AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Research and Applications

The new International Classification of Diseases 11th edition: a comparative analysis with ICD-10 and ICD-10-CM

Kin Wah Fung, Julia Xu, and Olivier Bodenreider (ID)

Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

Corresponding Author: Kin Wah Fung, MD, MS, MA, Lister Hill National Center for Biomedical Communications, Building 38A, Rm9S918, MSC-3826, National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894, USA (kwfung@nlm.nih.-gov)

ABSTRACT

Objective: To study the newly adopted International Classification of Diseases 11th revision (ICD-11) and compare it to the International Classification of Diseases 10th revision (ICD-10) and International Classification of Diseases 10th revision-Clinical Modification (ICD-10-CM).

Materials and Methods: : Data files and maps were downloaded from the World Health Organization (WHO) website and through the application programming interfaces. A round trip method based on the WHO maps was used to identify equivalent codes between ICD-10 and ICD-11, which were validated by limited manual review. ICD-11 terms were mapped to ICD-10-CM through normalized lexical mapping. ICD-10-CM codes in 6 disease areas were also manually recoded in ICD-11.

Results: Excluding the chapters for traditional medicine, functioning assessment, and extension codes for postcoordination, ICD-11 has 14 622 leaf codes (codes that can be used in coding) compared to ICD-10 and ICD-10-CM, which has 10 607 and 71 932 leaf codes, respectively. We identified 4037 pairs of ICD-10 and ICD-11 codes that were equivalent (estimated accuracy of 96%) by our round trip method. Lexical matching between ICD-11 and ICD-10-CM identified 4059 pairs of possibly equivalent codes. Manual recoding showed that 60% of a sample of 388 ICD-10-CM codes could be fully represented in ICD-11 by precoordinated codes or postcoordination.

Conclusion: In ICD-11, there is a moderate increase in the number of codes over ICD-10. With postcoordination, it is possible to fully represent the meaning of a high proportion of ICD-10-CM codes, especially with the addition of a limited number of extension codes.

Key words: ICD-11, ICD-10, ICD-10-CM, controlled medical vocabularies, medical terminologies

INTRODUCTION

The International Classification of Diseases (ICD) can be traced back over a century ago to the International List of Causes of Death (ICD-1) adopted by the International Statistical Institute in 1900 in Paris.[1,2] The classification was subsequently updated every decade. The update task was passed to the World Health Organization (WHO) in 1946, and the classification was renamed International Classification of

Diseases, Injuries, and Causes of Death to serve as the foundation for worldwide health trends and statistics. The update interval has lengthened considerably after ICD-9. ICD-10 was adopted in 1992, 17 years after ICD-9. The WHO started working on ICD-11 in 2007 with involvement of experts from over 90 countries. ICD-11 was adopted in May 2019 (27 years after ICD-10) by the World Health Assembly, to be effective for use from January 2022.[3–9] Over 2 dozen countries have

Table 1. Comparison of the Foundation Component and linearizations

Characteristic	Foundation component	Example	Linearizations	Example
Building block	Entity	Diaphragmatic hernia	Category	Diaphragmatic hernia
Identifier	URI	http://id.who.int/icd/entity/453532731	Code	DD50.0
Defining attributes	Description, body site, body system, causal mechanisms, synonyms, exclusions, signs and symptoms etc.	Description: A hernia occurs through the foramen in the diaphragm Synonyms: paraesophageal hernia, hiatus hernia, esophageal hiatus hernia, sliding hiatus hernia Exclusions: congenital diaphragmatic hernia, congenital hiatus hernia Body site: diaphragmatic structure (body structure), entire diaphragm (body structure)	Description, inclusions, exclusions	Description: A hernia occurs through the foramen in the diaphragm Inclusions: paraesophageal hernia Exclusions: Congenital diaphragmatic hernia (LB00.0), Congenital hiatus hernia (LB13.1)
Hierarchy	Multi-parenting	Parents: Non-abdominal wall hernia, Other diseases of the digestive system	Single-parenting	Parent: DD50 Non-abdominal wall hernia
Residual elements	None		Present	DD50.Y Other specified non-abdominal wall hernia, DD50.Z non-abdominal wall hernia, unspecified
Update frequency	Continuous		Periodic with official versioning	

Abbreviation: URI, universal resource identifier.

developed national extensions of ICD to suit their requirements. In the US, the Clinical Modification (CM) has been developed since ICD-9-CM to support morbidity coding for reimbursement and other purposes.[2,10] ICD-10-CM replaced ICD-9-CM in October 2015.

BACKGROUND

With every new ICD version, code syntax usually changes—presumably to avoid confusion with older versions. For example, the code for *Huntington disease* is G10 in ICD-10 and 8A01.10 in ICD-11. There is often expansion of the number of codes and some reorganization of the chapters. Apart from these usual changes, ICD-11 has 3 brand new features:[11,12]

1. Foundation Component. ICD-11 is built on an underlying knowledge base that holds all necessary information to generate the tabular list and alphabetical index for mortality and morbidity coding.[13] These derivatives are called "linearizations." It is also possible to generate alternative lists for different purposes (eg, specialty subsets, country specific modifications).

   The Foundation Component is a multidimensional collection of medical entities—diseases, disorders, injuries, external causes, signs, and symptoms. The entities are defined with attributes such as body site, body system, and causal mechanism. (Table 1) These entities are organized into hierarchies and multiparenting is allowed. When linearizations are derived from the Foundation Component, only single-parenting is allowed—an essential requirement in a statistical classification to avoid double counting. Categories in a linearization are derived from entities in the Foundation Component and are assigned ICD-11 codes. Not all entities acquire unique codes, as some entities may be merged into 1 category. Residual categories (eg, 'unspecified,' 'not elsewhere classified') are added to ensure that the categories are mutually exclusive and jointly exhaustive—another essential requirement of a statistical classification. The Foundation Component is updated in real time and linearizations are generated at fixed intervals (eg, yearly) and officially versioned.

2. Postcoordination. ICD-11 allows the combination of codes (called "cluster coding") to add additional detail to an existing code (called "stem code" or "precoordinated code"). Two kinds of postcoordination are allowed:

   a. Two or more stem codes (syntax: stemcode1/stemcode2/stemcode3, etc) for example, urinary tract infection due to extended spectrum beta-lactamase producing *Escherichia coli* = *GC08.0/MG50.27* (*GC08.0 Urinary tract infection, site not specified, due to Escherichia coli*; *MG50.27 Extended-spectrum beta-lactamase producing Escherichia coli*)

   b. Stem code(s) with 1 or more extension codes (syntax: stemcode1&extensioncode1&extensioncode2 etc.) for example, tuberculosis of prostate = *1B12.5&XA63E5* (*1B12.5 Tuberculosis of the genitourinary system*; *XA63E5 Prostate gland*).

3. Digital-friendly. Fully embracing the digital age, ICD-11 is accompanied by a host of online and digital resources. Online resources include browsers of the Foundation Component and various linearizations and a coding tool for the Mortality and Morbidity Statistics linearization (MMS).[14–16] Downloadable resources include maps between ICD-10 and ICD-11 and the MMS. Application programming interfaces (API) allow programmatic access to the Foundation Component, MMS, and ICD-10. There is also an online maintenance platform for collaborators in the update process.

We present a comparative analysis of ICD-11 in relation to ICD-10 and ICD-10-CM. Updating ICD to a new version is a nontrivial endeavor which incurs significant cost and has potential impact on longitudinal data comparability, as evidenced by various reports when the US moved from ICD-9-CM to ICD-10-CM.[17–26] The goal of the ICD-10 comparison is to provide a high-level view of the extent and pattern of changes. The comparison with ICD-10-CM is motivated by the possibility that the US could move from ICD-10-

CM directly to ICD-11 for morbidity coding, without creating a Clinical Modification. Before the WHO finalizes the licensing and copyright restrictions for national modifications, it is not clear whether ICD-11-CM can be created as usual. This study focuses on the differences between ICD-11 and its predecessors, highlighting the incremental changes as well as innovative features. It is not intended to be a comprehensive evaluation of ICD-11 as a medical terminology or its fitness for purpose. There have been studies on the differences between ICD-11 and ICD-10, but most are focused on specific disease areas (eg, mental health).[27,28] We believe that our study is the first broad-based comparison of ICD-11 to ICD-10 and ICD-10-CM.

MATERIALS AND METHODS

Data sources
We downloaded the following from the WHO ICD-11 website (Version 04/2019):

1. *Simple Tabulation* – ICD-11 codes, titles, and indexing terms in MMS
2. *MMS Linearization Tabulation* – similar to Simple Tabulation, with additional information (eg, kind of code [chapter, block, or category]) and depth in tree
3. *One Category ICD-10 to ICD-11 Map* – each ICD-10 code maps to only 1 ICD-11 code
4. *Multiple Categories ICD-10 to ICD-11 Map* – each ICD-10 code can map to multiple ICD-11 codes
5. *One Category ICD-11 to ICD-10 Map* – each ICD-11 code maps to only 1 ICD-10 code

We used the MMS browser and coding tool to look up individual codes. We used the API to collect additional information not in the downloadable files.

ICD-10 comparison
We focused on the first 25 chapters of the ICD-11 MMS linearization that aligned with the scope of ICD-10, using only precoordinated ICD-11 codes. We used the one-category ICD-10 to ICD-11 map to identify "chapter drift" (ie, codes moved to a chapter other than the main corresponding chapter). To quantify chapter drift, we defined a "chapter drift index" (CDI) for each ICD-11 chapter as the percentage of codes coming from ICD-10 chapters other than the main corresponding chapter. We identified equivalent codes between ICD-10 and ICD-11 by "round tripping," using the 2 one-category maps. We postulated that if an ICD-10 code mapped to a single ICD-11 code in the forward map, which mapped back to the same ICD-10 code in the backward map, then the 2 codes were likely equivalent. We manually reviewed some round trip maps.

ICD-10-CM comparison
Since no maps existed, we used 2 approaches to compare ICD-11 to ICD-10-CM, lexical matching and manual recoding. For lexical matching, we used the lexical tool LuiNorm from the Unified Medical Language System (UMLS) (2019 version) to normalize ICD-11 code names from chapters 1 to 25.[29,30] We matched the normalized names to UMLS concepts (version 2019AA) using the normalized English strings index (MRXNS_ENG).[31] Through the UMLS concepts, we matched to ICD-10-CM codes (2019 version). We ignored ICD-11 index terms and ICD-10-CM inclusion terms because they could be narrower in meaning. For example, the index terms for

Paratyphoid Fever included *Paratyphoid fever A* and *Paratyphoid fever B*. For manual recoding, we picked a convenient sample of 6 disease areas in ICD-10-CM that covered common conditions (diabetes, hypertension, pregnancy) and pathologies (infection, trauma, malignancy) and recoded them in ICD-11. For each ICD-10-CM code, we determined whether its meaning could be fully represented in ICD-11 with or without postcoordination. The recoding was done by 1 of the authors (JX, physician with extensive ICD knowledge).

RESULTS

ICD-10 comparison
Chapter structure, chapter drift and extent of change
ICD-11 had 28 chapters, 6 more than ICD-10. The last 3 chapters were outside the scope of ICD-10 and excluded from further analysis:

- *Chapter 26 Supplementary Chapter Traditional Medicine Conditions*
- *Chapter V Supplementary section for functioning assessment*
- *Chapter X Extension Codes* (for support of postcoordination)

  Among the first 25 chapters, 3 were new:

- *Chapter 4 Diseases of the immune system*
- *Chapter 7 Sleep-wake disorders*
- *Chapter 17 Conditions related to sexual health*

The other 22 chapters largely mirrored the chapters of ICD-10. However, some conditions could be moved to a chapter other than the main corresponding chapter (chapter drift). Figure 1 shows the degree of correspondence of codes by chapter. The rows are ICD-10 chapters and the columns are ICD-11 chapters. The number in each cell is the number of ICD-10 leaf codes in the one-category ICD-10 to ICD-11 map. Only leaf codes, which are the lowest level codes with no children, are allowed in coding. The largest numbers are found along the diagonal, meaning that the majority of codes remain in their main corresponding chapters. Three notable breaks in the diagonal pattern correspond to the new chapters 4, 7, and 17 (red arrows). Not surprisingly, many codes from the ICD-10 *Chapter III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism* end up in the new *Chapter 4 Diseases of the immune system*. The ICD-10 *Chapter V Mental and behavioral disorders* is the biggest contributor of codes to the new *Chapter 7 Sleep-wake disorders* and *Chapter 17 Conditions related to sexual health*.

We identified 7 ICD-11 chapters with CDI over 5% (Figure 1, last row). Among these were, not surprisingly, the 3 new chapters, since they did not correspond neatly to a single ICD-10 chapter (thus the need for a new chapter). The other 4 chapters were:

- *Chapter 1 Certain infectious or parasitic diseases*: some diseases used to be classified based on body location were now grouped under infectious diseases eg, *Bacterial meningitis* (previously under *Chapter VI Diseases of nervous system*), *Acute rheumatic myocarditis* (previously under *Chapter IX Diseases of circulatory system*), *Impetigo* (previously under *Chapter XII Diseases of skin and subcutaneous tissue*).
- *Chapter 8 Diseases of the nervous system*: much of the chapter drift was due to the movement of stroke, cerebral hemorrhage, and other cerebrovascular diseases (previously under *Chapter IX Diseases of the circulatory system*) to this chapter.
- *Chapter 14 Diseases of the skin*: some congenital conditions primarily affecting the skin (eg, *X-linked ichthyosis*, *Congenital leu-*

ICD-10	Ch1	Ch2	Ch3	Ch4	Ch5	Ch6	Ch7	Ch8	Ch9	Ch10	Ch11	Ch12	Ch13	Ch14	Ch15	Ch16	Ch17	Ch18	Ch19	Ch20	Ch21	Ch22	Ch23	Ch24	Ch25
I	725							3						1											
II		749	1	↓																					
III			104	53			↓														1				
IV	2			4	332		1	1				4		1		1					1				
V					370	8	6										21				1			1	
VI	16			2	2	3	7	285											1			16			
VII	1				1					248											1	4	3		
VIII											106											2			
IX	8			2	1			54				301		7							1	4			
X	8			1									221												
XI	8				1									405							4	10			
XII	18	3		2	2	1						6		253	2	2						7			
XIII	12		1	35	1			5		1				2	458		↓		1		1	23			
XIV	10	9												2		364	2	1			2	30			
XV																		434							
XVI								1				1		1						332					
XVII		1			1			2	1		1			14	1		6				589				
XVIII	4					1		4			1		1	4							309				
XIX				7				6			1			1							1	1212	23	1	
XX						5															10	2	1313		
XXI	1															1								620	
XXII	11																					19			
CDI	12%	2%	2%	50%	2%	3%	50%	22%	0%	1%	3%	2%	2%	9%	1%	2%	9%	0%	0%	2%	29%	0%	2%	0%	

Figure 1. Alignment of codes by chapter based on map from ICD-10 to ICD-11 (I to XXII: ICD-10 chapters, Ch1–Ch25: ICD-11 chapters, CDI: chapter drift index, see text for explanation).

konychia [previously under *Chapter XVII Congenital malformations, deformations and chromosomal abnormalities*]) were moved here.

- *Chapter 21 Symptoms, signs or clinical findings, not elsewhere classified*: some examples were *Cardiac arrest* (previously under *Chapter IX Diseases of circulatory system*), *Hematemesis* (previously under *Chapter XI Diseases of the digestive system*) and *Pain in joint* (previously under *Chapter XIII Diseases of the musculoskeletal system and connective tissue*). One possible reason for the chapter drift might be that these conditions could be caused by diseases outside the original ICD-10 chapter (eg, cardiac arrest could be due to diseases outside the circulatory system).

The last column was empty because *Chapter 25 Codes for special purposes* (similar to *chapter XXII* in ICD-10) was the placeholder for provisional codes to assign to new diseases of uncertain etiology. There were no maps to this chapter.

Table 2 shows the distribution of leaf codes by chapter. We merged *Chapter 3 Diseases of the blood or blood-forming organs* and *Chapter 4 Diseases of the immune system* to correspond to *Chapter III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism* in ICD-10. We excluded the new chapters 7 and 17, since they did not correspond to an ICD-10 chapter. Three chapters had the highest percentage of growth - *Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified* (217%), *Chapter III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism* (157%) and *Chapter VII Diseases of the eye and adnexa* (135%). However, the number of leaf codes did not fully reflect coverage and expressivity because of the possibility of postcoordination. Two chapters, *Chapter XIII Diseases of the musculoskeletal system or connective tissue* and *Chapter*

XX External causes of morbidity or mortality, actually had fewer codes in ICD-11 than ICD-10. With postcoordination, some precoordinated codes were no longer necessary. For example, there were 6 ICD-10 leaf codes for injury due to venomous animals (*X20 Contact with snakes and lizards, X21 Contact with venomous spiders, X22 Contact with scorpions, etc*). In ICD-11, there was only 1 leaf code *PA78 Unintentionally stung or envenomated by animal*, and the different animals could be postcoordinated by extension codes: *XE9H6 Venomous snake, XE6A7 Lizard, gecko, goanna, XE75L Spider and XE2EP Scorpion*, and many more.

Codes that have remained the same
There were altogether 14 622 ICD-11 leaf codes (chapters 1–25), a 38% increase over ICD-10 (10 607 leaf codes). By round tripping, we identified 4037 unique pairs of ICD-10 and ICD-11 leaf codes that were potentially equivalent. In 44% of the code pairs, the names of the codes were the same in ICD-11 and ICD-10. (Table 3, category A) We assumed these pairs to be truly equivalent and did no further review.

We reviewed 250 randomly selected code pairs from categories B, C, and D. Among the cases where the ICD-10 name was a substring of the ICD-11 name (category B), 88% were found to be equivalent. In many cases, the extra word in ICD-11 was "unspecified," which we ignored because it conferred no additional meaning. The remaining 12% were not equivalent (eg, *Central diabetes insipidus* and *Diabetes insipidus*), because the latter included nephrogenic diabetes insipidus. All the cases where the ICD-11 name was a substring of the ICD-10 name (category C) were found to be equivalent. The commonest extra word was "unspecified" or an eponym (eg, *Synovial cyst of popliteal space [Baker]*). In the remaining cases (category D), we found 93% of equivalence, such as *Candidiasis of vulva and vagina* and *Vulvovaginal candidosis*. Over-

Table 2. Extent of change by chapter

ICD-10 chapter		Corresponding ICD-11 chapter	ICD-10 leaf codes	ICD-11 leaf codes	% change
I	Certain infectious and parasitic diseases	1	783	835	6.6%
II	Neoplasms	2	759	1056	39.1%
III	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	3 & 4	165	424	157.0%
IV	Endocrine, nutritional and metabolic diseases	5	356	541	52.0%
V	Mental and behavioral disorders	6	407	718	76.4%
VI	Diseases of the nervous system	8	335	719	114.6%
VII	Diseases of the eye and adnexa	9	262	615	134.7%
VIII	Diseases of the ear and mastoid process	10	113	135	19.5%
IX	Diseases of the circulatory system	11	383	479	25.1%
X	Diseases of the respiratory system	12	234	296	26.5%
XI	Diseases of the digestive system	13	435	804	84.8%
XII	Diseases of the skin and subcutaneous tissue	14	342	615	79.8%
XIII	Diseases of the musculoskeletal system and connective tissue	15	545	364	−33.2%
XIV	Diseases of the genitourinary system	16	438	447	2.1%
XV	Pregnancy, childbirth, and the puerperium	18	434	437	0.7%
XVI	Certain conditions originating in the perinatal period	19	337	525	55.8%
XVII	Congenital malformations, deformities, and chromosomal abnormalities	20	619	1118	80.6%
XVIII	Symptoms, signs, and abnormal clinical and laboratory findings, not elsewhere classified	21	340	1078	217.1%
XIX	Injury, poisoning, and certain other consequences of external causes	22	1278	1668	30.5%
XX	External causes of morbidity and mortality	23	1372	842	−38.6%
XXI	Factors influencing health status and contact with health services	24	630	759	20.5%
XXII	Codes for special purposes	25	40	17	−57.5%
Overall			10 607	14 492*	36.6%*

*not including chapters 7 and 17.

Table 3. Candidate equivalent codes identified by round trip method

Category	Number of codes (%)	Examples		Manual review		
		ICD-10	ICD-11	Equivalent	Not equivalent	Total
A. Same name	1773 (44%)	A06.4 Amoebic liver abscess	1A36.10 Amoebic liver abscess	Not reviewed (assumed equivalent)		
B. ICD-11 name contains ICD-10 name	338 (8%)	A01.0 Typhoid fever	1A07.Z Typhoid fever, unspecified	29 (88%)	4 (12%)	33 (100%)
C. ICD-10 name contains ICD-11 name	146 (4%)	I73.1 Thromboangiitis obliterans [Buerger]	4A44.8 Thromboangiitis obliterans	16 (100%)	0 (0%)	16 (100%)
D. Others	1780 (44%)	Q69.2 Accessory toe(s)	LB78.3 Polydactyly of toes	187 (93%)	14 (7%)	201 (100%)
Total	4037 (100%)			232 (93%)	18 (7%)	250 (100%)

all, 93% of the reviewed cases were equivalent. If we projected these results to all the 4037 candidate equivalent pairs and assumed that all cases in which the ICD-10 and ICD-11 names were identical (category A) were equivalent, 96% of the candidate pairs would be truly equivalent. This confirmed the validity of the round trip method. Figure 2 shows the overlap between ICD-10 and ICD-11 based on equivalent codes identified by round tripping.

ICD-10-CM comparison

Lexical matching (to precoordinated codes only)

By normalized lexical matching through the UMLS, we managed to find 4059 pairs of matching ICD-11 and ICD-10-CM codes, covering 3294 unique ICD-11 codes (3211 leaf and 83 nonleaf) and 3985

unique ICD-10-CM codes (2366 leaf and 1619 nonleaf). The breakdown of these code pairs:

- 52 pairs: nonleaf codes in both ICD-10-CM and ICD-11
- 1596 pairs: ICD-11 leaf codes matched to ICD-10-CM nonleaf codes
- 66 pairs: ICD-11 nonleaf codes matched to ICD-10-CM leaf codes
- 2345 pairs: leaf codes in both ICD-10-CM and ICD-11

There were many more cases where an ICD-11 leaf code was matched to a nonleaf ICD-10-CM code compared to the other way around. This shows that ICD-10-CM is larger (71 932 vs 14 622 leaf codes) and more fine-grained than ICD-11. However, with postcoordination, it is possible to bridge some of these differences.

Manual recoding (to pre- and postcoordinated codes)
We selected 388 ICD-10-CM leaf codes from 6 disease areas—tuberculosis, skin cancer, diabetes mellitus (DM) type 2, hypertension, polyhydramnios, and fracture of thumb—and recoded them in ICD-11, using postcoordination when necessary. (Table 4)

a. Tuberculosis—Among the 51 codes from the block *A15—A19 Tuberculosis*, 23 could be fully represented by precoordinated ICD-11 codes. All remaining codes could be fully represented by postcoordination. For example, *A18.32 Tuberculous enteritis* could be recoded as *1B12.7 Tuberculosis of the digestive system & XA6452 Small intestine*

b. Skin cancer—There were 101 codes from the category *C44 Other and unspecified malignant neoplasm of skin*, of which only 5 could be fully represented by precoordinated codes. However, most of the remaining codes could be postcoordinated. For example, *C44.212 Basal cell carcinoma of skin of right ear and external auricular canal* could be recoded as *2C32.Z Basal cell carcinoma of skin, unspecified & XK9K Right & XA6ZY6 External Ear*. Four codes could not be fully represented (eg, *C44.81 Basal cell carcinoma of overlapping sites of skin*) because there was no ICD-11 extension code for 'overlapping sites of skin.'

c. Diabetes mellitus type 2—Among the 86 codes from the category *E11 Type 2 diabetes mellitus*, only 1 code had an equivalent precoordinated code. Sixty codes could be postcoordinated (eg, *E11.42 Type 2 diabetes mellitus with diabetic polyneuropathy* could be recoded as *5A11 Type 2 diabetes mellitus/8C03.0 Diabetic polyneuropathy*). Twenty-five codes could only be partially represented (eg, *E11.638 Type 2 diabetes mellitus with other oral complications*), because there was no ICD-11 code for 'other oral complications.'

d. Hypertension—there were 17 codes under the block *I10 – I16 Hypertensive diseases*, and 5 codes had precoordinated equiva-

lents. Nine codes could be postcoordinated (eg, *I11.0 Hypertensive heart disease with heart failure* could be recoded as *BA01 Hypertensive heart disease/BD1Z Heart failure, unspecified*). Three codes could only be partially represented (eg, *I15.0 Renovascular hypertension*) because of the lack of a code for 'renovascular disease.'

e. Polyhydramnios—there were 28 codes under *O40 Polyhydramnios* representing all possible combinations of trimester (first trimester, second trimester, third trimester, and unspecified trimester) and specific fetus affected in multiple pregnancy (eg, fetus 1, fetus 2). None of these codes could be fully represented because ICD-11 did not provide extension codes for trimester or specific fetus.

f. Fracture of thumb—there were 105 codes under the category *S62.5 Fracture of thumb* representing the combinations of 6 attributes: laterality (left, right, unspecified); location (proximal phalanx, distal phalanx, unspecified phalanx); type of fracture (open, closed); displacement (displaced, nondisplaced); healing (routine healing, delayed healing, nonunion, malunion); and episode of care (initial encounter, subsequent encounter, sequela). Postcoordination could represent all attributes except episode of care. As a result, none of the ICD-10-CM codes could be fully represented in ICD-11.

Overall, about 60% of ICD-10-CM codes we examined could be represented fully by pre- or postcoordinated ICD-11 codes. With the addition of 3 extension codes (for episode of care), the coverage would increase to 85%.

DISCUSSION

The development of ICD-11 has taken considerably longer than all its predecessors. This is probably because ICD-11 has embraced some brand-new features. The availability of maps in both directions is certainly helpful, and they are heavily used in our study. The introduction of the Foundation Component and postcoordination features will have significant impact on the update process, tooling, and coding practice. Their potential benefits are worth examining in more detail.

Benefits of the new features
Knowledge-based approach to terminology management
The Foundation Component provides a knowledge base that can facilitate the creation, maintenance, and quality assurance of ICD-11.[32] The Foundation Component can be considered an ontological basis on which ICD-11 is built. It helps ICD-11 to circumvent some inherent restrictions of a statistical classification (eg, single-parenting, residual categories) which are considered as undesirable features in modern terminologies.[33] The ability to generate multiple
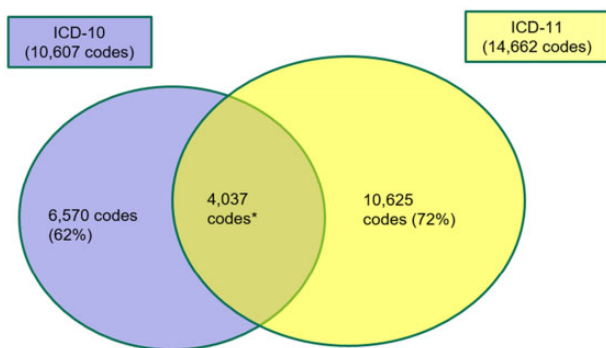

Figure 2. Overlap between ICD-10 and ICD-11. (*equivalent codes found by round tripping, not all manually validated)

Table 4. Recoding ICD-10-CM codes in ICD-11

Disease area	ICD-10-CM codes	Full representation with pre-coordination (%)	Full representation with postcoordination (%)	Partial representation only (%)
Tuberculosis	51	23 (45%)	28 (55%)	0 (0%)
Skin cancer	101	5 (5%)	92 (91%)	4 (4%)
DM type 2	86	1 (1%)	60 (70%)	25 (29%)
Hypertension	17	5 (29%)	9 (53%)	3 (18%)
Polyhydramnios	28	0 (0%)	0 (0%)	28 (100%)
Fracture of thumb	105	0 (0%)	0 (0%)	105 (100%)
Total	388	34 (9%)	189 (49%)	165 (43%)

Figure 3. Display of cerebrovascular diseases in 2 tree positions in the MMS browser.

linearizations from a single knowledge base for different use cases will help to improve the interoperability and comparability of data collected from disparate settings. The use of attributes in defining entities in the Foundation Component opens up the possibility of structural alignment with logically defined terminologies, such as SNOMED CT.[34–38]

Graceful evolution

The Foundation Component paves the way to more open, transparent, and traceable change management processes. Theoretically, with the continuous update model, ICD could undergo gradual, graceful evolution and avoid an abrupt change to a totally new version (ie, ICD-12)—a desirable feature of modern terminologies.[33] The Foundation Component can also help in achieving concept permanence—codes should not be renamed in ways that change their meaning, and codes could be inactivated but not deleted—a terminology desideratum to which historically the ICD classifications have not strictly adhered. With an underlying ontological structure to keep track of the definition and meaning of codes, it is theoretically easier to identify where concept permanence is being violated.

The Foundation Component can also help to alleviate the problems caused by the considerable amount of chapter drift in ICD-11. Chapter drift can cause problems in several ways. Coders may be unaware of the new location of a condition and use suboptimal codes in the original chapter. Code-based data retrieval (eg, cohort identification) may be missing data if data analysts are not aware of codes moved to a different chapter. Since ICD is a single-parent hierarchy, traditionally there is no easy way to show that a condition historically belonged to another chapter. With multi-parenting in the Foundation Component, it is possible to maintain the link of a moved code to its original hierarchy. The MMS browser can show codes in multiple tree positions, including those not used in the MMS linearization. For example, *Cerebrovascular diseases* has been moved to *Chapter 8 Diseases of the nervous system* but is still shown (as greyed-out entries) under *Chapter 11 Diseases of the circulatory system*. (Figure 3)

Increased expressivity

The number of precoordinated leaf codes in ICD-11 is only 38% higher than in ICD-10, but this does not fully reflect the increase in coverage or expressivity. With postcoordination, the number of precoordinated codes can even be reduced in some ICD-11 chapters without affecting the ability to encode certain conditions. The increased expressivity afforded by postcoordination could potentially obviate the need for national extensions, such as ICD-10-CM, which is created to capture additional details outside of the core ICD. Despite being only a quarter of the size of ICD-10-CM, ICD-11 is able to represent fully the meaning of 60% of ICD-10-CM codes we studied. By adding a limited number of extension codes, the coverage can be increased significantly. However, one caveat of postcoordination is the potential increase in coding variability: the same meaning can sometimes be expressed by different code combinations. Unlike SNOMED CT, ICD-11 does not use description logic for postcoordination. Therefore, there is no computational way to identify equivalence of coding. The ICD-11 coding tool provides some guidance by showing which codes are eligible for postcoordination and the allowable extension codes. It will be interesting to see whether this is adequate to ensure consistent postcoordination in practice.

Limitations and future work

We recognize the following limitations in this preliminary study. We did not perform any validation of the WHO maps we used in our analysis. We only manually reviewed a subset of the equivalent ICD-10 and ICD-11 codes we identified by the round trip method. The results of normalized lexical mapping between ICD-11 and ICD-10-CM have not been manually reviewed. The choice of the 6 disease areas for the recoding study was based on the clinical and terminological knowledge of the authors, and the results may not be generalizable to other disease areas. The recoding exercise was done by a single clinician terminologist and not independently validated. Our study should be considered a preliminary one, covering a small number of disease areas and ICD-10-CM codes which may not be representative. Moreover, we did not consider the usage frequencies of the codes, and it is possible that some unmappable codes were more heavily used than the mappable ones. More research is necessary to fully address the feasibility of replacing ICD-10-CM with ICD-11.

In the future, we will do a more comprehensive study that focuses on the comparison with ICD-10-CM involving frequently used codes from all chapters and addressing concerns that include inter-rater agreement in coding, postcoordination variability, and compatibility of coding guidelines between ICD-11 and ICD-10-CM.

CONCLUSION

Compared to ICD-10, ICD-11 has 38% more precoordinated leaf codes, but many more code combinations can be produced with postcoordination. In the exercise of recoding ICD-10-CM codes in ICD-11, we found that 60% of ICD-10-CM codes could be fully represented with precoordinated codes or postcoordination, with the potential of even higher coverage by the addition of a small number of extension codes.

FUNDING

AUTHOR CONTRIBUTIONS

KWF and OB conceived and designed the study. JX performed the manual recoding of ICD-10-CM codes. KWF performed the data analysis. KWF drafted the manuscript and all authors contributed substantially to its revision.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT

None declared

REFERENCES

1. World Health Organization. History of the development of the ICD. http://www.who.int/classifications/icd/en/HistoryOfICD.pdf. Accessed March 24, 2020.
2. Gersenovic M. The ICD family of classifications. *Methods Inf Med* 1995; 34 (01/02): 172–5.
3. World Health Organization. World Health Assembly Update, 25 May 2019: International Statistical Classification of Diseases and Related Health Problems (ICD-11). 2019. https://www.who.int/news-room/detail/25-05-2019-world-health-assembly-update. Accessed March 24, 2020.
4. Pocai B. The ICD-11 has been adopted by the World Health Assembly. *World Psychiatry* 2019; 18 (3): 371–2.
5. Editorial. ICD-11. *Lancet* 2019; 393 (10188): 2275.
6. Editorial. ICD-11: a brave attempt at classifying a new world. *Lancet* 2018; 391 (10139): 2476.
7. Jakob R. [ICD-11-Adapting ICD to the 21st century]. *Bundesgesundheitsbl* 2018; 61 (7): 771–7.
8. Jakob R, Üstün B, Madden R, et al. The WHO family of international classifications. *Bundesgesundheitsbl* 2007; 50 (7): 924–31.
9. Jette N, Quan H, Hemmelgarn B, et al. The development, evolution, and modifications of ICD-10: challenges to the international comparability of morbidity data. *Med Care* 2010; 48 (12): 1105–10.
10. Hirsch JA, Nicola G, McGinty G, et al. ICD-10: history and context. *AJNR Am J Neuroradiol* 2016; 37 (4): 596–9.
11. ICD-11 Reference Guide. https://icd.who.int/icd11refguide/en/index.html. Accessed March 24, 2020.
12. ICD-11 Home Page. https://icd.who.int/en/. Accessed March 24, 2020.
13. ICD-11 Foundation *Component* Browser. https://icd.who.int/dev11/f/en#/. Accessed March 24, 2020.
14. ICD-11 for Mortality and *Morbidity* Statistics Browser. https://icd.who.int/browse11/l-m/en. Accessed March 24, 2020.
15. ICD-11 Coding Tool *Mortality* and Morbidity Statistics (MMS). https://icd.who.int/ct11/icd11_mms/en/release. Accessed March 24, 2020.
16. ICD-11 Application *Programming* Interface (API). https://icd.who.int/icdapi. Accessed March 24, 2020.
17. Boyd AD, Li JJ, Burton MD, et al. The discriminatory cost of ICD-10-CM transition between clinical specialties: metrics, case study, and mitigating tools. *J Am Med Inform Assoc* 2013; 20 (4): 708–17.
18. Caskey R, Zaman J, Nam H, et al. The transition to ICD-10-CM: challenges for pediatric practice. *Pediatrics* 2014; 134 (1): 31–6.
19. Grief SN, Patel J, Kochendorfer KM, et al. Simulation of ICD-9 to ICD-10-CM transition for family medicine: simple or convoluted? *J Am Board Fam Med* 2016; 29 (1): 29–36.
20. Venepalli NK, Qamruzzaman Y, Li JJ, et al. Identifying clinically disruptive International Classification of Diseases 10th Revision Clinical Modification conversions to mitigate financial costs using an online tool. *J Oncol Pract* 2014; 10 (2): 97–103.
21. Wollman J. ICD-10: a master data challenge. *Health Manag Technol* 2011; 32 (7): 16, 20–1.
22. Bhuttar VK. Crosswalk options for legacy systems. Implementing near-term tactical solutions for ICD-10. *J AHIMA* 2011; 82 (6): 34–7.
23. Butler R, Bonazelli J. Converting MS-DRGs to ICD-10-CM/PCS. Methods used, lessons learned. *J AHIMA* 2009; 80 (11): 40–3.
24. Venepalli NK, Shergill A, Dorestani P, et al. Conducting retrospective ontological clinical trials in ICD-9-CM in the age of ICD-10-CM. *Cancer Inform* 2014; 13 (Suppl 3): 81–8.
25. Boyd AD, Li JJ, Kenost C, et al. Metrics and tools for consistent cohort discovery and financial analyses post-transition to ICD-10-CM. *J Am Med Inform Assoc* 2015; 22 (3): 730–7.
26. Boyd AD, Yang YM, Li J, et al. Challenges and remediation for patient safety indicators in the transition to ICD-10-CM. *J Am Med Inform Assoc* 2015; 22 (1): 19–28.
27. Gaebel W, Stricker J, Riesbeck M, et al. Accuracy of diagnostic classification and clinical utility assessment of ICD-11 compared to ICD-10 in 10

mental disorders: findings from a web-based field study. *Eur Arch Psychiatry Clin Neurosci* 2020; 270 (3): 281–9.

28. Tanno LK, Chalmers R, Bierrenbach AL, *et al*. Changing the history of anaphylaxis mortality statistics through the World Health Organization's International Classification of Diseases-11. *J Allergy Clin Immunol* 2019; 144 (3): 627–33.

29. McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. *Proc Annu Symp Comput Appl Med Care* 1994; 1994: 235–9.

30. UMLS Reference Manual: SPECIALIST Lexicon and Lexical Tools. https://www.ncbi.nlm.nih.gov/books/NBK9680/. Accessed March 24, 2020.

31. UMLS Reference Manual: Metathesaurus - Rich Release Format (RRF). https://www.ncbi.nlm.nih.gov/books/NBK9685/. Accessed March 24, 2020.

32. Cimino JJ, Clayton PD, Hripcsak G, *et al*. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *J Am Med Inform Assoc* 1994; 1 (1): 35–50.

33. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med* 1998; 37 (4-5): 394–403.

34. Mamou M, Rector A, Schulz S, *et al*. ICD-11 (JLMMS) and SCT inter-operation. *Stud Health Technol Inform* 2016; 223: 267–72.

35. Mamou M, Rector A, Schulz S, *et al*. Representing ICD-11 JLMMS using IHTSDO representation formalisms. *Stud Health Technol Inform* 2016; 228: 431–5.

36. Rodrigues J-M, Schulz S, Rector A, *et al*. Sharing ontology between ICD 11 and SNOMED CT will enable seamless re-use and semantic interoperability. *Stud Health Technol Inform* 2013; 192: 343–6.

37. Schulz S, Rodrigues J-M, Rector A, *et al*. What's in a class? Lessons learnt from the ICD - SNOMED CT harmonisation. *Stud Health Technol Inform* 2014; 205: 1038–42.

38. Rodrigues J-M, Robinson D, Della Mea V, *et al*. Semantic alignment between ICD-11 and SNOMED CT. *Stud Health Technol Inform* 2015; 216: 790–4.