



# Mapping the Korean National Health Checkup Questionnaire to Standard Terminologies

Ji Eun Hwang<sup>1,2</sup>, Hyeoun-Ae Park<sup>3</sup>, Soo-Yong Shin<sup>1,2,4</sup>

<sup>1</sup>Department of Digital Health, SAIHST, Sungkyunkwan University, Seoul, Korea

<sup>2</sup>Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon, Korea

<sup>3</sup>College of Nursing, Seoul National University, Seoul, Korea

<sup>4</sup>Center for Research Resource Standardization, Samsung Medical Center, Seoul, Korea

**Objectives:** An increasing emphasis has been placed on the integration of clinical data and patient-generated health data (PGHD), which are generated outside of hospitals. This study explored the possibility of using standard terminologies to represent PGHD for data integration. **Methods:** We chose the 2020 general health checkup questionnaire of the Korean Health Screening Program as a resource. We divided every component of the questionnaire into entities and values, which were mapped to standard terminologies—Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) version 2020-07-31 and Logical Observation Identifiers Names and Codes (LOINC) version 2.68. **Results:** Eighty-nine items were derived from the 17 questions of the 2020 health examination questionnaire, of which 76 (85.4%) were mapped to standard terms. Fifty-two items were mapped to SNOMED CT and 24 items were mapped to LOINC. Among the items mapped to SNOMED CT, 35 were mapped to pre-coordinated expressions and 17 to post-coordinated expressions. Forty items had one-to-one relationships, and 17 items had one-to-many relationships. **Conclusions:** We achieved a high mapping rate (85.4%) by using both SNOMED CT and LOINC. However, we noticed some issues while mapping the Korean general health checkup questionnaire (i.e., lack of explanations, vague questions, and overly narrow concepts). In particular, items combining two or more concepts into a single item were not appropriate for mapping using standard terminologies. Although it is not the case that all items need to be expressed in standard terminology, essential items should be presented in a way suitable for mapping to standard terminology by revising the questionnaire in the future.

**Keywords:** Standards, Patient Generated Health Data, Surveys and Questionnaires, Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), Logical Observation Identifiers Names and Codes (LOINC)

**Submitted:** March 23, 2021

**Revised:** July 11, 2021

**Accepted:** August 11, 2021

## Corresponding Author

Soo-Yong Shin

Department of Digital Health, SAIHST, Sungkyunkwan University, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea. Tel: +82-2-3410-1449, E-mail: sy.shin@skku.edu (<https://orcid.org/0000-0002-2410-6120>)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2021 The Korean Society of Medical Informatics

## I. Introduction

The 2020 survey results of the Korea Health Information Service showed that 85.7% of tertiary hospitals had introduced Electronic Health Records (EHRs) [1]. However, most of these systems only involved changing the manner of inputting content from handwriting to keyboard entry. Therefore, due to the widespread use of free text content, the utilization of EHR data remains unsatisfactory. To overcome these hurdles, many projects that implement structured EHRs based on standard terminology have been introduced [2,3].

Throughout the world, Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is widely used to represent clinical information. The SNOMED International website presents information about various SNOMED CT implementation experiences. The representative cases are the United Kingdom's National Health Service (UK NHS) e-referral service, the concept dictionary of the Columbia International eHealth Laboratory (CIEL), and the national death registry of India. General practitioners in the UK utilize the NHS e-referral service to refer patients [4]. They document patients' clinical information using SNOMED CT concepts [5]. CIEL released a dictionary-type mapping table between interface terminology and reference terminologies. SNOMED CT is the primary reference terminology among the various reference terminologies that are used, with other systems including the International Classification of Diseases-Tenth Revision (ICD-10) and Logical Observation Identifiers Names and Codes (LOINC) [6]. The e-death notes created by doctors from the All India Institutes of Medical Science use SNOMED CT [7]. Doctors complete death certificates or international death forms using the SNOMED CT concepts.

In addition to the above-described clinical data standardization initiatives, the importance of health-related data outside of hospitals has been increasing due to precision medicine. Data, including information on social determinants (race, ethnicity, education, housing, and employment) and patient-generated health data (PGHD), have proven to be an effective tool to improve people's health [8]. In particular, PGHD and patient-reported outcomes have been emphasized [9-11]. According to the Office of the National Coordinator for Health Information Technology (ONC), "PGHD are health related data created, recorded, or gathered by or from patients (or family members or other caregivers) to help address a health concern" [12]. PGHD are not confined to the data created while operating a health information system [13-15]; instead, PGHD also include questionnaires and data from smartphones, wearable devices, and activity trackers. Mapping PGHD is difficult due to the presence of a variety of sub-categories, such as health history, family history, and lifestyle, and requires more processes than mapping diagnoses, examinations, or operations. However, to our best knowledge, there are no published articles on mapping PGHD onto standard terminologies.

In Korea, most terminology standardization research has been conducted for secondary research use (i.e., a common data model), not for clinical practice. Nonetheless, if EHR content could be represented using standard terminology, re-

searchers would be able to significantly reduce the time and effort spent on secondary research. Of course, some research has been done to map terms of classification systems, such as the Korean Standard Classification of Diseases (KCD), the Korean Standard Terminology of Medicine (KOSTOM), or a list of covered/non-covered services in the National Health Insurance system to standard terminology. However, that research only corresponds to a small portion of terminology standardization research and does not include PGHD. Some articles have been published on the mapping of standardization terminology, but most of these studies were conducted in the framework of natural language processing research or did not publish the complete mapping results [16,17].

This study aimed to explore the possibility of using standard terminologies to represent PGHD for data integration.

## II. Methods

### 1. Standard Terminology

There are many clinical terminology systems, including the ICD, International Classification of Health Interventions (ICHI), SNOMED CT, LOINC, and the Unified Medical Language System (UMLS). We chose SNOMED CT and LOINC. SNOMED CT is the most comprehensive and widely used system of clinical terminology, and it covers clinical findings, procedures, body structures, and other domains, while LOINC is a terminology standard for identifying laboratory tests and other measurements.

Formerly known as the abbreviation of Systematized Nomenclature of Medicine Clinical Terms, but recently itself a licensed registered trademark, SNOMED CT uses two ways to represent clinical ideas: pre-coordinated expressions and post-coordinated expressions. Pre-coordinated expressions represent the meaning of individual concepts that are pre-defined by SNOMED CT. In contrast, post-coordinated expressions combine two or more concepts that are added to the meaning. We utilized both pre-coordinated expressions and post-coordinated expressions to increase the mapping rate. We used SNOMED CT version 2020-07-31 and LOINC version 2.68 for mapping.

### 2. Source Report Selection

As the first step in mapping PGHD onto standard terminology, we chose the Korean national health checkup questionnaire. Since Korea provides national health examinations to the entire population, the Korean government has access to a nationwide source of PGHD. Health checkup data can help in detecting diseases early and be used in various clinical

studies. Furthermore, according to a previous study, there is a gap between institutions even within the same country [18]. This means that applying the mapping results from a specific institution to other hospitals is very difficult. A nationwide questionnaire can reduce this limitation, as the mapping results can be easily applied to the electronic medical record system of any hospital.

In the initial period of the study, on April 29, 2020, we downloaded nine forms from the Korean National Health Insurance Service (NHIS) website: the general health checkup questionnaire, the health checkup questionnaires for infants and young children (for 4–6 months, for 9–12 months, for 18–24 months, for 30–36 months, for 42–48 months, for 54–60 months, and for 66–71 months), and the cancer screening checkup questionnaire. The general health checkup questionnaire was selected since it had the largest number of examinees and the highest inspection rate [19]. We used the English version of the Korean general health checkup questionnaire, which is shown in Supplement A.

### 3. Item Analysis and Term Extraction

We categorized questionnaire items into groups, entities, and values. We then analyzed the data type of each value. A group is a set of the same subjects. An entity denotes the health status of the respondents from the health checkup questionnaire. A value is an answer to the questionnaire. If there were any questions that contained more than one subject, we divided the item into several entities so that each entity would contain a single meaning. For example, in the questionnaire, “Have you ever been diagnosed by a doctor with any of the following diseases or are you currently taking any medication?” contains two meanings (Figure

1A). Therefore, we divided it into two: one was “history of diagnosis” and the other was “current medication status.” In another instance, the questionnaire asks, “Do you smoke cigarettes now?”; this was divided into six sub-questions: “smoking status,” “smoking period for current smoker,” “smoking amount for current smoker,” “smoking period for ex-smoker,” “smoking amount for ex-smokers,” and “cessation period for ex-smoker” (Figure 1B).

Figure 2 shows examples of item reorganization. Each entity is related to frequency, but values are related to status and frequency. To summarize, the values corresponding to two questions are mixed within one value-set. Here, as elsewhere, we separated the entities and reset the values. Then, we analyzed the data type of each value. All entities and values became sources for mapping.

### 4. Mapping between Standard Terminology and Clinical Terms

Figure 3 presents the entire mapping process. In phase 1, we first searched for a concept using keywords in the SNOMED CT browser (<https://browser.ihtsdotools.org>). Then, we selected a matching concept based on the semantic tag (step 1 of phase 1). If there was no pre-coordinated SNOMED CT concept match, we tried to express the term using post-coordination that complied with the Machine Readable Concept Model (MRCM) rules (step 2 of phase 1) [20]. If we were not able to express the term using SNOMED CT, we looked for the concept in LOINC (step 3 of phase 1). If LOINC did not have a concept that expressed the entire meaning of the concept, we tried to implement partial mapping with a pre- or post-coordinated expression using SNOMED CT (step 4 of phase 1). If we could not express the term with exact

Have you ever been diagnosed by a doctor with any of the following diseases or are you currently taking any medication?

	Diagnosis <sup>(a)</sup>		Medication therapy <sup>(b)</sup>	
	Yes	No	Yes	No
Brain stroke (paralysis)	Yes	No	Yes	No
Cardiac infarction/angina	Yes	No	Yes	No
...	...	...	...	...

Meaning of the item  
<sup>(a)</sup> History of diagnosis      <sup>(b)</sup> Current medication status      **(A)**

---

Do you smoke cigarettes now?

① I do <sup>(a)</sup>	A total of ___ years <sup>(b)</sup> An average of ___ cigarettes a day <sup>(c)</sup>
② I used to but not anymore <sup>(a)</sup>	A total of ___ years <sup>(d)</sup> Used to smoke ___ cigarettes a day ___ years since I quit <sup>(f)</sup> on average <sup>(e)</sup>

Meaning of the item  
<sup>(a)</sup> Smoking status      <sup>(b)</sup> Smoking period for current smoker      <sup>(c)</sup> Smoking amount for current smoker  
<sup>(d)</sup> Smoking period for ex-smoker      <sup>(e)</sup> Smoking amount for ex-smoker      <sup>(f)</sup> Cessation period for ex-smoker      **(B)**

Figure 1. Examples of split items. (A) History of diagnosis and current medication status. (B) Smoking status, cessation period, smoking period, and consumption amount.



We conducted target report selection, item analysis and term extraction, and the previously described steps of mapping between standard terminology and clinical terms from April to August 2020. Then, we received external expert advice between January and February 2021.

### III. Results

#### 1. Results of Item Analysis and Term Extraction

The general health checkup questionnaire consists of 17 questions. Table 1 shows a summary of the analysis results. Source terms were extracted from all groups. The total number of source terms was 89. The detailed results of the item analysis are shown in Supplement B.

#### 2. Results of Standardized Terminology Mapping

As shown in Table 2, 76 out of 89 items (85.4%) were mapped to standardized terms. Fifty-one items were mapped to

SNOMED CT and 24 items were mapped to LOINC. More specifically, 35 pre-coordinated expressions and 17 post-coordinated expressions were used. Over half of the post-coordinated expressions were associated with medical history items. All 24 items expressed with LOINC were entities. The relationship between items and standard terminologies was also categorized. One-to-one relationships, expressed as 1:1 in Table 2, refer to cases where one item matches one standard term, whereas one-to-many relationships, expressed as 1:N, represent cases where one item matches two or more standard terms. More than 82% of mapping cases had a 1:1 relationship. Table 3 shows an example of a post-coordinated expression.

Half of the one-to-many relationships were related to “current medication: no,” regardless of the disease. Both those who were “healthy” and those who replied “do not take medication though diagnosed” chose the answer “No.” Thus, we mapped these replies to “no” and “not applicable.” Three

Table 1. Summary of the analysis results

	Number of items	Examples
Group	4	Medical history, smoking and e-cigarettes, drinking, exercising
Entity	32	History of diagnosis, current medication status, family history, etc.
Value	57	“Brain stroke: yes,” “Brain stroke: no,” “Cardiac infarction/angina: yes,” “Cardiac infarction/angina: no,” “high blood pressure: yes,” “high blood pressure: no” etc.

Table 2. Results of mapping

Classification				Number of items	Total
Mapped	SNOMED	Pre-coordinated expression	1:1	25 (28.09)	35 (39.33)
			1:N	10 (11.24)	
	Post-coordinated expression	1:1	15 (16.85)	17 (19.10)	
		1:N	2 (2.25)		
LOINC	-	1:1	24 (26.97)	24 (26.97)	
		1:N	0 (0)		
Non-mapped				13 (14.61)	13 (14.61)
Total				-	89 (100)

Values are presented as number (% of total number).

SNOMED CT: Systematized Nomenclature Of Medicine Clinical Terms, LOINC: Logical Observation Identifiers Names and Codes.

Table 3. Example of post-coordination expressions: history of dyslipidemia

Source item	Expression
History of dyslipidemia: Yes	417662000  History of clinical finding in subject (situation)  : {246090004  Associated finding (attribute)  = 370992007  Dyslipidemia (disorder) }
History of dyslipidemia: No	443508001  No history of clinical finding in subject (situation)  : {246090004  Associated finding (attribute)  = 370992007  Dyslipidemia (disorder) }



out of 12 items that formed one-to-many relations were associated with “current medication status.” The remainder were related to “smoking experience.” Drug prescriptions for stroke, ischemic heart disease, and diabetes mellitus varied according to patients’ health status. Therefore, we mapped standardized terminologies that considered every possible drug. Six items about the lifetime experience of smoking (regardless of the kind of tobacco) were also mapped to one-to-many relationships. Even if an examinee answered “yes,” we could not infer the respondent’s current smoking status. Accordingly, we mapped this response to both “current smoker” and “ex-smoker.”

Medical history made up 69.2% of non-mapped cases, followed by smoking and e-cigarette use (23.1%) and exercising (7.7%). The detailed mapping results are shown in Supplement C.

## IV. Discussion

We tried to standardized term mapping using the Korean general health checkup questionnaire, which is PGHD in a broad sense, as the source. To achieve this goal, we used two terminology systems (SNOMED CT and LOINC), as each has its own distinct characteristics. Concerns have been raised regarding the use of a multi-standard terminology system in a single mapping table; however, this is not a substantial problem, because if the mapping table contains the information of the terminology system, computer-based systems can process that information appropriately.

We found that most items in the national health checkup questionnaire could be expressed using standard terminology. This study present helpful results based on earlier works. These studies conducted by us and used the previous Korean national health checkup questionnaires. Compared with a previous study that was produced with the same set of conditions and implemented post-coordinated expressions, there were differences in the mapping results. Table 4 shows

the results of the comparison.

First, we split and reorganized the items, as previously described in the Methods section. For this reason, the number of total items changed, but the content of the source template remained consistent. Second, we realized that we misinterpreted the question related to “medication therapy” in the previous study. The question aims to determine whether a person currently takes medication, not the person’s drug history. By correcting this error in the past study, the proportion of post-coordinated expressions declined and that of pre-coordinated expressions increased. Last, we changed the mapping process. In the 2019 study, we applied different methods according to each group and each item’s characteristics. However, the present study did not differentiate between groups to express the meanings. Therefore, the proportion of the items matched to the standard terminology system changed.

Apart from the research outcomes, this study found that there were some issues in the general health checkup questionnaire, as detailed below.

### 1. Ambiguous Item: “Other”

Two different concepts (“other” [including cancer] and “cardiac infarction/angina,” which is a subtype of ischemic heart disease) were expressed as single items. A standard terminology system, such as SNOMED CT, does not include nebulous concepts; these concepts are therefore indescribable in standard terminology. The combined concept might also be expressed using broad concepts, such as “312850006 [History of disorder (situation)].” However, doing so would be uninformative, as it would contain a comprehensive meaning. To ensure that the data can be utilized, clearly worded items should be organized in the questionnaire.

However, we also recognize that the purpose of general health checkups in Korea is to promote health through the prompt detection of cardiocerebrovascular diseases, such as hypertension and diabetes, and linkage of patients to

Table 4. Comparison of the present results to the previous study results (2019 vs. 2020)

Classification			Number of mapping items	
			2019	2020
Mapped	SNOMED	Pre-coordinated expression	18	35
		Post-coordinated expression	21	17
	LOINC	-	11	24
Non-mapped			23	13
Total			73	89

SNOMED CT: Systematized Nomenclature Of Medicine Clinical Terms, LOINC: Logical Observation Identifiers Names and Codes.

treatment and follow-up management. For this reason, two concepts, including a personal medical history of malignant neoplasm and other diseases and family medical history, have remained unchanged on the questionnaire over the last 10 years. However, to utilize the data of this questionnaire, the item should be revised.

**2. Who is Included in the Family Medical History?**

The family medical history is an important factor, as family members share many determinants of health, such as genes, dietary habits, economic levels, and residential environments. According to the National Cancer Institute (NCI) thesaurus, family medical history is a record of a patient’s background regarding the health and disease of blood relatives [21]. Other references, such as HealthWA [22], WebMD [23], and Healthline [24], only include blood relatives in the scope of family history.

The general health checkup questionnaires contain family history questions in the form “Has anyone in your family (parents and siblings) died from or gotten any of the following diseases?” The problem is that the general health checkup questionnaire misses non-blood relations. Most Koreans think of a blood relationship when they think of family. However, there are many non-blood relationships, such as those in adoptive families and step-families through remarriage.

As shown in Figure 4, there are two SNOMED CT concepts related to “family history of diabetes: yes.” One is “family history of diabetes” (SCTID: 160303001). The other is “family history of diabetes mellitus in first degree relative” (SCTID: 416855002). The values of the two concepts, corresponding to attributes including associated findings, the context of the findings, and the temporal context are equal; only the value of the subject relationship context varies (Figure 5), and this variation relates to family boundaries. The first question relates to a “person in family of subject” (SCTID: 444148008) and the other relates to a “first degree blood relative of subject” (SCTID: 444193000). If only information about natural family history was included, we would map this onto “family history of diabetes mellitus in first degree relative.”

However, we thought that the inclusion of information about family history including non-blood-relatives would also be important. Thus, we mapped this item onto “family history of diabetes,” and did the same for the other items with this structure.

Regardless of the specific results, taking a comprehensive range of real-world factors into consideration, the collected data hardly met international standards. Considering the respondents, the questionnaire forms need to be changed to reflect answers beyond just blood relatives to record specific information on all members of a family.

**3. Narrow Range**

According to Article 27-2 of the enforcement decree of the National Health Promotion Act [25], the classification of tobacco is as follows: cigarettes, electronic cigarettes, pipe tobacco, cigars, rolling tobacco, chewing tobacco, inhaling tobacco, waterpipe tobacco, and snuff. The Food and Drug Administration (FDA) also regulates the following tobacco products: cigarettes, roll-your-own tobacco products, smokeless tobacco products, electronic nicotine delivery systems (ENDS), cigars, pipe tobacco products, waterpipe tobacco products, and others [26]. The detailed items vary across countries.

The general health questionnaire contains questions about smoking status, such as “Have you ever smoked more than five packs of cigarettes (100 cigarettes) in your lifetime?” Tobacco products are not limited to cigarettes, but include other products such as cigars, pipe tobacco, and others excluded from that question. Although most smokers in Korea consume cigarettes [27,28], it is not recommended to ask about only a limited range of tobacco products.

**4. Doubts about the Value of Data**

A question about electronic cigarettes appeared for the first time in 2018. Since 2019, the questions have been divided by product types—heated tobacco products and liquid electronic cigarettes—to examine respondents’ smoking behavior and history (amount and duration of smoking).

Although revising the question to reflect trends is desir-

2. Has anyone in your family (parents and siblings) died from or gotten any of the following diseases?

...	...	...
Diabetes	Yes	No
...	...	...

- Family history of diabetes: yes
- ① 160303001 |Family history of diabetes mellitus (situation)
- ② 416855002 |Family history of diabetes mellitus in first degree relative (situation)

Figure 4. Questions related to family medical history.

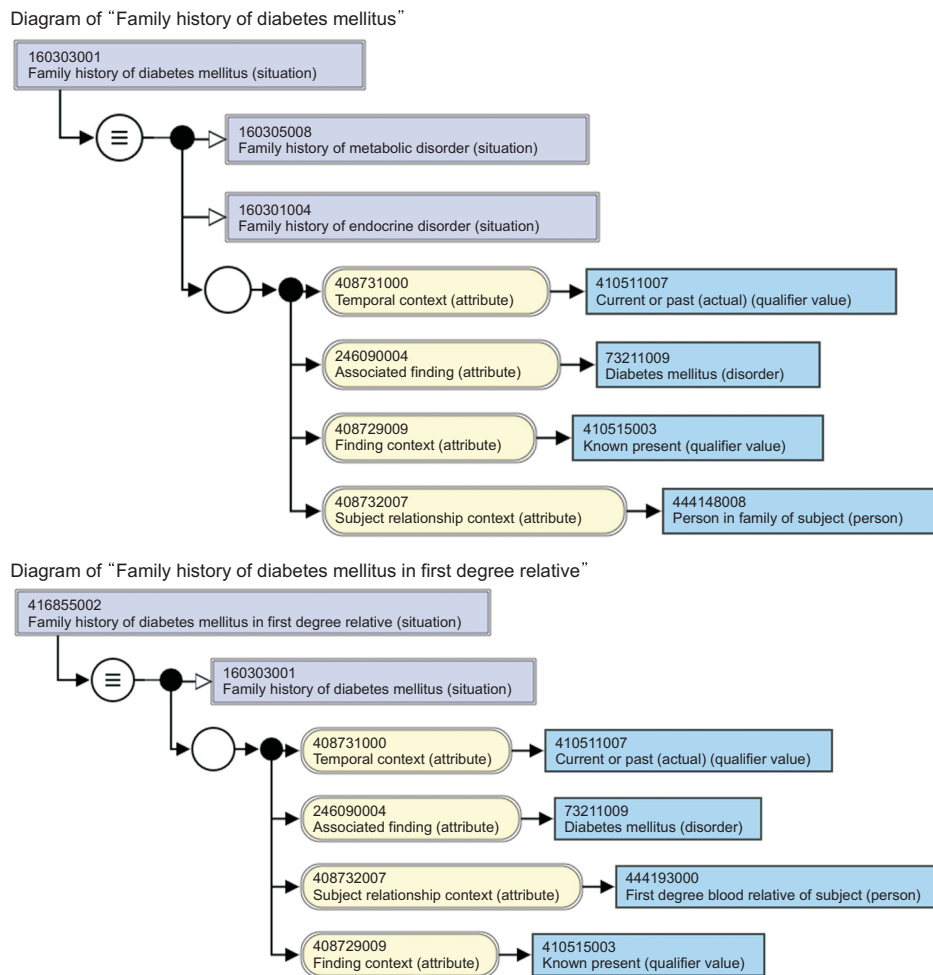


Figure 5. Diagram of SCTIDs 160303001 and 416855002.

able, there are doubts about the utilization of data from items related to liquid electronic cigarettes. Questions 6 and 6-1 in Supplement A ask about liquid electronic cigarettes in the general health checkup. The questionnaire asks respondents whether they have used liquid electronic cigarettes and the frequency of their use. These questions should be revised to collect information about the duration of smoking history and the amount of liquid electronic cigarettes consumed.

### 5. Understanding Implicit Meaning

As shown in question 7-1 in Supplement A, questions about drinking quantity (both typical drinking days and heavy drinking days) are designed to provide information on both the type of alcohol consumed and the units consumed. Therefore, we first attempted to map all alcoholic beverages and units using the descendant concepts of "alcoholic beverage" (SCTID: 53527002) and "unit of measure" (SCTID: 767524001). Eventually, we created standardized terminology mappings focused on the desired outcomes from the questions that did not just involve Korea-specific alcohol

types (e.g., *soju* and *makgeolli*), which are difficult to express using standardized terminology, but any liquor type.

This question ultimately aims to quantify the drinking volume and to identify personal risk. To do this, the value entered in this question is converted into the number of alcohol units. Therefore, each question about drinking quantity was mapped to the concepts "number of alcohol units consumed on typical drinking day" (SCTID: 443315005) and "number of alcohol units consumed on heaviest drinking day" (SCTID: 442547005).

### 6. Validity of the Questionnaire

The terms used in mapping related to exercising are items on the International Physical Activity Questionnaire (IPAQ). However, we did not confirm whether questions about exercising in the general checkup questionnaire were translations from English to Korean. As shown in Table 5, this study only considered their implications. If the research results are to be applied in practice, this point will have to be addressed.

In this study, as mentioned above, more items were repre-



Table 5. International Physical Activity Questionnaire (IPAQ) items corresponding to questions

Question in general health checkup	Used terms to map
8-1. How often do you do high intensity exercise (making you short of breath) per week? · Examples of high intensity exercise: Running, aerobics, fast bicycling, construction labor, carrying items using stairs, etc.	77583-3 [During the last 7 days, on how many days did you do vigorous physical activities like heavy lifting, digging, aerobics, or fast bicycling for at least 10 minutes at a time [IPAQ]]
8-2. How long do you do high intensity exercise (making you short of breath) per day?	77584-1 [How much time did you usually spend doing vigorous physical activities on one of those days [IPAQ]]
9-1. How often do you do moderate intensity exercise (making you slightly short of breath) per week? · Exclude exercise you have already written in Question 8. · Examples of moderate intensity exercise: Power walking, doubles tennis games, cycling at normal speed, carrying light items, cleaning, etc.	77585-8 [During the last 7 days, on how many days did you do moderate physical activities like carrying light loads, bicycling at a regular pace, or doubles tennis for at least 10 minutes at a time [IPAQ]]
9-2. How long do you do moderate intensity exercise (making you slightly short of breath) per day?	77586-6 [How much time did you usually spend doing moderate physical activities on one of those days [IPAQ]]

sented with standardized terminology than in prior studies. Nevertheless, non-mapped items still existed, and some of them can probably be mapped if the questionnaire is revised. We acknowledge that the standard terminologies used in our research do not include all clinical concepts and that not all items can be expressed using standard terminologies. However, we expect that if the questionnaire is revised so that one item corresponds to one meaning, an explanation is added, and the use of data is considered, it would be possible to generate reliable PGHD for future clinical research.

Nonetheless, we believe that our study made a noteworthy contribution by mapping PGHD, which is one of the key factors of health determinants, onto standard terminology for the first time in a novel manner. The mapped items, such as medical history, smoking and e-cigarettes, drinking, and exercising, can be used for both clinical practice and research as fundamental elements of factors used to evaluate personal health state. Previous studies have usually aimed to map clinical data from sources such as examinations for assessments, diagnoses, operations, or treatment procedures. In those cases, there are consistent patterns according to the scope—for instance, diagnoses use a hierarchy of disorders or findings, and examinations, operations, and procedures use hierarchies of procedures or regimes/therapies. However, since a questionnaire consists of various sub-categories, it is much more difficult to map onto standard terminology. Mapping of the questionnaire should consider an appropriate hierarchy (e.g., question-answer pairs) and answer value sets. Therefore, representing the questionnaire using standard terminology needs an additional step, such as item

analysis and term extraction.

Our study is part of a research program aiming to use SNOMED CT to represent and share clinical data. All readers are welcome to use the results of our research shown in Supplement C for their research, such as data-driven projects for the Ministry of Health and Welfare or a clinical oncology network for unifying electronic medical data of the National Cancer Data Center. We hope that this study will help encourage the implementation of standard terminologies.

## Conflict of Interest

Hyeoun-Ae Park is a member of the Editorial Board of Healthcare Informatics Research; however, she did not involve in the peer review evaluation and decision process of this article. Otherwise, no potential conflict of interest relevant to this article was reported.

## Acknowledgments

This research was supported by a grant of the Medical data-driven hospital support project through the Korea Health Information Service (KHIS), funded by the Ministry of Health & Welfare, Republic of Korea. This research was supported by a fund (No. 2021-ER1002-00) from Research of Korea Disease Control and Prevention Agency.

## ORCID

Ji Eun Hwang (<https://orcid.org/0000-0002-7920-0640>)

Hyeoun-Ae Park (<https://orcid.org/0000-0002-3770-4998>)

Soo-Yong Shin (<https://orcid.org/0000-0002-2410-6120>)

## Supplementart Materials

Supplementary materials can be found via <https://doi.org/10.4258/hir.2021.27.4.287>.

## References

1. Korea Health Information Service. 2020 Survey of the healthcare information [Internet]. Seoul, Korea: Korea Health Information Service; 2021 [cited at 2021 Sep 30]. Available from: [https://www.k-his.or.kr/board.es?mid=a10306040000&bid=0005&tag=&act=view&list\\_no=283](https://www.k-his.or.kr/board.es?mid=a10306040000&bid=0005&tag=&act=view&list_no=283).
2. Yun JH, Ahn SJ, Kim Y. Development of clinical contents model markup language for electronic health records. *Healthc Inform Res* 2012;18(3):171-7.
3. Park YR, Shin SY. Status and direction of healthcare data in Korea for artificial intelligence. *Hanyang Med Rev* 2017;37(2):86-92.
4. NHS Digital. NHS e-referral service [Internet]. London, UK: NHS; 2021 [cited at 2021 Sep 30] Available from: <https://digital.nhs.uk/services/e-referral-service>.
5. NHS Digital. Referring a patient: NHS e-referral service [Internet]. London, UK: NHS; 2021 [cited at 2021 Sep 30] Available from: <https://digital.nhs.uk/services/e-referral-service/document-library/referring-a-patient>.
6. SNOMED International. Case study: Columbia international ehealth laboratory concept dictionary for openMRS [Internet]. London, UK: SNOMED International; c2021 [cited at 2021 Sep 30]. Available from: <https://www.snomed.org/snomed-ct/case-studies/columbia-international-ehealth-laboratory-concept>.
7. SNOMED International. Case study: National death registry of India [Internet]. London, UK: SNOMED International; c2021 [cited at 2021 Sep 30]. Available from: <https://www.snomed.org/snomed-ct/case-studies/national-death-registry-of-india>.
8. Shin SY. Current status and future direction of digital health in Korea. *Korean J Physiol Pharmacol* 2019;23(5): 311-5.
9. Lai AM, Hsueh PS, Choi YK, Austin RR. Present and future trends in consumer health informatics and patient-generated health data. *Yearb Med Inform* 2017;26(1): 152-9.
10. Demiris G, Iribarren SJ, Sward K, Lee S, Yang R. Patient generated health data use in clinical practice: a systematic review. *Nurs Outlook* 2019;67(4):311-30.
11. Austin E, Lee JR, Amtmann D, Bloch R, Lawrence SO, McCall D, et al. Use of patient-generated health data across healthcare settings: implications for health systems. *JAMIA Open* 2019;3(1):70-6.
12. Office of the National Coordinator for Health Information Technology (ONC). What are patient-generated health data? [Internet]. Washington (DC): ONC; 2018 [cited at 2021 Sep 30]. Available from: <https://www.healthit.gov/topic/otherhot-topics/what-are-patient-generated-health-data>.
13. Figueiredo MC, Chen Y. Patient-generated health data: dimensions, challenges, and open questions. *Found Trends Hum Comput Interact* 2020;13(3):165-297.
14. Shapiro M, Johnston D, Wald J, Mon D. Patient-generated health data [Internet]. Research Triangle Park (NC): RTI International; 2012 [cited at 2021 Sep 30]. Available from: <https://www.rti.org/publication/patient-generated-health-data-white-paper/fulltext.pdf>.
15. Office of the National Coordinator for Health Information Technology (ONC). Patient-generated health Data [Internet]. Washington (DC): ONC; 2018 [cited at 2021 Sep 30]. Available from: <https://www.healthit.gov/topic/scientific-initiatives/patient-generated-health-data>.
16. Campbell WS, Karlsson D, Vreeman DJ, Lazenby AJ, Talmon GA, Campbell JR. A computable pathology report for precision medicine: extending an observables ontology unifying SNOMED CT and LOINC. *J Am Med Inform Assoc* 2018;25(3):259-66.
17. Batool R, Khattak AM, Kim TS, Lee S. Automatic extraction and mapping of discharge summary's concepts into SNOMED CT. *Proceedings of 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*; 2013 Jul 3-7; Osaka, Japan. p. 4195-8.
18. Seto R, Inoue T, Katayama S. Terminology gap in continuous care between acute and long-term care hospitals. *Stud Health Technol Inform* 2019;264:1771-2.
19. National Health Insurance Service. 2019 National health screening statistical yearbook [Internet]. Wonju, Korea: National Health Insurance Service; 2020 [cited at 2021 Sep 30]. Available from: <https://www.nhis.or.kr/nhis/together/wbhaec07000m01.do?mode=download&articleNo=10803730&attachNo=312981>.
20. SNOMED International. SNOMED CT machine readable concept model [Internet]. London, UK: SNOMED International; c2021 [cited at 2021 Sep 30]. Available from: <https://confluence.ihtsdotools.org/display/DOC->

- MRCM.
21. National Cancer Institute. NCI dictionaries: “Family medical history” [Internet]. Bethesda (MD): National Cancer Institute; c2021 [cited at 2021 Sep 30]. Available from: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/family-medical-history>.
  22. Government of Western Australia. Your family health history [Internet]. East Perth, Australia: Government of Western Australia; c2021 [cited at 2021 Sep 30]. Available from: [https://healthywa.wa.gov.au/Articles/U\\_Z/Your-family-health-history](https://healthywa.wa.gov.au/Articles/U_Z/Your-family-health-history).
  23. WebMD. Family health history: what you should know [Internet]. New York (NY): WebMD; 2021 [cited at 2021 Sep 30]. Available from: <https://www.webmd.com/a-to-z-guides/ss/slideshow-family-history>.
  24. Holland K. Family health history: why it’s important and what you should know [Internet]. San Francisco (CA): Healthline Media; 2017 [cited at 2021 Sep 30]. Available from: <https://www.healthline.com/health/family-health-history-day>.
  25. Korea Legislation Research Institute. Enforcement Decree of the National Health Promotion Act. Seoul, Korea: Korea Legislation Research Institute; 2021.
  26. Food and Drug Administration. Tobacco Products: principles for designing and conducting tobacco product perception and intention studies [Internet]. Silver Spring (MD): Food and Drug Administration; 2020 [cited at 2021 Sep 30]. Available from: [file:///C:/Users/c/Downloads/FDA-2019-D-4188-0002\\_attachment\\_1.pdf](file:///C:/Users/c/Downloads/FDA-2019-D-4188-0002_attachment_1.pdf).
  27. Rhee MS, Lee DW. Current profile of the tobacco industry in Korea. Jeju, Korea: Cooperation Center for Scientific Research Relative to Tobacco; 2007.
  28. Lee CM. Trends in tobacco use behaviors among Korean adults, 2013-2018: comparison with tobacco sales. *Public Health Wkly Rep* 2020;13(22):1561-72.