4; NQF0018), Cholesterol management (CMS347v0; NQF pending), and Smoking cessation (CMS138v4; NQF0028).<sup>7</sup> To assist in this effort, the Agency for Healthcare Research and Quality (AHRQ) supports the EvidenceNOW grant initiative. This initiative

provides support to thousands of primary care practices across seven regional cooperatives to reduce patients’ cardiovascular risk by improving adherence to the ABCS measures to 70 percent.<sup>8</sup> A mixed methods evaluation of EvidenceNOW is being conducted by Evaluating System Change to Advance Learning and Take Evidence to Scale (ESCALATES), and global technical assistance is being provided by a separate team.<sup>9</sup> As part of EvidenceNOW, a set of experienced informatics leads help practices understand how they can leverage technology to produce and improve the ABCS measures, including helping with the CQM definitions.<sup>10</sup>

For EvidenceNOW, the ABCS measures are defined by standardized electronic CQMs, where the CQM definition, logic, and pertinent concepts are encoded into Health Quality Measure Format (HQMF) and value sets.<sup>11</sup> Value sets are grouped codes from standard vocabularies (e.g., SNOMED CT, RxNorm, LOINC, and ICD-10-CM), that are found in Electronic Health Records (EHRs) and represent the clinical concepts of interest. Although the Aspirin, Blood Pressure, and Smoking Cessation measures have been used extensively, the measures previously used for Cholesterol Management were outdated and needed revisions with new criteria from the American Heart Association. This required the creation of the HQMF and value set for the Cholesterol measure. Although measures endorsed by CMS are stored on their website and the value sets held and updated by the National Library of Medicine in the Value Set Authority Center (VSAC), these value sets needed be created rapidly and no official measure existed.<sup>12,13</sup>

Because the Cholesterol measure had recently changed, the specific definition and value sets had not yet been created. Therefore, ESCALATES contracted a standard measure developer (Research or “Res.”) to produce these elements. Since this



measure was also planned to be used in value-based care programs, another standard measure developer created a separate value set for the Cholesterol measure using CMS' Measure Authoring Tool (CMS). This parallel process allowed for a natural comparison of the measure development process. As with any process using controlled terminologies applied to real health care, many issues may impact results. As Cimino warned in the "Desiderata for Controlled Terminologies", "concept orientation" requires using conceptual building blocks in order for the concept to be complete and cover multiple levels of detail.<sup>14,15</sup> He also discussed the inevitable risk for redundancy in controlled vocabularies and stated that coding information in multiple ways should generally be avoided.<sup>14</sup> Likewise, Chute et. al described the need for non-ambiguous, non-overlapping concepts in order to support "aggregate outcome analyses".<sup>13,16</sup> More recently, Winnenbun and Bodenreider found gaps in the "completeness" and "correctness" of CQM value sets on the VSAC, and suggested metrics for value set developers to identify common errors.<sup>17</sup> However, measure developers are not required to follow these precepts for approval; rather, the process requires more generic steps to creation, validation, and implementation of measures as outlined in the CMS Measures Management System Blueprint.<sup>18</sup> With the advent of MACRA and MIPS, the widespread reporting of CQM data continues to grow, with the goal of monitoring population health for those receiving health care services; facilitating more complete data in key areas is crucial for new payment incentives, and for understanding public health concerns.<sup>19</sup> Dependable value sets – with proper concept orientation, consideration of redundancy, completeness, and correctness - are central to this aim.

The parallel measure development process from EvidenceNOW and CMS led to subsequent

parallelism in Cholesterol measure implementations; some cooperatives were using the CMS value set and others were using the Res. value set in their implementations of the Cholesterol measure. Rector described the implementation challenges of medical terminologies and stated that: "[one] reason that medical terminology is hard is the complexity of clinical pragmatics ... and the need for testing the pragmatics of terminologies implemented in software."<sup>20</sup> In this study we sought to use this parallel measure development process to:

- (1) understand the common differences that may occur as part of measure definition and value set creation;
- (2) understand how variations in value sets change a measure's meaning through implementation differences in a CQM calculation registry; and
- (3) discuss alternative processes for CQM development and implementation.

*Our hypothesis was that the decisions made in the value set creation process would lead to differences in the measure populations and performance estimates.*

## Methods

We performed our study in two stages. We first identified the difference in specifications – unique identifiers, concepts, code groups, and coding systems – between the two value sets used to define the same Cholesterol measure. We then implemented the various versions in a quality measure calculation registry to understand how the differences in the value sets affected calculated prevalence of risk and measure performance.

### Measure specifications

The intent of the Cholesterol CQM is to measure the proportion of patients at high risk for cardiovascular

events – largely heart attack and stroke – who are currently prescribed or taking statin medication. The measure has three denominators that relate to risk: patients with atherosclerotic cardiovascular disease (ASCVD); those with low-density lipoproteins (LDLs) greater than 190 mg/dL; and patients with diabetes between 40 and 75 years of age with LDLs greater than 70 mg/dL. All three groups have high risk of initial or subsequent vascular events, and evidence shows that the use of statin medications lower those events.<sup>21</sup>

### Value set analysis

The EvidenceNOW team sent the Res. and CMS value sets to us as flat files. The two value sets contained different human-readable terms to represent convergent clinical concepts, so we first created categories by grouping the convergent concepts between the two sets. See Appendix A for a complete list of all the OID names contained in the Res. Value Set and those contained in the CMS value set, as well as a list of the study categories used for comparison. For example, the Res. value set contained a concept (assigned to one unique object identifier or “OID”) named ‘palliative care’, while the CMS value set contained four concepts (assigned to four distinct unique identifiers) with various names (e.g., ‘comfort measures’) that represented the same broader clinical concept. We assigned the four concepts in the CMS value set into a single category named ‘palliative care’ and compared it to the ‘palliative care’ concept from the Res. value set; in this example, both categories were defined by the same code group comprised of 11 codes (ICD9CM, ICD-10CM, and SNOMED CT). Next, we mapped each category to the specific measure criteria as defined by the HQMF (e.g., the category ‘ischemic vascular disease’ was mapped to the criteria for ‘denominator one’ which identifies patients with ASCVD). We also counted the unique object identifiers, or “OIDs” used to specify distinct

concepts, and the codes from standard vocabularies that define each category. Any identifiers that belonged to the code system “grouping” were excluded from our analysis. These conceptual groupings used a single grouping OID to represent a collection of individual OIDs; the individual OIDs within the groupings were used in our analysis rather than the singular grouping OID. The ASCVD denominator was the most complex – it includes the grouping of acute coronary syndromes, history of myocardial infarction, stable or unstable angina, coronary arterial revascularization, stroke or transient ischemic attack, and peripheral artery disease of atherosclerotic origin – and we explored specific code variations of the aggregate denominator and its sub-groups, and related them to taxonomy structure.

### Implementation and comparison

The Integrated Care Coordination Information System (ICCIS) is a clinical quality measure calculation registry.<sup>22</sup> ICCIS contains data mapped from a variety of EHRs from over 500,000 patients into a star database format that facilitates value set queries. Implementation of CQMs in ICCIS currently requires processing of human-readable CQM definitions and writing queries in Structured Query Language (SQL) to try to find the right patient data from the right clinical sources to calculate CQM performance. The Cholesterol measure HQMF included logic for patients who met the denominator criteria (patients with ASCVD, elevated LDLs, and diabetes), numerator criteria (patients taking a statin medication), and those excluded from the CQM calculation if they met the exceptions criteria (palliative care, pregnancy, statin allergy). After we received the specifications for the Cholesterol measure, we implemented a query in ICCIS following the specific criteria for each value set, and queried against the OIDs for the two different versions of the measure; the measure logic remained constant.



Initially, the global composite measures were implemented and a quality measure performance rate was measured. Then, the measure was divided into its three unique denominators and each was implemented individually. Next, denominator one (patients with ASCVD) was divided into aggregate clinical concepts to see which concepts, taxonomies, and individual codes were most responsible for the performance rate.

### Inclusion Criteria

The measures were implemented against data from five medium to large primary care clinics (e.g., family medicine, internal medicine) serving almost 47,000 patients, in urban, suburban, and rural regions in Oregon. The practitioners in each clinic used a fully functional EHR in their ambulatory work. We chose to implement the measures using data from these five clinics because they had suitable structured data in the ICCIS database to generate reliable quality measure performance rates. The three clinics excluded from our analysis did not have an adequate average number of medications per patient stored as structured data in the ICCIS database; therefore we could not have reliably queried whether the patients were taking a statin medication.

### Results

A synopsis of the vocabularies by measure criteria and category found in the Res. and CMS value sets is shown in Table 1.

Globally, there were more unique OIDs in the CMS value set (66) than the Res. value set (23); however, there were more unique codes in the Res. value set (4,302) than CMS (4,037). The OIDs between the two versions differed in their levels of granularity. The CMS value set contained more individual OIDs which represented specific concepts (e.g., individual OIDs for Hepatitis A and Hepatitis B, respectively) while the Res. value set had a fewer total number

of OIDs and these represented broader conceptual groupings (e.g., 1 OID for Liver Disease which included Hepatitis A and B). Some of the larger differences within each category are highlighted in Table 1.

Table 2 shows the performance rates for various implementations of the measure, as well as the number of patients who met the exceptions criteria. The composite measures had similar results, with the Res. returning 4,451 patients and 71.9 percent adherence, and CMS producing 4,204 patients with 72.9 percent adherence. Despite these similar rates, there was noticeable variation for the stand-alone implementations of 'ischemic vascular disease' (IVD) and 'myocardial infarction' (MI). The Res. implementation of IVD returned 677 more patients in its denominator than that of the CMS implementation of IVD, and had a 3.5 percent lower adherence rate. The Res. implementation of MI had a 7.5 percent better performance rate than the CMS version of MI, and the Res. denominator returned 110 patients versus 258 patients in the CMS version.

These differences can be explained by the codes that are distinct to one value set and not included in the other, as shown in Table 3. Across all categories, there were 224 ICD-10-CM codes (620 unique codes total) in the Res. value set that were not in the CMS value set, and 119 ICD-10CM codes (370 codes total) in the CMS value set that were not in that of Res. For IVD in particular, the Res. value set included 128 distinct ICD-10-CM codes and one code alone (I65.29) accounted for 4.1 percent of the patients included in the Res. composite denominator, and 8.1 percent of patients included when Res. denominator 1 is a stand-alone measure. Appendix B contains an example of the overlap of codes for the 'MI' category. Of the 72 codes that were distinct to one of the 'MI' value sets and not included in the other, a subset were used frequently and are included in Table 3.

**Table 1. Counts of Cholesterol measure vocabularies by criteria and category (excluding grouping OIDs)**

CRITERIA	CONCEPT	OIDS		CODES		TAXONOMIES	
		RES	CMS	RES	CMS	RES	CMS
Initial Pt Population	Encounters	6	8	32	42	SNOMED, CPT, HCPCS	SNOMED, CPT, HCPCS
Denominator 1	CABG	1	3	160	160	ICD-9/10, SNOMED	ICD-9/10, SNOMED
Denominator 1	Carotid Intervention	1	3	757	757	ICD-9/10, SNOMED	ICD-9/10, SNOMED
Denominator 1	Ischemic Vascular Disease	1	9	561	169	ICD-9/10, SNOMED	ICD-9/10, SNOMED
Denominator 1	Myocardial Infarction	1	3	62	90	ICD-9/10, SNOMED	ICD-9/10, SNOMED
Denominator 1	PCI	1	3	80	80	ICD-9/10, SNOMED	ICD-9/10, SNOMED
Denominator 1	Stroke	2	3	197	198	ICD-9/10, SNOMED	ICD-9/10, SNOMED
Denominator 2	Hypercholesterolemia	0	3	0	14	—	ICD-9/10, SNOMED
Denominator 2	LDL Test	1	1	9	9	LOINC	LOINC
Denominator 3	Diabetes	1	3	225	224	ICD-10, ICD-9, SNOMED	ICD-10, ICD-9, SNOMED
Exceptions	Breastfeeding	1	2	12	25	SNOMED	ICD-10, SNOMED
Exceptions	ESRD	1	3	6	6	ICD-9/10, SNOMED	ICD-9/10, SNOMED
Exceptions	Liver Disease	1	9	85	162	ICD-9/10, SNOMED	ICD-9/10, SNOMED
Exceptions	Medical Reason	1	0	19	0	SNOMED	—
Exceptions	Palliative Care	1	4	11	11	ICD-9/10, SNOMED	ICD-9/10, SNOMED
Exceptions	Pregnancy	1	3	1981	1981	ICD-9/10, SNOMED	ICD-9/10, SNOMED
Exceptions	Statin Allergen	1	3	35	38	RxNorm	RxNorm
Numerator	Statin RXNORM	1	3	70	71	RxNorm	RxNorm
<b>Aggregate</b>	<b>Full Value Set</b>	<b>23</b>	<b>66</b>	<b>4302</b>	<b>4037</b>	<b>ICD-10, ICD-9, SNOMED, RxNorm, LOINC, CPT, HCPCS</b>	<b>ICD-10, ICD-9, SNOMED, RxNorm, LOINC, CPT, HCPCS</b>



Table 2. CQM Performance Rates in Composite Measures, Aggregate Concepts, & Distinct Denominators

CRITERIA	VALUE SET NAME/CATEGORY	RES. NUMERATOR/DENOMINATOR (%)	CMS NUMERATOR/DENOMINATOR (%)	RES. EXCEPTIONS	CMS EXCEPTIONS
Composite Measure	Statin Therapy for the Prevention and Treatment of Cardiovascular Disease (CMS347v0)	3202/4451 (71.9%)	3057/4204 (72.7%)	327	356
Denominator 1	Ischemic Vascular Disease (including Stable/Unstable Angina & Artery Disease)	1802/2305 (78.2%)	1330/1628 (81.7%)	93	82
Denominator 1	Myocardial Infarction	96/110 (87.3%)	206/258 (79.8%)	2	13
Denominator 1	Stroke	466/684 (68.1%)	519/753 (68.9%)	31	42
Complete Denominator 1	Patients aged 21 years and older at the beginning of the measurement period with clinical ASCVD	1896/2487 (76.2%)	1767/2305 (76.7%)	100	118
Complete Denominator 2	Patients aged 21 years and older at the beginning of the measurement period with any fasting or direct laboratory test result of LDL-C ≥ 190 mg/dL	196/445 (44.0%)	196/440 (44.5%)	3	8
Complete Denominator 3	Patients aged 40 through 75 years at the beginning of the measurement period with Type 1 or Type 2 Diabetes with the highest fasting or direct laboratory test result of LDL-C 70-189 mg/dL in the measurement year or two years prior to the beginning of the measurement period	1441/1891 (76.2%)	1375/1782 (77.2%)	255	259

Table 3, below, provides the global prevalence of any of the distinct most prevalent ICD-10 and ICD-9 codes between the two versions.

The category ‘ischemic vascular disease’ (IVD) had the largest variation in OIDs and code groups, with the Res. value set containing 1 OID and 561 individual codes, while the CMS value set contained 9 OIDs and 169 individual codes. The Res. value set contained several codes that were not included in CMS; for example, it included ten codes from the I75 ICD-10CM series for different granularities of ‘atheroembolism’. Although the CMS value set contained fewer ICD-CM codes, it also included distinct codes that were not in the Res. version. Interestingly, the Res. value set included 42 codes from the I70 series for ‘atherosclerosis’; the CMS value set, meanwhile, included only 6 codes from the

I70 series – and none of these 6 codes overlapped with the code group included in the Res value set.

In addition to the ICD-CM codes, 75 percent of the IVD codes in Res. that are not in CMS were SNOMED codes. The CMS value set defined IVD with SNOMED codes related to ‘atherosclerosis and peripheral arterial disease’, ‘stable and unstable angina’, and ‘myocardial ischemia’ while the Res. included these and many more concepts including but not limited to ‘coronary arteriosclerosis’, ‘coronary artery disease’, and ‘coronary heart disease’. Figure 1 demonstrates the selected SNOMED codes for one parent (carotid artery occlusion) and six children. Only one child was in both versions, while the parent and three other children were in the Res. version. Two children related to “asymptomatic occlusion of carotid artery” were in neither version; these

**Table 3. Most Prevalent Distinct Code Details for Denominator 1 Clinical Concepts**

VALUE SET NAME/ CATEGORY	SOURCE	ICD-10CM CODE	PREVALENCE IN COMPOSITE MEASURE	ICD-9CM CODE	PREVALENCE IN COMPOSITE MEASURE
Stroke	CMS	G45.9	4.82%	—	—
		Z86.73 I61.9	1.98% 1.15%		
	Res.	G45.8	0.31%	431	0.93%
		I65.21	0.27%	430	0.72%
Ischemic Vascular Disease	CMS	I25.5	1.51%	-	-
		I65.29 I70.0	4.14% 1.39%	414.00	24.28%
Myocardial Infarction	CMS	I25.2	1.28%	435.9	2.59%
				410.90	2.57%
				410.70	2.26%
				412	1.58%
				410.0	0.51%
				410.40	0.20%
				410.10	0.18%
Res.	—	—	—	—	



codes represent incidental findings and might be appropriately excluded. There was no rationale provided for inclusion/exclusion of codes.

Finally, one EvidenceNOW cooperative reported a poor mapping to statin drugs as stored in the variety of EHRs from which they extracted the measures. Although the value sets are very similar in count (70 versus 71 RxNorm codes with complete overlap except for 1 medication), no clarification was made to address data mapped to differing level of the hierarchies. Figure 2 shows the RxNorm hierarchy with drug class at the top and generic clinical products and branded products at the bottom. We downloaded the value set for statins from RxMix, which is an application that provides RxNorm, RxTerms, NDF-RT (web service for accessing the current National Drug File), RxClass, and interaction application programming interfaces (APIs).<sup>23</sup> The Res. value set for the Cholesterol measure included 70 of the Semantic Clinical Drugs (SCDs) from the RxMix value set, while the CMS value set contained all 71 SCDs. There were neither Semantic Branded Drugs (SBDs) nor Generic Pack (GPCK) in either Cholesterol measure value set.

## Discussion

We identified key differences in the value set creation process for two similar definitions of the same CQM. Although this did not drastically affect the global measure performance, the two different versions did lead to variations in calculated prevalence of patients at high risk from key conditions, with 7.5 percent difference in performance rates for the aggregate clinical concept 'myocardial infarction', and 3.5 percent performance difference for the concept 'ischemic vascular disease'. Furthermore, there were 41.5 percent (677) more patients included in the aggregate concept 'IVD' as defined by the Res. value set, and 135 percent (148) more patients included in the concept 'myocardial infarction' as defined by the CMS value set. The CQM requirements of new payment systems have operationalized many vocabularies, but deciding how to interpret broad terms such as 'heart attack' and 'stroke' when creating specific sets of codes intended to represent these conditions lead to specific variation and confusion. The implications of these differences extend beyond administrative concerns around measurement, and also directly

**Figure 1. SNOMED Code Representation for Parent and Children of 'Carotid Artery Occlusion' in the Different Measures**

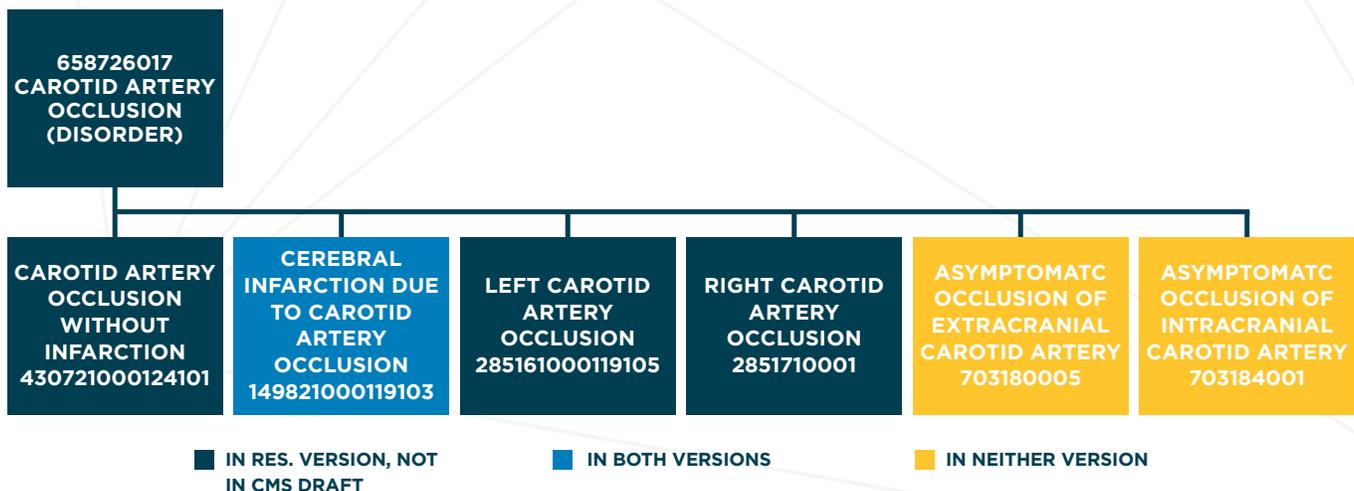
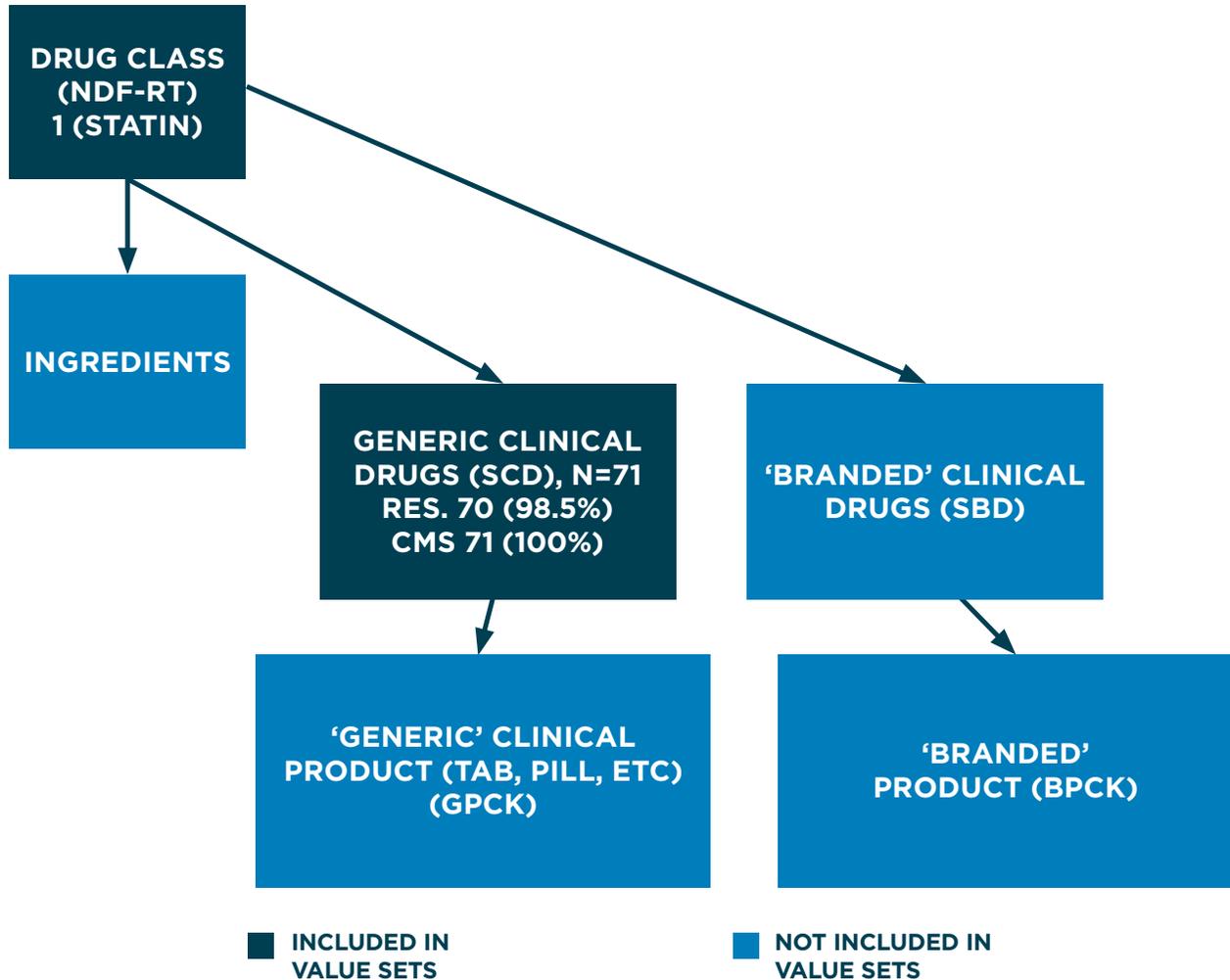


Figure 2. RxNorm Hierarchy



impact delivery of targeted patient care. As technical assistance providers, we know, anecdotally, that high functioning care teams use patient-level data from CQM reports to plan targeted quality improvement interventions for patients not meeting measures. Clinics depending on implementations of measures with inaccurate value sets will not be able to identify all of the relevant patients for these interventions.

It is difficult to make a judgment about whether the Res. or CMS value set is more clinically accurate based on the face-validity of the performance

rates alone. Upon careful perusal of the distinct code details we observed that the Res. value set's definition of 'MI' only included ICD-10-CM codes associated with an "initial encounter", whereas the CMS value set also included codes associated with "subsequent" and "unspecified" encounters. We could not think of any rationale for limiting the 'MI' subgroup to only relevant diagnoses at "initial encounters", and the Res. value set developers did not provide us with any criteria for which to base this decision; therefore, we inferred that the CMS



value set is more clinically accurate for the concept 'myocardial infarction'. For IVD, we observed more random variation in code differences, and observed that the same code groups represented different clinical concepts between the two value sets. For example, the Res. value set included 70 codes from the 'I63' ICD-10-CM series (which represents the concept 'cerebral infarction') and included these in its concept 'IVD', while the CMS value set included the 'I63' ICD-10-CM series in its concept 'stroke'. We could not make a clear determination about clinical accuracy for the IVD subgroup given this overlap and ambiguity of the codes and concepts included. No logic was provided by measure developers about the decisions made in value set development.

It is also difficult to evaluate the long-term impact of these value set differences on CQM implementations. One might hope that the coding differences we observed in this natural experiment might be resolved in further value set development. In fact, we learned retrospectively that the OID which represents 'MI' in the Res. value set was updated in 2016 to include "subsequent" and "unspecified" encounters making it identical to the 'MI' concept definition from the CMS value set. From this, we infer that the Res. OID was corrected to include the codes that were erroneously omitted. However, given our current process, we did not have any way of knowing about this update without manually searching the VSAC for this particular OID. Our results presented in this paper reflect the 2015 version of the 'MI' OID, and not the most updated version. Cimino identified the need for controlled vocabularies to gracefully change over time: "if ... a concept is changed in such a way as to alter its meaning, what happens to the ability of the aggregated patient data that are coded before and after the change?"<sup>14</sup> Moving forward, CQM implementers will need to overcome the challenge of inadvertently using antiquated value sets.

From discussions with the EvidenceNOW informatics leads and measure developers, we suggest a set of potential changes to the development and implementation process:

- 1) The measure development process should require significant validation with extant data sets, utilizing the substantial number of EHR mapped data sets existing now to determine the most accurate code sets to include. The developers of the Cholesterol measure conducted workflow assessments to determine whether the measure's data elements were integrated into a typical clinic's documentation workflow as structured data, but did not implement the measure directly against any clinical datasets.
- 2) Another suggestion is that when value sets need to be created for new measures, informaticians should be included in the creation to mitigate potential challenges with concept orientation, redundancy, completeness, and correctness. The developers may already be doing this, as they indicated that they have a team of clinical experts and coding experts who make decisions about the level of granularity of code groups to include/exclude.
- 3) Additionally, value set designation should be changed to clearly highlight developers' choices at different levels of the various hierarchies. An example of this would be indicating which RxNorm drug class is chosen, and the subsequent SCDs that come from that choice in a measure. The measure developers indicated that they make these decisions about inclusiveness on a case-by-case basis based on their understanding of the clinical concepts of interest.
- 4) Finally, system vendors, CQM implementers, the VSAC, and the NQF need to establish integrated processes for adequately disseminating updates to value sets. Collaboration between these groups is necessary to make sure that CQM implementations reflect the most recent - and most accurate -

versions of measure specifications. Promoting the use of APIs to the VSAC for updates, for example, could be a promising step to automate this process.

### Limitations

This study has two primary limitations. First, we analyzed the performance rates for the various versions of the measure against data in only five primary care clinics in Oregon. Thus, the practical significance of our findings is unclear given the small size of the study. Second, our research design is based on a case study of a single statin therapy clinical quality measure so the extent to which our findings are generalizable beyond this measure is unclear. Future analysis could be done against data from several other clinics to further assess the differences in the prevalence of the at-risk patient groups based on these distinct value sets.

The goal of this paper was to show differences between point estimates of measures when different definitions of key concepts were chosen by developers. Although we could test whether each difference is statistically different, the primary point is to show the difference itself. The IVD-only implementation of denominator 1 had the largest difference in the number of patients included; a chi-square test provides a p-value < .05, showing that the magnitude of difference can be large enough to be significant. A difference of 1-2 percent in CQM performance indicates that 1-2/100 additional patients would be given appropriate treatment and monitoring for their disease in the definition most aligned with the concepts intended to be measured. Depending on the measure set, this may be important clinically. In addition, we know the causality of the difference (the change in value sets), so detecting whether the change causes large or small differences is what is important.

### Conclusion

We discovered significant differences for the clinical concept 'ischemic vascular disease' – and surprisingly little difference in other concepts – in the new measure development decision-making for the same clinical measure. Thus, we conclude that the measure development process should require developers of value sets to provide the explicit criteria used in the choices made regarding inclusion and exclusion of concepts. This study revealed that it is very hard to understand decisions at different levels of the various hierarchies (e.g., SNOMED, RxNorm) because there is no specific guidance provided in current value sets. Without this information, we may be underspecifying broad medical terms. Therefore, when CQMs are implemented against clinical data, the ability to accurately identify patients at risk for key conditions may be impacted. Importantly, there is a need to evaluate value sets against clinical data before publishing for them for implementation. In addition, assumptions about temporality of the condition – whether they can be new, existing, and/or historical to qualify – made a significant difference in the value set creation. Endorsements are needed to encourage collaboration between system vendors, payers, informaticians, and clinicians so that integrated terminology sets can evolve.

### Acknowledgments

This project is supported by grant number R18HS023908 from the Agency for Healthcare Research & Quality (AHRQ).

### References

1. Zuvekas SH, Cohen JW. Fee-For-Service, While Much Maligned, Remains The Dominant Payment Method For Physician Visits. *Health affairs (Project Hope)*. 2016;35(3):411-4.
2. Nicholas LH, Dimick JB, Iwashyna TJ. Do hospitals alter patient care effort allocations under pay-for-performance? *Health services research*. 2011;46(1 Pt 1):61-81.



3. Dorr DA, Cohen AM, Williams MP, Hurdle J. From simply inaccurate to complex and inaccurate: complexity in standards-based quality measures. AMIA Annual Symposium proceedings AMIA Symposium. 2011;2011:331-8.
4. Berwick DM. Era 3 for Medicine and Health Care. *Jama*. 2016;315(13):1329-30.
5. Centers for Medicare and Medicaid Services, MACRA, <<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-MIPS-and-APMs.html>>; last accessed November 23, 2016.
6. Martin J. 5 Steps to Prepare for MACRA Now. *Texas medicine*. 2016;112(10):39-46.
7. Million Hearts, HHS, <<http://millionhearts.hhs.gov/>>; last accessed November 23, 2016.
8. Agency for Healthcare Research and Quality, EvidenceNOW, <<http://www.ahrq.gov/evidencenow/about/index.html>>last accessed November 23,2016.
9. Evaluating System Change to Advance Learning and Take Evidence to Scale (ESCALATES) ,<<https://escalates.org/>>; last accessed November 23,2016.
10. Healthy Hearts Northwest, H2N, <<http://healthyheartsnw.org/>>; last accessed November 23, 2016.
11. Chronaki C, Jaffe C, Dolin B. eMeasures: a standard format for health quality measures. *Studies in health technology and informatics*. 2011;169:989-91.
12. National Library of Medicine, Value Set Authority Center, <<https://vsac.nlm.nih.gov/>>; last accessed November 23, 2016.
13. Bodenreider O, Nguyen D, Chiang P, Chuang P, Madden M, Winnenburg R, et al. The NLM value set authority center. *Studies in health technology and informatics*. 2013;192:1224.
14. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of information in medicine*. 1998;37(4-5):394-403.
15. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *Journal of the American Medical Informatics Association : JAMIA*. 2006;13(3):277-88.
16. Chute CG, Cohn SP, Campbell JR, Group FTAHISBVW, the Computer-based Patient Records Institute Working Group on C, Structures. A Framework for Comprehensive Health Terminology Systems in the United States: Development Guidelines, Criteria for Selection, and Public Policy Implications. *Journal of the American Medical Informatics Association : JAMIA*. 1998;5(6):503-10.
17. Winnenburg R, Bodenreider O. Metrics for assessing the quality of value sets in clinical quality measures. AMIA Annual Symposium Proceedings. 2013;2013:1497-505.
18. CMS Measures Management System Blueprint v12.0, <<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/MMS-Blueprint.html>>; last accessed February 2, 2017.
19. Heisey-Grove D, Wall HK, Helwig A, Wright JS. Using electronic clinical quality measure reporting for public health surveillance. *MMWR Morbidity and mortality weekly report*. 2015;64(16):439-42.
20. Rector AL. Clinical terminology: why is it so hard? *Methods of information in medicine*. 1999;38(4-5):239-52.
21. Robinson JG. The future of lipid guidelines. *Current opinion in lipidology*. 2016;27(6):585-91.
22. Dale JA, Behkami NA, Olsen GS, Dorr DA. A multi-perspective analysis of lessons learned from building an Integrated Care Coordination Information System (ICCS). AMIA Annual Symposium proceedings AMIA Symposium. 2012;2012:129-35.
23. National Library of Medicine, RxMix, <<http://mor.nlm.nih.gov/RxMix/>>; last accessed November 23, 2016.

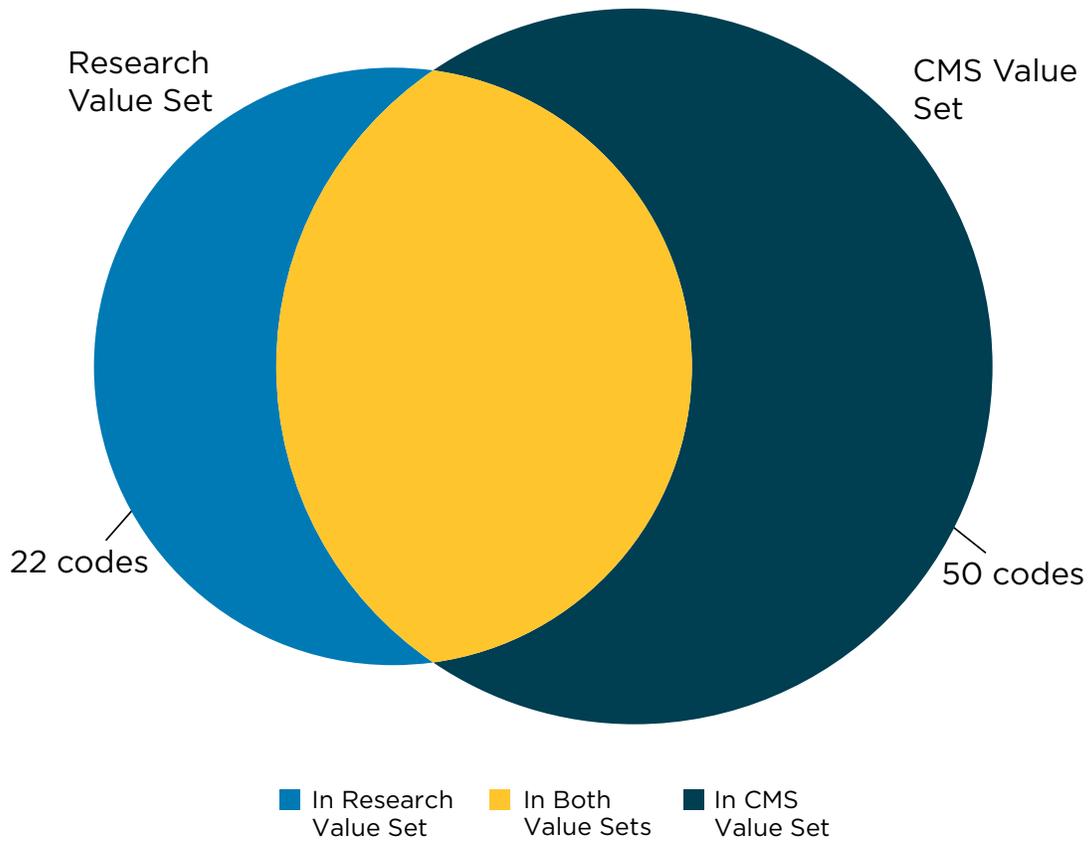
## Appendix A. List of OID Names and Other Identifiers

RESEARCH VALUE SET	CMS VALUE SET	STUDY CATEGORIES USED FOR COMPARISON
Acute Myocardial Infarction	Annual Wellness Visit	Breastfeeding
Annual Wellness Visit	Atherosclerosis and Peripheral Arterial Disease	CABG
Breastfeeding	Atherosclerosis and Peripheral Arterial Disease ICD10	Carotid Intervention
CABG Surgeries	Atherosclerosis and Peripheral Arterial Disease ICD9	Diabetes
Carotid Intervention	Atherosclerosis and Peripheral Arterial Disease Snomed	Encounters (including Office Visit, Annual Wellness, Preventive Care, Outpatient)
Diabetes	Breastfeeding	ESRD
End Stage Renal Disease	Breastfeeding ICD10	Hypercholesterolemia
Face-to-Face Interaction	Breastfeeding SNOMEDCT	Ischemic Vascular Disease (including Stable/Unstable Angina & Artery Disease)
Hemorrhagic Stroke	CABG Surgeries	LDL Test
Ischemic Stroke	CABG Surgeries ICD10PCS	Liver Disease (including Hepatitis A & Hepatitis B)
Ischemic Vascular Disease	CABG Surgeries ICD9	Medical Reason
LDL Test	CABG Surgeries SNOMED CT	Myocardial Infarction
Liver Disease	Carotid Intervention	Palliative Care (including Comfort Measures)
Medical Reason	Cerebrovascular disease, Stroke, TIA	PCI
Office Visit	Cerebrovascular disease, Stroke, TIA ICD10	Pregnancy
Palliative Care	Cerebrovascular disease, Stroke, TIA ICD9	Statin Allergen (including Rhabdomyolysis)
PCI	Cerebrovascular disease, Stroke, TIA SNOMED	Statin RXNORM
Pregnancy Dx	Comfort Measures	Stroke
Preventive Care Services - Established Office Visit, 18 and Up	Diabetes	
Preventive Care Services- Initial Office Visit, 18 and Up	End Stage Renal Disease	
Statin	Hepatitis A	
Statin Allergen	Hepatitis B	
	High intensity statin therapy	
	Hypercholesterolemia	
	Hypercholesterolemia ICD 9	
	Hypercholesterolemia ICD10	
	Hypercholesterolemia SNOMEDCT	
	Ischemic heart disease or coronary occlusion ICD9	



RESEARCH VALUE SET	CMS VALUE SET	STUDY CATEGORIES USED FOR COMPARISON
	Ischemic heart disease or coronary occlusion or rupture ICD10	
	Ischemic heart disease or coronary occlusion or rupture SNOMED	
	Ischemic heart disease or coronary occlusion, rupture, or thrombosis	
	LDL Test	
	Liver Disease	
	Liver Disease ICD10	
	Liver Disease ICD9	
	Liver Disease SNOMEDCT	
	Low intensity statin therapy	
	Moderate intensity statin therapy	
	Myocardial Infarction	
	Office Visit	
	Outpatient Consultation	
	Outpatient Encounters for Preventive Care	
	Palliative SNOMED	
	Palliative Care	
	Palliative care ICD 10	
	Palliative ICD 9	
	PCI	
	Pregnancy Dx	
	Pregnancy I10	
	Pregnancy I9	
	Pregnancy SNM	
	Preventive Care Services - Established Office Visit, 18 and Up	
	Preventive Care Services - Other	
	Preventive Care Services-Individual Counseling	
	Preventive Care Services-Initial Office Visit, 18 and Up	
	Rhabdomyolysis	
	Rhabdomyolysis ICD10	
	Rhabdomyolysis SNOMED	
	Stable and Unstable Angina	
	Stable and Unstable Angina ICD10	
	Stable and Unstable Angina ICD9	
	Stable and Unstable Angina SNOMED	
	Statin Allergen	

**Appendix B. Overlap of Codes Between Research Value Set and CMS Value Set for the Category 'Myocardial Infarction'**





## Appendix C. Acronym Glossary

CMS - Centers for Medicare and Medicaid Services

MIPS - Merit-based Incentive Payment System

MACRA - Medicare Access and CHIP Reauthorization Act

CQM - Clinical quality measure

ABCS - : Aspirin when appropriate (CMS164v5; NQF0068), Blood pressure control (CMS165v4; NQF0018), Cholesterol management (CMS347v0; NQF pending), and Smoking cessation (CMS138v4; NQF0028).

AHRQ - Agency for Healthcare Research and Quality

ESCALATES - Evaluating System Change to Advance Learning and Take Evidence to Scale

HQMF - Health Quality Measure Format

SNOMED CT - Systematized Nomenclature of Medicine Clinical Terms

ICD-10-CM - International Classification of Diseases Tenth Revision

ICD-9-CM - International Classification of Diseases Ninth Revision

EHR - Electronic health record

VSAC - Value Set Authority Center

Res. - Research value set created by standard measure developer for ESCALATES and EvidenceNOW

ASCVD - Atherosclerotic cardiovascular disease

LDL - Low-density lipoprotein

OID - Unique Object Identifier (used in VSAC to assign identifier to concepts)

ICCIS - Integrated Care Coordination Integration System

SQL - Structured Query Language

LOINC - Logical Observation Identifiers Names and Codes

CPT - Current Procedural Terminology

HCPCS - The Healthcare Common Procedure Coding System

IVD - Ischemic vascular disease

MI - Myocardial infarction

NDF-RT - The National Drug File - Reference Terminology

API - Application programming interface

SCD - Semantic clinical drug

SBD - Semantic branded drug

GPCK - Generic pack

---