MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220040

Mapping Korean National Health Insurance Reimbursement Claim Codes for Therapeutic and Surgical Procedures to SNOMED-CT to Facilitate Data Reuse

Hannah Kang[a] RN. MSN., Hyeoun-Ae Park[a] RN. PhD. FAAN

[a] *College of Nursing, Seoul National University, Seoul, South Korea*

Abstract

South Korea has a public and single-payer system for healthcare services based on fee-for-service payments. The National Health Insurance (NHI) reimbursement claim codes are used by all healthcare providers for reimbursement. This study mapped NHI reimbursement claim codes for therapeutic and surgical procedures to the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) to facilitate semantic interoperability and data reuse for research. The Source codes for mapping were 2,500 reimbursement claim codes for therapeutic and surgical procedures such as surgery, endoscopic procedures, and interventional radiology. The target terminology for mapping was the 'Procedure' hierarchy of the international edition of SNOMED-CT released in July 2019. We translated Korean terms into English, clarified their meaning, extracted characteristics of the source codes, and mapped them to pre-coordinated concepts. If a source concept was not mapped to a pre-coordinated concept, we mapped it to a post-coordinated expression. The mapping results were validated internally using dual independent mapping and group discussion by trained terminologists, and by two physicians with experience of SNOMED-CT mapping. Out of 2,500 source codes, 1,298 (51.9%) codes were mapped to pre-coordinated concepts, and 1,202 (48.1%) codes were mapped to post-coordinated expressions. The mapping of the NHI reimbursement claim codes for therapeutic and surgical procedures to SNOMED-CT is expected to support clinical research by facilitating the utilization of health insurance claim data.

Keywords:

Systematized Nomenclature of Medicine, Health Information Interoperability, Health Insurance Reimbursement

Introduction

Electronic medical record systems have been widely adopted in South Korea: in 100% of tertiary hospitals, 94.0% of general hospitals, and 90.0% of other hospitals [1]. However, most of them were developed in-house and so are not standardized in terms of vocabulary/ terminology, content, transport, privacy and security, or identifiers. This contrasts with the National Health Insurance (NHI) reimbursement claim codes provided by Health Insurance Review and Assessment Services (HIRA) being used in all institutions for claims related to medical procedures. The NHI reimbursement claim data are valuable in that they are collected nationwide through a single-payer system. As of 2019 there were 9,879 NHI reimbursement claim codes in 19 categories, including therapeutic and surgical procedures, basic medical examinations, evaluation procedures, diagnostic medical imaging, dental services, palliative care, emergency care, anesthesia, traditional medicine, meals, and medications [2]. However, it is difficult to use claim data in research with the current type of code system since it does not have a hierarchical structure [3]. Mapping code systems can be helpful for linking between clinical records and clinical knowledge resources, and also integrating data from multiple heterogeneous sources. It would therefore be helpful for data reuse if the meaning of each code is clearly defined and it is mapped to a global standard terminology in a hierarchical manner for supporting research [4].

Terminology standards are controlled terms and definitions used to express concepts in a manner that is clear to the recipients of information, and they use structured vocabularies and terms that represent concepts [5]. The Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) is a globally used reference terminology that has extensive coverage, with the January 2021 release containing 354,448 active concepts. SNOMED-CT supports post-coordination to represent additional detail of the meaning of clinical concepts by combining existing concepts. The mapping of SNOMED-CT to ICD-9 (ninth revision of the International Classification of Diseases), ICD-10-CM (tenth revision of the International Classification of Diseases, Clinical Modification), and LOINC (Logical Observation Identifiers Names and Codes) facilitates data integration. South Korea joined SNOMED International as the 39th member in August 2020 and designated the Korea Health Information Service as the national release center.

In this study, we mapped therapeutic and surgical procedure codes of NHI reimbursement claim codes to SNOMED-CT, to facilitate the reuse of health insurance claim data for clinical research.

Methods

Material

Source codes

The source codes utilized in this study comprised 2,500 active codes for therapeutic and surgical procedures covered by insurers for treatment and surgery excluding dental services in 2019. The codes for therapeutic and surgical procedures include those for procedures such as general treatment, endoscopy, emergency treatment, solid organ transplantation, interventional radiography, dialysis, cast, and surgery.

An NHI reimbursement claim code consists of five digits: one or two alphabet letters and three or four numbers. Each code has Korean and English terms, and there are cases where English terms are not provided.

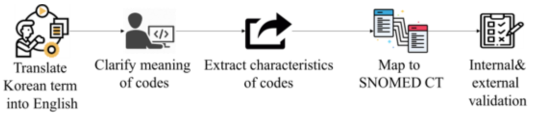
Target codes

The international edition of SNOMED-CT released on July 31, 2019 was used as the target terminology. The target concepts in this study were restricted to the 'Procedure' top-level hierarchy.

Mapping Method

The mapping method consisted of steps for clarifying the meaning of the code, mapping, internal and external validation, and presentation of the map, as shown in Figure 1.

Figure 1– Research process



Translate Korean terms into English

When English terms were not provided with the source codes, Korean terms were translated into English terms.

Clarify the meaning of the term

When the English name of a source code was ambiguous, the meaning of the code was clarified by using the following information provided by the HIRA portal (https://biz.hira.or.kr/):

1. Text definition of the procedure
2. Target, means, and action of the procedure based on the International Classification of Health Interventions (ICHI) published by the World Health Organization.
3. Additional information, such as when to use the procedure, a detailed description of the procedure, and a case study of the procedure.

If the meaning of the source was not clear even though the above information, we consulted an expert in the field.

Extract characteristics of source codes

We extracted characteristics such as the procedure site, approach, device, substance, focus disorder, and intent of the procedure.

Mapping to SNOMED-CT

Mapping Personnel

Four registered nurses who majored in nursing informatics and with experience in SNOMED-CT mapping were recruited. They each had at least 2 years of clinical experience in tertiary hospitals. They had received education about the component, concept model, expressions, and mapping of SNOMED-CT, and had performed SNOMED-CT mapping in a previous research project that involved mapping local diagnoses, procedures, and clinical findings from around 12,000 surgical records. The mapping was conducted in two teams, with two mappers assigned to each team. Half of the source codes were assigned to each team.

Mapping Principles

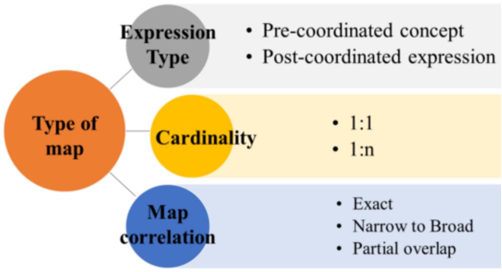
The basic mapping principles used in this study were as follows:

1. Semantic mapping (not lexical mapping) was performed.
2. The source codes were mapped to the concepts of 'Procedure' hierarchy.
3. The map correlation was classified as exact, narrow to broad, or broad to narrow as provided by SNOMED-CT.
4. When a source code was not mapped to a pre-coordinated concept, it was mapped to a post-coordinated expression to represent the characteristics of the code.
5. When a source code included multiple procedures, they were separated for the mapping process.

- Map Types

The map types were classified by the expression type, map cardinality, and map correlation, as shown in Figure 2. The expression type was classified as either pre-coordinated concepts or post-coordinated expressions. When a source code mapped to a pre-coordinated concept and post-coordinated expression, the map was classified as a 1:n map. The map cardinality was based on how many target concepts were used for a single source code. The map correlation based on the target concept was semantically exact, or broader or narrower than the source code.

Figure 2– Map types



Mapping Flow

We followed the mapping flow and classified the mapping type as shown in Figure 3.

First, mappers searched the SNOMED-CT browser for a lexically matched pre-coordinated concept, and when semantic equivalence was confirmed the map was classified as 'Exactly mapped to a pre-coordinated concept.'

Second, when there was no lexical match, we looked for a semantic match with the pre-coordinated SNOMED-CT concept using synonyms, related terms, broader terms, or narrower terms.

We used typical synonyms for medical terms; for example, we used 'surgery' instead of 'operation' when searching for semantically matched concept for 'Operation of nystagmus,' and this was mapped to the pre-coordinated concept '|Nystagmus surgery (procedure)|' and classified as 'Exactly mapped to a pre-coordinated concept.' As another example, 'removal,' 'correction,' 'stomach,' 'cerebral,' and 'hepato,' were considered as synonyms for 'extraction,' 'repair,' 'gastric,' 'intracranial,' and 'liver' respectively.

Related terms, broader terms, and narrower terms were also used to find a semantic match using the browse-up and browse-down approach. When there is no equivalent pre-coordinated concept, we looked for a partially overlapped pre-coordinated concept. If there were no partially overlapping pre-coordinated concepts, the source concept was not mapped.

When there was a partially overlapping concept, and the concept was narrower than the source code, the map was classified as 'Broad to narrow mapped to a pre-coordinated concept.'

When there was a broader pre-coordinated concept, we checked if post-coordination could be performed to express characteristics of the source code. If there was no attribute–value pair available to express the characteristics, the source concept was mapped to the focus pre-coordinated concept, and classified as 'Narrow to broad mapped to a pre-coordinated concept;' for example, 'hyperbaric oxygen therapy over 2 hours' was mapped to '|Hyperbaric oxygen therapy (procedure)|.'

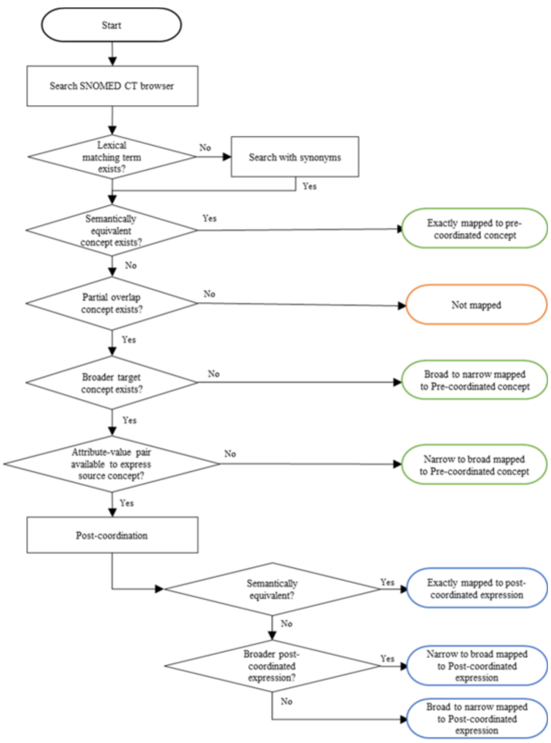
If there were attribute–value pairs available to express the characteristics of the source concept, they were post-coordinated as close as possible with a focus concept and one or more attribute–value pairs. When we post-coordinated a focus concept with attribute–value pairs, we searched attribute and value concepts while considering SNOMED-CT rules with the MRCM (Machine Readable Concept Model) browser; for example, ‘Angioplasty of blood vessel (End-to-End Anastomosis) - By Thoracotomy’ was mapped to |Angioplasty of blood vessel (procedure)|: {|Method (attribute)| = |End-to-end anastomosis - action (qualifier value)|, |Surgical approach (attribute)| = |Transthoracic approach (qualifier value)|)} with extracted characteristics.

When the post-coordinated expression was semantically equivalent to the source code, the map was classified as ‘Exactly mapped to a post-coordinated expression.’ When the post-coordination expression had a broader meaning than the source code, the map was classified as ‘Narrow to broad mapped to a post-coordinated expression.’ When the post-coordinated expression had a narrower meaning than the source code, the map was classified as ‘Broad to narrow mapped to a post-coordinated expression.’

For post-coordinated expressions, the frequency of use of attributes was analyzed in order to understand the characteristics of the source codes.

When a single source code included multiple procedures, each of them was followed using the mapping process and mapped separately. The map in this case was classified as a 1:n map in terms of cardinality; for example, the source code ‘Release of scar contracture and flap operation’ was mapped to the two SNOMED-CT concepts of ‘|Relaxation of scar contracture of skin (procedure)|’ and ‘|Skin flap operation (procedure)|,’ and categorized as a 1:n map.

Figure 3– Mapping flow



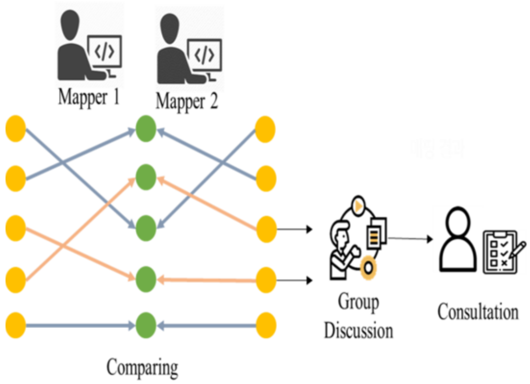
Internal and external validation

The therapeutic and surgical procedure codes to SNOMED-CT map were validated internally and externally, as shown in Figure 4.

For internal validation, the mapping results of the two mappers were compared within each team after they had independently performed mapping. When the two mappers agreed with the maps, the maps were considered internally valid. If they disagreed with the maps, the maps were discussed in group meetings attended by the project manager and other researchers who were not involved in the mapping process. Group meetings were held throughout the study period and were attended by an average of five researchers with experience of SNOMED-CT mapping. The most appropriate mapping for the map for which there was disagreement was discussed and then decided. Consistency with other maps was also considered in this process.

For external validation, some of the maps were extracted and validated by two physicians with experience of SNOMED-CT mapping but who had not participated in the mapping process. They either agreed with the extracted map or suggested a new map.

Figure 4 – Map validation process



Presentation of the map

The final map was provided in an Excel spreadsheet along with a ‘read me’ text file. The columns in the spreadsheet comprised source code, Korean terms, English terms, map correlation types, pre-coordinated SNOMED-CT concepts (ID and fully specified name), post-coordinated focus concepts (ID and fully specified name) with attribute concepts (ID and fully specified name), and value concepts (ID and fully specified name) and post-coordinated expressions. When a post-coordinated expression had more than one attribute–value pair, each attribute–value pair was represented in a row. When a single source code was mapped to multiple target concepts, it was presented as a separate row. The ‘read me’ file described the data format, the source code, the target code including version information, and precautions when using the map file.

Results

Mapping result

The result of mapping between the therapeutic and surgical procedure codes and SNOMED-CT is presented in Table 1. Out of 2,500 source codes, 1,298 (51.9%) codes were mapped to pre-coordinated concepts, and 1,202 (48.1%) codes were mapped to post-coordinated expressions. Out of 2,500 source codes, 2,057 (82.3%) codes were mapped to one target concept, and 443 (17.7%) codes were mapped to many target concepts. Regardless of mapping expression or cardinality, semantically 1,589 (63.6%) source codes were exactly mapped to SNOMED-CT pre-coordinated concepts or post-coordinated expressions.

Table 1– Mapping result

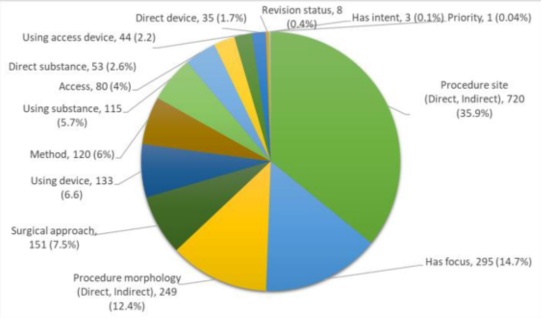
Mapping type			**No. (%)**
Pre-coordinated concept	1:1	Exact	651 (26.0)
		Narrow to broad	395 (15.8)
	1:n	Exact	147 (5.9)
		Narrow to broad	100 (4.0)
		Partial overlap	5 (0.2)
Subtotal			1,298 (51.9)
Post-coordinated expression	1:1	Exact	666 (26.6)
		Narrow to broad	344 (13.8)
		Partial overlap	1 (0.04)
	1:n	Exact	125 (5.0)
		Narrow to broad	62 (2.5)
		Partial overlap	4 (0.2)
Subtotal			1,202 (48.1)
Total			2,500 (100)

Analysis of attribute–value pairs for post-coordinated expressions

There were 2,007 attribute–value pairs used for 1,202 post-coordinated expressions, corresponding to an average of 1.7 coordinated attribute–value pairs. There were 14 types of attributes used for post-coordination as well as 4 subtypes, as shown in Figure 5.

The most commonly used attribute type was 'Procedure site.' This was used 720 times, including in the subtype attributes 'Procedure site - direct' and 'Procedure site - indirect.' The next most commonly used attribute type was 'Has focus,' while the third most commonly used one was 'Procedure morphology.' In the case of 'Procedure morphology,' both the subtype attributes of 'Direct morphology' and 'Indirect morphology' were used. The 'Surgical approach,' 'Using device,' 'Method,' and 'Using substance' types were most often used next.

Figure 5 – Frequency of attributes used for post-coordinated expressions



The maximum number of attribute–value pairs used to express a single code was three; for example, 'Angioplasty of blood vessel with artificial graft using surgical patch by thoracotomy' was expressed using '*|Angioplasty of blood vessel (procedure)| : {(|Using substance (attribute)| = |Artificial graft - material (substance)|, |Using device (attribute)| = |Surgical patch (physical object)|, |Surgical approach (attribute)| = |Transthoracic approach (qualifier value)|)}*.'

Discussion

In this study, the following challenges were encountered due to the characteristics of the health insurance reimbursement claim codes. First, the Korean and English terms of the source codes were ambiguous, so that we had to clarify the exact meaning with additional information. Although the current reimbursement claim codes contain structural information referring to the ICHI, the exact meaning of the procedure is not reflected in the code name. Therefore, preprocessing of the source codes is required to utilize the data. Second, similar types of procedures were assigned to one code if the premiums paid were the same for the convenience of claiming, which required post-coordinations. According to the mapping result, SNOMED-CT post-coordinated expressions were required at 48.1% to express the source codes. We found that the 'Procedure site' attribute type was the most frequently used; for example, the source code 'Hypospadias repair of penile structure, perineal structure, or glans penis structure' includes the three procedure sites of 'Penile structure,' 'Perineal structure,' and 'Glans penis structure' for a single procedure. To express all three procedure sites, we used the 'Procedure site - Direct' attribute and body structure values as pairs.

Therefore, the result of this study can facilitate the reuse of health insurance claim data for clinical research by clearly and structurally expressing the meaning of the codes with SNOMED-CT pre-coordinated concepts and post-coordinated expressions.

The considerations of mapping discovered through this research process are as follows. First, basic principles for mapping must be established and shared within the team for consistency. Next, in the mapping process, not only terminology experts with background knowledge are needed, but also advisors with expertise in the mapping area are needed to accurately understand the source concepts and map them semantically. Finally, systematic validation for the mapping result should be established. As the number of mapping increases, more efforts will be required to assure the quality of the mapping.

Conclusions

The findings of this study are expected to facilitate the reuse of research on widely used therapeutic and surgical procedure codes other than for insurance claim purposes.

Future research should map the domestic codes and local terms used in various fields to SNOMED-CT. Also, education about consistent mapping and the establishment of a national terminology management system is needed to extend the mapping and use of SNOMED-CT.

Acknowledgment

This study was funded by the Korea Health Information Service.

References

[1] Korea Health Information Service, *2020 Health and Medical Informatization Survey*, Korea Health Information Service, 2021.

[2] National Insurance Review and Assessment Service and National Health Insurance Service, *National Health Insurance Statistical Yearbook,* National Insurance Review and Assessment Service and National Health Insurance Service, 2019.

[3] Hwang EJ, Park H-A, Sohn SK, Lee HB, Choi HK, Ha S, Kim HJ, Kim TW, &Youm W., *Mapping Korean EDI Medical Procedure Code to SNOMED-CT*, MedInfo. (2019), 178-182.

[4] RANDORFF, Højen A., ROSENBECK, Gøeg K., *SNOMED-CT implementation. Mapping guidelines facilitating reuse of data.* Methods of information in medicine, **51.6** (2012), 529.

[5] The Korean Society of Medical Informatics, *Healthcare Big data Standardization Roadmap and Application Strategy Establishment Research*, Korea Health Information Service, 2020.

[6] Sedano, F. F., Cuadrado, M. T., Rebolledo, E. G., Clemente, Y. C., Balazote, P. S., & Delgado, A. G. *Implementation of SNOMED-CT to the medicines database of a general hospital*. Stud Health Technol Inform **148** (2009), 123-30.

[7] Ministry of Health and Welfare, *Project report of Developing and Spreading the Certification Standard of EMR System*, Ministry of Health and Welfare, 2019.

[8] Shin S., *Introduction of a certification system to advance electronic medical records (EMR),* Health Insurance Review & Assessment Service, 2018.

[9] SNOMED International, SNOMED-CT Editorial Guide, SNOMED International, 2019.

[10] National E-Health Transition Authority Ltd, SNOMED-CT-AU Mapping Guidelines v2.0, Australian Digital Health Agency, 2014.

[11] Korea Health Industry Development Institute, *Electronic Medical Record System Certification Criteria Review and Inspection Guidelines*, Korea Health Industry Development Institute. 2019.

Address for correspondence

Hyeoun-Ae Park, address: College of Nursing, Seoul National University, 103 Dahakro, Jongno-gu, Seoul, 03080, South Korea.

email: hapark@snu.ac.kr. phone: (82)10-3786-3284. fax: (82)2-765-4102.