ICD-10 OMOP ADOPTION: MAPPING INCONSISTENCIES, ASSESSMENT OF POTENTIAL DATA LOSS AND THE SOLUTION TO THESE PROBLEMS

Polina Talapova MD[1], Dmitry Dymshyts MD[1], Christian Reich MD, PhD[2]

1 - Odysseus Data Services, Cambridge MA USA; 2 - IQVIA, Cambridge MA USA

BACKGROUND

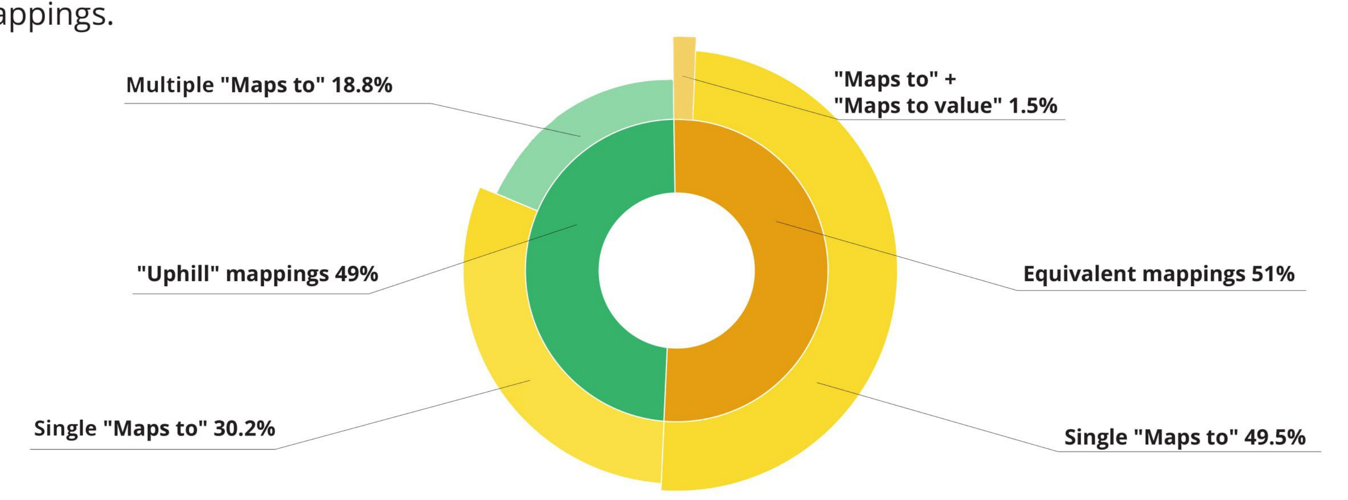
The 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) is known to be the most widely used medical classification worldwide. Many countries adapt ICD-10 for their health systems, and some have made modifications. In the OMOP Standardized Vocabularies, diseases (called Conditions) are represented by SNOMED-CT, a very rich and comprehensive ontology of medical entities with a fully developed description logic. But despite its size, SNOMED-CT does not cover all semantic entities of ICD-10, and mapping inconsistencies can arise.

In this study, we are focusing on potential causes of those inconsistencies and how they can be detected and rectified. Also, in order to simplify the work with this large corpus of information, we seek to familiarize researchers and scientists with proposed ways of ICD-10 mappings. This has an increased relevance with an addition of European country-specific ICD-10 extensions to the ICD-10 WHO.

The quality of the mapping today

Currently, ICD-10 to SNOMED mappings are implemented via “Maps to” and “Maps to value” relationships. “Maps to” can stand for either a full equivalence or an “uphill” mapping (a mapping to a more general semantic category) meanwhile “Maps to value” points to a value concept, indicating of a measurement or an observation.

For each ICD-10 concept, we categorized a type of its current relationship, counted the number of equivalents represented by a single “Maps to” and a “Maps to + Maps to value” combination, and nonequivalents represented by a single “uphill” (Single “Maps to”) and several “uphill” (Multiple “Maps to”) mappings.


Multiple "Maps to" 18.8%
"Maps to" +
"Maps to value" 1.5%
"Uphill" mappings 49%
Equivalent mappings 51%
Single "Maps to" 30.2%
Single "Maps to" 49.5%

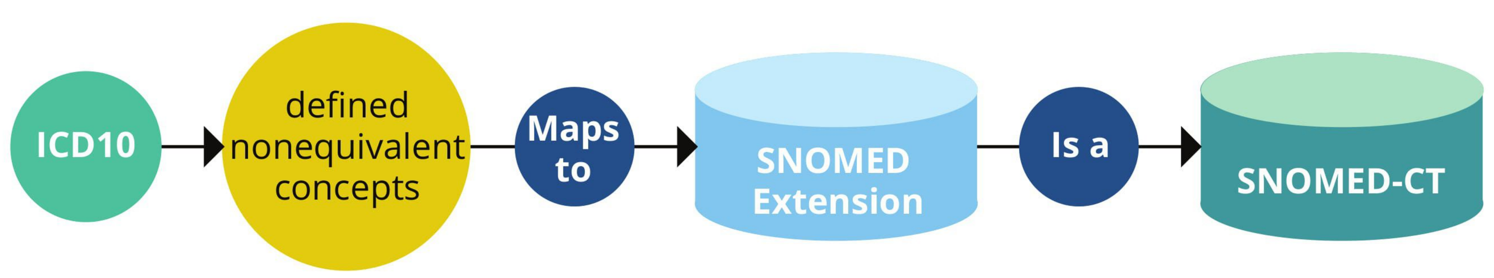
30.2% of ICD-10 concepts having a single “Maps to” relationship are used for mappings to non-equivalent SNOMED concepts. In total, slightly less than half of ICD-10 codes do not have equivalent mappings. This could be of concern with respect to data loss. Cohort definitions using standard concepts become more complicated, if the task is to recreate the exact same population that would have resulted from utilizing ICD-10 source codes.

To address that problem, we previously suggested adding SNOMED Extension concepts. These carry the full semantic equivalent of the source codes, but participate in the SNOMED hierarchy and use the same SNOMED attributes: Any ICD-10 concept that does not have a SNOMED equivalent will be paired with such a SNOMED Extension. As a result, all non-equivalent mappings will be distinguished from equivalent ones as follows:

- Equivalent mappings are represented through “Maps to” relationships from the source to the standard concepts.
- Nonequivalent mappings are represented through “Is a” relationships from SNOMED Extension concepts to concepts from the higher-level SNOMED corpus. The SNOMED Extension concepts are created for each source concept without an equivalent map. They are linked to each other through “Maps to” relationships.

SOLUTION

As a result, each ICD-10 source concept will have a “Maps to” relationship to either a SNOMED or a SNOMED Extension concept, embedding it into the overall Disease hierarchy. To accomplish that, we categorized all problematic cases leading to data loss in ICD-10 to SNOMED mappings and proposed the solution for each category.

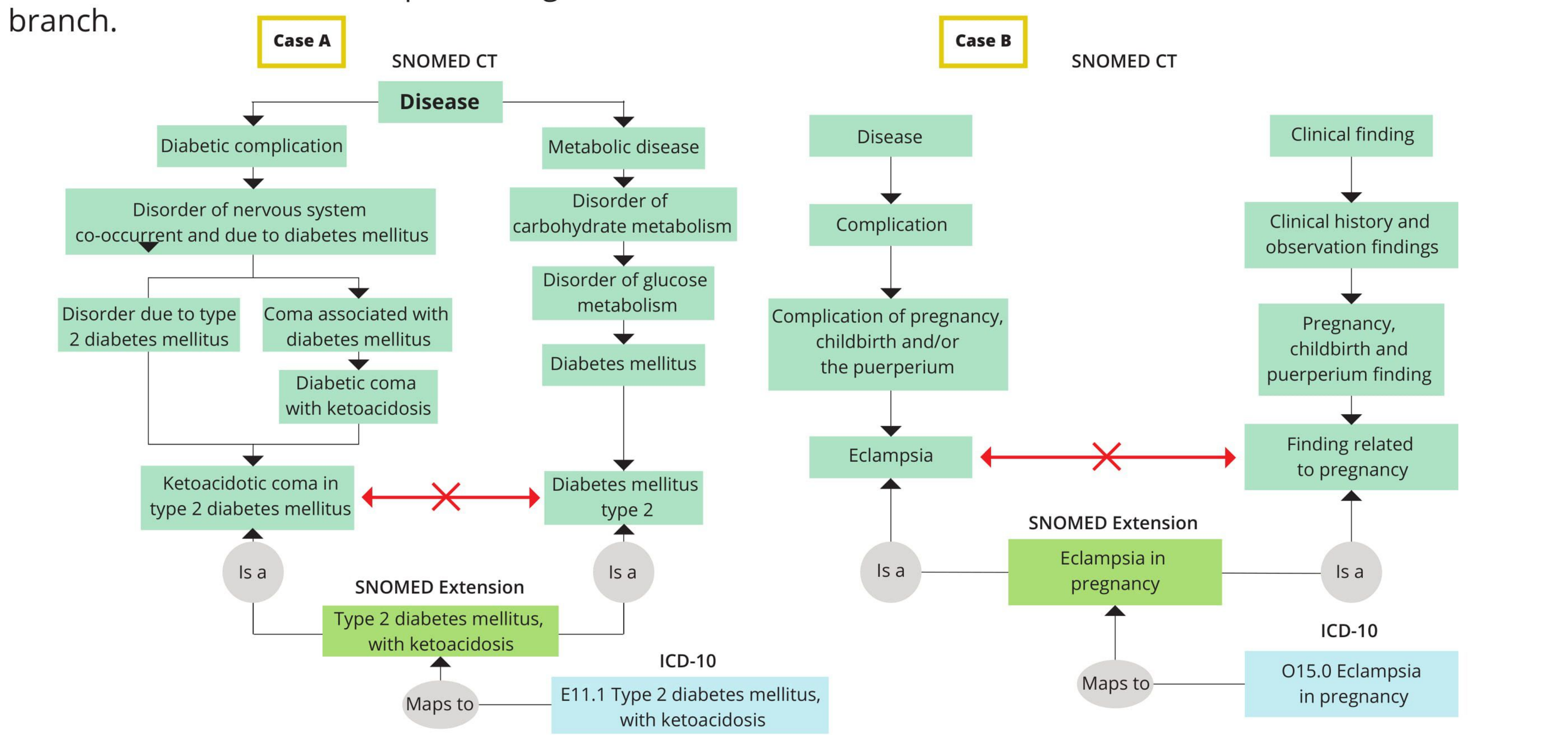

ICD10
defined nonequivalent concepts
Maps to
SNOMED Extension
Is a
SNOMED-CT

1. Hierarchical inconsistency between ICD-10 and SNOMED-CT

For example, in ICD-10 **diabetic complications** are a part of the disorder of **Diabetes Mellitus**, while in SNOMED the **diabetic complications** and **Diabetes Mellitus** proper reside in different hierarchical branches (Case A).

Another example are ICD-10 codes indicating **complications of pregnancy**, placed into the chapter “Pregnancy, childbirth and the puerperium” with a semantic relation to **the fact of a pregnancy**, while in SNOMED pregnancy denoting conditions and their complications are in separate hierarchical branches (Case B).

The SNOMED Extension representing them will have hierarchical "Is a" connections to either SNOMED branch.


Case A
SNOMED CT
Disease
Diabetic complication
Metabolic disease
Disorder of nervous system co-occurrent and due to diabetes mellitus
Disorder of carbohydrate metabolism
Disorder due to type 2 diabetes mellitus
Coma associated with diabetes mellitus
Disorder of glucose metabolism
Diabetic coma with ketoacidosis
Diabetes mellitus
Ketoacidotic coma in type 2 diabetes mellitus
Diabetes mellitus type 2
Is a
Is a
SNOMED Extension
Type 2 diabetes mellitus, with ketoacidosis
ICD-10
Maps to
E11.1 Type 2 diabetes mellitus, with ketoacidosis
Case B
SNOMED CT
Disease
Clinical finding
Complication
Clinical history and observation findings
Complication of pregnancy, childbirth and/or the puerperium
Pregnancy, childbirth and puerperium finding
Eclampsia
Finding related to pregnancy
SNOMED Extension
Is a
Eclampsia in pregnancy
Is a
ICD-10
Maps to
O15.0 Eclampsia in pregnancy

2. Missing ICD-10 concept equivalents

The following categories leading to the problem are distinguished:

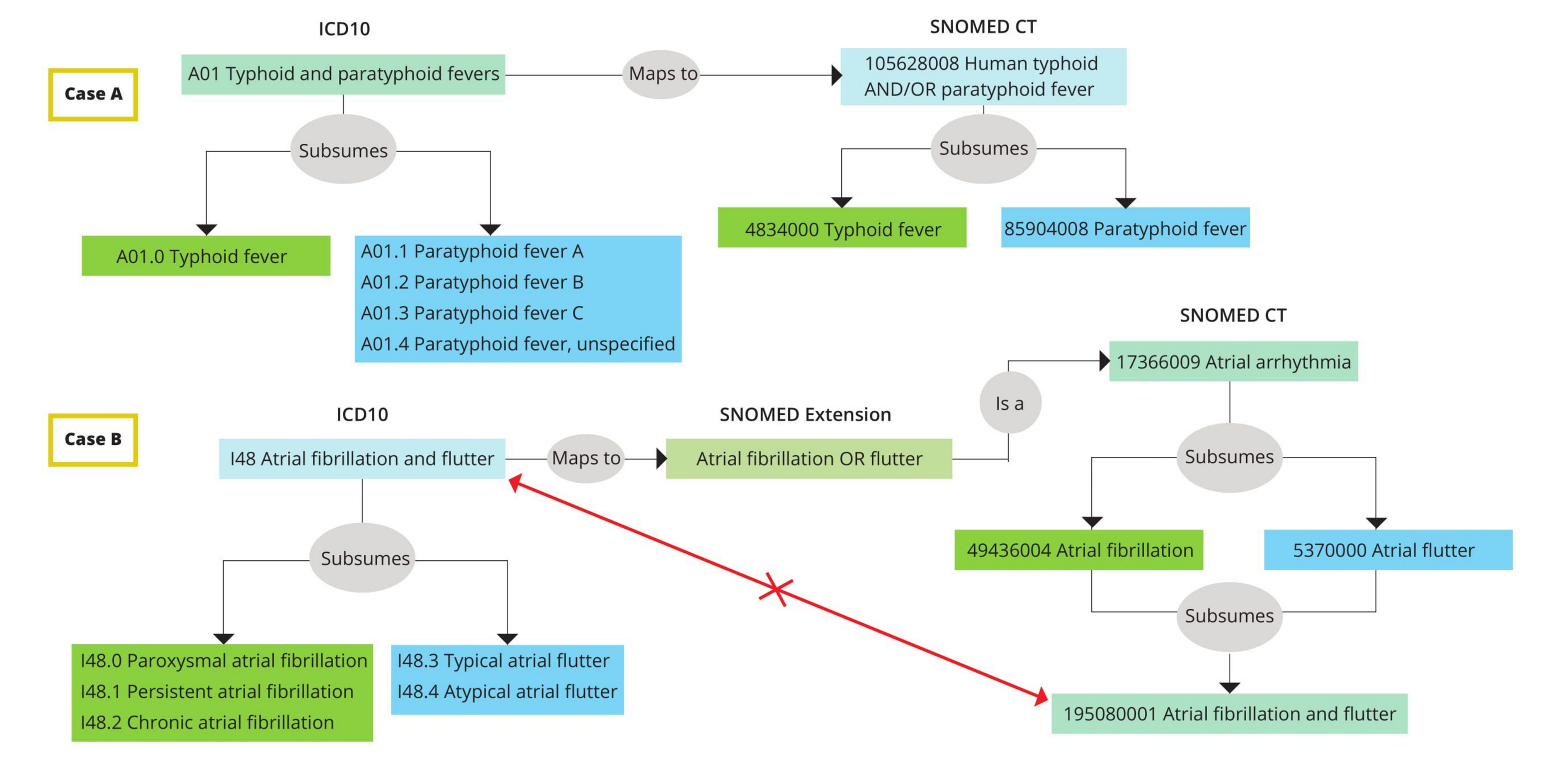
- **Vague formulation of ICD-10 concepts**
- **Excessive granularity of ICD-10 concepts**
- **Classification difference in comparison with SNOMED-CT**

category	icd_code	icd_name	relationship_id_1		relationship_id_2	snomed_concept_code	snomed_concept_name
Vague formulation	H18.3	Changes in corneal membranes	Maps to	the SNOMED Extension equivalent	Is a	15250008	Disorder of cornea
	A28.9	Zoonotic bacterial disease, unspecified	Maps to		Is a	87628006	Bacterial infectious disease
	M99.9	Biomechanical lesion, unspecified	Maps to		Is a	928000	Disorder of musculoskeletal system
Excessive granularity: mapping to a one SNOMED concept	B33.4	Hantavirus (cardio-)pulmonary syndrome [HPS] [HCPS]	Maps to		Is a	359761005	Disease due to Hantavirus
	D46.5	Refractory anaemia with multi-lineage dysplasia	Maps to		Is a	109996008	Refractory anemia
	I12.9	Hypertensive renal disease without renal failure	Maps to		Is a	38481006	Hypertensive renal disease
	M81.50	Idiopathic osteoporosis, multiple sites	Maps to		Is a	3345002	Idiopathic osteoporosis
Excessive granularity: mapping to more than one SNOMED concepts	S61.1	Open wound of finger(s) with damage to nail	Maps to		Is a	125653000	Open wound of finger
					Is a	275320004	Injury of nail
	I26.0	Pulmonary embolism with mention of acute cor pulmonale	Maps to		Is a	49584005	Acute cor pulmonale
					Is a	59282003	Pulmonary embolism
Classification difference	I48	Atrial fibrillation and flutter	Maps to		Is a	17366009	Atrial arrhythmia
	G06.0	Intracranial abscess and granuloma	Maps to		Is a	363176004	Inflammatory disorder of head
	M40	Kyphosis and lordosis	Maps to		Is a	64217002	Curvature of spine
	U82.1	Resistance to methicillin	Maps to		Is a	713728002	Drug resistance to antibacterial agen

A particular case of a classification difference is the array of ICD-10 codes such as “Condition A AND Condition B”.

The true meaning of such concepts can be interpreted as either “Condition A **AND** Condition B” or “Condition A **OR** Condition B”, that is confirmed by the structure of ICD-10. In the ICD-10 hierarchy, “Condition A AND Condition B” concepts with the genuine value of **“AND”** mostly do not have descendants. On the contrary, those with the true meaning of **“OR”** subsumes separate “Condition A” and “Condition B”. In turn, in SNOMED-CT, **“AND”** concepts should have ancestors: “Condition A” and “Condition B”, while **“OR”** concepts should have them as descendants.

Being based on such comparative analysis, the mapping happens to either an existing SNOMED analogue with the appropriate hierarchy (Case A) or to the equivalent SNOMED Extension concept, which goes “uphill” to the most common logical SNOMED-CT ancestor via “Is a” (Case B).


ICD10
SNOMED CT
Case A
A01 Typhoid and paratyphoid fevers
Maps to
105628008 Human typhoid AND/OR paratyphoid fever
Subsumes
Subsumes
A01.0 Typhoid fever
A01.1 Paratyphoid fever A
A01.2 Paratyphoid fever B
A01.3 Paratyphoid fever C
A01.4 Paratyphoid fever, unspecified
4834000 Typhoid fever
85904008 Paratyphoid fever
SNOMED CT
17366009 Atrial arrhythmia
Case B
ICD10
SNOMED Extension
Is a
I48 Atrial fibrillation and flutter
Maps to
Atrial fibrillation OR flutter
Subsumes
Subsumes
49436004 Atrial fibrillation
5370000 Atrial flutter
Subsumes
I48.0 Paroxysmal atrial fibrillation
I48.1 Persistent atrial fibrillation
I48.2 Chronic atrial fibrillation
I48.3 Typical atrial flutter
I48.4 Atypical atrial flutter
195080001 Atrial fibrillation and flutter

The solution we propose is creating of SNOMED Extension equivalent concepts with “Is a” relationships to existing SNOMED-CT for those cases where equivalence concepts do not exist.

3. ICD-10 codes indicating values of Measurements or content of Observations

These typically are not diseases in the direct sense:

1. Conditions indicating abnormal levels of tests
2. Encounter for procedure and aftercare
3. Conditions resulting from the use of devices
4. Need for immunization
5. History of a condition
6. Family history of a condition
7. History of medical treatment
8. Status of transplant/prosthesis presence
9. Persons encountering procedures, not carried out

To preserve the meaning of these concepts, “Maps to” will be complemented with a comprehensive set of “Maps to value” relationships. The following semantic groups falling into this category were defined.
In the table below a red font color stands for values of measurements or observed historical facts, while blue font color points a proper fact of a measurement or an observation.

case #	icd_code	icd_name	relationship_id	snomed_concept_code	snomed_concept_name	domain_id
1	R77.0	Abnormality of albumin	Maps to	26758005	Albumin measurement	Measurement
			Maps to value	263654008	Abnormal	Spec Disease Status
2	Z47.0	Follow-up care involving removal of fracture plate and other internal fixation device	Maps to	390906007	Follow-up encounter	Observation
			Maps to value	72010008	Removal of internal fixation device	Procedure
3	Y71	Cardiovascular devices associated with adverse incidents	Maps to	269691005	Medical accidents to patients during surgical and medical care	Observation
			Maps to value	303476002	Cardiovascular equipment	Device
4	Z23.2	Need for immunization against tuberculosis [BCG]	Maps to	723620004	Requires vaccination	Observation
			Maps to value	42284007	BCG vaccination	Procedure
5	Z86.4	Personal history of psychoactive substance abuse	Maps to	417662000	History of clinical finding in subject	Observation
			Maps to value	91388009	Psychoactive substance abuse	Condition
6	Z84.1	Family history of disorders of kidney and ureter	Maps to	416471007	Family history of clinical finding	Observation
			Maps to value	443820000	Disorder of kidney and/or ureter	Condition
7	Z92.2	Personal history of long-term (current) use of other medicaments	Maps to	416940007	Past history of procedure	Observation
			Maps to value	266713003	Long-term drug therapy	Procedure
8	Z96.3	Presence of artificial larynx	Maps to	416940007	Past history of procedure	Observation
			Maps to value	27926003	Construction of artificial larynx	Procedure
9	Z53.2	Procedure not carried out because of patient's decision for other and unspecified reasons	Maps to	416237000	Procedure not done	Observation
			Maps to value	105480006	Refusal of treatment by patient	Observation

4. Concepts with additional coding information

According to the ICD-10 coding guidelines [2], such ICD-10 concepts can be grouped by their meaning:

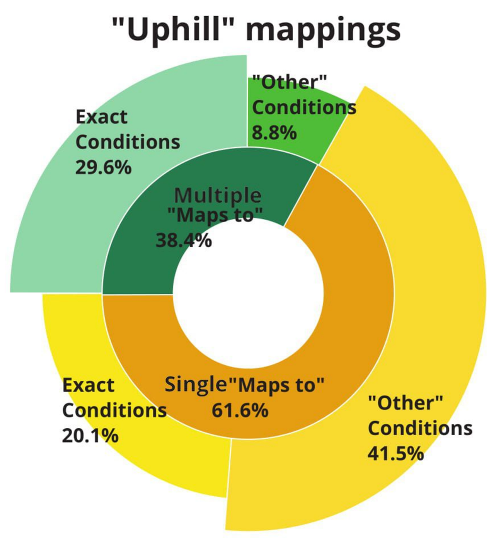
Group A contains concepts in which proper diagnosis is not fully defined: “**Not otherwise specified”** or **“unspecified”** indicates, that information in the medical record is insufficient to assign a more specific code. **“Other and unspecified”** represents mixed category of concepts: sub-types of diagnosis not encoded in the main chapter and unspecified types of the main chapter diagnosis. **“Classified elsewhere”** indicates the code, that can never be used as a principal diagnosis and it requires additional code with underlying condition. Hence, it is unspecified by itself. Thus, “Maps to” relationship is recommended to apply, because SNOMED has these general categories as well.

icd_code	icd_name	relationship_id	snomed_concept_code	snomed_concept_name
A80.3	Acute paralytic poliomyelitis, **other and unspecified**	Maps to	240460008	Acute paralytic poliomyelitis
E16.9	Disorder of pancreatic internal secretion, **unspecified**	Maps to	111560008	Disorder of pancreatic internal secretion
F99	Mental disorder, **not otherwise specified**	Maps to	74732009	Mental disorder
G59	Mononeuropathy in diseases **classified elsewhere**	Maps to	128189008	Mononeuropathy

Group B includes concepts indicating the absence of an exact diagnosis in the ICD-10 hierarchy: “Not elsewhere classified" or “Other (disorder/site)” or “Other specified” denotes, that the information in the medical record provides a detail for which a specific code does not exist, but we definitely know that it is not a condition defined by “sibling” concepts.

icd_code	icd_name	relationship_id_1		relationship_id_2	snomed_concept_code	snomed_concept_name
E88.1	Lipodystrophy, **not elsewhere classified**	Maps to	the SNOMED Extension equivalent	Is a	71325002	Lipodystrophy
N81.8	**Other** female genital prolapse	Maps to		Is a	73998008	Prolapse of female genital organs

So, these “other” concepts should be treated as exact semantic entity and require SNOMED Extension creation. See their distribution on the diagram at the right.


"Uphill" mappings
"Other" Conditions 8.8%
Exact Conditions 29.6%
Multiple "Maps to" 38.4%
Exact Conditions 20.1%
Single"Maps to" 61.6%
"Other" Conditions 41.5%

5. ICD-10 codes indicating malignant neoplasms

According to the ICD-10 guidelines [3] information declared, all malignant neoplasms not mentioned as secondary or unknown, are considered to be primary. Therefore, mappings from such “Malignant neoplasm” are pointed at “Primary malignant neoplasm”. However, these ICD-10 concepts can either have appropriate SNOMED equivalence (Group A) or not.

group	icd_code	icd_name	relationship_id	snomed_concept_code	snomed_concept_name
A	C34.1	Malignant neoplasm: Upper lobe, bronchus or lung	Maps to	372135002	Primary malignant neoplasm of upper lobe, bronchus or lung
	C50.2	Malignant neoplasm: Upper-inner quadrant of breast	Maps to	373089009	Primary malignant neoplasm of breast upper inner quadrant

For those, which do not have SNOMED analogues (Group B), SNOMED Extension concepts will be created and mapped to the SNOMED-CT concepts (denoting an original cancer site) through “Is a” relationships.

group	icd_code	icd_name	relationship_id_1		relationship_id_2	snomed_concept_code	snomed_concept_name
B	C22.4	Malignant neoplasm: Other sarcomas of liver	Maps to	The SNOMED Extension equivalent titled as "Primary malignant neoplasm of..."	Is a	254601002	Sarcoma of liver
					Is a	95214007	Primary malignant neoplasm of liver
	C47.2	Malignant neoplasm: Peripheral nerves of lower limb, including hip	Maps to		Is a	188324003	Malignant neoplasm of peripheral nerves of lower limb, including hip
					Is a	422736007	Primary malignant neoplasm of peripheral nerve
	C69.6	Malignant neoplasm: Orbit	Maps to		Is a	187906008	Malignant neoplasm of orbital bone
					Is a	93824007	Primary malignant neoplasm of head

CONCLUSION

In summary, the way to provide both improvement of mapping quality and prevention of data loss is the creation of the SNOMED Extension and the embedding into the existing SNOMED hierarchy through “Is a” relationships. This approach will solve the problem of concept mapping with a loss of semantic content. Other ontologies such as ICD-10-CM, ICD-9-CM etc. show similar problems. Therefore, the expected outcome: all ICD concepts will always be mapped to SNOMED CT or SNOMED Extension as one-to-one.

References:
1. Dymshyts D, Ostropolets A, Reich C, Belenkaya R. 2018. SNOMED Extension - Filling Gaps for Comprehensive OMOP Condition Vocabulary [Poster]. OHDSI Symposium.
2. Centers for Medicare and Medicaid Services [CMS]. ICD-10 Official Guidelines for Coding and Reporting [Internet]. 2019 [cited 2019 Feb 10]. Retrieved from: https://www.cms.gov/Medicare/Coding/ICD10/Downloads/2019-ICD10-Coding-Guidelines-.pdf
3. World Health Organization [WHO]. Chapter II: Neoplasms (C00-D48) [Internet]. 2003 [cited 2019 Feb 11]. Retrieved from: http://apps.who.int/classifications/apps/icd/icd10online2003/fr-icd.htm?kc00.htm