# Synergism between the Mapping Projects from SNOMED CT to ICD-10 and ICD-10-CM

## Kin Wah Fung[1], MD, MS, MA; Junchuan Xu[2], MD, MS

**[1]National Library of Medicine, Bethesda, MD [2]Lockheed Martin Corporation, Bethesda, MD**

kwfung@nlm.nih.gov|xujc@nlm.nih.gov

## Abstract

*Two mapping projects are currently underway, creating maps from SNOMED CT to ICD-10 and ICD-10-CM respectively. Even though the two projects belong to different organizations, there has been a lot of synergism between them. The ICD-10-CM map project heavily re-used the mapping methodology, tools and map data developed in the ICD-10 map project. An algorithm was derived to generate candidate ICD-10-CM map records from the ICD-10 map. We evaluated the algorithm in 5,264 SNOMED CT concepts common to both maps. 4,317 ICD-10-CM candidate maps could be generated from the ICD-10 map and 3,341 (77%) of the generated maps agreed with the published map. By priming the mapping process with candidate maps generated from the other map project, significant saving in time and effort in future phases of the two projects can be anticipated. The reasons for the discordance between the generated map and published map were also analyzed.*

## Introduction

The need for mapping between controlled terminologies arises commonly to enable the re-use of data encoded in one terminology for another purpose that requires a different encoding system – the "code once, use multiple times" principle. [1-5] A myriad of medical terminologies exist with varying degree of overlap in their scope of coverage. Even terminologies with significant domain overlap can differ substantially in granularity, structure and organizing principles, usually as a result of the different purposes that they are designed for. A case in point is SNOMED CT (Systematized Nomenclature of Medicine--Clinical Terms) and the ICD (International Classification of Disease) classifications, which are very different even though they are both primarily related to the domain of findings, diseases and disorders. SNOMED CT is considered as the most comprehensive, multilingual clinical terminology in the world. [6] Since its inception in 2007, the International Health Terminology Standards Development Organisation (IHTSDO, owner of SNOMED CT) has doubled its number of member countries from nine to eighteen. SNOMED CT is slowly gaining momentum as the international clinical terminology standard. On the other hand, ICD-10 is a medical classification.[7] The root of ICD can be traced back to the International List of Causes of Death created 150 years ago. [8] ICD is endorsed by the World Health Organization (WHO) to be the international standard diagnostic classification for epidemiology, health management and clinical purposes. The current version of ICD is ICD-10 which was first published in 1992. Apart from reporting national mortality and morbidity statistics to WHO, many countries use ICD-10 for reimbursement and healthcare resource allocation. To better suit their national needs, several countries have created national extensions to ICD-10, including ICD-10-AM (Australia), ICD-10-CA (Canada) and ICD-10-CM (U.S.).[9-11]

A collaborative project between IHTSDO and WHO was started in 2008 to create a map between SNOMED CT and ICD-10. The aim of the map is to support semi-automated generation of ICD-10 classification data from a clinical record encoded in SNOMED CT. [12] The expected benefits of the map include:

- Re-use of clinical data for additional statistical purposes
- Rapid submission and response to national reporting requirements
- Saving time and improving efficiency for the coding professional
- Improved accuracy and reproducibility of code mapping
- Promulgation of widespread comparable epidemiologic and statistical data
- Cost saving for IHTSDO member countries which maintain ICD-10 derivative product maps

One of the stated goals of the map is to facilitate the development of maps to national extensions of ICD-10. A preview release of the SNOMED CT to ICD-10 Crossmap was published in September 2011. [13]

A separate project led by the U.S. National Library of Medicine (NLM) was started in 2010 to create a map between SNOMED CT and ICD-10-CM. At present, the U.S. is still using ICD-9-CM for reimbursement purpose. A transition date to ICD-10-CM on October 2013 has been announced, but currently a postponement of this date is under consideration by the CMS (Centers for Medicare & Medicaid Services). It is anticipated that the public availability of a map from SNOMED CT to ICD-10-CM will promote the use of SNOMED CT as the primary clinical terminology in the EHR (Electronic Health Record) by enabling the secondary generation of ICD-10-CM codes for administrative purposes.

Despite the big difference in size (14,000 codes in ICD-10 vs. 68,000 codes in ICD-10-CM), ICD-10 and ICD-10-CM have very similar structure and organizing principles. The foundation of ICD-10-CM is generally based on ICD-10. New codes are added as refinements to the ICD-10 codes in areas where additional specificity is considered necessary. Occasionally, there are more significant differences between the two. One such example is the deletion of the whole category *E14 Unspecified diabetes mellitus,* because diabetes mellitus of unspecified type is classified as *E11 Type 2 diabetes mellitus* in ICD-10-CM. Sometimes a new category is added in ICD-10-CM e.g. *K68 Disorders of retroperitoneum* which is not present in ICD-10. Both classifications are published as an alphabetical index and a tabular listing using very similar notations and conventions (e.g. essential and non-essential modifiers, exclusions and inclusions). The recommended coding procedure in both is essentially the same – first look up the term in the alphabetical index and then confirm the code in the tabular listing by checking the inclusion, exclusion and other notes. It is therefore not surprising that very similar methods can be used in mapping from SNOMED CT to ICD-10 and ICD-10-CM. This is also the reason for the high degree of synergism between the two mapping projects. Chronologically, the ICD-10 map was started first, and the ICD-10-CM map re-used a lot of the resources developed in the ICD-10 project. These resources included mapping methodology, tools and data. This paper describes the ways in which the two projects can benefit each other. We also report on a study of the efficacy of an algorithm we derived to generate candidate ICD-10-CM map records from the ICD-10 map.

**Three kinds of synergy**

1. Re-use of mapping methodology

The creation of a map from SNOMED CT to ICD-10 or ICD-10-CM is to create a bridge between two very different artifacts: a clinical terminology and a classification.[14, 15] It is not always possible to find a one-to-one correspondence between a SNOMED CT concept and an ICD code.[16] In some cases, more than one ICD code is required to fully encode the meaning of a SNOMED CT concept. For example, the SNOMED CT concept *301011002 Escherichia coli urinary tract infection* requires two ICD-10 codes for proper encoding: *N39.0 Urinary tract infection, site not specified* and *B96.2 Escherichia coli as the cause of diseases classified elsewhere*. In addition, there are cases in which alternative ICD codes exist for the same SNOMED CT concept, and which one to use depends on the patient's demographic characteristics (e.g. age, gender) and existing co-morbidities. For example, *73430006 Sleep apnea* can map to either *P28.3 Primary sleep apnea of newborn* or *G47.3 Sleep apnea*, depending on the patient's age. To allow the map to reflect the coding guidelines in ICD-10, a rule-based map structure was used. Alternative ICD map targets are represented as "map rules", and related map rules are grouped into "map groups". At run-time, the map rules are evaluated in a prescribed order, taking into consideration patient contextual information and co-morbidities. Each map group will resolve to at most one ICD-10 code in the final result. The same rule-based approach has been adopted in the ICD-10-CM map, using the same type of rules (age and gender rules) to handle patient context, and the same method to handle exclusion notes.[12]

| Map group | Rule order | Map rule | Target code | ICD description |
|---|---|---|---|---|
| 1 | 1 | IF AGE >= 29 DAYS | L08.82 | Omphalitis not of newborn |
| 1 | 2 | DEFAULT | P38.9 | Omphalitis (of newborn) without hemorrhage |
| 2 | 1 | IFA 403841009 \| staphylococcal omphalitis of newborn (disorder) | B95.8 | Unspecified staphylococcus as the cause of diseases classified elsewhere |
| 2 | 2 | IFA 403843007 \| streptococcal omphalitis of newborn (disorder) | B95.5 | Unspecified streptococcus as the cause of diseases classified elsewhere |
| 2 | 3 | DEFAULT | NULL | NULL |

Table 1. ICD-10-CM Map for the SNOMED CT concept "Omphalitis"

Table 1 illustrates the rule-based map using the ICD-10-CM map for the SNOMED CT concept *239095007 Omphalitis* as an example. It consists of two map groups and five map rules. In map group 1, the first rule is evaluated based on the patient's age. If the patient is not a neonate (defined as from birth to 28 days), the map target is *L08.82 Omphalitis not of newborn*. Otherwise, the default code is *P38.9 Omphalitis (of newborn) without hemorrhage*. If the infective agent is specified, as in two of the descendants of the source SNOMED CT concept *(403841009 staphylococcal omphalitis of newborn* and *403843007 streptococcal omphalitis of newborn)*, an additional ICD-10-CM code is needed. This triggers the creation of map group 2, in which the first two map rules point to the appropriate additional codes. If neither map rule 1 or 2 is satisfied, the second ICD-10-CM code is not needed. This is represented by the NULL map target in map rule 3.

Due to the additional level of specificity in ICD-10-CM, we have to make some modifications in the methodology. In ICD-10-CM, most injury-related diagnostic codes require episode of care information (e.g. initial encounter, subsequent encounter or sequela) for proper coding. Most SNOMED CT concepts do not specify episode of care. Without this information, no valid target codes can be assigned. We circumvent this difficulty by adding placeholder codes (replacing the $7^{th}$ character of the ICD-10-CM code by "?"), which allow us to identify the target code cluster. We add the map advice "EPISODE OF CARE INFORMATION NEEDED" to alert users that additional information will be needed to arrive at a valid ICD-10-CM code. Another difference between ICD-10-CM and ICD-10 is the specification of laterality (e.g. fracture of right hip), which is only used in ICD-10-CM. In most SNOMED CT concepts, laterality is not specified. In those cases, we can still map to a valid ICD-10-CM code for "unspecified laterality" as a default map target. We add a special map advice "CONSIDER LATERALITY SPECIFICATION" to alert the user that further specification of laterality is possible in ICD-10-CM. For similar reasons, two new map advices "CONSIDER TRIMESTER SPECIFICATION" and "CONSIDER WHICH FETUS IS AFFECTED BY THE MATERNAL CONDITION" are added for the possibility of further specifying the trimester of pregnancy, or a particular fetus in multiple pregnancy.

As a built-in mechanism for quality assurance, both projects employed dual independent mapping to reduce the variability in manual mapping. Every concept was mapped concurrently by two map specialists, who were terminology experts with knowledge in both terminology systems, and had received special training for the mapping projects. Only maps that were identical were accepted as final. Any discordance was reviewed by a third specialist, and escalated to wider discussion if necessary. The only exception to the dual mapping process was when an existing, reliable source of maps (which we called "legacy maps") was available. In those cases, we compared the legacy map to a manually-created map. Only when the two maps were different was it necessary for mapping by a second map specialist. Because of this, having an existing source of legacy map would save time by avoiding a second mapping in some cases. In the ICD-10 map, we used two sources of legacy maps: an existing map owned by the U.K, and synonymy maps identified through the UMLS (Unified Medical Language System). [2, 3] In the ICD-10-CM map, we used three legacy map sources: ICD-10-CM simple maps included in some of the Convergent Medical Terminology contents donated by Kaiser Permanente, [17] synonymy maps from the UMLS, and algorithmically generated ICD-10-CM maps based on the available ICD-10 maps from the ICD-10 map project (see 3. below).

2. Re-use of mapping tools

Three kinds of tools were developed for the mapping projects:

**a. ICD index viewer**
The search for a target code in either ICD-10 or ICD-10-CM began with the index. For ICD-10, there was already an electronic version of the index developed by the WHO with adequate searching functionality. However, there was no readily available ICD-10-CM index viewer suitable for the mapping project. We custom-built an ICD-10-CM index viewer, primarily modeled on the ICD-10 index viewer, with some enhancements for more efficient searching (Figure 1). Direct level-specific search in the first three levels of the index was added, so that the user could reach the code *I50.9 congestive heart failure* in one search. Other enhancements included: wild card search words, tooltip hints showing the full index hierarchy, spell check and universal search of the whole index.
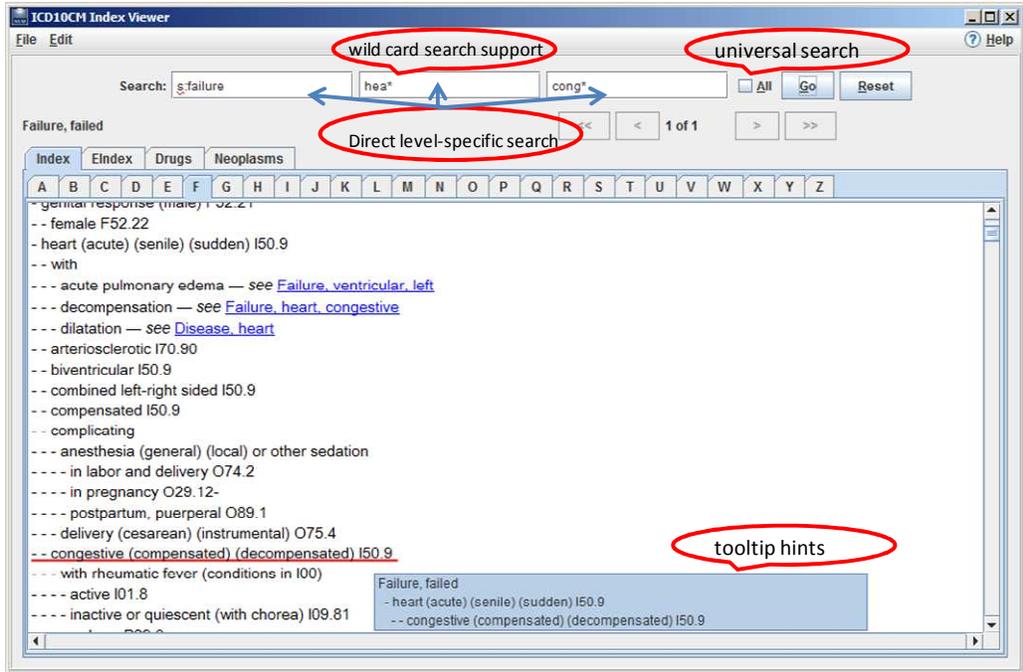
Figure 1. Screenshot of the ICD-10-CM index viewer

### b. Mapping tool

A new mapping tool was specially developed for the ICD-10 map project. It served two main functions: communicating with the map data repository and creating map records. Map specialists could connect with the repository, check out assigned batches, and check them back after mapping. In the mapping view, the user could explore the SNOMED CT hierarchy, view the defining relationships of the source SNOMED CT concept, check the ICD-10 codes in the tabular view (showing all coding notes, inclusions and exclusions), and create the map record for the source concept and its descendants, with the appropriate map rules and map advices. (Figure 2)
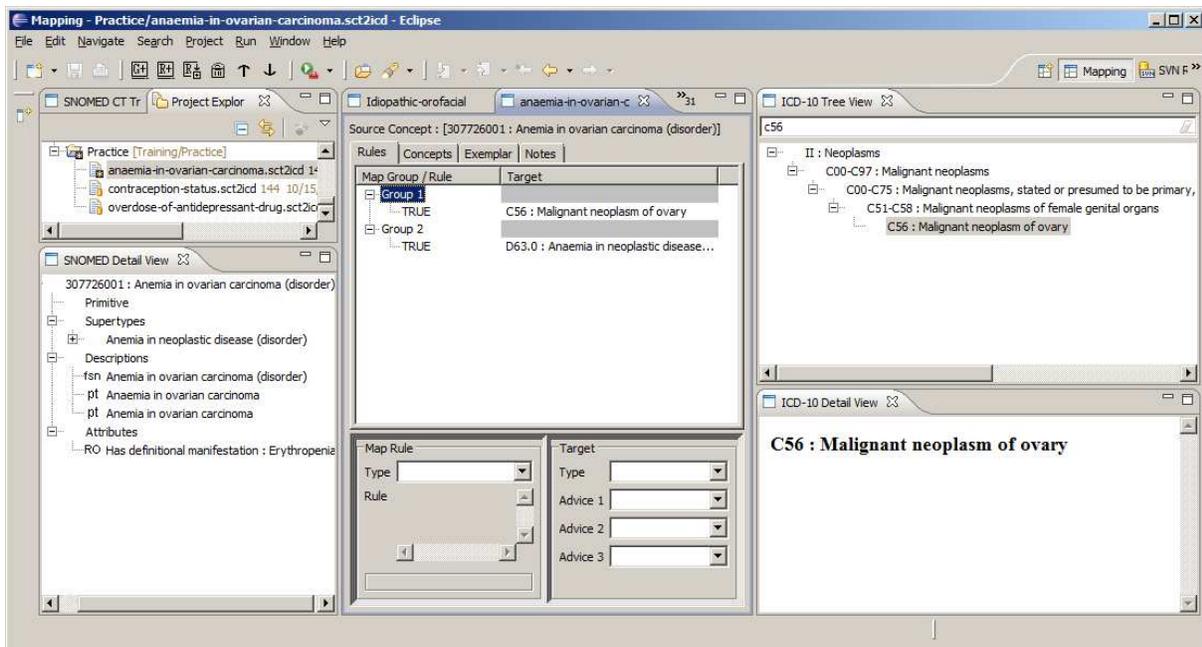


Figure 2. Screenshot of mapping tool (mapping view)

The same tool was used in the ICD-10-CM map project, with some minor modifications. Since the tool required loading of ICD-10 data in the ClaML (Classification Markup Language) format, the ICD-10-CM data had to be converted to the same format before loading into the tool. The map advices for the two maps were different and so the pick list for map advice field had to be modified.

**c. Map repository and workflow management tools**

For the ICD-10 project, a map repository was created to store the mapping data. The mapping tool communicated with the repository through secure Internet connection. In the repository, concepts to be mapped were algorithmically divided into batches (each containing up to 25 concepts). The batching program would try to keep concepts that were close to each other in the SNOMED CT hierarchy in the same batch as far as possible. Batches were assigned to map specialists for mapping. For the ICD-10-CM project, a similar repository was installed on an NLM server. Initially, the handling of workflow and batch assignment was done manually, which led to some data issues (e.g. typos in folder names). To circumvent this problem, we developed a set of web-based tools to automate the process. Automated scripts on the repository kept track of the status of the batches, ran comparison between candidate maps, and generated work statistics reports. These tools will improve workflow management in the next phase of the ICD-10 map project.

3. Re-use of mapping data

For concepts that already had a finalized map from the ICD-10 project, we derived an algorithm to generate a candidate ICD-10-CM map to use as legacy maps in the ICD-10-CM project. When a map target was valid for coding (codes that were at the lowest level of the hierarchy, or the "leaf nodes") in both ICD-10-CM and ICD-10, we simply used the same code as the map target. When an ICD-10 code was a non-leaf node in ICD-10-CM (therefore not valid for coding), we tried to find the valid "unspecified" descendant of that ICD-10-CM code. The full algorithm is as follows:

1. Exclude ICD10 codes that are not found in ICD-10-CM at any level e.g. *E14 unspecified diabetes mellitus*
2. ICD10 codes that are valid codes (leaf nodes) in ICD-10-CM – use directly as target e.g. *K70.0 Alcoholic fatty liver* is a valid code in both ICD-10 and ICD-10-CM
3. Otherwise, generate default ICD-10-CM targets as follows, using 7th digit placeholder character ("?", see Re-use of mapping methodology above) and padding digits ("x" as required in ICD-10-CM for some 7-digit codes) where necessary:
    a. For a 3 digit code (e.g. *R92 Abnormal and inconclusive findings on diagnostic imaging of breast*) – give up because there is usually no "unspecified" code at the 4th digit level
    b. For a 4 digit code
        i. Ending in .8 (e.g. *H10.8 Other conjunctivitis*) – propagate to .89 (e.g. *H10.8 Other conjunctivitis* becomes *H10.89 Other conjunctivitis*), further propagation to .899 if necessary (don't use .890)
        ii. Ending in other digits .x (e.g. *M75.5 Bursitis of shoulder*)
            1. Use .x0 (e.g. *M75.5 Bursitis of shoulder* becomes *M75.50 Bursitis of unspecified shoulder*) if present and valid; if present but not valid
                a. Use .x09 (e.g. *S31.1 Open wound of abdominal wall without penetration into peritoneal cavity* becomes *S31.109? Unspecified open wound of abdominal wall, unspecified quadrant without penetration into peritoneal cavity, episode of care unspecified*)
                b. If no .x09 - give up
            2. If no .x0 – give up (don't use .x9)
    c. For a 5 digit code .xy, use .xy9 (e.g. *S92.30 Fracture of unspecified metatarsal bone(s)* becomes *S92.309? Fracture of unspecified metatarsal bone(s), unspecified foot*)

**Methods**

At the start of the ICD-10-CM project, we implemented the above algorithm in our mapping environment to generate legacy ICD-10-CM map records from all the available maps from the ICD-10 map project. At that time, the ICD-

10 map was only about half-way done. To evaluate the full benefit that the algorithm would have brought, we re-applied the algorithm when the ICD-10 map was finished. We compared programmatically these generated map records with the published ICD-10-CM maps in the preview release of February 2012.[18] For the purpose of this study, we only looked at those concepts that were considered mappable to ICD-10 and ICD-10-CM, and excluded those which were out-of-scope or considered unmappable. These mappable concepts all had a non-null default map target in map group 1. For simplicity, we only compared the default map target in map group 1, and ignored any map rules, map advice and other map groups. For cases in which the generated ICD-10-CM maps were different from the published maps, we manually reviewed a sample for the reasons for discordance.

**Results**

Among the 19,119 SNOMED CT concepts that were mapped in the ICD-10 project, 18,538 had a non-null default map target in map group 1. The algorithm was able to identify a valid ICD-10-CM target code in 13,681 concepts (74% of 18,538). The preview release of the ICD-10-CM map contained 7,277 concepts, of which 7,230 concepts had a non-null default map target in map group 1. Among the 7,230 concepts, 5,264 concepts were also found in the ICD-10 map and 4,317 concepts had a candidate map generated from the ICD-10 map. Out of the 4,317 generated maps, 3,341 (77%) agreed with the published maps.

We analyzed the performance of the algorithm according to the characteristics of the ICD-10 codes. The results are summarized in Table 2. The majority (71%) of the generated maps came from ICD-10 target codes that were also valid ICD-10-CM codes. This group also had the highest rate of agreement (88%) with the published ICD-10-CM map. Among those ICD-10 codes that were not valid ICD-10-CM codes, the 4-digit codes resulted in 27% of the generated maps. The 4-digit codes ending in .8 were mostly "catch-all" codes (also called "residual classes") with the words "other" or "not elsewhere classified" in their names. They had the lowest level of agreement with the published maps. Most ICD-10 codes contained three or four digits. The 5th digit subclassification was mainly used in musculoskeletal conditions to specify the site of involvement (e.g. 1 – shoulder region, 2 - upper arm), and in injury codes to add more details (e.g. 0 - closed fracture, 1 - open fracture). The 5-digit codes made up only a small fraction of the generated maps, but had a high rate of agreement (79%).

We manually reviewed a random sample of 100 maps among the 976 maps in which the generated map did not agree with the published map (the discordant maps). The distribution of the reviewed maps according to the ICD-10 code type is shown in Table 3, which affirms that the composition of the reviewed sample closely resembles that of all discordant maps.

| Type of ICD-10 code | Generated map targets (%) | Agreed with published maps (%) |
|---|---|---|
| Also valid in ICD-10-CM | 3,078 (71%) | 2,698 (88%) |
| 4-digit ending in .8 | 192 (4%) | 67 (35%) |
| 4-digit not ending in .8 | 974 (23%) | 518 (53%) |
| 5-digit | 73 (2%) | 58 (79%) |
| Total | 4,317 (100%) | 3,341 (77%) |

Table 2. Characteristic of ICD-10 codes and the agreement of the generated maps with the published map

| Type of ICD-10 code | Discordant map targets (%) | Discordant map targets reviewed (%) |
|---|---|---|
| Also valid in ICD-10-CM | 380 (39%) | 37 (37%) |
| 4-digit ending in .8 | 125 (13%) | 10 (10%) |
| 4-digit not ending in .8 | 456 (47%) | 50 (50%) |
| 5-digit | 15 (2%) | 3 (3%) |
| Total | 976 (100%) | 100 (100%) |

Table 3. Composition of all discordant maps and the reviewed sample

The reasons for the discordance between the generated and published maps fell into four categories:

**1. Additional detail in SNOMED CT concept: 63 concepts**
These were cases in which there was additional detail in the SNOMED CT concept that was not included in the ICD-10 target code, but that additional detail could be captured in ICD-10-CM. For example, the SNOMED CT concept *83291003 Cor pulmonale* was mapped to *I27.9: Pulmonary heart disease, unspecified* in ICD-10. However, in ICD-10-CM, there were more detailed diagnostic codes available for this particular condition. So even though the same code I27.9 was also valid in ICD-10-CM, the correct target of this map was *I27.81 Cor pulmonale.*

**2. Structural difference between ICD-10 and ICD-10-CM: 2 concepts**
These were cases in which there were structural differences between ICD-10 and ICD-10-CM which led the algorithm to generate the wrong code in ICD-10-CM. One example was the map from the SNOMED CT concept *414564002 Kyphosis deformity of spine*. In ICD-10, this was mapped to *M40.29: Other and unspecified kyphosis*, in which the $5^{th}$ digit "9" represented the site of involvement ("9" means "site unspecified"). The correct map target in ICD-10-CM should be *M40.209 Unspecified kyphosis, site unspecified*, but the code generated by the algorithm was *M40.299 Other kyphosis, site unspecified*. In this case, the $5^{th}$ digit in the ICD-10-CM code was not specifying the site of involvement. Another example was the map for the SNOMED CT concept *23148009 Undifferentiated attention deficit disorder*. In ICD-10, this was mapped to *F98.8 Other specified behavioral and emotional disorders with onset usually occurring in childhood and adolescence* which had the inclusion note for "Attention deficit disorder without hyperactivity". However, in ICD-10-CM, the specific inclusion note under F98.8 was deleted. Instead, all cases of attention-deficit hyperactivity disorders were classified under *F90*. In addition, the name of the category *F90* was changed from "Hyperkinetic disorders" in ICD-10 to "Attention-deficit hyperactivity disorders" in ICD-10-CM. So the correct target code in this case was *F90.8: Attention-deficit hyperactivity disorder, other type.*

**3. Different interpretation of the meaning of terms: 29 concepts**
A certain level of judgment was sometimes involved in creating the maps, particularly when the ICD indexes did not give an exact index trail leading to a specific target code. One example was the map for the SNOMED CT concept *247404004 Allodynia* (meaning painful response to a stimulus which does not normally provoke pain). The term "allodynia" was not found in either of the indexes. In the ICD-10 map, this was mapped to *R20.8 Other and unspecified disturbances of skin sensation*, while in the ICD-10-CM map it was mapped to *R20.3 Hyperesthesia*. Another example was *235918000 Obstruction of biliary tree*, which was mapped to *K83.8 Other specified diseases of biliary tract* in ICD-10 and *K83.1 Obstruction of bile duct* in ICD-10-CM.

**4. Different order of target codes: 6 concepts**
The guidelines for the order of codes when multiple codes are used could be different for ICD-10 and ICD-10-CM. Since we were only looking at the first map group, some cases were wrongly identified as discordant because the generated map target was located in a different map group. For example, in the map for the SNOMED CT concept *234171009 Drug-induced hypotension*, the map target *I95.2 Hypotension due to drugs* was put in map group 1 in the ICD-10 map. The same target code was used in the ICD-10-CM map, but it was put in map group 2.

**Discussion**

It will be an additional incentive to use SNOMED CT as the primary terminology in the EHR if the coded data can be re-used for purposes other than clinical care. The output of ICD codes is an important function of many EHRs. Providing a map from SNOMED CT to the ICD-10 will help to promote the adoption of SNOMED CT internationally. Similarly, mapping from SNOMED CT to national extensions of ICD-10 will encourage the uptake of SNOMED CT nationally. Since the national extensions are structurally and semantically based on the international core, it is logical to employ the same method to map them. As we have demonstrated in the two mapping projects, there was heavy re-use of methods, tools and data between the ICD-10 and ICD-10-CM map projects. In addition, there was also sharing of documentation, training materials and mapping expertise.

Mapping is a labor-intensive process. Even for trained mappers, it took on average 10 minutes to map a SNOMED CT concept to ICD-10 or ICD-10-CM. In addition, we have decided to employ dual independent mapping which we think is necessary as a built-in quality assurance mechanism to ensure that the maps are consistent and reproducible. This significantly increased the mapping effort required. If we had to map every one of the 20,000 concepts (the full scope of Phase 1 of the ICD-10 map project) twice, it would have taken 6,700 hours (over 40 months) to create the

candidate maps alone, not to mention the time required to resolve the differences between candidate maps. It was therefore very important to find ways to cut down the mapping workload. We found that when we had a reliable pre-existing (or legacy) map, the average number of maps required for a concept was 1.3 (instead of 2) maps, not an insignificant saving. To maximize the number of legacy maps for the ICD-10-CM map project, we have developed an algorithmic way to generate candidate ICD-10-CM maps from the ICD-10 map. The ICD-10 map was not completely done when the ICD-10-CM project began, so we were not able to benefit from the full set of the ICD-10 Phase 1 map. The results reported in this study pertain to the theoretical benefit that the ICD-10-CM map could have derived from the full ICD-10 map, had it been available from the start.

It was encouraging to find that the algorithm was able to generate a candidate ICD-10-CM map in 74% of mappable concepts, and that 77% of the generated maps actually agreed with the final map. The rate of agreement would be even higher if we included those cases in which the agreement was not recognized because of the order of the map groups. The bulk of the generated maps originated from codes that were valid in both ICD-10 and ICD-10-CM. Not surprisingly, these were also the cases that had the highest rate of agreement with the published maps. When a disease code was a leaf node in both ICD-10 and ICD-10-CM, it usually meant that ICD-10-CM did not add refinement to that code and so the same code was likely to be the correct map target for that SNOMED CT concept. On the other hand, in 30% of the generated maps, the ICD-10 map target was not a valid code in ICD-10-CM. Most of these maps were generated from a 4-digit ICD-10 code. Among the maps generated from 4-digit ICD-10 codes, the ones originating from codes ending in .8 had a lower rate of agreement with the published maps, compared to those ending in other digits. This was understandable because in ICD-10, codes ending in .8 (the "catch-all" codes with "other" or "not elsewhere classified" in their descriptions) were heterogeneous classes that were often targets for refinement in ICD-10-CM, where more specific codes were added either as siblings or descendants. For example, the SNOMED CT concept *2055003 Recurrent erosion of cornea* was mapped to *H18.8 Other specified disorders of cornea* in ICD-10. In ICD-10-CM, under *H18.8* there was a leaf node *H18.839 Recurrent erosion of cornea, unspecified eye* which was a better target code than *H18.899: Other specified disorders of cornea, unspecified eye*, the candidate target code generated by the algorithm. Despite the relatively low precision for the .8 codes, it would not be advisable to exclude them. In the way that the generated maps were used, the wrong maps would be discarded on comparison with the manually-created map, and the concept would be mapped by a second map specialist, which would have happened anyway without the generated map, so no real harm was done. In other words, the recall of the algorithm was more important than the precision in this situation.

The most common reason for discordance between the generated ICD-10-CM candidate and the published maps was that the SNOMED CT concept contained additional information not covered in the ICD-10 map target, but was covered by a more specific code in ICD-10-CM. Since SNOMED CT was much bigger and had a finer granularity than either ICD-10, it was inevitable that in some cases, the best available map for a SNOMED CT concept did not cover its full meaning. This will remain to be an intrinsic limitation of the algorithm because there is no easy way to identify such cases and to suggest alternative target codes.

So far, we have focused on the use of the ICD-10 map data to prime the ICD-10-CM map project because the maps were created in that order. Going forward, in subsequent phases of the two map projects, we can envisage that the ICD-10-CM map data can also be used to prime the ICD-10 map project. Generating ICD-10 candidate maps from ICD-10-CM should be even easier to do. Most of the time, one only needs to truncate the ICD-10-CM code sequentially until a valid ICD-10 code is reached. There can be some situations in which this will not work. For example, there are some ICD-10-CM categories that are not present in ICD-10 e.g. *K68 Disorders of retroperitoneum*. In addition, there are a few situations in which ICD-10 is actually more granular that ICD-10-CM. In such cases, a leaf node in ICD-10-CM is a parent node in ICD-10. An example is the code *R52 Pain, unspecified* which is a leaf node in ICD-10-CM but has descendants in ICD-10 (e.g. *R52.0 Acute pain, R52.1 Chronic intractable pain*). Generally, these special cases are few and far between. To assess the feasibility of the reverse generation of ICD-10 candidate maps, we did a simple comparison of the same map datasets. In the 5,264 concepts in this study which had both ICD-10 and ICD-10-CM maps, 4,197 concepts (80%) had ICD-10 target codes that could be generated simply by truncation of the ICD-10-CM target codes. This suggests that the yield in the reverse direction will also be quite high, but this needs to be confirmed in subsequent phases of the ICD-10 project.

Whether the findings in this study can be generalized to other national ICD-10 extensions (e.g. ICD-10-CA, ICD-10-AM) remains to be seen. Theoretically, these extensions bear similar relations to ICD-10, and it is possible that candidate maps to the extensions can be generated from the ICD-10 maps in a similar algorithmic way. If that is the

case, it will be interesting to see whether the generated maps will agree with the actual maps to the same extent as in ICD-10-CM.

**Conclusion**

The NLM's ICD-10-CM map project re-used a lot of the resources developed for the IHTSDO's ICD-10 map project. These resources included mapping methodology, documentation, tools and map data. Re-use of the map data was possible through algorithmic generation of candidate ICD-10-CM map records from the ICD-10 map. A high proportion of the generated map records agreed with the published ICD-10-CM map. This could lead to significant saving in mapping effort. Going forward, the two projects can mutually learn and benefit from each other in their future phases of activity.

**References**

1.      Bowman S. Coordinating SNOMED-CT and ICD-10. J AHIMA 2005;76:60-1.

2.      Fung KW, Bodenreider O. Utilizing the UMLS for semantic mapping between terminologies. AMIA Annu Symp Proc 2005:266-70.

3.      Fung KW, Bodenreider O, Aronson AR, et al. Combining lexical and semantic methods of inter-terminology mapping using the UMLS. Medinfo 2007;12:605-9.

4.      Cimino JJ, Barnett GO. Automated translation between medical terminologies using semantic definitions. MD Comput 1990;7:104-9.

5.      Dolin RH, Huff SM, Rocha RA, et al. Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies. J Am Med Inform Assoc 1998;5:203-13.

6.      International Health Terminology Standards Development Organisation, owner of SNOMED CT http://www.ihtsdo.org/.

7.      International Classification of Diseases (ICD), World Health Organization http://www.who.int/classifications/icd/en/.

8.      History of the development of the ICD, World Health Organization, http://www.who.int/classifications/icd/en/HistoryOfICD.pdf.

9.      International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), National Center for Health Statistics, http://www.cdc.gov/nchs/icd/icd10cm.htm.

10.     ICD-10-AM International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification, http://nccc.uow.edu.au/icd10am/icd10am/index.html.

11.     ICD-10-CA, Canadian Institute for Health Information website http://www.cihi.ca.

12.     Mapping SNOMED CT to ICD-10 Technical Specifications, Mapping Special Interest Group, IHTSDO, http://www.ihtsdo.org/develop/documents/activities-to-support-crossmaps/icd-10-crossmaps/preview-release/.

13. IHTSDO ICD-10 Crossmap Preview Release, http://www.ihtsdo.org/news/article/article/ihtsdo-icd-10-crossmap-preview-release/.

14. Chute CG. Clinical classification and terminology: some history and current observations. J Am Med Inform Assoc 2000;7:298-303.

15. Ingenerf J, Giere W. Concept-oriented standardization and statistics-oriented classification: continuing the classification versus nomenclature controversy. Methods Inf Med 1998;37:527-39.

16. Imel M. A closer look: the SNOMED clinical terms to ICD-9-CM mapping. J AHIMA 2002;73:66-9; quiz 71-2.

17. Convergent Medical Terminology (CMT) donated by Kaiser Permanente, http://www.nlm.nih.gov/research/umls/Snomed/cmt.html.

18. SNOMED CT to ICD-10CM Map, Preview Release, http://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm.html.