ICD11 MMS (April 2017 frozen release) to SNOMED CT Equivalence Table

Summary of methodology, format & results within the scope of the ICD11 Mortality Morbidity & Statistics linearization

Date: 20171205

Version: 0.01

Status: Draft

SNOMED International
Registered in England and Wales • Company Registration Number 9915820
Reg. address: One Kingdom Street • Paddington Central • London W2 6BD • United Kingdom
Tel: +44 (0) 203 755 0974 • info@snomed.org • www.snomed.org

SNOMED International is the trading name of the
International Health Terminology Standards Development Organisation,
a private company limited by guarantee

Amendment History

Version	Date	Editor	Comments
0.01	20171205	Phil Brown, David Robinson	First draft for comment

Approvals

Version	Date	Approver	Comments

Future Review Timetable

Review date	Responsible owner	Comments

Table of Contents

1 Introduction

1.1 Purpose

This document describes the processes and outcomes of the work to align the World Health Organization, International Classification of Diseases 11th Revision (ICD11) categories to SNOMED CT content; this includes the addition of approximately 3,300 concepts[1] as a result of this work.

This is a working document as there are a number of uncertainties around timescales to complete tasks not yet commenced that are dependent on activities outside this project; which impact on resources and deadlines: It will be updated in the light of experience as work proceeds and concludes.

The results of this work can be found in 2 accompanying data files

scticd11equivalence_201704mms.tsv (suitable for import into databases/applications)

scticdequivalence_201704mms.xlsx (suitable for human viewing)

1.2 Scope

The scope of this matching exercise is limited to the content of the linearization for Mortality and Morbidity Statistics (MMS): This is a subset of the larger ICD11 foundation layer which includes more codes, from which other subsets may be derives, e.g. for Primary Care.

1.3 Audience

SNOMED International Content Team

SNOMED International Mapping Team

West Coast Informatics

[1] This includes additions for ICD11 rubrics that were subsequently removed from the MMS and some concepts added via SIRS in the early stages of the project.

2 Overview

The aim of this work has been to identify equivalent concepts in ICD11 and SNOMED CT, adding new content where necessary. Originally it was intended to develop a common ontology between the ICD11 foundation layer and SNOMED CT, but this approach was abandoned and the scope of the matching has subsequently been limited to ICD11 content within the April 2017 frozen MMS supplied by the WHO.

While the results of this exercise will provide input into work to develop a map from SNOMED CT to ICD11, it is not a map, The direction is from ICD11 to SNOMED CT so we have used the word 'match' as an abbreviation for 'equivalence match'.

The first release of this table contains provisional data, as there are a number of potential outstanding impacts:

- changes to SNOMED CT data within SCA prior to the January 2018 release
- changes to WHO data in a subsequent MMS release
- feedback from end users of the table
- feedback from external reviewers
- response from WHO to queries

3 Methodology

3.1 Initial matching

The matching process was undertaken using a MySQL database hosted by SNOMED International. ICD11 rubrics and descriptions were inspected to determine whether each ICD11 code (representing a classification category):

- existed within SNOMED CT
- required addition to SNOMED CT
- should be excluded as a non-clinical concept e.g. navigational (grouping heading), complex clauses such as either/or/excluding; or
- required clarification from WHO or domain experts

Lexical matching was used but every generated match was also manually inspected, as a number of errors were possible including:

- synonym errors within SNOMED CT
- different concept types within SNOMED CT e.g. disorder vs. morphology

3.2 Review & Quality Assurance

3.2.1 Internal Review

Every match within the scope of the MMS, whether it involved additions to SNOMED CT, existing SNOMED CT content or a decision not to add was further reviewed at the end of the process for quality assurance and to take account of changes that occurred in the course of the project including:

- changes within SNOMED CT
- changes of editorial policy for SNOMED CT
- changes to ICD11 terms or descriptions

Any proposed new content was also reviewed by members of the authoring team within the SCA environment.

3.2.2 External Review

Domain experts, in selected areas, were recruited where specialized knowledge was required to clarify the meaning of certain ICD11 rubrics (see Glossary). 165 rubrics, with suggested solutions, were offered for specialist opinion [2]. The experts advised whether the proposed match or new relationship was acceptable or provided feedback for an alternative solution.

In 71 cases the meaning of the ICD11 rubric was unclear, these were summarised into 30 distinct queries and fed back to the WHO; all these issues were ratified by the WHO as requiring further clarification and sent on to the appropriate sub-domain panels. Feedback from the WHO is still outstanding and once feedback is received the suggested changes will be implemented

3.2.3 Quality Assurance

Over one hundred SQL based quality assurance (QA) rules have been developed during the course of the matching process covering data quality issues such as:

- validity of SNOMED CT codes
- use of proximal primitive concepts within the modelling
- appropriate use of attributes & value ranges
- US & GB spellings
- coverage of the MMS
- accuracy of cardinality
- accuracy of match types

The nature of SNOMED CT is that concepts can be retired or changed and this table is derived from a snapshot of SNOMED CT at the time is was exported, so while every effort has been made to run all the relevant QA rules prior to export, subsequent changes to SNOMED CT may not be reflected in this table. Equally ICD11 is subject to change, however any ICD11 data is derived from the April 2017 MMS.

[2] This excludes rubrics routinely submitted for review at the start of the project

4 ICD11 SNOMED CT Equivalence Table

4.1 Structure

The ICD11 description has not been included, as the MMS file, released by the WHO, does not contain this field. We also no longer have a mechanism for importing data from ICD11, other than the MMS files; and given the changing nature of ICD11 we are unable to validate the accuracy of descriptions in our database.

Field	Data Type	Description
FoundationId	Integer	The final part of the foundation layer id for the ICD11 category, stripped of the 'http'
MMS_Code	String	Alphanumeric string allocated by WHO in the MMS files
ICD_Rubric	String	Rubric allocated by WHO to the ICD11 code (as of April 2017 but subject to possible change by the WHO)
SCTID	Integer	ConceptId for the matched SNOMED CT concept
SCT_Terms	String	All of the terms for the matched SNOMED CT concept (in pipe separated format)
Match_Type	String	Indicates the nature of the match or the reason why a match is absent (listed in section 4.3)
Completeness	String	Indicates the completeness of the match from ICD11 to SCT whether the match can be considered full (F) or partial (P)
Match_no	Integer	Number of the match (for one to one matches this is one)
Cardinality	String	Indicates whether the match is: • one to one (1:1) • one to many (1:M) • many to one (M:1) • many to many (M:N)
Comments	String	Reason for absent match and other warnings
MMS_Other	String	Other MMS codes that have been matched to the SCT concept where a match is many to one, or many to many the (in pipe separated format)

Table 1: Equivalence table structure

4.2 Content

All clinically relevant concepts derived directly from the ICD11 rubrics that are in scope are present within SCT (except for those rubrics awaiting feedback from WHO; specialist opinion or editorial policy). This incudes some but not all 'index terms' which were outside the scope of this project and still in flux within ICD11.

4.2.1 Version Control

As ICD11 is not yet stable, a reference point to the April 2017 frozen MMS has been used for MMS codes as this was the last release of data available for import. There are known to have been changes to codes and content since that date.

While references have been made to the current browser for definitions and index terms

https://icd.who.int/dev11/l-m/en

For MMS codes, the April 2017 frozen (draft for QA) browser has been used for alignment:

http://apps.who.int/classifications/icd11/frozen-2017-04-02/l-m/en

A re-alignment will need to be undertaken when a final, stable MMS is available.

4.2.2 Concept Types

Valid concept types for a map from SNOMED CT to ICD11 are:

- Clinical finding
- Disorder
- Event
- Situation

Within chapter XXIV (*Factors influencing health status and contact with health services*) there are a number of 'reasons for encounter' included which are outside these four allowable concept types. These include:

- Procedures
- Regime/therapy
- Person

A decision was made not to add any more of these expressed as clinical findings as has been done in the past; however matches to 'out of scope concepts', in terms of mapping to ICD-11 (Procedures, Regime/therapy; and Person) have been made and new content added where appropriate to improve SNOMED CT content. These concepts are flagged as being inappropriate for mapping purposes (Match_type P, see 4.2.1) and also highlighted in the Comments field.

4.2.3 Provisional data

At the time of releasing test data, in December 2017, a number of concepts are in the SCA having been added but have not yet been formally issued in the January 2018 release.

Provisional data has been generated from the SCTIDs that are returned in the import process from our database into SCA with the proposed fully specified name. These will be flagged in the Comments field as 'PROVISIONAL - scheduled for January 2018 release'.

4.2.4 Residual matches

Residual categories ('Other X' or 'X unspecified') have the same FoundationId as the equivalent non-residual ICD11 rubric, with the MMS code having an 'Y' or 'Z" appended. These categories will have the same match as the non-residual, which is in line with a decision made early in the project to match to the non-residual SNOMED CT equivalent. For example:

1A03 *Intestinal infections due to Escherichia coli*

and its residual categories:

1A03.Y *Intestinal infections due to other specified Escherichia coli*

1A03.Z *Intestinal infections due to Escherichia coli, unspecified*

are all matched to the SCTID 111839008 *Intestinal infection caused by Escherichia coli (disorder)* (also see ***Many to one (M:1)*** in section 4.2.9).

4.2.5 Chapters included

The following chapters are included in the equivalence table

I Infectious diseases

II Neoplasms

III Diseases of the blood and blood-forming organs

IV Disorders of the immune system

V Endocrine

VI Conditions related to sexual health

VII Mental and behavioural disorders

VIII Sleep-wake disorders

IX Diseases of the nervous system

X Diseases of the eye and adnexa

XI Diseases of the ear and mastoid process

XII Diseases of the circulatory system

XIII Diseases of the respiratory system

XIV Diseases of the digestive system

XV Diseases of the skin

XVI Diseases of the musculoskeletal system and connective tissue

XVII Diseases of the genitourinary system

XVIII Pregnancy

XIX Certain conditions originating in the perinatal and neonatal period

XX Developmental anomalies

XXI Symptoms

XXII Injury

XXIV Factors influencing health status and contact with health services

4.2.6 Chapters excluded

The following chapters are excluded from the equivalence table

XXIII External causes of morbidity and mortality

XXV Codes for special purposes

XXVII Extension Codes

4.2.7 Match_Type

This column indicates the nature of the match where present, or the status of the match if not present. The full set of values is listed in Table 2.

Value	Description
A	Match absent, reason stated in *Comments field*
B	Block (see Glossary) in ICD11 MMS with no code and no match
C	'Certain specified' stated in ICD rubric but not in SNOMED CT concept terms
E	'Elsewhere' referenced in ICD rubric but not in SNOMED CT concept terms
H	Chapter (see Glossary) in ICD11 MMS with no code and no match
L	Concept AND lexical match - one term in the SNOMED CT concept matches the ICD rubric (and concept is equivalent)
M	Concept match - no exact lexical match, but SNOMED CT concept & ICD category considered to be the same
N	'Not otherwise specified' stated in ICD rubric but not in SNOMED CT concept terms
O	'Other' stated in ICD rubric but not in descriptions and with a MMS code not ending in 'Y' or 'Z' associated with SNOMED CT concept

Value	Description
P	Match out of scope e.g. procedure, regime/therapy (as these are no longer being added as reasons for encounter as findings)
Q	Awaiting clarification of meaning from the WHO
R	Residual match where 'other' or 'unspecified' stated in ICD rubric and with an MMS code ending in 'Y' or 'Z' but not in descriptions associated with SNOMED CT concept [3]
S	'Specified' stated in ICD rubric but not in descriptions associated with SNOMED CT concept
T	Partial match - elements of ICD11 concept implicit but not in SNOMED CT concept which cannot be modeled
U	'Unspecified' stated in ICD rubric but not in descriptions and with a MMS code not ending in 'Y' or 'Z' associated with SNOMED CT concept, can be considered to be a concept match [4]
V	Awaiting clarification of editorial approach in SNOMED CT
W	'Without' stated in ICD rubric but not in descriptions associated with SNOMED CT concept
X	'Excluding' or 'Except' stated in ICD rubric but not in descriptions associated with SNOMED CT concept
Y	Awaiting addition as part of Orphanet work
Z	Awaiting external review

Table 2: Match_type values

4.2.8 Completeness

The completeness of the match from ICD11 to SNOMED CT is flagged as to whether the match was Full (F) or Partial (P) – this latter situation was commonly due to the ICD11 rubric containing classification clauses which are unrelated to clinical language e.g. 6B4E.3 *Other specified or multiple specified psychoactive substance dependence* was a partial match to *Psychoactive substance dependence* (disorder).

Where the Cardinality is greater then one i.e. M:1 and M:N (see section 4.2.8) the completeness relates to the combined match e.g. 1B70 *Bacterial cellulitis, erysipelas and lymphangitis* has a full match (completeness flag F) because it is matched to three concepts which fully encapsulates its meaning (*Lymphangitis caused by bacterium* (disorder); *Bacterial cellulitis* (disorder); and *Erysipelas* (disorder)).

[3] These will have the same match as the equivalent non-residual MMS code

[4] Where a rubric contains 'other' and 'unspecified' or 'specified', 'other' takes precedence as it has more of an impact on the quality of the match

4.2.9 Cardinality

This column identifies the numeric relationship between ICD11 and SNOMED CT for a particular match or set of matches.

There are four values for cardinality in the equivalence table:

One to one (1:1)

Where possible matches were achieved from one MMS code to one SNOMED CT concept identifier (SCTID) e.g.

1A07 *Typhoid fever* has a single (lexical) match to SNOMED CT Concept 4834000 with terms *Typhoid fever (disorder)* | *Infection by Salmonella Typhi* | *Typhoid fever*

And no other MMS code matches to this SNOMED CT concept

One to many (1:M)

Some complex ICD11 codes required a 'one to many' match, particularly where a disjunctive 'or' clause is present e.g.

LA8C.31 *Congenital anomaly of descending thoracic* ***or*** *abdominal aorta*, requires a match to

724435004 *Congenital anomaly of descending thoracic aorta*
AND
724436003 *Congenital anomaly of abdominal aorta*

Many to one (M:1)

More than one MMS code matches to the same target SNOMED CT concept

In some cases this is restricted to a combination of residual and equivalent non-residual codes (see 4.2.4). A decision was made in the project not to add new 'other' and 'unspecified' concepts to SNOMED CT, and to strip out the residual components from the ICD11 to match, e.g.

1A03 *Intestinal infections due to Escherichia coli*
1A03.Y *Intestinal infections due to other specified Escherichia coli*
1A03.Z *Intestinal infections due to Escherichia coli, unspecified*

All match to the SNOMED CT concept

111839008 *Intestinal infection due to Escherichia coli*

NB the match_type of each match may vary; in the above example M for the first match; and R for the last two as they contain 'residual' *other* or *unspecified* clauses.

There is also a minority of cases where 'many to one' matches do not (exclusively) involve residuals e.g.

1D50 *Influenza due to seasonal identified influenza virus*
1D51 *Influenza due to identified zoonotic or pandemic influenza virus*
1D53 *Influenza, virus not identified*

All match to the SNOMED CT concept

6142004 *Influenza*

In this instance the ICD11 codes have more detail than can be defined or modeled in SNOMED CT, but there may be other reasons, which are discussed in the next section.

Many to many (M:N)

One MMS code matches to more than one SNOMED CT concept but one or more of those SNOMED CT concepts may me the target of more than one MMS code. In many cases this is a combination of residual and disjunctive ICD11 MMS codes e.g.

2E80 Benign neoplasm of middle ear or respiratory system
2E80.Y Other specified benign neoplasm of middle ear or respiratory system
2E80.Z Benign neoplasm of middle ear or respiratory system, unspecified

All match to the SNOMED CT concepts

92218002 Benign neoplasm of middle ear
255166003 Benign neoplasm of respiratory system

However in some cases there are other reasons e.g.

2B81 Ew*ing sarcoma, primary site*

Is matched to 2 SNOMED CT concepts a

307608006 *Ewing sarcoma of bone*
447951009 *Ewing sarcoma of soft tissue*

- as these are the known primary sites.

However 447951009 *Ewing sarcoma of soft tissue* is also the target of a more detailed MMS code:

2B81.4 *Ewing sarcoma of soft tissue*

The implications of many to many matches are discussed in the next section.

5 Utility & Application

5.1.1 Direct mapping

The *Equivalence table* can be used to generate a candidate map from a SNOMED CT concept to an ICD11 MMS code and the confidence of this map is generally dependent on the match *Cardinality*:

1:1 & 1:M
> can be used to automatically create a mapping from SNOMED CT to ICD11

M:1 and M:N
> can be used to automatically create multiple candidates for mapping from SNOMED CT to ICD11 - but requires a check as to which alternative is optimal.

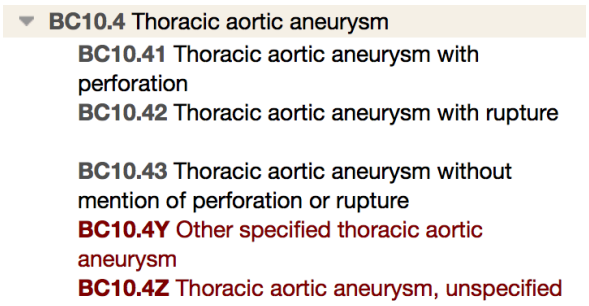
This is the application for which the work and table was designed and generated. As a value added deliverable the relationships within the *Equivalence table* might also potentially have utility outside of this primary purpose (see next section), although this functionality would require further scoping.

5.1.2 Hierarchical mapping

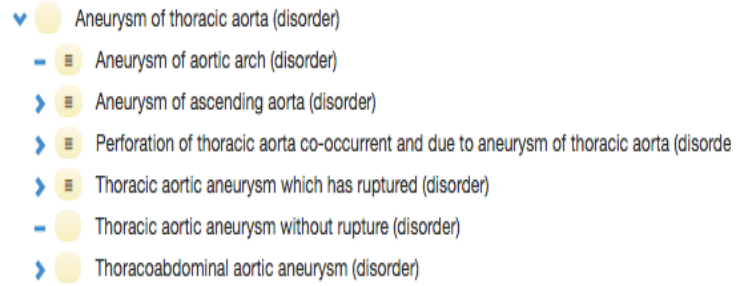
The *Equivalence table* theoretically could also be used to produce candidate maps from descendants of a chosen SNOMED CT concept 'node' - although the utility of this would require testing.

Assuming that the descendant hierarchy of a 'node' concept in SNOMED CT is semantically correct - the MMS code (or its derived residual categories i.e. 'Other * - Y code' or '* unspecified' – Z code) to which the SCT node is matched, are candidate codes for all descendants; unless a descendant SCT concept appears in the table (in which case it's direct match takes precedence). For example consider the following ICD11 category and SNOMED CT hierarchy:

ICD11



SNOMED CT



The SNOMED CT node *Aneurysm of thoracic aorta* is matched to the ICD11 code BC10.4Z (*Thoracic aortic aneurysm, unspecified*), this would be the candidate MMS code for all descendants unless a descendant has an explicit match – In this case the third & fourth child already have a match to BC10.42 & the fifth child to BC10.43 & the sixth to BC10.6Z *Thoracoabdominal aortic aneurysm, unspecified* - Thus all these preexisting matches

take precedence – this leaves *Aneurysm of aortic arch* & *Aneurysm of ascending aorta* as having a candidate map of BC10.Z – And when looking at the index terms this is confirmed. N.B. the alternative map to these would be BC10.4Y (Other specified) which was manually rejected.

Ideally these assumptions would be enhanced by also using:

- ICD inclusion terms, which explicitly state terms that index the particular MMS code; and
- ICD exclusion codes, where specific instances of disorders are specifically excluded from the ICD notion

Currently both the inclusion and exclusion terms for each MMS code are not available in electronic format: However if the *Equivalence table* is applied from 'leaf to trunk' then these exclusions should be partly detectable.

6 Glossary of Terms

The following table contains the definition of any terms used within this document.

Term	Definition
Block	A medium level ICD11 rubric representing a group of codes within ICD11, but with no code in the MMS
Chapter	A high level ICD11 rubric representing a chapter within ICD11, but with no code in the MMS
Descriptions	WHO description accompanying an ICD11 code and rubric
Match	A match from an ICD11 code to an equivalent SNOMED CT concept [or multiple concepts where a one to many match has been identified]
MMS	ICD-11 for Mortality and Morbidity Statistics
SCA	Single Concept Authoring environment
SCTID	Numeric SNOMED CT Identifier
Rubric	WHO term for an ICD11 code
WHO	World Health Organization, responsible for ICD11

Appendix A Summary Statistics

Total number of rows in table	**17,555**
Concept match	5320
Residual match	4202
Lexical match	4038
Match absent, reason stated in Explanation field	2068
Block - no match required	654
Miscellaneous additional statement in ICD rubric	322
Partial match, elements of ICD11 concept implicit not in SNOMED CT	197
Match out of scope e.g. procedure, regime/therapy, person	129
'Unspecified/unclassified/not specified/' in ICD rubric	117
Awaiting external review	115
Awaiting clarification of meaning from WHO	70
'Certain specified' in ICD rubric	68
Exclusion in ICD rubric/definition	66
Awaiting clarification of editorial approach in SNOMED CT	65
Elsewhere in ICD rubric	54
'Without' in ICD rubric	25
Chapter - no match required	23
'NOS' in ICD rubric	15
'Specified' in ICD rubric	6
Awaiting addition as part of Orphanet work	1
Total number of matches to SCT	**14,557**
1:1 matches	6,777
1:M matches	1,117
M:1 matches	5,128
M:N matches	1,535