BACKGROUND

The 11th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-11) is known to be the most widely used medical classification worldwide. Many countries adopt ICD-10 for their health systems, and some have made modifications. In the OMOP Standardized Vocabulary (OSV), 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 these inconsistencies and how they can be detected and rectified. Also, in order to simply the work with this large corpus of information, we have focused on five leading hospitals in the USA, which have provided detailed data. This has increased our accuracy with an addition of European country-specific ICD-10 extensions to the ICD-10 WHO.

The quality of the mapping the ICD-10 CT to SNOMED-CT is not high. Currently, ICD-10 to SNOMED mappings are implemented via “Maps to” and “Maps to value” relationships. “Maps to” is defined as either a full equivalent concept, or a “uphill” mapping (a mapping to a more general semantic category) that aligns with “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 numbers of equivalents represented by a single “Maps to” and a “Maps to” or “Maps to value” combination, and non-equivalents represented by a single “uphill” (Single “Maps to”) and several “uphill” (Multiple “Maps to”) mappings.

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 this, we categorized all problematic cases leading to data loss in ICD-10 to SNOMED mappings and proposed the solution for each category.

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-CT diabetic complications and Diabetes Mellitus proper reside in different hierarchical branches (Case A1). 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-CT pregnancy denoting conditions and their complications are in separate hierarchical branches (Case B2).

The SNOMED Extension representing them will have Hierarchical “is a” connections to either SNOMED branch.

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**

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 is with the genuine value of “OR”, mostly do not have descendants. On the contrary, they both the true meaning of “OR” subsume 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.

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. Encounters for pain or fever
3. Conditions resulting from the use of devices
4. Need for immunization
5. History of a condition

To preserve the meaning of these concepts, “Maps to” will be complemented with a comprehensive set of “Maps to value” relationships. The following groups of concepts falling into these categories were defined:

- Conditions indicating abnormal levels of tests
- Encounters for pain or fever
- Conditions resulting from the use of devices
- Need for immunization
- History of a condition

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 a category of concepts. “Other” is the type of diagnosis not encoded in the main chapter and “unspecified” is the type 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 CT has these general categories as well.

- **Group B** includes concepts indicating the absence of an exact diagnosis in the ICD-10 hierarchy: “Not elsewhere classified” or “Other (disorder-titled)” 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 “is a” relationships.

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

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 neoplasms” are pointed at “Primary malignant neoplasm”. However, these ICD-10 concepts can either have appropriate SNOMED equivalence (Group A) or not.

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.

In summary, the way to provide both improvement of mapping quality and prevent 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 IC-CM, ICD-9-CM etc. have similar problems. Therefore, the solution will be mapped to SNOMED CT or SNOMED Extension as one-to-one.