
Kin Wah Fung, MD, MS, MA; Julia Xu, MD, PhD, Olivier Bodenreider, MD, PhD

Affiliations: National Library of Medicine, Bethesda, MD, USA

Correspondence and reprints:

Kin Wah Fung
Building 38A, Rm9S918, MSC-3826
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894

Telephone: 301 – 827 5001
Fax: 301 – 496 0663
Email: kwfung@nlm.nih.gov

Keywords: ICD-11, ICD-10, ICD-10-CM, Controlled medical vocabularies, Medical terminologies

Word count: 3,832
ABSTRACT

Objective: To study the newly adopted ICD-11 and compare it to ICD-10 and ICD-10-CM.

Materials and Methods: Data files and maps were downloaded from the WHO website and through the APIs. 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 six disease areas were also manually recoded in ICD-11.

Results: Excluding the chapters for traditional medicine, functioning assessment and extension codes for post-coordination, 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 4,037 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 4,059 pairs of possibly equivalent codes. Manual recoding showed that 60% of ICD-10-CM codes could be fully represented in ICD-11 by pre-coordinated codes or post-coordination.

Conclusion: In ICD-11, there is a moderate increase in the number of codes over ICD-10. With post-coordination, 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.
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. 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. 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. Over two dozen countries have developed national extensions of ICD to suit their requirements. In the U.S., the Clinical Modification (CM) has been developed since ICD-9-CM to support morbidity coding for reimbursement and other purposes. 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 three 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. These derivatives are called “linearizations”. It is also possible to
generate alternative lists for different purposes e.g., specialty subsets, country specific modifications.

The Foundation Component is a multi-dimensional 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 multi-parenting 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 one category. Residual categories (e.g., ‘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 (e.g., yearly) and officially versioned.

2. Post-coordination – ICD-11 allows the combination of codes (called “cluster coding”) to add additional detail to an existing code (called “stem code” or “pre-coordinated code”). Two kinds of post-coordination are allowed:

a. Two or more stem codes (syntax: stemcode1/stemcode2/stemcode3 etc.) e.g., 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 one or more extension codes (syntax: stemcode1& extensioncode1&extensioncode2 etc.) e.g., tuberculosis of prostate = 1B12.5&XA63E5 (1B12.5 *Tuberculosis of the genitourinary system*; XA63E5 *Prostate gland*).

Table 1. Comparison of the Foundation Component and linearizations

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Foundation Component</th>
<th>Example</th>
<th>Linearizations</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building block</td>
<td>Entity</td>
<td>Diaphragmatic hernia</td>
<td>Category</td>
<td>Diaphragmatic hernia</td>
</tr>
<tr>
<td>Identifier</td>
<td>URI (universal resource identifier)</td>
<td><a href="http://id.who.int/icd/entity/453532731">http://id.who.int/icd/entity/453532731</a></td>
<td>Code</td>
<td>DD50.0</td>
</tr>
<tr>
<td>Defining attributes</td>
<td>Description, Body site, body system, causal mechanisms, synonyms, exclusions, signs and symptoms etc.</td>
<td>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)</td>
<td>Description, inclusions, exclusions</td>
<td>Description: A hernia occurs through the foramen in the diaphragm Inclusions: paraesophageal hernia Exclusions: Congenital diaphragmatic hernia (LB00.0), Congenital hiatus hernia (LB13.1)</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>Multi-parenting</td>
<td>Parents: Non-abdominal wall hernia, Other diseases of the digestive system</td>
<td>Single-parenting</td>
<td>Parent: DD50 Non-abdominal wall hernia</td>
</tr>
<tr>
<td>Residual elements</td>
<td>None</td>
<td>Present</td>
<td>Present</td>
<td>DD50.Y Other specified non-abdominal wall hernia, DD50.Z Non-abdominal wall hernia, unspecified</td>
</tr>
<tr>
<td>Update frequency</td>
<td>Continuous</td>
<td>Periodic with official versioning</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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-
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 non-trivial endeavor which incurs significant cost and has potential impact on longitudinal data comparability, as evidenced by various reports when the U.S. 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 U.S. could move from ICD-10-CM directly to ICD-11 for morbidity coding, without creating a Clinical Modification. There have been studies on the differences between ICD-11 and ICD-10, but most are focused on specific disease areas (e.g., mental health). 27, 28 We believe that our study is the first systematic comparison of ICD-11 to ICD-10 and ICD-10-CM.

MATERIALS AND METHODS

1. 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 including kind of code (chapter, block or category), depth in tree, whether code is a residual category and whether code is a “leaf code” (the lowest level code that does not have any children, only leaf codes can be used for coding).
3. **One Category ICD-10 to ICD-11 Map** – each ICD-10 code maps to only one 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 one ICD-10 code

We used the MMS browser and MMS coding tool to look up individual codes. There was no downloadable version of the Foundation Component and ICD-10. We used the API to collect all available information on the Foundation Component, MMS and ICD-10. Information about post-coordination was only available through the API and not the downloadable files.

2. **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 pre-coordinated ICD-11 codes. By analyzing which chapter of ICD-11 an ICD-10 code was mapped to in the one-category map, we identified correspondences among chapters, as well as “chapter drift” (i.e., 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 used both one-category maps to find equivalent codes between ICD-10 and ICD-11. Maps were not always between equivalent codes (e.g., narrow-to-broad map). We postulated that if an ICD-10 code mapped to a single ICD-11 code in the forward map, and that ICD-11 code mapped back to the same ICD-10 code in the backward map (“round-
tripping”), there was a high likelihood that the two codes were equivalent. We manually reviewed a sample of the round-trip maps to see if they were truly equivalent.

3. **ICD-10-CM comparison**

Since no maps existed, we used two approaches to compare ICD-11 to ICD-10-CM, lexical matching and manual recoding. For lexical matching, we used the text normalization tool luinorm from the UMLS (2019 version) to normalize the names of ICD-11 codes (leaf and non-leaf codes) from chapters 1 to 25 of MMS.\textsuperscript{29,30} We matched the normalized names to UMLS concepts (version 2019AA) using the normalized English strings index (MRXNS_ENG).\textsuperscript{31} From the UMLS concepts, we identified the ICD-10-CM codes (2019 version) that were included in that concept. We ignored ICD-11 index terms because some of them were narrower in meaning. For example, the index terms for *Paratyphoid Fever* included *Paratyphoid fever A* and *Paratyphoid fever B*. For the same reason, we ignored ICD-10-CM inclusion terms in the UMLS.

For manual recoding, we chose six disease areas in ICD-10-CM that represented some 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 by in ICD-11 with or without post-coordination, or only partially represented even with post-coordination.

**RESULTS**

**ICD-10 comparison**

1. **Chapter structure, chapter drift and extent of change**
ICD-11 had 28 chapters, six more than ICD-10. The last three 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 post-coordination)

Among the first 25 chapters, three 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 codes in the one-category ICD-10 to ICD-11 map. 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 behavioural 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 seven ICD-11 chapters with chapter drift index (CDI) over 5% (figure 1, last row). Among these were, not surprisingly, the three new chapters, since they did not correspond neatly to a single ICD-10 chapter (thus the need for a new chapter). The other four 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 e.g., *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, e.g., *X-linked ichthyosis*, *Congenital leukonychia* (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 e.g., 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 XXIII 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 3 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 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 post-coordination. 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 post-coordination, some pre-coordinated codes were no longer necessary. For example, there were six 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 one leaf code PA78 Unintentionally stung or envenomated by animal, and the different animals can be post-coordinated by extension codes: XE9H6 Venomous snake, XE6A7 Lizard, gecko, goanna, XE75L Spider and XE2EP Scorpion, and many more.
Table 3. Extent of change by chapter

<table>
<thead>
<tr>
<th>ICD-10 chapter</th>
<th>Corresponding ICD-11 chapter</th>
<th>ICD-10 leaf codes</th>
<th>ICD-11 leaf codes</th>
<th>% change</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Certain infectious and parasitic diseases</td>
<td>1</td>
<td>783</td>
<td>835</td>
</tr>
<tr>
<td>II</td>
<td>Neoplasms</td>
<td>2</td>
<td>759</td>
<td>1056</td>
</tr>
<tr>
<td>III</td>
<td>Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism</td>
<td>3 &amp; 4</td>
<td>165</td>
<td>424</td>
</tr>
<tr>
<td>IV</td>
<td>Endocrine, nutritional and metabolic diseases</td>
<td>5</td>
<td>356</td>
<td>541</td>
</tr>
<tr>
<td>V</td>
<td>Mental and behavioural disorders</td>
<td>6</td>
<td>407</td>
<td>718</td>
</tr>
<tr>
<td>VI</td>
<td>Diseases of the nervous system</td>
<td>8</td>
<td>335</td>
<td>719</td>
</tr>
<tr>
<td>VII</td>
<td>Diseases of the eye and adnexa</td>
<td>9</td>
<td>262</td>
<td>615</td>
</tr>
<tr>
<td>VIII</td>
<td>Diseases of the ear and mastoid process</td>
<td>10</td>
<td>113</td>
<td>135</td>
</tr>
<tr>
<td>IX</td>
<td>Diseases of the circulatory system</td>
<td>11</td>
<td>383</td>
<td>479</td>
</tr>
<tr>
<td>X</td>
<td>Diseases of the respiratory system</td>
<td>12</td>
<td>234</td>
<td>296</td>
</tr>
<tr>
<td>XI</td>
<td>Diseases of the digestive system</td>
<td>13</td>
<td>435</td>
<td>804</td>
</tr>
<tr>
<td>XII</td>
<td>Diseases of the skin and subcutaneous tissue</td>
<td>14</td>
<td>342</td>
<td>615</td>
</tr>
<tr>
<td>XIII</td>
<td>Diseases of the musculoskeletal system and connective tissue</td>
<td>15</td>
<td>545</td>
<td>364</td>
</tr>
<tr>
<td>XIV</td>
<td>Diseases of the genitourinary system</td>
<td>16</td>
<td>438</td>
<td>447</td>
</tr>
<tr>
<td>XV</td>
<td>Pregnancy, childbirth and the puerperium</td>
<td>18</td>
<td>434</td>
<td>437</td>
</tr>
<tr>
<td>XVI</td>
<td>Certain conditions originating in the perinatal period</td>
<td>19</td>
<td>337</td>
<td>525</td>
</tr>
<tr>
<td>XVII</td>
<td>Congenital malformations, deformations and chromosomal abnormalities</td>
<td>20</td>
<td>619</td>
<td>1118</td>
</tr>
<tr>
<td>XVIII</td>
<td>Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified</td>
<td>21</td>
<td>340</td>
<td>1078</td>
</tr>
<tr>
<td>XIX</td>
<td>Injury, poisoning and certain other consequences of external causes</td>
<td>22</td>
<td>1278</td>
<td>1668</td>
</tr>
<tr>
<td>XX</td>
<td>External causes of morbidity and mortality</td>
<td>23</td>
<td>1372</td>
<td>842</td>
</tr>
<tr>
<td>XXI</td>
<td>Factors influencing health status and contact with health services</td>
<td>24</td>
<td>630</td>
<td>759</td>
</tr>
<tr>
<td>XXII</td>
<td>Codes for special purposes</td>
<td>25</td>
<td>40</td>
<td>17</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>10607</td>
<td>14492*</td>
<td></td>
</tr>
</tbody>
</table>

*not including chapters 7 and 17

2. Codes that have remained the same

There were altogether ICD-11 14,622 leaf codes (chapters 1 – 25), a 38% increase over ICD-10 (10,607 leaf codes). By round-tripping, we identified 4,037 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 4, 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 e.g., *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 e.g., *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*. Overall, 93% of the reviewed cases were equivalent. If we projected these results to all the 4,037 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.
Table 4. Candidate equivalent codes identified by round-trip method

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of codes (%)</th>
<th>Examples</th>
<th>Manual review</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICD-10</td>
<td>ICD-11</td>
<td>Equivalent</td>
</tr>
<tr>
<td>A. Same name</td>
<td>1773 (44%)</td>
<td>A06.4 Amoebic liver abscess</td>
<td>1A36.10 Amoebic liver abscess</td>
</tr>
<tr>
<td>B. ICD-11 name contains ICD-10 name</td>
<td>338 (8%)</td>
<td>A01.0 Typhoid fever</td>
<td>1A07.Z Typhoid fever, unspecified</td>
</tr>
<tr>
<td>C. ICD-10 name contains ICD-11 name</td>
<td>146 (4%)</td>
<td>I73.1 Thromboangiitis obliterans [Buerger]</td>
<td>4A44.8 Thromboangiitis obliterans</td>
</tr>
<tr>
<td>D. Others</td>
<td>1780 (44%)</td>
<td>Q69.2 Accessory toe(s)</td>
<td>LB78.3 Polydactyly of toes</td>
</tr>
<tr>
<td>Total</td>
<td>4037 (100%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ICD-10-CM comparison

1. Lexical matching (to pre-coordinated codes only)

By normalized lexical matching through the UMLS, we managed to find 4,059 pairs of matching ICD-11 and ICD-10-CM codes, covering 3,294 unique ICD-11 codes (3,211 leaf and 83 non-leaf) and 3,985 unique ICD-10-CM codes (2,366 leaf and 1,619 non-leaf). The breakdown of these code pairs:

- 52 pairs: non-leaf codes in both ICD-10-CM and ICD-11
- 1,596 pairs: ICD-11 leaf codes matched to ICD-10-CM non-leaf codes
- 66 pairs: ICD-11 non-leaf codes matched to ICD-10-CM leaf codes
- 2,345 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 non-leaf code ICD-10-CM code compared to the other way round. This shows that ICD-10-CM is larger (71,932 vs. 14,622 leaf codes) and more fine-grained than ICD-11. However, with post-coordination, it is possible to bridge some of these differences.

2. Manual recoding (to pre- and post-coordinated codes)

We selected 388 ICD-10-CM leaf codes from six disease areas - tuberculosis, skin cancer, diabetes mellitus (DM) type 2, hypertension, polyhydramnios and fracture of thumb, and recoded them in ICD-11, using post-coordination when necessary. (Table 4)

<table>
<thead>
<tr>
<th>Disease area</th>
<th>ICD-10-CM codes</th>
<th>Full representation with pre-coordination (%)</th>
<th>Full representation with post-coordination (%)</th>
<th>Partial representation only (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuberculosis</td>
<td>51</td>
<td>23 (45%)</td>
<td>28 (55%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Skin cancer</td>
<td>101</td>
<td>5 (5%)</td>
<td>92 (91%)</td>
<td>4 (4%)</td>
</tr>
<tr>
<td>DM type 2</td>
<td>86</td>
<td>1 (1%)</td>
<td>60 (70%)</td>
<td>25 (29%)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>17</td>
<td>5 (29%)</td>
<td>9 (53%)</td>
<td>3 (18%)</td>
</tr>
<tr>
<td>Polyhydramnios</td>
<td>28</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>28 (100%)</td>
</tr>
<tr>
<td>Fracture of thumb</td>
<td>105</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>105 (100%)</td>
</tr>
<tr>
<td>Total</td>
<td>388</td>
<td>34 (9%)</td>
<td>189 (49%)</td>
<td>165 (43%)</td>
</tr>
</tbody>
</table>
a) Tuberculosis – Among the 51 codes from the block A15 – A19 Tuberculosis, 23 could be fully represented by pre-coordinated ICD-11 codes. All remaining codes could be fully represented by post-coordination. 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 five could be fully represented by pre-coordinated codes. However, most of the remaining codes could be post-coordinated. 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 e.g., 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 one code had an equivalent pre-coordinated code. Sixty codes could be post-coordinated e.g., 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 e.g., 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 five codes had pre-coordinated equivalents. Nine codes could be post-coordinated e.g., 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 e.g., 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 (e.g. 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 six 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). Post-coordination 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 post-coordinated ICD-11 codes. With the addition of three 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 post-coordination 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

1. 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. The ability to generate multiple 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.

2. 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 “graceful evolution” and avoid an abrupt change to a totally new version (i.e. ICD-12) – one of the desiderata of modern terminologies.

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 (e.g., 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)

3. Increased expressivity

The number of pre-coordinated 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 post-coordination, the number of pre-coordinated codes can even be reduced in some ICD-11 chapters without affecting the ability to encode certain conditions.

Before WHO finalizes the licensing and copyright restrictions for national modifications, it is not clear whether ICD-11-CM can be created as usual. The increased expressivity afforded by post-coordination could potentially obviate the need for national extensions. 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. One caveat of post-coordination is that it can increase the variability in coding because the same meaning can sometimes be expressed by different code combinations. Unlike SNOMED CT, ICD-11 does not use description logic for post-coordination. 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 post-coordination and the
allowable extension codes. It will be interesting to see whether this is adequate to ensure consistent post-coordination 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 six disease areas for the re-coding study was based on the clinical and terminological knowledge of the authors, and the results may not be generalizable to other disease areas. The re-coding exercise was done by a single clinician terminologist and not independently validated.

In future, we plan to study the value of the Foundation Component, especially in the alignment of ICD-11 with other terminologies. We also plan to do more study on post-coordination and other aspects of the feasibility of replacing ICD-10-CM with ICD-11, including the alignment of definitions, coding guidelines and inclusion/exclusion criteria.

CONCLUSION

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

Acknowledgements

The authors would like to thank Robert Jakob of the World Health Organisation for providing
information about ICD-11 and answering our questions. We thank the U.S. National
Committee on Vital and Health Statistics for the opportunity to participate in the review of
ICD-11. We also thank Donna Pickett from the U.S. National Center for Health Statistics for
the insightful discussions.

Funding Statement

This research was supported in part by the Intramural Research Program of the NIH, National
Library of Medicine.

Competing Interests Statement

The authors do not have competing interests.

Contributorship Statement

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.

References


List of figures:

Figure 1. Alignment of codes by chapter based on map from ICD-10 to ICD-11 (ICD-10 chapters I to XXII, ICD-11 chapters ch1 to ch25, CDI - chapter drift index, see text for explanation)

Figure 2. Overlap between ICD-10 and ICD-11

Figure 3. Display of Cerebrovascular diseases in two tree positions in the MMS browser