Automated Methods for Transforming Clinical Phenotype Definitions and Quality Measure Value Sets from ICD-9-CM to ICD-10-CM

Kin Wah Fung, MD, MS, MA1, Rachel Richesson, PhD2, Michelle Smerek3, Kathy Pereira, DNP2, Beverly B. Green, MD, MPH4, Ashwin Patkar, MD3,5, Megan Clowse, MD5, Alan Bauck6, Olivier Bodenreider, MD, PhD1

1National Library of Medicine, Bethesda, MD, 2Duke University School of Nursing, Durham, NC, 3Duke Clinical Research Institute, Durham, NC, 4Group Health Research Institute, Seattle, WA, 5Duke University School of Medicine, Durham, NC, 6Center for Health Research, Kaiser Permanente Northwest, Portland, OR

Abstract
The migration of healthcare data from ICD-9-CM to ICD-10-CM in late 2015 will impact research and quality monitoring activities that rely on these codes. We used the CMS General Equivalency Mappings (GEMs) to transform a convenience sample of 32 ICD-9-CM-based phenotypes from three pragmatic trials to ICD-10-CM, using 4 different mapping strategies. We also mapped 202 dually-coded value sets from the Value Set Authority Center (VSAC). We measured the performance of the mapping strategies by clinical review and comparing the map-generated value sets with the original ICD-10-CM value sets. Forward backward map yielded the highest F-1 score overall, but more aggressive mapping strategies are sometimes warranted. Different mapping methods yielded different results and can introduce errors. Therefore, pragmatic trials and other activities using GEMs to transform code sets should clearly specify their approach, ensure that the same methodology is used across sites, and validate the semantic equivalence of the transformation.

Introduction
Large-scale multi-site observational research studies and pragmatic clinical trials utilize clinical data, including diagnosis data that is encoded with the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM), collected by health systems as a byproduct of patient care. The national mandate for health systems to migrate to ICD-10-CM in October 2015 will impact a number of research and quality monitoring activities that rely on these codes, and multi-year studies will need to deal with a mixture of ICD-9-CM and ICD-10-CM encoded data. In the context of pragmatic clinical trials and clinical quality measurement, we explore the use of publicly available mapping files to convert clinical phenotype definitions and value sets from ICD-9-CM to ICD-10-CM, and compare the outcome of different approaches.

Background
The tremendous costs associated with traditional clinical trials limits their use to address the majority of clinical questions and treatment decisions that are based upon insufficient evidence.[1-4] Further, the limited generalizability inherent in clinical trials has stimulated interest in alternative research models, including observational research and pragmatic trials, to support patient-centered outcomes research.[5, 6] These alternative research models depend upon access to electronic health record (EHR) data. The HMORN and other networks are using electronic healthcare and claims data to support understanding of disease.[7, 8] While the use of electronic healthcare claims data has supported observational studies for decades, the growing adoption and sophistication of EHRs has the potential to support research activities such as cohort selection or randomization and enable prospective interventional studies. [9, 10] The routine use of EHR data is part of the vision of the learning healthcare system and is becoming feasible with the widespread adoption and meaningful use of EHRs in healthcare systems.[11]

Pragmatic trials are those conducted in real world conditions and in cooperation with healthcare systems.[6] The NIH Health Care Systems Research Collaboratory is funded by the NIH Common Fund to strengthen the national capacity to implement cost-effective large-scale research studies that engage health care delivery organizations as research partners, with the assumption that this will make research results more relevant to providers and ultimately patients.[12] The Collaboratory includes a number of pragmatic trial demonstration projects that are multi-site, often cluster randomized, intervention studies.[13] The use of EHRs to support trial activities, including the identification
of patient cohorts with precise clinical attributes, is an important component of this vision and the next generation of clinical trials.

Clinical quality measurement is another example of the re-use of EHR data for secondary purposes. As part of the CMS (Centers for Medicare & Medicaid Services) Meaningful Use of EHR program, certified systems have to demonstrate the electronic submission of data for some selected clinical quality measures. Value sets are defined lists of codes from coding systems that allow automatic computation of the numerator and denominator of a quality measure. To support this effort, NLM launched the Value Set Authority Center (VSAC) in 2012 to provide access to all official versions of value sets.[14, 15]

At present, the ICD-9-CM coding system is one of the very few types of healthcare data that is implemented broadly enough to support multi-site clinical research or quality measurement. Therefore, many existing cohort or phenotypic definitions and quality measure value sets are defined with ICD-9-CM codes. However, as of October 2015, all health systems will be changing to ICD-10-CM. The ICD-10-CM is not an incremental version change from ICD-9-CM but a radical transformation, involving major changes not only in the size of the terminology, but in the organization, granularity, and semantics (or meaning) of terms.[16] The more than 68,000 possible terms in ICD-10-CM more than quadruple the 14,000 terms in ICD-9.

The GEMs (General Equivalent Map) are created and maintained by the CMS and the Centers for Disease Control and Prevention (CDC), and serve as a tool for the conversion of data from ICD-9-CM to ICD-10-CM and ICD-10-PCS and vice versa.[17] The GEMs are often also referred to as “crosswalks” since they provide important information linking codes from one system with codes in the other system[18]. Users are cautioned against using the GEMs for actual coding as they have not been completely validated for clinical use. Among the listed applicable use cases in the GEM documentation are the conversion of ICD-9-CM-based data for quality measures and research.[18] There are four GEMs between ICD-9-CM and ICD-10-CM, covering diagnostic and procedure codes separately, and in either the forward (ICD-9-CM to ICD-10-CM) or backward (ICD-10-CM to ICD-9-CM) direction. Using the GEMs is not straightforward and caution is advised.[19-22] The directionality of GEMs allows various approaches for their use for translation of ICD-9-CM to ICD-10-CM codes. There are studies of the impact of ICD-10-CM transition in various healthcare settings. [23-25] To the best of our knowledge, no one has explicitly compared the application of different mapping approaches in the contexts of phenotype definitions or quality measures.

Pragmatic trials that use ICD-9-based phenotype definitions (i.e., sets of ICD-9-CM codes collectively representing a clinical condition used for research purposes) to identify research cohorts, characterize risk factors, and define outcomes will need to ensure the equivalence of ICD-9-CM and ICD-10-CM code sets after healthcare organizations transition to ICD-10. Further, the integrity of research, by definition, requires a complete understanding of any data transformations that take place. The number of clinical phenotypes that will be needed to support the growing number of pragmatic trials is so large that converting phenotype definitions from ICD-9-CM to ICD-10-CM will require automated methods that can be validated and replicated.

Methods

In this study, we examined four different approaches for using the GEMs to map from ICD-9-CM diagnostic codes to ICD-10-CM (figure 1). In increasing order of aggressiveness, the approaches are:

1. **Simple forward map (SFM)**, which uses the forward GEM map from ICD-9-CM to ICD-10-CM.
2. **Forward backward map (FBM)**, which uses the forward plus backward map, treating the backward map as if it is from ICD-9-CM to ICD-10-CM.
3. **Secondary map (SM)**, which is based on FBM, and identifies secondary ICD-9-CM codes (B in figure 1) that are associated with the primary source ICD-9-CM code (A), and uses them to identify additional ICD-10-CM targets (Y).
4. **Tertiary map (TM)**, which identifies tertiary ICD-9-CM codes (C) that are two steps removed from the primary ICD-9-CM code, and uses them to identify additional ICD-10-CM targets (Z).
Figure 1. Four mapping methods to use the forward and backward General Equivalent Maps

Users of the GEM often find that they need to apply the forward and backward maps iteratively to obtain mappings that would otherwise be missed. According to Boyd et al in [26], 36% of the ICD-9-CM codes are involved in so-called “convoluted” mappings, meaning that they are not simple one-to-one, one-to-many, or many-to-one maps. In those complex cases, repeated application of the forward and backward maps will discover more and more potential ICD-10-CM targets. In the online transition tool provided by Boyd’s group, the mappings suggested correspond to the SM in our study. We added the TM to see whether including tertiary ICD-9-CM codes would be helpful.

To evaluate the performance of the four mapping methods, we applied them to lists of ICD-9-CM codes used in two scenarios: clinical research and quality measure.

1. Clinical research scenario

Using a convenience sample of 32 phenotypes (developed to identify research cohorts, characterize risk factors, or define outcomes) from three different pragmatic trials\(^1\) that were defined solely by ICD-9-codes, we collated a list of ICD-9-CM codes. Non-leaf codes (codes that are not at the lowest level and not valid for coding) used in the definitions were expanded to the leaf level.

The ICD-9-CM codes were then converted to ICD-10-CM codes using the four mapping methods based on the 2014 release of the GEMs, and the resulting ICD-10-CM codes were reviewed by clinical experts. One generalist nurse practitioner (KP) and a MD domain expert for each trial (BG, AP and MC) reviewed the phenotype name and generated ICD-10-CM maps to determine if each ICD-10-CM code semantically “fit” into the assigned phenotype grouping, based on their understanding of that phenotype and its intent. For example, for the phenotype “active alcohol abuse” the reviewer was asked to look at the ICD-10-CM codes and determine (yes or no) if those codes were appropriate or inclusion in that heading. Reviewers were provided the original phenotype definition (i.e., the set of ICD-9-CM codes that constitute the target condition) as a reference on the same review sheet.

The reviewers were instructed not to evaluate the completeness of the ICD-10-CM set, but rather to look at the reasonableness of each code in the map-generated ICD-10-CM based phenotype definition. To limit the scope of the evaluation, the experts were not asked to search for additional ICD-10-CM codes that were not on the lists. To give a theoretical estimate of recall, we assumed that the most aggressive map (TM) contained all the correct ICD-10-CM codes. To shorten the list of ICD-10-CM codes that the experts needed to look at, we algorithmically rolled up codes to their parents iteratively, as long as the total number of codes to review was reduced. For example, if the list

\(^1\) Collaborative Care for Chronic Pain in Primary Care (PPACT), Strategies and Opportunities to Stop Colorectal Cancer in Priority Populations (STOP CRC), and A Pragmatic Trial of Population-Based Programs to Prevent Suicide Attempt
contained “M47.10, M47.11, M47.12, M47.13, M47.15, M47.16”, which were all children of M47.1; it would be converted into “M47.1 EXCEPT M47.14” because M47.14 was the only children of M47.1 not included in the list.

2. Quality measure scenario

In the VSAC, we identified all value sets for 2014 Clinical Quality Measures that were dually-defined by both ICD-9-CM and ICD-10-CM code lists. We applied the four mapping methods to the ICD-9-CM codes, and evaluated the map-generated ICD-10-CM code lists, using the original ICD-10-CM code lists as the gold standard.

IBM SPSS © for Windows (version 21) was used for statistical analysis.

Results

1. Clinical research scenario

Among the 3 pragmatic trials, there were 32 cohort definitions with 3 – 161 (mean 19.5) ICD-9-CM codes per definition (table 1). There were altogether 536 unique ICD-9-CM codes, all of which were mappable by all four mapping methods.

<table>
<thead>
<tr>
<th>Demonstration project</th>
<th># definitions</th>
<th># of ICD-9-CM codes/definition (mean)</th>
<th>SFM</th>
<th>FBM</th>
<th>SM</th>
<th>TM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic pain</td>
<td>6</td>
<td>14 – 161 (68.7)</td>
<td>18 – 140 (66.7)</td>
<td>36 – 1060 (374.7)</td>
<td>80 – 1138 (450.0)</td>
<td>80 – 1231 (542.0)</td>
</tr>
<tr>
<td>Suicide prevention</td>
<td>23</td>
<td>4 – 41 (7.9)</td>
<td>1 – 130 (9.1)</td>
<td>2 – 323 (30.4)</td>
<td>2 – 340 (135.6)</td>
<td>2 – 372 (183.1)</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>3</td>
<td>3 – 14 (9.3)</td>
<td>3 – 77 (31.3)</td>
<td>3 – 89 (35.3)</td>
<td>3 – 115 (44.0)</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>32</td>
<td>3 – 161 (19.5)</td>
<td>1 – 140 (19.9)</td>
<td>2 – 1060 (95.0)</td>
<td>2 – 1138 (185.1)</td>
<td>2 – 1231 (237.4)</td>
</tr>
</tbody>
</table>

Table 1. Distribution of ICD-9-CM and ICD-10-CM codes in the original and map-generated definitions.

By our roll-up algorithm, we managed to reduce the number of ICD-10-CM codes that needed to be reviewed from 7,596 to 2,118. Based on the review, the recall, precision and F-1 score was calculated for each definition per mapping method. The average performance for each project and across all projects is shown in table 2. Note that recall was only an approximation because the reviewers were not asked to look for false negatives, and we assumed that TM already contained all correct codes. Overall, the F-1 scores for SFM, FBM, SM and TM were 0.34, 0.67, 0.62 and 0.60 respectively, and the difference was statistically significant (one-way ANOVA, F=5.749, p=0.001).

<table>
<thead>
<tr>
<th>Demonstration Project</th>
<th>SFM recall</th>
<th>F-1</th>
<th>FBM recall</th>
<th>F-1</th>
<th>SM recall</th>
<th>F-1</th>
<th>TM recall</th>
<th>F-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic pain</td>
<td>0.39</td>
<td>0.40</td>
<td>0.76</td>
<td>0.65</td>
<td>0.96</td>
<td>0.70</td>
<td>1.0</td>
<td>0.55</td>
</tr>
<tr>
<td>Suicide prevention</td>
<td>0.22</td>
<td>0.28</td>
<td>0.62</td>
<td>0.64</td>
<td>0.86</td>
<td>0.56</td>
<td>0.96*</td>
<td>0.50</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>0.71</td>
<td>0.68</td>
<td>1.0</td>
<td>0.93</td>
<td>1.0</td>
<td>0.91</td>
<td>1.0</td>
<td>0.78</td>
</tr>
<tr>
<td>Overall</td>
<td>0.30</td>
<td>0.34</td>
<td>0.68</td>
<td>0.67</td>
<td>0.89</td>
<td>0.62</td>
<td>0.97*</td>
<td>0.53</td>
</tr>
</tbody>
</table>

* not 1.0 as expected because in one definition all ICD-10-CM codes were rated as incorrect so the recall was 0.0

Table 2. Average and overall performance of the maps for the 3 projects.

2. Quality measure scenario

In the VSAC, there were 202 quality measure value sets dually-coded in ICD-9-CM and ICD-10-CM. For each value set, we used the four mapping methods to generate ICD-10-CM code lists from the ICD-9-CM codes. Among
the 5,545 unique ICD-9-CM codes, only 2 codes were not mappable by all four mapping methods. The distribution of ICD-9-CM and ICD-10-CM codes in the original value sets and map-generated value sets is shown in table 3.

<table>
<thead>
<tr>
<th>Original value sets</th>
<th>Map-generated ICD-10-CM value sets, ICD-10-CM codes/set (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dually-coded value sets (total 202)</td>
<td>SFM</td>
</tr>
<tr>
<td>ICD-9-CM codes/set (mean)</td>
<td>ICD-10-CM codes/set (mean)</td>
</tr>
</tbody>
</table>

Table 3. Composition of value sets.

The performance of the mapping methods is shown in table 4. The overall average F-1 scores for the four maps SFM, FBM, SM and TM were 0.68, 0.92, 0.87 and 0.85 respectively, and the difference was statistically significant (one-way ANOVA, F=40.889, p<0.0005).

<table>
<thead>
<tr>
<th>SFM</th>
<th>FBM</th>
<th>SM</th>
<th>TM</th>
</tr>
</thead>
<tbody>
<tr>
<td>recall</td>
<td>prec</td>
<td>F-1</td>
<td>recall</td>
</tr>
<tr>
<td>Average of all value sets</td>
<td>0.62</td>
<td>0.94</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Table 4. Average performance of the four mapping methods on the quality measure value sets.

Since the size of the value sets varied considerably, we also examined the performance of the FBM (the overall best performing map) in relation to the size of the map-generated value sets (table 5). Generally, the best performance was observed among smaller value sets of 100 codes or less. Overall, there was a statistically significant negative correlation between the size of value set and F-1 score (two-tailed Kendall’s tau correlation coefficient = -0.283, p<0.0005).

<table>
<thead>
<tr>
<th>Size of value set</th>
<th># value sets</th>
<th>Avg recall</th>
<th>Avg precision</th>
<th>Avg F-1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 10</td>
<td>88</td>
<td>0.96</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>10 - 100</td>
<td>91</td>
<td>0.97</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>100 - 1000</td>
<td>13</td>
<td>0.93</td>
<td>0.75</td>
<td>0.79</td>
</tr>
<tr>
<td>More than 1000</td>
<td>10</td>
<td>0.81</td>
<td>0.58</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Table 5. Performance of the FBM mapping method according to size of the value set.

Discussion

In this study, we compared four progressively more aggressive methods: simple forward map (SFM), forward backward map (FBM), secondary map (SM) and tertiary map (TM), of using the GEMs to translate ICD-9-CM to ICD-10-CM codes used in phenotype definitions and quality measures. There is a clear difference in performance between the methods, and each method has its strengths and weaknesses. The least aggressive map (SFM) makes use of the forward GEM only. It consists of maps from 13,409 ICD-9-CM codes (92% of 14,567 available ICD-9-CM codes) to 16,614 ICD-10-CM codes (24% of 69,823 available ICD-10-CM codes). Since the forward map only covers a small percentage of the codes on the ICD-10-CM side, many ICD-10-CM codes are not reachable by the forward map. This explains the low recall of the SFM. The FBM combines the forward map with the backward map used in the reverse direction. Altogether, the FBM covers 13,478 (93%) of ICD-9-CM codes and 69,154 (99%) of ICD-10-CM codes.
The SM casts the net wider by identifying ICD-10-CM targets that are not directly mapped to the primary ICD-9-CM code in the GEMs, but are related through a secondary ICD-9-CM code. While this does not increase the overall coverage of ICD-9-CM or ICD-10-CM codes, it identifies additional map targets that would otherwise be missed. For example, if we start from the ICD-9-CM code 648.82 Abnormal glucose tolerance of mother, delivered, with mention of postpartum complication, the FBM finds the ICD-10-CM code O99.815 Abnormal glucose complicating the puerperium. However, since O99.815 is also a map target for 648.84 Abnormal glucose tolerance of mother, postpartum condition or complication, this makes 648.84 a secondary ICD-9-CM code for 648.82, which in turn leads to 3 secondary ICD-10-CM targets: O24.430 Gestational diabetes mellitus in the puerperium, diet controlled, O24.434 Gestational diabetes mellitus in the puerperium, insulin controlled, O24.439 Gestational diabetes mellitus in the puerperium, unspecified control, all valid targets for the primary ICD-9-CM code (figure 2).

Figure 2. Example of finding additional map targets through secondary ICD-9-CM and ICD-10-CM codes

However, the SM also brings in irrelevant map targets. This often happens in conditions pertaining to more than one disease. For example, starting from the ICD-9-CM code 716.80 Other specified arthropathy, site unspecified the FBM finds E08.618 Diabetes mellitus due to underlying condition with other diabetic arthropathy as one of the targets. E08.618 leads to 249.80 Secondary diabetes mellitus with other specified manifestations, not stated as uncontrolled, or unspecified as a secondary ICD-9-CM code, which further leads to ICD-10-CM targets like E10.621 Type 1 diabetes mellitus with foot ulcer, totally unrelated to arthropathy.

Judging from the F-1 score (a balanced measure between recall and precision), the FBM is the best performing in both scenarios. In the quality measure scenario, the FBM performs particularly well in smaller (< 100 codes) value sets. It seems that there is not much to gain in going to the more aggressive maps since the increase in recall is minimal, while precision suffers. In the phenotype definition case, the FBM also has the best F-1 score but the margin over the more aggressive maps is much smaller. From our limited sample, it seems that the FBM works better with well-defined, homogeneous conditions (e.g., colorectal cancer) as compared to less well-defined, heterogeneous conditions (e.g. chronic pain). In the latter situation, it may be worthwhile to adopt a more aggressive mapping approach (such as SM) because of the considerable improvement in recall (e.g., for chronic pain, recall increases from 0.76 to 0.96). The downside of doing that will be the increase of workload for expert reviewers, since the number of target codes almost doubles from FBM to SM. Some relief is offered in that the number of codes to be reviewed can be significantly reduced (by 72% in our study) by rolling up the codes to their parents.

Due to the large volume of ICD-9-CM codes used in research, quality measure and other use cases, transforming them to ICD-10-CM is an enormous task. Automated methods of translation will be a welcome help. Automatic translation is also likely to improve consistency in code selection. For the quality measure value sets, automatic translation can help to create ICD-10-CM code lists based on the ICD-9-CM codes, if they have not been dually-coded already. Otherwise, it can help in the quality assurance of existing or future dually-coded value sets. However, as illustrated in this study, the results of automatic translation are not perfect, and validation by human review is still necessary. Even with this limitation, automated translation will reduce the review workload by helping to identify relevant codes and exclude irrelevant ones.
There are potential ways to improve the performance of the maps that are worth exploring in future. There is additional information contained in the GEMs, such as flags for approximate or exact maps, and indicators of combination codes. The added information can possibly be exploited to refine the mapping algorithms. Another possible strategy is chapter-level refinement. Boyd et al. showed that the mapping relationships for codes from different ICD-9-CM chapters varied considerably.[15] This is because the difference between ICD-9-CM and ICD-10-CM is not uniform across all medical specialties. Chapters that do not change radically may require a less aggressive mapping approach. Outside the use of GEMs, two additional mapping resources may be worth considering. Firstly, there are two published maps from SNOMED CT to ICD-9-CM and from SNOMED CT to ICD-10-CM. It may be possible to map from ICD-9-CM to ICD-10-CM using SNOMED CT as an intermediary. In fact, SNOMED CT should be considered an alternative coding system in phenotype definition. This is already happening in quality measurement as many value sets are already defined in SNOMED CT, which is a better terminology for clinical content than ICD.[27] Also, with the Meaningful Use initiative, SNOMED CT codes will become more ubiquitous. Secondly, the UMLS is another resource commonly used for inter-terminology mapping. Mapping relations between ICD-9-CM and ICD-10-CM can be discovered by exploring the synonymy and other relationships within the UMLS. This can be used to corroborate or supplement the maps derived from the GEMs.

In addition to creating ICD-10-CM phenotype definitions from scratch, we have also explored the possibility that some quality measure value sets can be re-used for phenotype definition, because there is considerable similarity between a phenotype definition and the criterion to define a specific patient population for quality measurement. Since hundreds of value sets are already defined in ICD-10-CM, this possibility of re-use is worth considering. The challenge is how to identify a match between phenotype definitions and quality measure value sets. We have experimented with the use of the similarity score (Jaccard coefficient) between the ICD-9-CM codes in the phenotype definitions and the value sets. Using a similarity score of 0.9 as the cut-off, we found one value set for “Malignant neoplasm of colon” which matched the phenotype definition for “Colon Cancer”. All 13 ICD-10-CM codes in the value set were evaluated as correct by clinical experts.

In the future, this kind of “cross-fertilization” between various secondary uses of clinical codes will become more important and will encourage healthcare organizations to participate in pragmatic trials and nationally coordinated biomedical and health services research, in projects such as HMORN and PCORnet. The Phenotype Knowledge Base (PheKB)[28] and other repositories of phenotypes, should consider partnerships with VSAC and investigate formal linkages between research phenotypes and quality measurement value sets. The use of common value sets for clinical research and quality measurement can enable the generation of evidence from healthcare organizations and facilitate the vision of learning healthcare.

We note the following limitations in our study. The Collaboratory demonstration projects we used were a convenience sample and are not representative of all pragmatic trials. The phenotype definitions in this study were developed to support a number of tasks for very specific research studies and might not be generalizable or appropriate for other research or quality measurement use cases related to those conditions. Further, the phenotype definitions have not been vetted as national standards. Reviews by clinical experts have not been independently corroborated.

**Conclusion**

Although national reference mappings and tools exist to support ICD-9-CM to ICD-10-CM conversion, there are different approaches to using them. These different approaches yield different sets of ICD-10-CM codes, suggesting that migration of ICD-9-CM value sets to ICD-10-CM is vulnerable to variation unless specific approaches are specified. Each approach included some level of error or misclassification, and for some areas of research it will be important to validate GEMs mappings and how they are applied in the context of value set migrations. Variation in the migration of phenotype definitions can impact the consistency of definition of cohorts and data collection over time, and potentially impact study findings if not addressed.
Acknowledgements

This work was supported by the National Library of Medicine intramural research program, the National Institutes of Health (NIH) Common Fund, through a cooperative agreement (U54 AT007748) from the Office of Strategic Coordination within the Office of the NIH Director. The views presented here are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. The authors wish to thank researchers from demonstration pragmatic trials of the NIH Collaboratory as well as members of the Phenotype, Data Standards, and Data Quality Core for their support of this work. We are grateful to the following projects for developing their phenotypes and sharing them for this study: Strategies and Opportunities to Stop Colorectal Cancer (STOP CRC, Gloria Coronado, PI; NIH UH3CA188640), Pragmatic Trial of Population-Based Programs to Prevent Suicide Attempt (Gregory Simon, PI; NIH 4UH3MH007755-02), and Collaborative Care for Chronic Pain in Primary Care (PPACT, Lynn DeBar, PI; NIH 4UH3NS088731-02). Thanks also to Steve Emrick of NLM for pulling data from VSAC.

References