

## Eliciting the Intension of Drug Value Sets – Principles and Quality Assurance Applications

Nathan J. Bahr<sup>a</sup>, Scott D. Nelson<sup>b</sup>, Rainer Winnenburger<sup>c</sup>, Olivier Bodenreider<sup>d</sup>

<sup>a</sup>Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA,

<sup>b</sup>Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>c</sup>Stanford Center for Biomedical Informatics Research, Stanford, CA, USA

<sup>d</sup>National Library of Medicine, Bethesda, MD, USA

### Abstract

Value sets (VSs) used in electronic clinical quality measures are lists of codes from standard terminologies (“extensional” VSs), whose purpose (“intension”) is not always explicitly stated. We elicited the intension for the 09/01/2014 release of extensional medication value sets by comparison to drug classes from the October 2014 release of RxClass. Value sets matched drug classes if they shared common ingredients, as evidenced by Jaccard similarity score. We elicited the intension of 80 extensional value sets. The average Jaccard similarity was 0.65 for single classes and 0.80 for combination classes, with 34% (27/80) of the value sets having high similarity scores. Manual review by a pharmacist indicated 51% (41/80) of the drug classes selected as the best mapping for a value set matched the intension reflected in that value set name. This approach has the potential for facilitating the development and maintenance of medication value sets.

### Keywords:

Vocabulary, Controlled, RxNorm, Algorithms

### Introduction

#### Clinical quality measures and value sets

Healthcare professionals and hospitals must report clinical quality measures (CQM) in order to qualify for additional payments under the Electronic Health Records (EHR) Incentive Programs [1]. CQMs define proportion, ratio, or continuous variable measures that help track the quality of services provided within the healthcare system [2]. To calculate CQMs, healthcare organizations must use codes from standard terminologies in EHR data to identify target populations that meet a given set of criteria. For example, the *beta-blocker therapy for left ventricular systolic dysfunction (LVSD)* measure is defined as the proportion of patients receiving beta-blockers among patients diagnosed with LVSD.

Value sets are lists of codes from standard terminologies (including RxNorm for drugs) that are used to identify concepts such as *beta-blockers* and *LVSD*, which can be used to select appropriate patient cohorts. The National Library of Medicine’s (NLM) Value Set Authority Center (VSAC) is responsible for the validation and delivery of value sets.

#### Extensional vs. intensional value set definition

In practice, most value sets are defined by their extension (i.e., the list of their codes), rather than by their intension (i.e., the properties common to all codes in the value set). The intension of the value set is often only reflected in its name. However, the value set name itself is insufficient to unambiguously define the list of its members. For example, the value set *Beta Blocker Therapy Ingredient* is used in a cardio-vascular CQM,

and beta-blockers used as anti-glaucoma agents would be irrelevant in this value set.

Intensional definitions are critical for the maintenance of value sets. For example, when a new beta-blocker becomes available on the market, it should be added to the beta-blockers value set. If the value set is defined in reference to the drug class (i.e., intensional definition), the new drug will be added to the value set automatically. In contrast, if the list of beta-blockers is established by a pharmacist (i.e., extensional definition), it will need to be periodically revisited to reflect new drugs.

For most value sets currently in use to support CQMs, value set developers have created sets of codes from standard terminologies (e.g., RxNorm for drugs), without explicit documentation of why specific codes were selected. These extensional value sets pose a challenge for validation and maintenance.

### Objectives

The objective of this work is to elicit the intension of medication value sets in reference to drug classes. More specifically, we have observed that the value set names generally correspond to: 1) a single drug class, like *Statin*; 2) a drug class in the context of a given disease, like *Antibiotic Medications for Pharyngitis*, and 3) multiple drug classes, like *Ace Inhibitor or ARB Ingredient*. Based on this observation, we propose to use single drug classes, as well as combinations (i.e., intersections and unions) of drug classes to specify an explicit structured intension for medication value sets. A secondary objective is to evaluate the quality of medication value sets by comparison to drug classes. We investigated the 09/01/2014 release of medication value sets for eligible practitioners and hospitals, from the NLM Value Set Authority Center (VSAC), located at <https://vsac.nlm.nih.gov/>. These value sets were created against the October 2014 release of RxNorm.

### Related work

Most work on clinical quality measures (CQMs) has focused on their potential for electronically tracking and improving delivery of care [3], issues in validating results from CQMs [4], and accuracy and completeness issues of some CQMs [5; 6]. In this investigation, we specifically focus on the quality of value sets used in CQMs.

In previous work, Winnenburger and Bodenreider [7] assessed the quality of disease value sets by comparing them to the disease classes in the source from which they were derived (e.g., SNOMED CT or ICD10-CM). They hypothesized that concepts in the value set were rooted in one or more ancestor concepts and that these ancestor concepts represent the intension. The extension for a reference value set could then be constructed as the root concepts along with their descendant concepts. For example, the value set *left ventricular systolic dysfunction* would ideally include the concept *left ventricular*

*systolic dysfunction* and its descendants in SNOMED CT. These techniques provided a framework for evaluating the quality of a value set from a known, structured intension.

We take a similar approach for medication value sets. Of note, while disease value sets can be analyzed in reference to the disease hierarchy provided by SNOMED CT or ICD10-CM, medication value sets require drug classification systems external to RxNorm, namely ATC, MeSH, NDF-RT and DailyMed (described below). Moreover, drug classification systems organize drugs according to various dimensions (e.g., indications, mechanism of action) and multiple combined criteria may be needed to fully characterize a medication value set.

### RxNorm and drug classification systems

*RxNorm* [8] is a standardized nomenclature for medications produced and maintained by the NLM. It provides drug concepts and relations among them. RxNorm concepts are also linked to various drug classification systems through RxNorm's companion resource, RxClass. In this investigation, we leveraged the RxNorm and RxClass application programming interfaces (APIs), available at <https://rxnav.nlm.nih.gov/>. More specifically, we used the RxNorm API, to map various kinds of drug entities to ingredients (e.g., the brand name *Lipitor* to *Atorvastatin*), as drug classification systems generally reference ingredients. We used the RxClass API to associate ingredients with drug classes.

The following drug classification systems were used as a source of drug classes. The *Anatomical Therapeutic Chemical (ATC)* [9] classification system is maintained by the World Health Organization (WHO) for pharmaco-epidemiology purposes. Each ingredient is associated with one or more ATC class. For example, *Atorvastatin* is a member of the class *HMG CoA reductase inhibitors*.

The *Medical Subject Headings (MeSH)* [10] is a controlled vocabulary produced and maintained by the NLM for the indexing and retrieval of the biomedical literature. Its drug descriptors are linked to *Pharmacologic Action (PA)* descriptors which describe mechanisms of action and therapeutic uses. For example *Atorvastatin* has the following pharmacologic actions: *anticholesteremic agents* and *hydromethylglutaryl-CoA reductase inhibitors*.

The *National Drug File Reference Terminology (NDF-RT)* [11] is developed by the Department of Veterans Affairs (VA) Veterans Health Administration and associates ingredients with different pharmacological classes, including chemical structure and diseases for which the drug is indicated. For example, *Atorvastatin* is a member of the disease class *Hypercholesterolemia* (among others).

Finally, *DailyMed* [12] associates ingredients with different pharmacological classes, including the Food Drug Administration's Established Pharmacological Classes (EPC), mechanism of action (MoA), and physiologic effect (PE). Although these associations are also defined in NDF-RT, we used DailyMed because it represents a more authoritative source. For example, *Atorvastatin* is a member of the EPC class *HMG-CoA Reductase Inhibitor*, and the MoA class *Hydroxymethylglutaryl-CoA Reductase Inhibitors*.

### Methods

Our approach to eliciting the intension of medication value sets can be summarized as follows (Figure 1). We establish sets of ingredients from drug value sets and from drug classes, and we compare lists of ingredients between value sets and

drug classes. Finally, we perform a quantitative and qualitative evaluation of the elicited intensions.

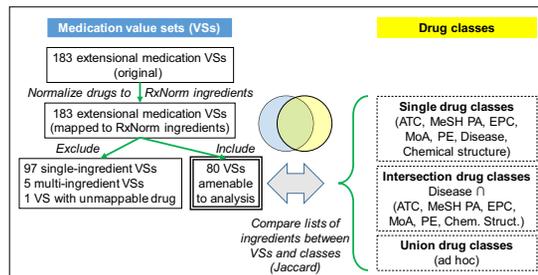


Figure 1 - Overview of the methods.

### Establishing sets of ingredients from drug value sets

Medication value sets from VSAC can contain various kinds of RxNorm drug entities, including ingredients (e.g., *Carvedilol*) and clinical drugs (e.g., *Carvedilol 25 MG Oral Tablet*), as well as brand names (e.g., *Coreg*), specific salts and esters (e.g., *carvedilol phosphate*) and other kinds of drug entities. We leveraged the RxNorm API to map the various kinds of drug entities to their corresponding ingredient to simplify the analysis. We excluded multi-ingredient drugs, because the corresponding single-ingredient drugs tend to be listed in the value sets and also because multi-ingredient drugs are not represented consistently across drug classification systems.

### Establishing sets of ingredients from drug classes

We leveraged the RxClass API to find the list of RxNorm drug members for drug classes from ATC, MeSH, NDF-RT, and DailyMed. As was done for drug entities from the value sets, we ignored multi-ingredient drugs and mapped the various kinds of drug entities to their corresponding ingredient.

In addition to the (single) classes found in drug classification systems, we created combination classes to represent the sets of drugs reflected in value set names. Namely, we created two types of combination classes: 1) intersection classes, where each single class is intersected with each of the disease classes (attempting to approximate value sets, such as *Antibiotic Medications for Pharyngitis*); 2) union classes, where multiple single classes are merged (attempting to approximate value sets, such as *Ace Inhibitor or ARB Ingredient*). We created the intersection classes systematically for each single class. In contrast, we created union classes corresponding to the best match for each value set (by finding the single class that is most similar to a given value set, and then the best single class that is most similar to the drugs not covered at earlier steps).

### Comparing ingredients between value sets and drug classes

We used the Jaccard coefficient to measure the similarity between value sets and (single or combination) drug classes based on their normalized ingredients. The Jaccard coefficient for a value set and a drug class,  $J(V, C)$ , computes similarity as the ratio between the number of ingredients common to the value set and the drug class,  $|V \cap C|$ , over the total number of ingredients in the value set and the drug class,  $|V \cup C|$ . The Jaccard coefficient ranges from 0 (no similarity) to 1 (exact match). The (single or combination) drug class with the highest Jaccard score is considered to best reflect the intension of the value set.

$$JC(V, C) = \frac{|V \cap C|}{|V \cup C|}$$

Figure 2 presents an example, in which the value set *Beta Blocker Therapy* is evaluated against the MeSH class *Adrenergic beta-Antagonists* intersected with the NDF-RT disease class *Hypertension*. The red dotted line indicates that there are 14 ingredients in the combination class. The white circle indicates that 12 ingredients are in both the value set and combination class. This results in a Jaccard score of 0.86 (12/14). The intension for *Beta Blocker Therapy* value set could then be interpreted as *Adrenergic beta-Antagonists* used for *Hypertension*.

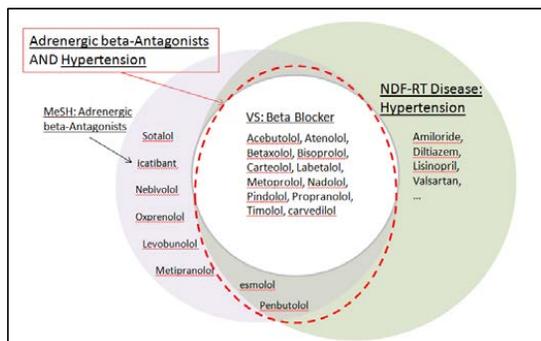


Figure 2 - The Jaccard metric is used to assess the equivalence between value sets and drug classes. The Beta Blocker Therapy value set has an initial match with Adrenergic beta-Antagonists from MeSH. This match is further refined by intersecting the MeSH class with the NDF-RT disease class Hypertension.

## Evaluation

We performed a quantitative and qualitative evaluation to assess the fit and validity of elicited intensions.

**Quantitative.** The quantitative evaluation assesses the overall similarity between value sets and (single or combination) drug classes. More specifically, we simply compute the average of the best Jaccard score for each value set-drug class pair. To assess the contribution of the combination classes (intersection and union classes defined earlier), we compared the averages obtained under the following strategies:

1. When using only single classes
2. When using single classes and intersection classes
3. When using single classes, intersection classes and union classes

We conducted a one-way ANOVA (with repeated measures) to compare the effect of these different strategies on eliciting the intension based on the Jaccard score. A Tukey *post hoc* analysis was performed to identify which strategies were significantly different (pairwise). The statistical analysis was completed using STATA 13 (StataCorp. 2013. College Station, TX).

Qualitative. The qualitative evaluation assesses the extent to which the (single or combination) drug class matches the intension reflected in the value set name. An expert pharmacist (SDN), who had not been involved with the development of the methods, analyzed the drugs listed in the class and in the value set for the class identified as the best match for each value set. More specifically, the pharmacist was asked to answer two main questions for each value set-drug class pair:

1. Do the drugs listed in the value set correspond to the intension reflected in the value set name?
2. Do the drugs listed in the (single or combination) class correspond to the intension reflected in the value set name?

Additionally, the pharmacist was asked whether there were missing or extraneous drugs in the value set, in the best-matching class, or in both.

## Results

### Establishing sets of ingredients from drug value sets

As shown in Figure 1, there were 183 extensional medication value sets. Ninety-seven were excluded because they contained only one ingredient and could be trivially mapped to the Chemical structure drug class restricted to this ingredient. Five were excluded because their extensions were composed entirely of multi-ingredient drugs. One contained one ingredient that could not be mapped to a drug class. The remaining 80 value sets were analyzed and contained 468 distinct ingredients after mapping to RxNorm.

### Establishing sets of ingredients from drug classes

Table 1 shows the number of single and combination classes for each source. The NDF-RT disease classes (reflecting drug indications) were combined with other drug classes resulting in approximately 4 million intersections. We iteratively generated the union of classes that had some equivalence to value sets based on the Jaccard score. This resulted in approximately 100 union candidates. The 6519 single classes contained 2957 distinct ingredients after mapping to RxNorm.

Table 1 - Number of drug classes in each source.

Type	Source	# classes
Single	ATC	882
	MeSH Pharmacological Action	350
	DailyMed Established Pharmacological Class	431
	DailyMed Mechanism of Action	348
	DailyMed Physiological Effect	239
	NDF-RT Disease	1434
	NDF-RT Chemical Structure	2835
Combination	NDF-RT Disease intersected with all single drug classes	Approximately 4 million
	Union of multiple drug classes	Approximately 100

### Comparing ingredients between value sets and drug classes

We compared each of the 80 value sets to all single drug classes, resulting in 521,520 comparisons, from which we selected the best value set-drug class pair match. We found that there was no single source that best described all value sets. However, many of the top matches appear to come from ATC and DailyMed EPC, followed by NDF-RT Physiologic effect and NDF-RT Mechanism of action.

Similarly, we identified the best match for each value set and all intersection drug classes (over 320 million comparisons). Here again, intersections between Disease classes and classes from ATC and DailyMed EPC provided most of the best matches. The intersection of Disease classes and classes from NDF-RT Chemical structure also contributed many of the best matches. Finally, we determined the best union classes for each value set.

## Evaluation

**Quantitative.** We examined the distribution of the Jaccard scores for best matches between the 80 value sets and drug

classes, using three mapping strategies: single classes only; single classes or intersection classes; single classes or intersection classes or union classes. A larger proportion of the value sets obtain better Jaccard scores under the last two strategies compared to single classes.

More specifically, the average Jaccard score of single classes was 0.65, with 23% (18/80) of the value sets having high similarity (0.9-1) with a drug class. Adding disease intersections increased the average Jaccard score to 0.79 with 34% (27/80) of the value sets having high similarity with a single drug class or a drug class intersected with a disease class. Adding union classes provided a very small performance gain, only increasing the average Jaccard score to 0.80. The union mostly affected value sets that did not have a good match to a single drug class or single drug class intersected with a disease class.

The one-way ANOVA test showed that the different strategies resulted in statistically significant differences in mean Jaccard score,  $F(2, 158) = 30.02$ ,  $p < 0.005$ . The *post hoc* Tukey test indicated that both types of combination classes performed significantly better than single classes only ( $p = 0.024$  and  $p = 0.009$ ). However, there was no significant difference between the two types of combination classes ( $p = 0.934$ ).

Table 2 - Examples of elicited intensions (original value set, single or combination drug class, Jaccard score, and pharmacist comments).

Value Set → Drug class [Pharmacist comments]	JC
<b>Leukotriene modifiers</b> → Leukotriene Antagonists [No comments, exact match.]	1.00
<b>IV or IM Beta-lactams</b> → Cefotaxime [Value set missing a lot of drugs.]	0.29
<b>IV PO Quinolone Antipneumococcal</b> → Pneumonia, Bacterial, AND Quinolones [Would have been more valid to filter on route.]	1.00
<b>Beta Blocker Therapy for LVSD</b> → Tachycardia, Supraventricular AND Adrenergic beta-1 Receptor Antagonists [Value set only contains oral forms of the drugs and there are differences in the salt forms.]	0.40
<b>Beta-Lactams for Allergy Determination</b> → beta-Lactams OR other beta-lactam antibacterials OR Beta-lactamase sensitive penicillins [Value set had a lot of missing SCDs. Drugs that are not antibiotics, but inhibit beta-lactams not in ingredients.]	0.80
<b>BH Mood Stabilizer Medication RxNorm</b> → Phenothiazines with piperazine structure OR Dibenazepines OR Fructose OR Other antipsychotics [Value set is poorly defined and has a lot of missing drugs.]	0.27

**Qualitative.** Table 2 shows examples of value sets with elicited intensions.

**Value set extension vs. value set name.** Eighty percent (64/80) of the value sets had ingredients that were consistent with the intension reflected in the value set name. Sixteen percent (13/80) of the value sets had extraneous ingredients and 46% (37/80) had missing ingredients.

**Drug class extension vs. value set name.** Fifty-one percent (41/80) of the (single or combination) drug classes selected as the best mapping for a value set matched the intension reflected in that value set name. Twenty-eight percent (22/80) of the best-matching drug classes had extraneous ingredients and 43% (35/80) had missing ingredients. In thirty-one percent (25/80) of the cases, there were missing ingredients in both the drug class and value sets.

**Other issues.** We found several duplicate value sets, such as *BH Antidepressant Medication - SSRI Antidepressants RxNorm* and *SSRI Antidepressants*. The issue of duplicate

value sets had already been reported by Winnenbourg et al [7] for disease and procedure value sets, so it was not surprising to find duplicates in medication value sets. We also found some naming inconsistencies, such as with the qualifiers used to refer to injectable drugs in value set names, which included *IV, parenteral, Injectable, and for IV Administration*.

## Discussion

### Applications

This work presents a framework for eliciting an explicit form of the intension in reference to a normative drug source, such as the ATC, MeSH, NDF-RT and DailyMed drug classification systems. The drug classes from these sources allow us to compare logical groupings of drugs and derive a reference extension that can be used to validate the ingredients in the value set.

There are two main applications for this framework.

The primary application is to derive operational intensional definitions for medication value sets in order to facilitate their maintenance. Once a value set has been associated with a (single or combination) class, value set developers can rely on the corresponding drug classification systems for the maintenance of the value set. In practice, instead of relying on experts for checking whether drugs should be added to or removed from a given value set, value set developers would only need to check the members of the drug class the value set has been mapped to when new versions of the drug classification system become available. Operational intensional definitions greatly facilitate the maintenance and reliability of drug value sets. We discussed our findings with representatives of the NLM Value Set Authority Center (VSAC), where these techniques could be implemented to facilitate the development and ongoing maintenance of the value sets.

A secondary benefit of comparing value sets to drug classes is that it provides an opportunity for quality assurance. Our expert pharmacist identified potential errors, specifically missing drugs and extraneous drugs, in almost two thirds of the value sets. Here again, the availability of such a framework will likely benefit the quality assurance of medication value sets in VSAC. Of note, similar errors were identified in drug classes and could be reported to the developers of the corresponding drug classification systems.

Finally, this work informed the visualization of drug classes in RxClass (<https://mor.nlm.nih.gov/RxClass/>), the tool we developed for browsing and comparing drug classes associated with RxNorm.

### Failure analysis

We were able to elicit an appropriate intension for about half of the value sets, while the other half corresponded to value sets with a mixture of drugs belonging to many different, but similar classes. For example, the value set *IV Quinolones Used For Prophylaxis for Hysterectomy and Colon Surgery* best matched the intersection between two drug classes, *Quinolone Antimicrobial* and *Sinusitis*. Antimicrobial drugs have many therapeutic uses, so in this case, we elicited the wrong intension that provided a good match (treatment of sinusitis vs. prophylaxis for hysterectomy and colon surgery).

In other cases, the value sets contained a complicated mixture of drugs. For example, the *Anti-Hypertensive Pharmacologic Therapy* value set contained 67 distinct ingredients. The best match consisted of a combination class resulting from the intersection of *Hypertension* with the top-level class *Established Pharmacologic Class (EPC)*, i.e., all drugs with any EPC class. Interestingly here, the expected best match was to the single

class *Hypertension* itself, not an intersection class. However, this occurred because several drugs used to treat hypertension were missing from the value set, such as *Methyldopa*, *Guanethidine*, and *Bumetanide*. This case illustrates not so much a failure of our approach to associating value sets with drug classes than a quality issue with the value set.

We observed that some of these value sets were restricted by dose form (e.g., *IV Quinolones*, which restricts antibiotics from the quinolone class to their intra-venous forms). Because drug classes list ingredients as their members, a drug class, such as *QUINOLONE ANTIBACTERIALS* in ATC would retrieve the appropriate ingredients, but would not distinguish between intra-venous and other forms of these ingredients. For these classes, RxNorm could be used to filter ingredients for which there exist clinical drugs for specific dose forms. Failure to implement this feature in our framework resulted in extraneous ingredients in some value sets. For example, *QUINOLONE ANTIBACTERIALS* in ATC lists the ingredient *Ofloxacin* as a member, for which there exist no injectable forms. Similarly, the dose of the drug often had contextual implications. For example, some value sets only listed heparin flushes, rather than the therapeutic dose. This suggests that value set developers created some value sets with specific dose form groups in mind.

Finally, there were a few cases, where we were unable to achieve a good match because the requisite information was often outside the scope of the terminology. For example, certain drugs or doses listed in the terminology are not legally prescribed in the U.S., such as *Phenprocoumon* and *Dicumarol*.

#### Limitations and future work

There are a few limitations regarding the materials used in this study. First, multi-ingredient drugs were not included in the analysis, because the corresponding single-ingredient drugs tend to be listed in the value sets and also because multi-ingredient drugs are not represented consistently across drug classification systems. However, this restriction resulted in eliminating five value sets composed entirely of multi-ingredient drugs. Second, the therapeutic intent of medications is often dependent on medication route and dose. As discussed in the failure analysis, we could easily restrict ingredients based on intended route, when a route restriction was expressed in the value set name. However, we would probably not be able to account for specific doses (such as heparin flushes), because this information was not made explicit in value set names. Third, this study is based on older versions of the value sets, RxNorm and drug classification systems, because the qualitative analysis was performed against these specific versions. Nevertheless, the methods are generalizable to newer versions.

#### Conclusion

We proposed an approach for eliciting the intension of medication value sets by comparing the list of ingredients in these value sets to single or combination drug classes derived from drug classification systems, such as ATC, MeSH, NDF-RT and DailyMed. With this approach, we were able to find drug classes that match the value sets with 0.79 Jaccard similarity on average. This approach has the potential for facilitating the development and maintenance of medication value sets. We also discussed its benefits in terms of quality assurance.

#### Acknowledgements

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine (NLM). During this project, Dr. Nelson was supported by the VA Advanced Fellowship Program in Medical Informatics of the Office of Academic Affiliations, Department of Veterans Affairs. The views, findings, and conclusions expressed in this report are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs, the Food and Drug Administration, or the National Library of Medicine.

#### References

- [1] Meaningful Use Definition and Meaningful Use Objectives of EHRs | Providers & Professionals | HealthIT.gov., Available: <https://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives>, Accessed: December 5 2016
- [2] CMS.gov, Available: <https://www.cms.gov>, Accessed: December 5 2016
- [3] P.C. Tang, M. Ralston, M.F. Arrigotti, L. Qureshi, and J. Graham, Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures, *J Am Med Inform Assoc* **14** (2007), 10-15.
- [4] F. Eisenberg, C. Lasome, A. Advani, R. Martins, P. Craig, and S. Sprenger, A study of the impact of meaningful use clinical quality measures, *American Hospital Association* (2013).
- [5] K.S. Chan, J.B. Fowles, and J.P. Weiner, Review: electronic health records and the reliability and validity of quality measures: a review of the literature, *Med Care Res Rev* **67** (2010), 503-527.
- [6] L.M. Kern, S. Malhotra, Y. Barron, J. Quaresimo, R. Dhopeswarkar, M. Pichardo, A.M. Edwards, and R. Kaushal, Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study, *Ann Intern Med* **158** (2013), 77-83.
- [7] R. Winnenburg and O. Bodenreider, Metrics for assessing the quality of value sets in clinical quality measures, *AMIA Annu Symp Proc* **2013** (2013), 1497-1505.
- [8] RxNorm, Available: <https://www.nlm.nih.gov/research/umls/rxnorm/>, Accessed: December 5 2016
- [9] WHOCC - ATC/DDD Index, Available: [https://www.whooc.no/atc\\_ddd\\_index/](https://www.whooc.no/atc_ddd_index/), Accessed: December 5 2016
- [10] Medical Subject Headings, Available: <https://www.nlm.nih.gov/mesh>, Accessed: December 5 2016
- [11] S.H. Brown, P.L. Elkin, S.T. Rosenbloom, C. Husser, B.A. Bauer, M.J. Lincoln, J. Carter, M. Erlbaum, and M.S. Tuttle, VA National Drug File Reference Terminology: a cross-institutional content coverage study, *Stud Health Technol Inform* **107** (2004), 477-481.
- [12] DailyMed, Available: <https://dailymed.nlm.nih.gov/dailymed>, Accessed: December 5 2016

#### Address for correspondence

Olivier Bodenreider, MD, PhD  
8600 Rockville Pike, 38A/09S904, Bethesda, MD 20894  
Phone Number: (301) 827-4982  
E-mail: [olivier.bodenreider@nih.gov](mailto:olivier.bodenreider@nih.gov)