Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220462

Mapping Korean National Health Insurance Pharmaceutical Claim Codes to SNOMED CT

Hyesil JUNG[a] , Hyeoun-Ae PARK[b,1], Hannah KANG[b], MinSun LEE[b], Sumi SUNG[b] and SeulKi PARK[c]
[a] *Office of eHealth Research and Business, Seoul National University Bundang Hospital, South Korea*
[b] *College of Nursing, Seoul National University, South Korea*
[c] *Department of nursing, Daejeon University, South Korea*

Abstract. The objective of this study was to map pharmaceutical claim codes to SNOMED CT and thereby facilitate multicenter collaborative research and improve semantic interoperability. The claim codes were mapped to SNOMED CT using rule-based automated and manual methods. The maps were internally validated by terminologists and a pharmacist. Finally, 80% of all claim codes were mapped to the concepts of Pharmaceutical/biologic product hierarchy in SNOMED CT. Of them, 50.6% of the codes were exactly mapped to one clinical drug branch concept.

Keywords. Systematized Nomenclature of Medicine Clinical Terms, Semantic interoperability, Pharmaceuticals, National Health Insurance Reimbursement

1. Introduction

Most electronic medical record (EMR) systems in South Korea have been developed in-house and use their own medical vocabulary/terminology to record clinical events or procedures. Thus, the systems cannot exchange data with unambiguous, shared meaning. Although every EMR system uses National Health Insurance (NHI) claim codes [also called electronic data interchange (EDI) codes] to claim a reimbursement in the fee-for-service system, NHI claim codes cannot be used as reference terminology due to limited granularity, non-polyhierarchy, lack of concept identifier (ID) version control, use of a semantic concept ID, a non-unique ID, and a lack of formal definitions [1–2]. Thus, researchers typically map the NHI claim codes to standard terminology in multicenter collaborative studies.

One research team attempted to incorporate the NHI claim codes into Observational Medical Outcomes Partnership (OMOP) vocabulary [2]. However, the pharmaceutical claim codes were not connected to standard terminology such as RxNorm or Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), limiting their use in clinical research.

[1] Corresponding Author, Hyeoun-Ae PARK, College of Nursing, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul, 03080, The South Korea; E-mail: hapark@snu.ac.kr

South Korea adopted SNOMED CT as the national standard terminology when the country became the 39th member of SNOMED International in August 2020. SNOMED CT is a globally used reference terminology with extensive coverage: it contains 350,936 active concepts and 1,521,274 descriptions in 19 top-level hierarchies [3]. One of the top-level hierarchies is Pharmaceutical/biologic products with 23,215 active concepts and 170,042 descriptions.

The objective of this study was to map Korea's NHI pharmaceutical claim codes to SNOMED CT and thereby facilitate multicenter collaborative research and improve semantic interoperability.

2. Methods

2.1. Mapping materials

The source data utilized in this study were 22,610 unique pharmaceutical codes covered by the NHI in South Korea published on August 1, 2019. The data included a 9-digit EDI identifier (product ID), the Korean brand name, the manufacturer, the WHO Anatomical Therapeutic Chemical Classification (ATC) code, and the ATC name. Missing information for pharmaceuticals that could not be identified from the EDI code was obtained from the FirstDIS Ltd. database, which is a repository containing information related to the pharmaceuticals used in South Korea.

The target data were restricted to the concepts in 'Pharmaceutical/biologic products', 'Substance', 'Organism', or 'Physical object' top-level hierarchies in the SNOMED CT international edition released on July 31, 2019. If no appropriate concepts appeared in the international edition, we used the Argentinian and United States extensions released on November 30, 2019.

2.2. Mapping methods

Figure 1 presents the process used to map pharmaceutical claim codes to SNOMED CT concepts and validate the results. Creating the maps was an iterative process of rule development and application to search for (map) the optimal target concept.

A rule-based mapping system was developed utilizing Snowstorm, which provides the terminology server API for the SNOMED international browser, including the international edition and about 14 local extensions. Snowstorm is built on top of Elasticsearch and applies the term-based search using multiple prefix, any order, matching. The first part of one or more words input can be used to match the descriptions of concepts. If a description term does not exactly match all the prefixes in the search term, it will be excluded from the results.

When the mapping system provides candidates that match the search terms, the first author reviewed the candidates and selected the optimal SNOMED CT concept among them. If no suitable concept was found among the candidates, she searched for the appropriate concept and mapped them manually. Pharmaceuticals consisting of three or more ingredients were also manually mapped to SNOMED CT.

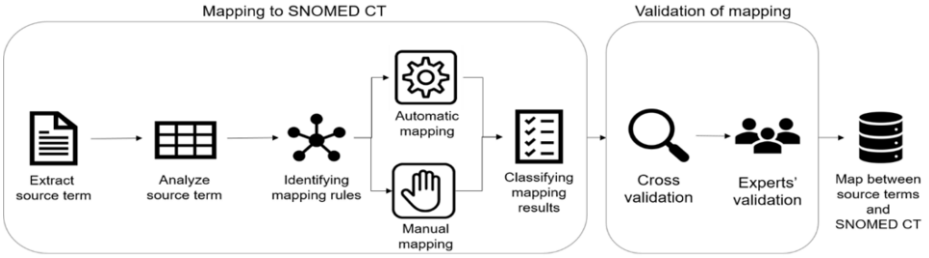

Figure 1. Research process

2.2.1. Creating the maps

2.2.1.1. Developing rules for mapping

We analyzed the source data and extracted the ingredients (base ingredient and active ingredient), strength (numerator value and numerator units), dose form, and route of administration information into a structured data format.

We randomly selected 600 source data and manually mapped them to SNOMED CT by combining the pharmaceutical information extracted in the previous step to develop rules for automatic mapping. First, we combined ingredient(s), strength, dose form, and administration route and searched whether the relevant SNOMED CT concept existed in the clinical drug branch of the Pharmaceutical/biologic product hierarchy. If the relevant concept did not exist in the clinical drug branch, we only combined ingredient(s) and administration route to search for the concept in the medicinal product form branch. Figure 2 presents the flow of combining structured elements (information) from the source data and mapping to the relevant concepts within the SNOMED CT hierarchy.

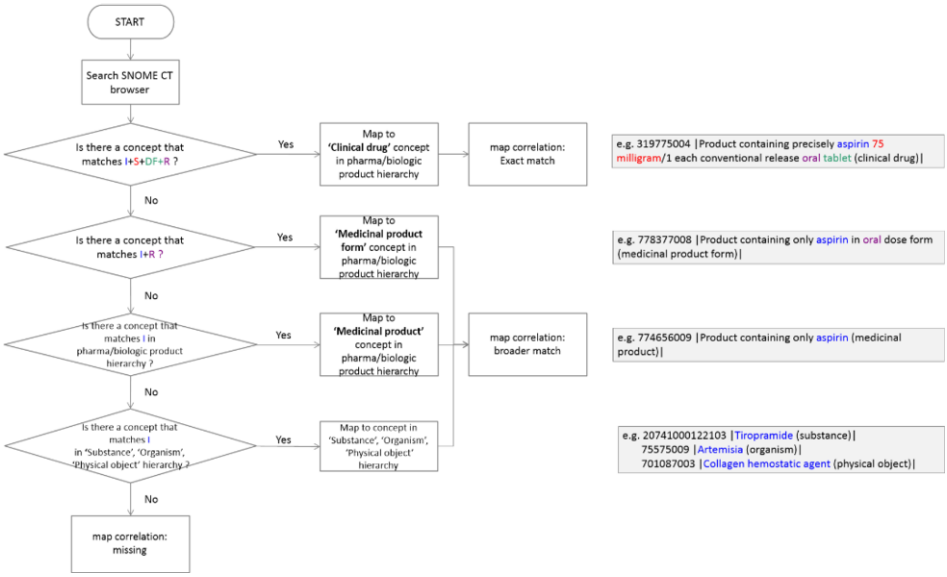

Figure 2. Mapping rules for combining structured elements (information) from source data
I-ingredient; S-strength; DF-dose form; R-route of administration

2.2.1.2. Search strategies

Three types of drug-specific normalization rules [4-5] for managing the variation in clinical drug names were applied to search for appropriate target concept; (1) expanding abbreviated words (2) reformatting specific parts of the drug name (3) removing salt modifiers in ingredient names.

For example, shortened dose forms such as 'tab', 'susp', and 'cap' were expanded to 'tablet', 'suspension' and 'capsule', respectively. In terms of reformatting, ‘syrup’ and ‘cutaneous emulsion’ were normalized to ‘oral suspension’ and ‘cutaneous lotion’, respectively. ‘Oral suspension’ and 'oral liquid’ were changed to ‘oral solution’. And when an ingredient’s strength value was less than 1, we converted the strength unit (e.g. 0.1 g →100 mg). The ingredient names of clinical drugs sometimes have the salt modifiers and sometimes do not. We removed the salt modifiers from the ingredient name when there was no appropriate target concept with the salt modifiers.

The map correlations were classified as ‘exact’ match, ‘broader’ match, and ‘missing’ according to the extent of coverage of the target concept compared to the source code [5-6]. If a target concept was consistent with the source code in terms of ingredients, strength, dose form, and administration route, the source code had the ‘exact’ match. If a target concept partially matched the source code, the source code had a ‘broader’ match. If there was no ingredient for the source code in the target concept, the source code had ‘missing’ as the map correlation.

Map cardinality was defined as ‘1:1’and ‘1: N’, according to the number of target concepts mapped to a single source code.

2.2.2. Validation

The maps were validated by another mapper with expert experience in SNOMED CT mapping. When the two mappers agreed on the mapping results, the results were considered internally valid. If they disagreed, the results were discussed in group meetings attended by the project manager and other researchers who were not involved in the mapping process. Any concepts not agreed upon during the group discussion or with the ‘missing’ correlations were validated by a pharmacist with experience in SNOMED CT and Identification of Medicinal Product mapping experience.

The final maps consisted of a source code, the map correlation type, the SNOMED CT concept identifier, fully specified name, hierarchy, and pre-coordinated expression.

3. Results

There was good agreement (79.9%) between mappers in the process of mapping pharmaceutical claim codes to SNOMED CT concepts. A pharmacist externally validated 250 concepts that were not agreed upon throughout the group discussion.

Of the 22,610 source data, almost 80% were mapped to the concepts of Pharmaceutical/biologic product hierarchy, and 11,453 (50.6%) of these codes were exactly mapped to one (1:1 map) ‘clinical drug’ branch concept. In total, 11,151 (49.3%) of the claim codes mapped to a broader concept(s), and 1233 of these codes were mapped to two or more (1:N map) target concepts. However, five pharmaceuticals, including Revanex, Noltec, Ulistin, and Godex did not map to any target concept.

4. Discussion

To develop mapping rules, we analyzed the source codes and extracted pharmaceutical information including the ingredients, strength, dose form, and administration route. First, all information (elements) was combined to search for the relevant target concept. If there was no relevant concept, elements were gradually omitted and mapped to a broader target concept. Almost 50% of the source codes were mapped to broader concepts. In particular, multi-ingredient pharmaceuticals with strength specified for each ingredient were challenging [5]. When conducting research in which the strength and dose form of pharmaceuticals are important, broad concepts make it impossible to provide information on strength and dose form, which may lower the reliability and accuracy of the study. As mentioned by Zhou et al., one possible solution is post-coordinate two or more SNOMED CT concepts that are semantically representative of a single claim code [5].

The five claim codes that were not mapped to any SNOMED CT concept were for pharmaceuticals used only in South Korea or certain countries such as Japan, China, or India. They were pharmaceuticals with ilaprazole, revaprazan, and ulinastatin as ingredients. The Korean extension to be developed soon may include these concepts.

5. Conclusions

The results of mapping Korea's NHI pharmaceutical claim codes to SNOMED CT concepts will facilitate semantic interoperability and collaborative research. The mapping rules and methodologies developed in this study can be used to map between expired or newly created pharmaceutical claim codes and SNOMED CT concepts.

Acknowledgments

This research was supported by the Korea Health Information Service and a National Research Foundation of Korea grant funded by the Ministry of Science and ICT (no. NRF-2021R1A2C1091261).

References

[1] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med. 1998 Nov;37(4-5):394–403.

[2] Seong Y, You SC, Ostropolets A, Rho Y, Park J, Cho J, Dymshyts D, Reich CG, Heo Y, Park RW. Incorporation of Korean electronic data interchange vocabulary into observational medical outcomes partnership vocabulary. Healthc Inform Res. 2021 Jan;27(1):29–38.

[3] SNOMED International [Internet]. General Release Statistics. Available from: https://browser.ihtsdotools.org/qa/#/generalReleaseStatistics

[4] Peters L, Kapusnik-Uner JE, Bodenreider O. Methods for managing variation in clinical drug names. AMIA Annu Symp Proc. 2010 Nov;13:637-41.

[5] Zhou L, Plasek JM, Mahoney LM, Chang FY, DiMaggio D, Rocha RA. Mapping Partners Master Drug Dictionary to RxNorm using an NLP-based approach. J Biomed Inform. 2012 Aug;45(4):626-33.

[6] Merabti T, Massari P, Joubert M, Sadou E, Lecroq T, Abdoune H, Rodrigues JM, Darmoni SJ. An automated approach to map a French terminology to UMLS. Stud Healht Technol Inform. 2010;160:1040-4.