

Handling Age Specification in the SNOMED CT to ICD-10-CM Cross-map

Junchuan Xu MD, MS¹, Kin Wah Fung MD, MS, MA²

¹Lockheed Martin Corporation, Bethesda, MD

²National Library of Medicine, Bethesda, MD

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

ABSTRACT

A SNOMED CT-encoded problem list will be required to satisfy the Certification Criteria for Stage 2 “Meaningful Use” of the EHR incentive program. ICD-10-CM will be replacing ICD-9-CM as the reimbursement code set in the near future. Having a cross-map from SNOMED CT to ICD-10-CM will promote the use of SNOMED CT as the primary problem list terminology, while easing the transition to ICD-10-CM. This rule-based map will support semi-automatic generation of ICD-10-CM codes from SNOMED CT-encoded data. Among the different types of rules, the age rule is used to handle age-specific code assignment in ICD-10-CM. To supplement the manual process of creation of age rules, a special QA process was implemented to flag maps that were potentially missing age rules. The QA flagged 342 concepts for review (out of 7,277), of which 172 concepts (50.3%) were true positives. Without the special QA, many of the age rules would have been missed.

INTRODUCTION

SNOMED CT and the Problem List

The problem list has become a key feature in electronic health records (EHR). The origin of the problem list can be traced back to 1968 when Larry Weed¹ first introduced the concept of a “Problem Oriented Medical Record” (POMR). A typical problem list includes information like medical diagnoses, abnormal clinical findings, laboratory results, social problems and risk factors. The problem list serves as a summary of the patient’s problems. It provides a concise overview of patient’s health conditions, particularly of patients in aging population suffering from multiple chronic illnesses. The problem list is also an effective means of communication among healthcare providers. It makes the patient’s current problems and important co-morbidities readily available to each provider for patient care.

Coded problem list data, as opposed to free-text narrative problem statements, are more useful in driving other functions of the EHR^{2,3,4}, such as clinical decision support⁴. Among candidates of standard terminologies that can be used as coding schema in electronic problem list, SNOMED CT is the most comprehensive⁵. The recently announced NPRM (Notice of Proposed Rule Making) for Stage 2 of Meaningful Use includes specific emphasis on standard health data exchange⁶, and SNOMED CT is proposed in the EHR Certification Criteria as the terminology standard for the problem list⁷.

SNOMED CT to ICD-10-CM mapping

One common requirement of the problem list in an EHR is to support the generation of administrative codes for reimbursement and statistical purposes. The administrative coding system used to report patient health conditions in US is the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). ICD-9-CM will soon be replaced by its successor, ICD-10-CM, which will provide richer and more up-to-date content^{8,9}. The initial announced date for the transition was Oct 1, 2013. This deadline for implementation is currently under review by the Center of Medicare & Medicaid Services (CMS), with a possible postponement to Oct 2014. The migration from ICD-9-CM to ICD-10-CM will have great impact on healthcare information systems¹⁰. ICD codes are used not only for analytic and reporting on morbidity/mortality statistics, reimbursement, and some clinical decision support, they are also used in many business processes like health care pricing, risk prediction, case management planning and many other critical business functions.

To support the use of SNOMED CT as the primary problem list terminology, and to facilitate the transition to ICD-10-CM, the National Library of Medicine (NLM) is undertaking a project to create a cross-map from SNOMED CT to ICD-10-CM. The anticipated benefits of the map include: re-use of clinical data for generation of administrative codes, and improved efficiency and reproducibility of ICD-10-CM coding ¹¹.

Rule-based mapping

In the creation of the SNOMED CT to ICD-10-CM cross-map (hereafter referred to as the “Map”), it is not always possible to find a one-to-one correspondence between a SNOMED CT concept and an ICD-10-CM code. One SNOMED CT concept may require multiple ICD-10-CM codes to fully represent its meaning. Sometimes, the same SNOMED CT concept can be mapped to several alternative ICD-10-CM codes, depending on patient context and co-morbidities. So the Map needs to be a rule-based map to provide the flexibility and logic required in its usage. One factor that often affects the choice of ICD-10-CM codes is patient age. The age rule is one type of rule to handle cases in which one SNOMED CT concept can map to different ICD-10-CM codes depending on the age of the patient. For example, the SNOMED CT concept “Urinary tract infection” is mapped generally to N39.0: *Urinary tract infection, site not specified*. But if the problem occurs in newborn, the proper code should be P39.3: *Neonatal urinary tract infection*. The rule-based Map can be embedded in the problem list of the EHR to support real-time, interactive generation of ICD-10-CM codes based on SNOMED CT encoded clinical problems. This is known as the I-MAGIC (Interactive Map-Assisted Generation of ICD-10-CM codes) use case. A typical scenario will be as follows. At the end of a clinic encounter, the clinician enters the patient’s problems in the problem list, and the entries are automatically encoded in SNOMED CT in the background. Based on the SNOMED CT-encoded problems and additional patient context information (gender and age) from the EHR, the Map-enabled application will output a list of candidate ICD-10-CM codes. When necessary, the clinician will be prompted for additional input to decide between alternative target codes or to refine the default codes. Figure 1 is a flow-chart showing how the age rule will help to determine the appropriate target code.

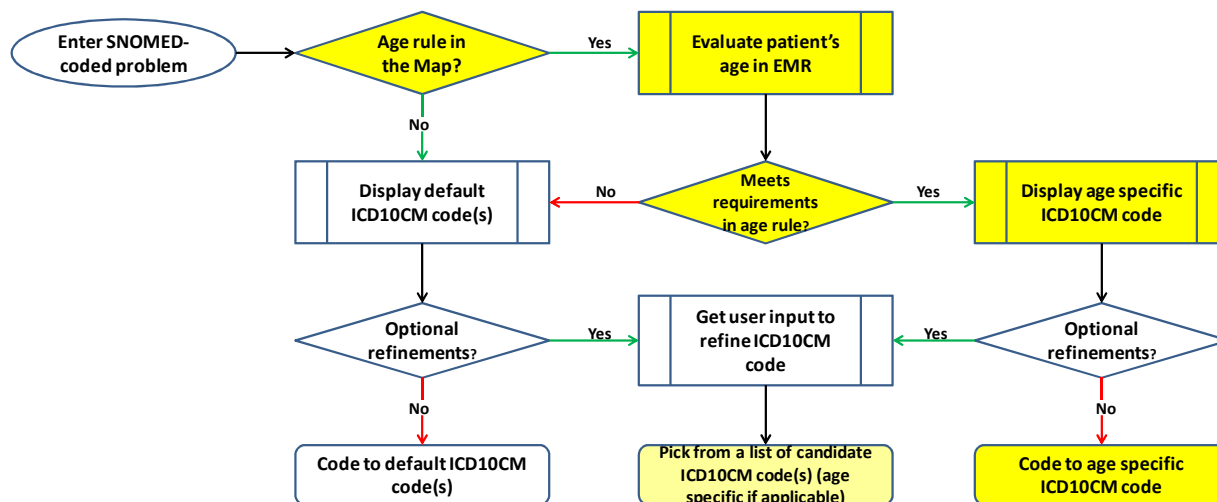


Figure 1: The use of age rule to determine the appropriate target ICD-10-CM code (highlighted boxes are specifically related to the age rule).

Age rule creation

In the creation of the Map, map specialists follow the standard procedure of coding, as recommended in the ICD-10-CM documentation. Firstly, the ICD-10-CM alpha index is searched using lookup terms. The identified codes are then confirmed by checking the tabular list of ICD-10-CM. The indication for an age rule usually comes from the alpha index. An age specific entry is found under the main entry as an indented entry (Figure 2). This is usually accompanied by a corresponding entry in the exclusion notes of the tabular list (Figure 3). For “Urinary tract infection”, one of the exclusion note for code N39.0: *Urinary tract infection, site not specified* is “Excludes1: neonatal urinary tract infection P39.3”. However, there are cases in which the age specification is only listed in the alpha index, without a corresponding exclusion note in tabular list.

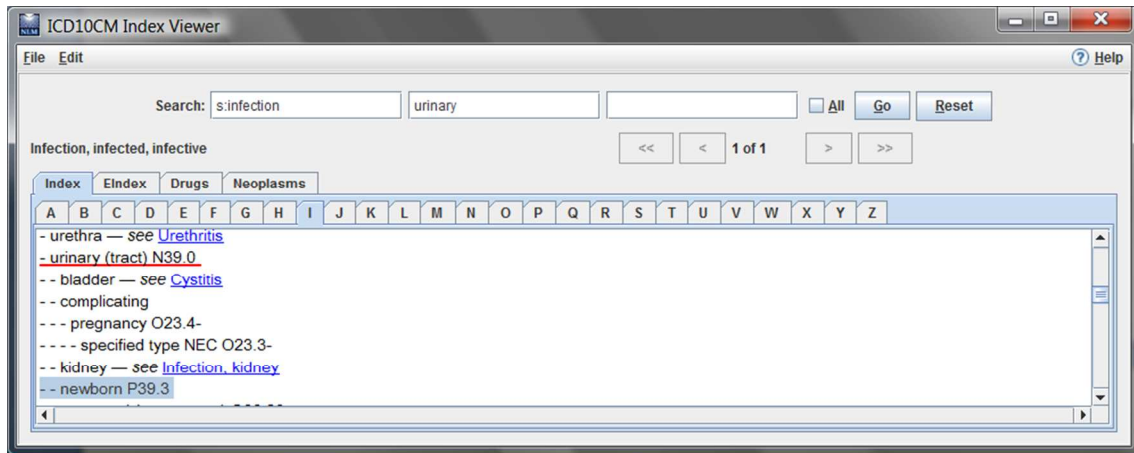


Figure 2: Screen shot of the ICD-10-CM Index Viewer, a custom-built browser for this mapping project, showing age specification listed in alpha index

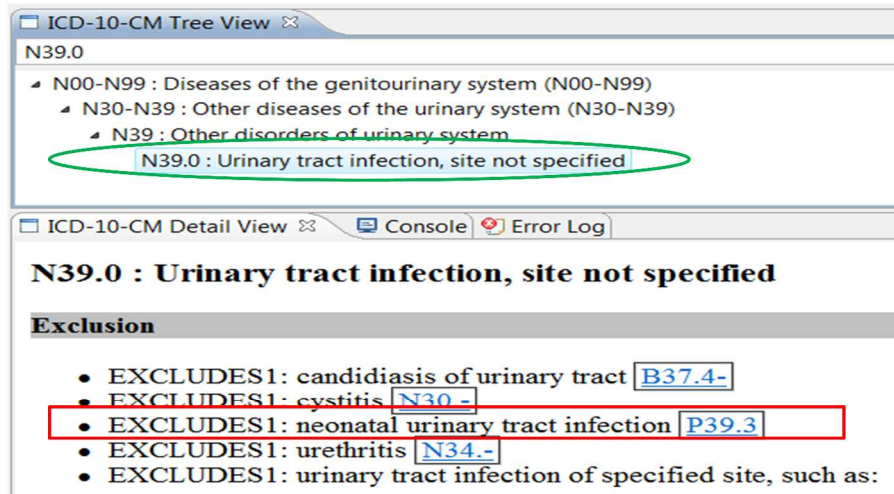


Figure 3: Screen shot of the ICD-10-CM tabular list in a custom-built mapping tool for this mapping project, showing exclusion note corresponding to age specification listed in alpha index in Figure 2

The age specification usually refers to a phase of life (e.g. neonate, adulthood). To convert it to a machine-processable age rule, we need to define the age ranges for these phases of life. The following list of reference age ranges was established and confirmed with ICD-10-CM experts (Table 1).

Age rule category	Definition	Age range
Perinatal	22 weeks of gestation to 7 days of life	age < 8.0 days
Neonatal	Birth to 28 days of life	age < 29.0 days
Infant (infancy)	Birth to less than 1 year of age	age < 1.0 year
Juvenile	1 year of age to less than 18 years of age	1.0 years =< age < 18.0 years
Adolescence	1 year of age to less than 18 years of age	12.0 years =< age < 18.0 years
Childhood	Less than 18 years of age	age < 18.0 years
Adult	18 years of age and older	18.0 years =< age
Pre-senile	less than 65 years of age	age < 65.0 years
Senile	65 years of age and older	65.0 years =< age

Table 1: Age ranges for age rule

According to the above definition, the age rule of the map for concept “Urinary tract infection” is represented as in Table 2.

Rules	Logic	Target ICD code	Target code description
Age Rule	IF AGE < 29.0 days	P39.3	Neonatal urinary tract infection
Default	ELSE	N39.0	Urinary tract infection, site not specified

Table2: Representation of the age rule in the map for “Urinary tract infection”

Sometimes a SNOMED CT concept is already specific to a particular age group e.g. “Neonatal cerebral ischemia”. In this case, the map to the ICD-10-CM code P91.0: *Neonatal cerebral ischemia* is final and no age rule is necessary. On the other hand, if the source SNOMED CT concept is “Cerebral ischemia”, it will need a newborn age rule leading to the target code P91.0, while the default target is I67.8: *Other specified cerebrovascular diseases*.

Soon after the beginning of the mapping work, we noticed that manually picking up concepts that required age rules was not carried out in a consistent manner by the map specialists. This manual process could be difficult because sometimes the age specification may be buried in a long list of indented entries in the alpha index. Besides, the proportion of concepts that required an age rule was relatively small. So age rules could be easily omitted when the maps were being created. For this reason, we developed an algorithmic way of identifying concepts that might potentially require age rules to supplement the manual process. In this paper, we describe this process and evaluate its performance.

METHODS

We use the following steps (the “age rule QA”) to add missing age rules to mapped concepts:

1. Identify all ICD-10-CM codes that can be refined by age-related modifiers.

We search the whole ICD-10-CM alpha index for age-related modifiers at all levels, based on a list of age-specifying keywords including neonate, newborn, perinatal, childhood, infant, juvenile, adolescence, senile etc. For example,

Uremia, uremic N19
- **newborn** P96.0

In this case, “newborn” is the age-related modifier, and the code in the level immediate above it (N19) is the code being refined. So N19 will be added to the list of ICD-10-CM codes that can be refined by age-related modifiers. Sometimes, the modifier contains additional information which is not age-related. For example,

Speech
- defect, disorder, disturbance, impediment R47.9
- - psychogenic, in **childhood and adolescence** F98.8

F98.8 can only be satisfied when the speech defect is also “psychogenic”. Even though “childhood” and “adolescence” are age-specifying keywords, we do not include the code being refined (R47.9) in our list.
2. Flag all mapped concepts that use any of the ICD-10-CM codes identified in 1. as target codes, but do not already have age rules in their maps
3. Manually review all flagged concepts and add age rules if appropriate
4. If, after review, an age rule is considered unnecessary in a flagged concept, it is labeled as “reviewed for age rule” so that the same concept will not show up again in future runs of the age rule QA.

To evaluate the performance of the age rule QA, we ran it on all mapped concepts that were to be published in the preview release of the Map in February 2012. We analyzed the distribution of the resulting age rules for the different age ranges. We also analyzed the reason why some flagged concepts ended up not requiring an age rule (false positives).

Results

The age rule QA was run on 7,277 mapped concepts, and identified 342 concepts to be potentially requiring an age rule. After manual review, age rules were added to 172 concepts (true positive = 50.3%). The other 170 concepts did not need an age rule (false positive = 49.7%).

Concepts requiring age rule (true positives)

The breakdown of the true positive concepts by the specified age range is shown in table 3. The three most common categories were: 123 concepts (71.5%) mapped to newborn codes, 18 concepts (10.5%) to infant codes, and 16 Concepts (9.3%) to childhood codes.

Age rule category	Age rule counts	Percentage over all age rules	Percentage over all concepts flagged for age rule
Newborn	123	71.5%	36.0%
Infancy	18	10.5%	5.3%
Childhood	16	9.3%	4.7%
Juvenile	6	3.5%	1.8%
Senile	6	3.5%	1.8%
Adult	2	1.2%	0.6%
Adolescent	1	0.6%	0.3%
Total	172	100.0%	50.3%

Table 3: Counts and percentage of map records with age context per age specification category

Some examples of these age rules are shown in table 4.

Age rule category	SNOMED CT concept	Default ICD-10-CM code	Age-specific ICD-10-CM code
Newborn	Hypertrophy of breast	N62 : Hypertrophy of breast	P83.4 : Breast engorgement of newborn
	Infection of skin	L08.9 : Local infection of the skin and subcutaneous tissue, unspecified	P39.4 : Neonatal skin infection
Infancy	Intoxication with Clostridium botulinum toxin	A05.1 : Botulism food poisoning	A48.51 : Infant botulism
	Rumination	R11.10 : Vomiting, unspecified	F98.21 : Rumination disorder of infancy
Childhood	Pica	F50.8 : Other eating disorders (inclusion: Pica in adults)	F98.3 : Pica of infancy and childhood
	Gender identity disorder	F64.9 : Gender identity disorder, unspecified	F64.2 : Gender identity disorder of childhood
Juvenile	Dermatitis herpetiformis	L13.0 : Dermatitis herpetiformis	L12.2 : Chronic bullous disease of childhood (inclusion: Juvenile dermatitis herpetiformis)
	Rheumatoid arthritis	M06.9 : Rheumatoid arthritis, unspecified	M08.00 : Unspecified juvenile rheumatoid arthritis of unspecified site
Senile	Fatigue	R53.83 : Other fatigue (Inclusion: Fatigue NOS)	R54 : Age-related physical debility (Inclusion: Senile debility)
	Degenerative brain disorder	G31.9 : Degenerative disease of nervous system, unspecified	G31.1 : Senile degeneration of brain, not elsewhere classified
Adult	Failure to gain weight	R62.51 : Failure to thrive (child) (inclusion: Failure to gain weight)	R62.7 : Adult failure to thrive
	Stuttering	F80.81 : Childhood onset fluency disorder (inclusion: Stuttering NOS)	F98.5 : Adult onset fluency disorder
Adolescent	Conduct disorder	F91.9 : Conduct disorder, unspecified	F91.2 : Conduct disorder, adolescent-onset type

Table 4: Examples of age rule for each category

Concepts not requiring age rule (false positives)

Upon manual review, 170 (49.7%) of the flagged concepts turned out to be false positives, due to the following reasons (Table 5):

1: The source SNOMED CT concept contained modifiers that do not match the age-specific ICD code. For example, SNOMED concept “Senile dementia” was mapped to the ICD-10-CM code F03: *Unspecified dementia (inclusion note: Senile dementia NOS)*. F03 was included in the list of target codes that potentially required an age rule because it had a subentry in the alpha index with the age related modifier *infantile, infantilis* under the main term **Dementia**, pointing to the alternative code F84.3: *Other childhood disintegrative disorder*. The alternative ICD-10-CM code clearly did not apply to the source SNOMED CT concept, so this age rule was not appropriate.

2: Multiple locations in the alpha index: An ICD-10-CM code could occur at multiple locations in the index, and the age-related specification was found in some locations but not others. The age rule QA included all ICD-10-CM codes with any age-specific modification anywhere in the index. So some map records might be falsely identified as requiring an age rule.

For example, “Cyclical vomiting – psychogenic” was mapped to F50.8 based on the following index trail. There was no age specification for this condition since there was no age specific subentry at this index position

Cyclical vomiting — see also Vomiting, cyclical G43.A09
- psychogenic F50.8

However, the code F50.8 *Other eating disorders* also occurred under the entry for eating disorder, because it covered “Pica in adults” according to the inclusion note in ICD-10-CM tabular list. In this other index location, there was an age-specific modifier

Disorder (of) — see also Disease
- eating (adult) (psychogenic) F50.9
- - pica F50.8
- - - childhood F98.3

With this index entry, F50.8 was included in our QA algorithm. However, this age specific code of F98.3 “Pica of infancy and childhood” did not apply to the SNOMED CT concept of psychogenic cyclical vomiting. So an age rule was not added to the map.

3: Questionable entries in the ICD-10-CM index: There were some cases in which we did not agree that the age-specific code should be used as an alternative to the main term. Here are a couple of examples. If we followed the index trail for “Heart disease”, it would lead us to I51.9: *Heart disease, unspecified*. However, the indented subentry of “senile” seemed to suggest that heart disease in the elderly should be coded to Myocarditis.

Alpha index entry for this example:
Disease, diseased — see also Syndrome
- heart (organic) I51.9
- - senile — see Myocarditis

Another example was concept “Anemia”, which was mapped to D64.9: *Anemia, unspecified*. An age specific subentry was listed under main term “Anemia” for childhood type:

Anemia D64.9
- childhood D58.9

However, the age specific code in this subentry was D58.9: *Hereditary hemolytic anemia, unspecified*. Anemia occurring in childhood could be due to many causes, and restricting it to “Hereditary hemolytic anemia” seemed questionable. So an age rule was not created for this map record.

4: Age specific code was the same as default code:

SNOMED concept “Paralysis” was mapped to code G83.9: *Paralytic syndrome, unspecified*. There was a subentry for senile paralysis in alpha index:

Paralysis, paralytic G83.9
- senile G83.9

The age rule QA flagged the concept “Paralysis” as candidate. But the age specific target code was no different from the default target. So an age rule was unnecessary for this map.

False positive category	Concept flagged for age rule	Default target code	Index subentry triggering age rule QA	Age specific target	Result
Conflict with SNOMED CT modifier	Senile dementia	F03 : Unspecified dementia (inclusion: Senile dementia NOS)	Dementia F03 - infantile , infantilis F84.3	F84.3 : Other childhood disintegrative disorder (Inclusion: Dementia infantilis)	Age rules were not created
Multiple positions in alpha index	Cyclical vomiting – psychogenic	F50.8 : Other eating disorders (index entry: Cyclical vomiting - psychogenic F50.8)	Disorder (of) - eating - - pica F50.8 - - - childhood F98.3	F98.3 : Pica of infancy and childhood	
Questionable index entry	Heart disease	I51.9 : Heart disease, unspecified	Disease, diseased - heart (organic) I51.9 - - senile — see Myocarditis	I51.4 : Myocarditis , unspecified	
Identical target as default	Paralysis	G83.9 : Paralytic syndrome, unspecified	Paralysis, paralytic G83.9 - senile G83.9	G83.9 : Paralytic syndrome, unspecified (same as default code)	

Table 5: Examples for reasons that concepts identified by the age rule QA but ended up not requiring an age rule in the Map (false positives)

The breakdown of the false positive cases is shown in table 6. False positives due to conflicting SNOMED modifiers and ICD codes with multiple index positions together accounted for two third of all cases. The third most common reason was questionable index entries.

False positive category	Candidate concept counts	Percentage over all the false positives	Percentage over all concepts flagged for age rule
Conflict with SNOMED CT modifier	63	37.1%	18.4%
Multiple positions in alpha index	51	30.0%	14.9%
Questionable index entry	45	26.5%	13.2%
Identical target as default	11	6.5%	3.2%
Total	170	100.0%	49.7%

Table 6: Distribution of the concepts identified by the age rule QA but ended up not requiring age rule in the Map (false positives)

Among the SNOMED CT concepts that were not picked up by the age rule QA, only 20 concepts contained age rules in their original maps. This means that a large proportion ($172/(172+20) = 90.0\%$) of concepts would have

been missing age rules without the age rule QA. Overall, for the 7,277 concepts included in the preview publication, a total 192 concepts (2.6%) contained age rules. However, due to a small glitch when the Map data was generated, the age rule was missing for 18 concepts in the released data. This has been subsequently corrected in the final release of the Map.

Discussion

Many differences between SNOMED CT and ICD-10-CM stem from the fact that SNOMED CT is a clinical terminology, while ICD-10-CM is a medical classification. One main use of the ICD group of classifications internationally is to generate epidemiological and population health statistics. Patient age is often an important consideration in such use cases, because the same disease occurring in different age groups may have different healthcare and resource allocation implications. Therefore, it is often the case that the same disease condition is assigned with different ICD codes for different age groups. To map from SNOMED CT to ICD-10-CM is to bridge the gap between a clinical terminology and a classification. One of the challenges is to properly reflect these coding guidelines and conventions in ICD-10-CM. We have designed the Map to be a rule-based map to fulfill this requirement. Even though only a small percentage (about 3%) of the concepts ended up requiring age rules, we think that without these age rules, the Map will be incomplete and less useful.

In our mapping project, trained map specialists who are knowledgeable in both SNOMED CT and ICD-10-CM are responsible for creating the maps. All map specialists follow a standard mapping protocol, which involves the following steps: understanding the meaning of a SNOMED CT concept by checking the SNOMED CT hierarchy and defining relationships, locating the target code by searching main term and relevant modifiers in the ICD-10-CM alpha index, confirming the target code with the code description and inclusion notes in the tabular list, and adding map rules and map advices indicated by exclusion notes or additional notes, where appropriate. This is a complicated and mostly manual process. The learning curve is steep. Age rules are particularly challenging because the age-related modifiers may be buried deep in a long list of indentations under the main term and can be easily overlooked. The exclusions in the tabular list will be an additional reminder but they are not always there. To cater for individual variations in mapping, the project already has a built-in quality assurance mechanism. Every concept has to be mapped by two map specialists independently, and only map records that agree are promoted to publication. Even with the additional check, we found a significant number of cases with age rule missing at the early stage of mapping process. This was the incentive for creating the age rule QA. It turned out to be a big help which identified the bulk of the age rules in the published Map. In the subsequent phase of the project, we will be running the age rule QA periodically while the mapping work is in progress.

So far we have encountered very few cases in which the age rule QA failed to pick up maps requiring an age rule (false negatives). One example is “Bronchitis”. The age-specific subentry in the alpha index for this condition is “- *in those under 15 years age — see Bronchitis, acute*”, which was not picked up by the age rule QA because it does not contain any of the age-related keywords. However, this kind of cases is generally rare and it is our impression that the recall of the age rule QA should be quite high. The precision of the algorithm is 50.3% (172/342). One way it can be improved is to exclude cases where the age-specific code is the same as the default code (see above - reason 4 under false positives) as they generally do not need age rules. This modification will slightly improve the precision to 52.0% (172/331). There were only 170 concepts being false positive for age rule, out of the 7,277 concepts in preview publication. Since they did not constitute a heavy workload for the project as a whole, we do not consider the false positives to be a major issue at this point. All the false positive cases will be flagged as “reviewed” so that they will not show up again in subsequent QA runs.

One unexpected beneficial outcome of the age-rule QA is the discovery of some questionable entries in the ICD-10-CM index which could be real errors. These made up 26% of the false positive cases. We will compile a list of such cases and submit it to NCHS for review.

One limitation of this study is that the results were based on the judgment of one terminology expert and had not been independently validated. The data in this study was generated by applying the age rule QA to only about half of

the concepts that were in scope for this phase of the project. Finally, the concepts that were not picked up by the QA checks have not been systematically reviewed to see if there are false negatives.

Conclusion

The age rule is the method employed in the SNOMED CT to ICD-10-CM Cross-map to address age-related code refinement in ICD-10-CM. The age rule QA identified 342 out of 7,277 concepts which potentially required age rules, among these 50.3% turned out to be true positives. Without this QA, a large proportion of age rules in the published Map would have been missed.

Acknowledgements

This work was supported by the Intramural Research Program of the National Institutes of Health and the National Library of Medicine.

References

1. Weed LL, Zimny NJ. The problem-oriented system, problem-knowledge coupling, and clinical decision making. *Phys Ther.* 1989 Jul;69(7):565-8.
2. Kuperman G, Bates DW. Standardized coding of the medical problem list. *J Am Med Inform Assoc.* 1994 Sep-Oct;1(5):414-5.
3. Warren JJ, Collins J, Sorrentino C, Campbell JR. Just-in-time coding of the problem list in a clinical environment. *Proc AMIA Symp.* 1998:280-4.
4. Fraser G. Problem list coding in e-HIM. *J AHIMA.* 2005 Jul-Aug;76(7):68-70.
5. Mantena S, Schadow G. Evaluation of the VA/KP problem list subset of SNOMED as a clinical terminology for electronic prescription clinical decision support. *AMIA Annu Symp Proc.* 2007 Oct 11:498-502.
6. NPRM for Stage 2 of Meaningful Use Available <http://www.healthdatamanagement.com/news/EHR-meaningful-use-Stage-2-details-44055-1.html>
7. Subject: HHS-OS-2010-0030; Comments of Integrating the Healthcare Enterprise (IHE) USA, Incorporated (FR Doc. 2010-31159) January 17, 2011
8. Brinjikji W, Kallmes DF. Hospitalization costs for endovascular and surgical treatment of unruptured cerebral aneurysms in the United States are substantially higher than medicare payments. *AJNR Am J Neuroradiol.* 2012 Jan;33(1):49-51. Epub 2011 Oct 27.
9. Goldstein L. Hospital revenues in critical condition. *Healthc Financ Manage.* 2011 Sep;65(9):74-9.
10. Bosanko-Cera R. The I10 impact. Preparing for ICD-10 in physician practices *J AHIMA.* 2011 Nov-Dec;82(11):38-40; quiz 41.
- 11: Mapping SNOMED CT to ICD-10-CM Technical Specifications
http://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm.html
- 12: Mapping from SNOMED CT to ICD-10 and ICD-10-CM
http://www.nlm.nih.gov/research/umls/mapping_projects/mapping_2012himss.pptx