The discriminatory cost of ICD-10-CM transition between clinical specialties: metrics, case study, and mitigating tools

Andrew D Boyd,1,2,3 Jianrong ‘John’ Li,1,2,4 Mike D Burton,1,2,4 Michael Jonen,2,6,7 Terry Vanden Hoek,2,6,7 Andrew D Boyd,1,2,3 Jianrong ‘John’ Li,1,2,4 Mike D Burton,1,2,4 Michael Jonen,2,6,7

ABSTRACT
Objective Applying the science of networks to quantify the discriminatory impact of the ICD-9-CM to ICD-10-CM transition between clinical specialties.

Materials and Methods Datasets were the Center for Medicaid and Medicare Services ICD-9-CM to ICD-10-CM mapping files, general equivalence mappings, and statewide Medicaid emergency department billing.

Results We identified five mapping motif categories: identity, class-to-subclass, subclass-to-class, convoluted, and no mapping. Convoluted mappings indicate that multiple ICD-9-CM and ICD-10-CM codes share complex, entangled, and non-reciprocal mappings. The proportions of convoluted diagnoses mappings (36% overall) range from 5% (hematology) to 60% (obstetrics and injuries). In a case study of 24,008 patient visits in 217 emergency departments, 27% of the costs are associated with convoluted diagnoses, with ‘abdominal pain’ and ‘gastroenteritis’ accounting for approximately 3.5%.

Discussion Previous qualitative studies report that administrators and clinicians are likely to be challenged in understanding and managing their practice because of the ICD-10-CM transition. We substantiate the complexity of this transition with a thorough quantitative summary per clinical specialty.

Conclusions Post-transition, successful management of frequent diseases with convoluted mapping network patterns is critical. The http://lussierlab.org/transition-to-ICD10CM web portal provides insight in linking onerous diagnoses arising from mappings between ICD-9-CM and ICD-10-CM.

INTRODUCTION
The World Health Organization (WHO) released the International Classification of Diseases V10 (ICD-10) in 1990. While the rest of the world transitioned to ICD-10 (~14,000 codes) in the late 1990s, the USA will be transitioning from International Classification of Disease 9th Revision Clinical Modification (ICD-9-CM; 14,567 codes) to ICD-10-CM (~68,000 codes) as of 1 October 2014. Using the Center for Medicare and Medicaid Services (CMS) mapping tables, the American Medical Association (AMA) predicts implementation costs of US$83,000 to US$2.7 million per practice.1

Fundamentally, changing the controlled billing terminology impacts our capacity to compare, contrast, manage, and plan future needs during the transition to the new coding set, ICD-10-CM. These concerns were also voiced when the US government transitioned from ICD-8, ICDA-8, and H-ICDA-2 in 1979.2

As encoding into these terminologies is usually performed manually or semi-automatically, there is a potential impact on the overall accuracy. The ICD-10-CM coding system contains three times the number of codes, which requires using an entirely new coding organization, or significantly restructuring the relationships between codes. In other words, memorized codes, training, and coding-support software need to start afresh. Some commercial software have been proposed to bridge this transition, but there are limited details on their capabilities.3 Training materials have been provided by a number of organizations. However, the material is either at the planning stage or more qualitative. Few provide specific analytic tools to identify high value challenges.3–5

We hypothesized that network models6 can without bias identify problematic ICD-9-CM to ICD-10-CM mapping patterns (mapping motifs) and quantify their proportions per clinical specialty. We further hypothesized that these mapping motifs can clarify and quantify the administrative and financial impact arising from the ICD-10-CM implementation in clinical datasets. In this report, we quantify unaddressed ambiguities and redundancies arising from mappings between ICD-9-CM and ICD-10-CM. We establish that the mappings of a high proportion of the ICD-9-CM to ICD-10-CM mappings are entangled in complex mapping motifs that have the potential to induce inaccuracies and reporting errors. Using a case study of emergency departments’ Medicaid data, we demonstrate how a substantial proportion of non-reciprocal or abstruse mappings have the potential to disrupt billing and clinical practice.

METHODS
An overview of the methodology appears in supplementary figure S1 (available online only). Data integration and analyses are detailed in sections A–E and table 1. The research project was approved by the University of Illinois Institutional Review Board (id#2012-0150).
Construction of bidirectional mapping network from unidirectional maps of CMS–GEM

CMS–general equivalence mappings (GEM) files provide distinct directional mapping tables from ICD-9-CM to ICD-10-CM and from ICD-10-CM to ICD-9-CM because the mappings are not necessarily reciprocal. From the CMS mapping tables described in table 1, we created a bipartite network consisting of two types of nodes (ICD-9-CM and ICD-10-CM codes) and their directed relationships (arrow pointing in the direction of the mapping) (figures 1 and 2A; tables 2 and 3). This was loaded as a large table in MySQL V5.0.18 (table 4). Of note, approximately 14 000 ICD-9-CM codes directionally map to only approximately 18 000 of the approximately 68 000 ICD-10-CM codes, while nearly all 68 000 ICD-10-CM codes map to ICD-9-CM codes. Therefore, the two mapping tables are required in order to query patient data coded in ICD-10-CM, and to compare these to ICD-9-CM coded data (table 1, 2012_I9gem.txt and 2012_I10gem.txt). In each CMS table, mapping is unidirectional, either ICD-9-CM to ICD-10-CM (table 1, 2012_I9gem.txt) or the reverse (table 1, 2012_I10gem.txt). In the CMS map of ICD-9-CM to ICD-10-CM maps (2012_I9gem.txt file), each row represents a single ICD-9-CM to ICD-10-CM mapping. We then calculate the number of distinct ICD-10-CM codes associated with each ICD-9-CM code and vice versa. This leads to a model of the cardinality of these relationships represented on the vertical axis in figure 1: as one-to-one (1→1), many-to-one (M→1), one-to-many (1→M), and none. We similarly modeled the CMS map of ICD-10-CM to ICD-9-CM (2012_I10gem.txt file) to produce the columns of figure 1. As the bidirectional mappings are not necessarily reciprocal, their complex combinations have been systematically detailed in figure 1, where each axis describes the type of unidirectional mapping (vertical axis: ICD-9-CM to ICD-10-CM; horizontal axis: ICD-10-CM to ICD-9-CM) and where each of their combinations generate a bidirectional mapping motif (cells of the matrix).

We further combined unidirectional mappings from ICD-9-CM to ICD-10-CM to those of ICD-10-CM to ICD-9-CM, thus creating a complex bidirectional network. The complex network of bidirectional mapping is illustrated in figure 2A using Cytoscape V2.8, where the blue nodes are ICD-9-CM codes, the purple nodes are ICD-10-CM codes, and the arrows represent mapping between two codes.

Decomposition of the bidirectional mapping network in bidirectional mapping motifs and defining bounded versus unbounded mapping motifs

The combination of these two directions of mapping (ICD-9-CM to ICD-10-CM and vice versa) can be synthetized as bidirectional mapping motifs. For example, the simplest mapping motif corresponds to one ICD-9-CM mapped to one ICD-10-CM (figure 1 mauve background). On the other hand, some of the most complex mapping motifs arise as illustrated in the other combinations of rows and columns. We systematically indexed each type of mapping motif via specialized SQL queries for each matrix cell (provided with the database, table 4). Of note, a single ICD-9-CM code mapped to a single ICD-10-CM code from file 2012_I9gem.txt (table 1) is used as a computational seed (seed) for each mapping motif calculation with the exception of J-IV, for which there are no mappings to an ICD-10-CM code. This seed is presented in figure 1 as: (1) a single large blue circle (primary ICD-9-CM code); (2) a single large purple circle (primary ICD-10-CM code); and (3) the corresponding relationship (arrow) between the two. For each seed, the first step of the mapping motif construction consists in finding the additional non-primary ICD-9-CM codes that map to the primary ICD-10-CM codes using the 2012_I9gem.txt file, as well as in finding if they map to other non-primary ICD-10-CM codes (defined here as secondary ICD-9-CM codes and represented by small purple circles in figure 1). In addition, non-primary ICD-10-CM codes targeted by the primary ICD-9-CM code of the seed are discovered at this step (these are defined as secondary ICD-10-CM codes). Of note, there may be no mapping to an ICD-10-CM code for a primary ICD-9-CM code. The second step consists in mapping the primary ICD-10-CM code of the seed, as well as the secondary ICD-10-CM codes (discovered in the first step) back to the ICD-9-CM code using the 2012_I10gem.txt file. Within this step, we also identify all non-primary ICD-10-CM codes mapping to the primary ICD-9-CM code (these are also defined as secondary ICD-10-CM codes). These ICD-10-CM to ICD-9-CM maps can point to the primary ICD-9-CM code of the seed and/or additional non-primary ICD-9-CM codes (also defined as secondary ICD-9-CM codes represented by small blue circles in figure 1). The third and fourth steps consist in repeating steps one and two to determine whether or not the mapping motif identified in the first two steps is limited to these relationships (bounded mapping motifs) or the motif keeps propagating in the network (unbounded mapping motifs represented by dashed lines pointing out of the mapping motif in figure 1). Of note, many ICD-10-CM codes are not targeted by the computational seeds that originate from the 2012_I9gem.txt maps. An additional step consists in identifying ICD-10-CM codes with no mapping to ICD-9-CM codes in the 2012_I10gem.txt file. Each cell in the figure 1 matrix corresponds to one mapping motif. Our analysis identified 37 distinct mapping motifs. We then quantified the number of distinct seed ICD-9-CM codes in each of the remaining mapping motifs, and organized the results as quartiles (figure 1 bar graphs).

Table 1 Datasets

<table>
<thead>
<tr>
<th>Descriptions</th>
<th>Abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-10-CM release (2012) release</td>
<td>ICD-10-CM</td>
</tr>
<tr>
<td>2010 Emergency departments statewide Medicaid billing data for all patients with University of Illinois as primary home; 24 008 patient visits in 217 emergency departments</td>
<td>IHC–ED</td>
</tr>
</tbody>
</table>

Three datasets were used. Twenty-two per cent of the Illinois Health Connect, Emergency Department (IHC–ED) care was delivered at University of Illinois Hospital, and the remainder of the data were generated from 217 other facilities. An expert curator reviewed 100 randomly selected Center for Medicaid and Medicare Services (CMS)–general equivalence mappings (GEM) maps and observed one error (95% CI 0.2% to 5.0% precision).
An alternative approach, addressed briefly in the Results and Discussion sections, consists in seeding the relationship using the ICD-9-CM to ICD-10-CM mapping rather than the ICD-9-CM to ICD-10-CM mapping.

Organization of mapping motifs into mapping categories and complexity

In order to provide reports to a non-informatician audience (eg, clinicians and administrators) (figures 1 and 2B, table 4), we aggregate the mapping motifs further into five bidirectional mapping categories: identity (mauve background; figure 1); class-to-subclass (blue background; figure 1); subclass-to-class (yellow background; figure 1); convoluted (pink background, dashed arrows; majority of matrix cells); no mapping in either direction (gray background; figures 1 and 2B, examples of motifs). In figure 2B, the percentage of distinct primary ICD-9-CM (seed of each distinct motif) is reported for each mapping category (when an ICD-9-CM code is primary in multiple motifs, it is counted in the most complex motif, thus each ICD-9-CM code is counted only once). By definition, the identity, class-to-subclass, subclass-to-class, and no mapping categories are composed exclusively of bounded motifs and are thus...
bounded mapping categories (table 2). The convoluted category is defined by exclusion as motifs that are more complex than the four other categories. Furthermore, the convoluted category is the only one comprising unbounded mapping motifs (represented with dashed lines in figure 1). However, some bounded mapping motifs may be the recipient of mappings (see Methods section: Calculation of mapping motifs’ entanglement).

In the tables and figures, each ICD-9-CM code is counted only once according to the highest complexity of its associated mapping motif category (table 2, definition of complexity). Indeed, an ICD-9-CM code may be considered primary in multiple seeds, each of which has an associated mapping motif classified in one of five mapping motif categories.

Calculation of mapping motifs’ entanglement—a higher level of complexity

Entanglement between mapping motifs occurs when either mapping motifs are unbounded and point into other motifs or when other mapping motifs point into a bounded mapping motif (not represented in figure 1 for simplicity) (table 3). An example of entanglement between a bounded mapping motif and an unbounded one is provided in the supplementary methods (available online only). Therefore, unentangled mapping motifs provide straightforward transitions from ICD-9-CM to ICD-10-CM because they are bounded (do not point to other motifs), and no other mapping motifs point to them.

Summarization of ICD-9-CM codes in clinical classes

Clinical classes were constructed from the topmost ICD-9-CM hierarchies, which served as a basis to calculate the impact of the mapping to ICD-10-CM codes across clinical specialties (figure 3). The range of ICD-9-CM codes is reported beside each clinical class (figure 3). Each ICD-9-CM code was previously assigned a corresponding mapping category (see Methods section: Organization of mapping motifs into mapping categories and complexity), as well as a clinical class, from which the proportions were calculated and shown as horizontal bars (figure 3; color coding legend). In addition, the total count of ICD-9-CM codes and of ICD-10-CM codes are reported, as well as the ratio of the count of ICD-10-CM to ICD-9-CM to determine the importance of the change between versions of ICD. All novel ICD-10-CM clinical classes were manually mapped to their equivalent ICD-9-CM clinical classes for comparative purposes.

Case study calculations

All primary ICD-9-CM codes of emergency department encounters (Illinois Health Connect, emergency department dataset, table 1) were tallied and assigned to a mapping category of identity, class-to-subclass, subclass-to-class, convoluted, or no mapping (Methods section: Organization of mapping motifs into mapping categories and complexity) (figure 4). By design, for simplicity of the network and interpretation, no secondary ICD-9-CM codes associated with encounters were included.
As of 2013, we had a complex mapping network of approximately 6000 nodes (a graph whose nodes are clustered in two disjoint sets (in this manuscript the sets are ICD-9-CM and ICD-10-CM nodes). Every relationship connects one node of a set with one of the other set.

In this manuscript, a computational seed corresponds to a single ICD-9-CM code mapped to a single ICD-10-CM code in the 2012_I9gem.txt file. It is used as an input for each calculation of the motifs.

In the 2012_I9gem.txt file, it is used as an input for each calculation of the motifs.

Network analysis
Bounded/unbounded mapping motif
A bounded mapping motif is a motif from which all the relationships originating from it are constrained to the motif. Conversely, in an unbounded mapping motif, the relationships propagate in the network, out of the motif.

Complexity, mapping complexity
In the context of the clinic–administrative transition to ICD-CM-10, we arbitrarily defined an ordinal scale of complexity for mapping motif categories; from less to more complex: identity, one-to-many, many-to-one, convoluted, or no mapping.

Entanglement
A relationship is a mapping between two nodes (ie, two ICD codes).

Descriptive statistics of the ICD-9-CM to ICD-10-CM mapping network
In summary, we simplified a complex mapping network of approximately 80 000 ICD codes and approximately 100 000 mappings to a network of five types of mapping categories of approximately 14 000 motifs and approximately 6000 relationships. Only approximately 60% of the motifs are easily understood (untangled; tables 3 and 4).

The entire bidirectional mapping network is composed of 23 912 mappings of 14 567 ICD-9-CM to 16 604 ICD-10-CM codes and 78 840 converse mappings from 69 833 ICD-10-CM to 11 603 ICD-9-CM codes, of which only 4123 are reciprocal (figure 2A; also available as a scalable version, table 4). Thirty-seven distinct mapping motifs were predicted from the systematic combination of unidirectional mappings from ICD-9-CM to ICD-10-CM and the converse map, but only 28 mapping motifs contained actual mappings (figure 1; table 3).
database and SQL queries provided—see middle line of table 4). Furthermore, these mapping motifs could be synthetized as five mapping categories for which the proportion of associated distinct motifs are described in figure 2B (the number of primary ICD-9-CM codes represents each motif, figure 1, Methods). We first report that convoluted motifs account for 36% of the network, with a potential impact on transitioning well-defined clinical conditions and their management (figure 2B). Forty-two per cent (6158) of all ICD-9-CM codes are entangled in more than one mapping motif (table 3).

Of note, an alternative method to calculate the motifs could proceed using a seed relationship starting from ICD-10-CM and mapped to ICD-9-CM. The network of figure 2B remains the same. Only the proportion of motifs would differ in figure 1, but the matrix remains with the same axes and cells (not shown). The corresponding proportions of mapping categories are: 0.3% identity, 1% class-to-subclass, 10% subclass-to-class, 87% convoluted, and 1% no mapping (reported from the count of ICD-10-CM codes).

Impact of mapping motifs on clinical specialties

By distributing these mapping motifs into clinical classes, we have estimated the proportion of increasingly more complex mapping (figure 3). From the proportion of convoluted mapping motifs (figure 3, pink bars), we determined that hematology and oncology are poised for easy transition, while

![Clinical Classes (ICD-9-CM Range) vs Percentage of Mapping Category](image)

**Figure 3** Discrimination by clinical specialty. Furthermore, clinical specialty is unequally impacted as shown with the percentage of ICD-9-CM codes per mapping category (color coding of the bars from figure 2B, column 5). Clinical classes with a larger proportion of convoluted network motifs and higher ICD-10-CM to ICD-9-CM codes ratios are most likely to be affected by the transition. Mapping categories range from simple (identity) to convoluted, and are used as a proxy to estimate the impact of ICD-10-CM transition to clinical practice. Convoluted and no mapping will incur disproportionally more costs than simple motifs of mappings due to the inability to compare clinical practice before and after transition using ICD codes. In addition, a ratio was calculated comparing the number of total codes per clinical class (figure 3, rightmost column [#ICD-10-CM]/[#ICD-9-CM]). ‘Injury and poisoning’s’ outstandingly high ratio is highlighted in yellow.

<table>
<thead>
<tr>
<th>Resource sharing work product</th>
<th>Use case or targeted audience</th>
<th>Description or content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tables of mapping motifs and categories (.xls format)</td>
<td>Rapid reuse in software developed by health information technologists and informaticians.</td>
<td><a href="http://lussierlab.org/transition-to-ICD10CM/ICD-9%E2%80%9310-Transl-Cat.xlsx">http://lussierlab.org/transition-to-ICD10CM/ICD-9–10-Transl-Cat.xlsx</a>, eight columns headers</td>
</tr>
<tr>
<td>SQL database of mapping motifs and categories</td>
<td>Lookup of SQL queries and specific results by health system analysts strategically to improve health system operations and plan transition to ICD-10-CM.</td>
<td><a href="http://lussierlab.org/publication/Motif_table_SQLcode/DB">http://lussierlab.org/publication/Motif_table_SQLcode/DB</a> name, 38 distinct queries, one table, 324913 rows, five columns</td>
</tr>
</tbody>
</table>

Available at http://lussierlab.org/transition-to-ICD10CM.
obstetrics, psychiatry, and emergency medicine (poisoning) will be among the most challenged. Furthermore, 42% of infectious
disease code mappings remain convoluted, which will impact
most specialties. In addition, harder to transition ICD-10-CM
to ICD-9-CM code ratios greater than five are found in muscu-
loskeletal, injury, and poisoning clinical classes (figure 3, right
columns).

Shared resources
We computed ICD-9-CM codes, their mapping category, their
corresponding ICD-10-CM codes, and their unique subnetwork
identifier (table 4, second line). We also created a web portal to
help mitigate the mapping challenges for IT personnel, clini-
cians, and administrators (table 4, last line). On the web portal,
the mapping from ICD-9-CM to ICD-10-CM codes and
inversely is provided for user-defined lists of ICD-9-CM codes
with options for a tabular output (text files) and a dynamic
network visualization as a web portal.

Case study
In a case study of 24 008 encounters from 14 472 patients of
217 emergency departments in Illinois for calendar year 2010
(table 1, Illinois Health Connect, emergency department), we
calculated the impact of conversion to ICD-10-CM codes on
the visits and costs (figure 4A). A total of 59 846 ICD-9-CM
codes was affiliated with these encounters. On average, 27% of
the costs are attributed to ICD-9-CM codes for which the asso-
ciated mapping motifs to ICD-10-CM are convoluted. As illus-
trated in figure 4B, abdominal pain displays a cost of over US
$500 000/year (1.8% of emergency department billing in
Illinois) mapping to many ICD-10-CM codes and back to other
ICD-9-CM codes. Gastroenteritis is similarly convoluted (1.7%
of emergency department billing, figure 4C).

DISCUSSION
By using exclusively CMS data for determining the complexity
of mappings, this study was designed with simplification and
clarification of this complexity using network topology trans-
formation from ICD-level relationship to motif level relation-
ships, and reports straightforward metrics to facilitate the
interpretation of results by a broader community of administra-
tors and clinicians.

Contribution of network metrics to ‘change management’ of
terminologies
Full network analyses have yielded valuable insight into the plei-
ropy of genes between different diseases and have been applied to clinical claims coded in ICD codes for discovery of
patterns. In addition, a number of simplified networks such as directed acyclic graph (DAG) theory-based approaches have

Figure 4  Case study: identifying ICD-10-CM conversion challenges in 24 000 clinical encounters in 217 emergency departments. (A) The
convoluted mapping categories correspond to approximately 27% of the emergency department (ED) costs, encounters and codes, increasing the
risk of inaccuracies and errors and has significant implications on the data reliability pre and post-ICD-10-CM transition; 31% of the billed ED codes
were convoluted and corresponded to 28% of visits and 27% of costs, while 56% of codes were the less complex mapping motifs (blue and purple)
which correspond to 57% of encounters and 60% of costs. Interestingly, there was a 3.6% decrease of ED payments for encounters coding to
convoluted mapping category and an increase of 5.2% for those associated with less complex mapping categories. There is no inherent
inconsistency of the payment variations because complexity of mapping from ICD-9-CM to ICD-10-CM is not associated with the amounts of
diagnoses payments. (B) Example of convoluted mapping in the ED: ‘Abdominal pain’ with associated cost data. Of note, Center for Medicare and
Medicaid Services mapping confounds mappings of male and female genital symptoms (ICD-9-CM) with abdominal pain location (ICD-10-CM).
Post-transition, gender-specific information will be required in addition to the ICD codes for inventory management of speculum. (C) Example of
convoluted mapping in the ED: ‘diarrhea’ and ‘non-infection gastroenteritis’ are confounded in ICD-10-CM with implication for infectious disease
protocols and inventories (eg, culture sampling, disposable isolation supplies).
been applied to the hierarchical system of a controlled termin-
ology\textsuperscript{12} (eg, segmentation service, partition tools, etc).
However, these systems were not designed for full graph ana-
lyses or bipartite graphs. Furthermore, change management
approaches to controlled terminologies, developed by clinical
informaticians and ontologists, have been focused on semantics
and DAG or tree-like hierarchies within a version of a termin-
ology\textsuperscript{5} \textsuperscript{13} \textsuperscript{14} \textsuperscript{15}. The bipartite network revealed by the transition
from ICD-9-CM codes to ICD-10-CM codes requires a different
paradigm, as it cannot be simplified to a DAG\textsuperscript{16} nor defined
exclusively with the desiderata of controlled terminology\textsuperscript{17}. Furthermore, the well-established formalisms of ambiguity and
redundancy\textsuperscript{14} of terms within a terminology are insufficient to
describe unbounded and convoluted motifs. Indeed, the
class-to-subclass mapping category (23\% of ICD-9-CM codes,
figure 2B) can be viewed as both a mapping to potentially
redundant ICD-10-CM codes for an ICD-9-CM code, as well as an
ambiguous ICD-9-CM code disambiguated in multiple
coding to ICD-10-CM. The converse applies to the
subclass-to-class mapping category (12\% of ICD-9-CM codes,
figure 2B). While a focus in semantics is sufficient to address a
small number of changes between two updates of a terminology,
we show here how the significant changes imparted by the
ICD-9-CM to ICD-10-CM transition require additional struc-
tural metrics that characterize the complexity of bipartite net-
works of mappings. In our framework, the cardinality of
relationships is initially described according to the direction of
mapping (see Methods section: Construction of bidirectional
mapping network from unidirectional maps of CMS–GEM):
ICD-9-CM→ICD-10-CM and then ICD-10-CM→ICD-9-CM.
Reciprocal cardinalities are 1→M and 1←M of which a subset is
bounded and labeled as the ‘class-to-subclass’ category. Similar
reciprocal cardinalities are observed for the bounded mapping
categories labeled as ‘identity’ and ‘subclass-to-class’. The com-
bination of two non-reciprocal cardinalities produces exclusively
unbounded motifs: M→1 and 1→M, as well as 1←M and
M←1. This principle has a practical application: one can deter-
mine a priori that some pairing of cardinalities for an
ICD-9-CM code will obligate unbounded motifs. As cardina-
lities can be calculated in simple tables, there is no need to con-
struct or analyze a network to determine that a subset of
unbounded motifs would arise from combinations of mappings
forward and back for each ICD-9-CM code. It follows that
unbounded motifs correspond to unconstrained definitions or
undetermined meanings in semantics. This network composi-
tion principle is also scalable to the translation of terms
between any two terminologies (eg, applicable to the unified
medical language system).\textsuperscript{15}

\textbf{Lessons learned from other ICD transitions and our study}
The WHO ICD-9 (7000 terms) transition to ICD-10 (14 400
terms) met with the following challenges. In a Swiss analysis,
co-morbidity coded in the simpler ICD-10 that required 5 years
of coding sensitivity (recall), improved from 37\% to 43\% using
detailed chart abstraction, which was attributable to the coders’
‘learning curve’.\textsuperscript{18} Of note, the authors do not mention
pre-ICD-10 coding accuracies. Thirty-two diagnoses assessed
from billing data in another study comprising 4008 randomly
selected charts were re-coded in ICD-9-CM and compared to
the billed ICD-10-CA (Canadian enhancement to ICD-10). The
authors report a low sensitivity for all conditions in both coding
systems (9–72\%), worsening in seven diagnoses in
ICD-10-CA.\textsuperscript{19} Within the field of the Centers for Disease
Control and Prevention (CDC) public health in the USA, over
two million decedents were coded in both the WHO of ICD-9
and ICD-10 (6969 and 14 199 codes, respectively).\textsuperscript{20} The
authors observe inconsistencies in outcomes when coded as
ICD-9 versus ICD-10, with sensitivity as low as 26\% for some
categories of death. They conclude that there is a substantial
impact of this transition on relative risk estimates.\textsuperscript{20} They rec-
ommend recording cause of death in ICD-9 to avoid bias during
the transition to ICD-10. Based on 1852 671 individuals
recoded from ICD-9 to ICD-10 from the 1996 national vital
statistics reports of the CDC, substantial discrepancies in death
were attributed to the differences in coding scheme.\textsuperscript{21} For
example, septicemia was 20\% more likely to be selected in
ICD-10 than in ICD-9, adding over 3000 additional cases.
Conversely, bronchitis was 60\% less likely to be selected in
ICD-10. From the study, the following diseases demonstrated
discontinuity: septicemia, influenza, pneumonia, Alzheimer’s
disease, nephritis, nephrotic syndrome, and nephrosis.

Unsurprisingly, our results corroborate these previous reports of
considerable disruption in reporting clinical data post-ICD transi-
tion. Indeed, the transition to ICD-10-CM is far more challenging
than those reported for ICD-10. Here, we substantiate that 36\%
of ICD-9-CM code mappings are convoluted and have no straight-
forward correspondence in ICD-10-CM (figures 1 and 2). Indeed,
the convoluted motifs are so complex that substantial discontinu-
ities in reporting patient diseases are expected. Furthermore, clin-
ical specialties will be affected unequally, some with a proportion
less than 25\% and some with a proportion over 75\%.

\textbf{Minimizing disruption in reporting diagnoses post-transition}
To create longitudinal reports from data coded in ICD-9-CM
and then ICD-10-CM, mapping maps will be required (a ‘cross-
walk’). The AMA recommends ‘the direction that you crosswalk
the data will depend on how much of the data is in one code set
or the other.’\textsuperscript{22} However, others and our team have shown such
unsophisticated mapping is likely to contribute to significant discontinuities in convoluted motifs. For 2014, as 9 months will
have been coded in ICD-9-CM before the transition, the AMA
states: ‘it will be easier to crosswalk the ICD-10 codes back to
ICD-9 in order to compare all of the data together.’\textsuperscript{22} However,
we have shown that the number of convoluted mappings when
seeding the mapping from ICD-10-CM increases to 87\% (see
Results, after table 3). With a quarter of the calendar year 2014
coded in ICD-10-CM, this disruption is likely to be substantial,
not to mention that the coding from ICD-10-CM to ICD-9-CM
has less than 1\% identity motifs and is missing 2964 ICD-9-CM
codes as target. With the AMA strategy, 21\% of ICD-9-CM
codes would contain not a single patient in the last quarter of
2014 (a trend to zero). Of note, Nadkarni and Darer\textsuperscript{23} have also
identified limitations with concept matching software to trans-
late ICD-9 to SNOMED. They recommend the use of ‘query
expansion strategies’. Here, we follow such advice and propose
a more comprehensive crosswalk involving an educated use of
bidirectional mappings and entanglement annotations. For
example, reports could be stratified into two parts until the
meaning of the new trends is understood: the 60\% ICD-9-CM
codes associated with non-entangled motifs would map without
discontinuity and should be immediately interpretable, the
remaining 40\% of ICD-9-CM codes require additional work in
future studies. For example, it would be useful to provide
coding policies that would allow including the parts of
ICD-10-CM codes involved in entangled patterns in a step-wise

incremental fashion in order to control the discontinuity in disease groups. However, this may go against government or AMA policies. An alternative straightforward approach could be to conduct double coding (ICD-9-CM and ICD-10-CM) for the entangled ICD-9-CM codes and compare motifs in ICD-9-CM and ICD-10-CM in the final reports of the medical system or clinics, such as graph-pruning strategies to subsets offering reasonable coverage.24 However, dual coding is cost-prohibitive as coding to ICD-10-CM codes may require additional patient information that is available in patient charts but unobtainable from the historical ICD-9-CM claims. To mitigate the costs of double billing, we provide web portal tools, files, and charts to assess the risk profile per clinical condition, and to identify minimally affected ICD-9-CM codes (eg, transition motifs of identity or class-to-subclass; tables 2 and 3 and figure 3).

Future studies and limitations

In future work, we plan to report the metrics of particularly intricate ICD-9-CM and ICD-10-CM mapping motifs using additional properties such as centrality. The motifs could have been reported differently. For example, we could have reported the hubs and the bottlenecks of the networks;11 25 however, we believe the insight gathered would not necessarily have translated into action plans for the non-informaticians. We have provided tools to identify the problems; however, identifying strategies to mitigate the complexity of the mapping with practical solutions is likely to be more useful and is one of our next planned steps. This issue is particularly important because studies report that some ICD-10 codes are complex and difficult and frustrating to use.26

Additional analyses are warranted to understand the combination of primary and secondary ICD-9-CM codes to create a patient centered transition to ICD-10-CM coding; for simplicity, this additional analysis was omitted. We plan to provide guidelines to compare the data reported in ICD-10-CM with historical data reported in ICD-9-CM. In future studies, sophisticated semantics leveraging the unified medical language system could provide deeper insight into the transition to ICD-10-CM.15 27

CONCLUSION

The case study informs us that converting primary encounters with ICD-9-CM to ICD-10-CM codes will be convoluted for 28% of emergency room encounters (0–100%) depending on the clinical class), potentially impacting staff (utilization, workflow, division of labor, etc), supply management, and clinical revenue. The top two ICD-9-CM codes with convoluted mapping account for approximately 3.5% of Illinois emergency room visits (figure 4B). Comparable observations can be derived for all clinical departments, and are likely to vary considerably across clinical specialties and individual practices, justifying the requirement for customized mitigating tools. Furthermore, training of personnel and management resources of clinical specialties should focus on the frequently used and complex mapping motifs to ensure a successful transition to ICD-10-CM, which can readily be assessed via web portal tools.

Contributors

YAL, ADB, JJL, MI, and IZ conceived and designed the experiments. ADB, JIL, MDB, and RQL performed the experiments. ADB, JIL, MDB, MI, VI, IA, RQL, IL, NB, SBB, TVH, and YAL analyzed and evaluated the data. ADB, JIL, YAL, ADB, NB, and MI designed and evaluated the web portal. YAL, JIL, MB, NB, SBB, and TVH contributed reagents/materials/analysis tools. YAL, ADB, JIL, MI, IZ, and MG wrote the paper. YAL conceived and directed the project. YAL had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. ADB, JIL, MDB, and MI contributed equally

Funding

ADB and YAL are supported in part by the Center for Clinical and Translational Sciences of the University of Illinois (NIH 1UL1RR029879-01, NIH/NCATS UL1TR000050), the Institute for Translational Health Informatics of the University of Illinois at Chicago and the Office of the Vice-President for Health Affairs of the University of Illinois Hospital and Health Science System.

Competing interests

None.

Ethics approval

The case study was approved as a de-identified study exempt of consent by the Institutional Review Board of the University of Illinois at Chicago (IRB protocol #2012-0150).

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

Comprehensive computed motifs and data are available from the web portal: http://flussierlab.org/transition-to-ICD10CM.

Open Access

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES

23 Nadkarni PM, Darer JA. Migrating existing clinical content from ICD-9 to SNOMED. J Am Med Inform Assoc 2010;17:602–7.
Call for Papers

Special Issue on Big Data in Healthcare and Biomedical Research

The collection, annotation, organization, integration, analysis, and sharing of big, complex, and sensitive data from healthcare and biomedical research has recently become the emphasis of several government initiatives in the USA and abroad. While the informatics community has been dealing with many of these issues for several years, the dissemination of new approaches and solutions may be scattered in several different journals, conference proceedings, and web sites. The purpose of this special focus issue is to highlight the most innovative approaches and solutions (through Research and Applications articles, Brief Communications, and Case Reports), and place them in context of the work that has been previously published (through Reviews and Perspectives). User-friendly tutorials that help non-specialists better understand the methods commonly used in machine and statistical learning, as well as in integrative approaches related to common data models, ontologies and standards are also welcome.

Topics of Interest

Possible topics include, but are not limited to:

- Scalable systems for data integration across sites
- Standards-based representation and modeling of data and algorithms
- Information retrieval that combines literature and patient data sources
- Collaborative filtering
- Adaptive learning
- Distributed multiparty computation
- Evaluation strategies

Authors should make sure to place their work in the context of biomedical research or healthcare, and to carefully review the relevant literature. Research articles, case studies, and brief communications should describe clear evaluation strategies and quantitative or qualitative results, and discuss how results could be generalized to other settings. Preference will be given to systematic Reviews. Perspectives should provide consensus of a group of experts who are highly experienced in the topic, and should demonstrate command of the existing literature, particularly recent literature whose synthesis would be new to our readers. Open-source software code and data should be submitted, as well as data when appropriate.

Important Dates

July 1, 2013 Manuscript submission deadline
August 30, 2013 (expected) Initial decisions sent to authors
September 30, 2013 (expected) Revised manuscript submission deadline
November 30, 2013 (expected) Final decisions sent to authors

Submission and Peer Review Process

To ensure consideration in the special issue, authors should note in a cover letter that their submission is for the “Special Issue on Big Data”. Detailed information for online submission to JAMIA is available via http://jamia.bmj.com. All manuscripts will be subject to the rigorous JAMIA peer-review process. Manuscripts that are considered within scope and meet quality expectations will be typically reviewed by two experts for scientific merit. Assistance of a native English speaker is highly recommended prior to submission. Authors should format and structure their manuscripts according to the guidelines specified at: http://jamia.bmj.com/site/about/guidelines.xhtml. Accepted articles may appear in print or in an online JAMIA issue.

Questions Regarding the Issue

Please direct any questions regarding the special issue or submissions to machado@ucsd.edu