

Proposal: Patient-Generated Health Data (PGHD) conventions for the OMOP Common Data Model

V1.2, Feb 2023

Seoul National University Bundang Hospital (SNUBH)

Healthcare ICT Research Center (HIRC)

버전	날짜	재개정 이력
V1.0	Mar, 2021	V1.0 작성
V1.2	Feb, 2023	OMOP-CDM Vocab update 사항 반영 (키, 체중, 혈압, 심박수, 혈당) TYPE_CONCEPT_ID update Measurement Table 추가

이 연구는 2019년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임 (No.20004927, CDM 기반 바이오헬스데이터 통합 플랫폼 고도화 및 검증 기술개발)

Table of Contents

I. PGHD의 OMOP CDM 변환 소개	4
II. PHR Application 라이프로그 데이터의 OMOP CDM 변환	6
1) 다양한 데이터 수집 방법	6
2) 다양한 데이터 속성	7
3) 데이터 재가공을 통한 새로운 데이터 생성 및 변환	7
2. 표준 용어 매핑	7
1) 변환 대상 PGHD 종류	7
2) PGHD 항목/속성/단위에 대한 표준 용어 매핑	8
3) CDM 변환 규칙	12
4) PGHD data의 CDM 변환 결과	14
3. Data Quality Rules	14
1) Data Quality Rule 필요성	14
2) Data Quality Rule 정의	14
III. 건강검진설문지 데이터의 CDM 변환	16
1. 변환 대상 데이터 소개	16
2. 표준용어 매핑	17
1) 환자 기본정보	17
2) 건강정보: 환자 투약 정보, 가족력 및 과거력	18
3) 생활습관 설문지	24
4) 정신건강 설문지	36
3. Data Quality Rules	41
1) Data Quality Rule 필요성	41
2) Data Quality Rule 개발	41

I. PGHD의 OMOP CDM 변환 소개

본 가이드라인은 Patient-generated Health Data (PGHD)를 OMOP (Observational Medical Outcomes Partnership) CDM (Common Data Model)으로 변환을 위한 규칙 및 표준 용어 매핑 정보를 제공하고자 한다.

PGHD는 ONC (Office of the National Coordinator)에서 정의하는 바와 같이 환자 또는 가족, 보호자에 의해 생성, 기록, 수집되는 건강 관련 데이터로, 일상생활 중 생성 가능한 모든 정보들을 일컫는 용어이며 최근 이러한 데이터를 수집 활용하고자 다양한 기기 및 서비스들이 출시되고 있다.

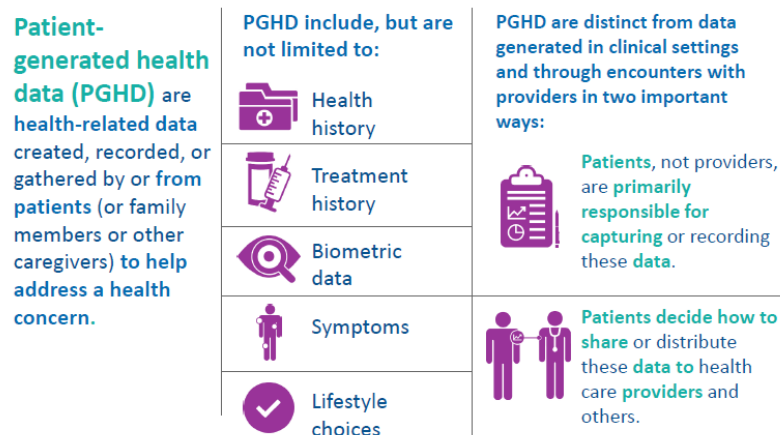


Figure 1. PGHD 범위

* Source: PGHD page on ONC website, www.healthit.gov

이러한 PGHD 중 본 가이드라인은 의료기관에서 건강검진 과정에서 수집되는 건강검진 설문지 중 생활습관과 관련된 문항과 일상생활 중 건강관리를 위해 사용하는 애플리케이션을 통해 수집되는 라이프로그 데이터를 대상으로 OMOP CDM 변환 가이드라인을 제안하고자 한다.

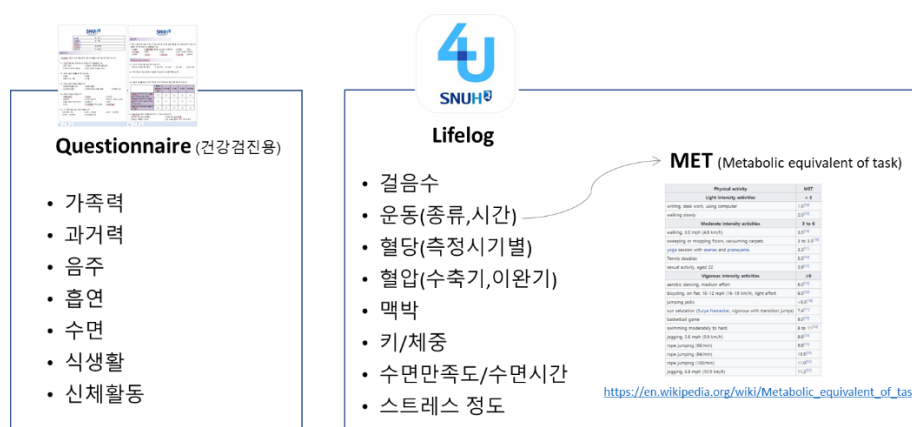


Figure 2. CDM 변환 대상

실제 데이터 변환을 검증하기 위해 사용한 데이터는 분당서울대학교병원에서 환자가 사용하는 PHR(Personal Health Record) 애플리케이션인 Health4U를 통해 수집된 데이터와 건강검진 설문지 데이터이다.

PGHD의 CDM 데이터 변환은 OBSERVATION과 MEASUREMENT 테이블을 이용한다. 이때 Figure 2.의 변환 대상에 매핑된 CONCEPT_ID의 DOMAIN에 따라 OBSERVATION과 MEASUREMENT 중 변환을 수행한다. PGHD 측정(관찰) 및 설문 결과 관련하여, PGHD 사용 기기 및 앱에 대한 정보는 DEVICE_EXPOSURE 테이블에 저장하고 응답한 설문지에 대한 정보는 SURVEY_CONDUCT 테이블을 활용하여 저장하며 OBSERVATION 테이블과의 연결을 통해, PGHD 수집 방법 및 결과 데이터를 모두 변환할 수 있도록 하였다.

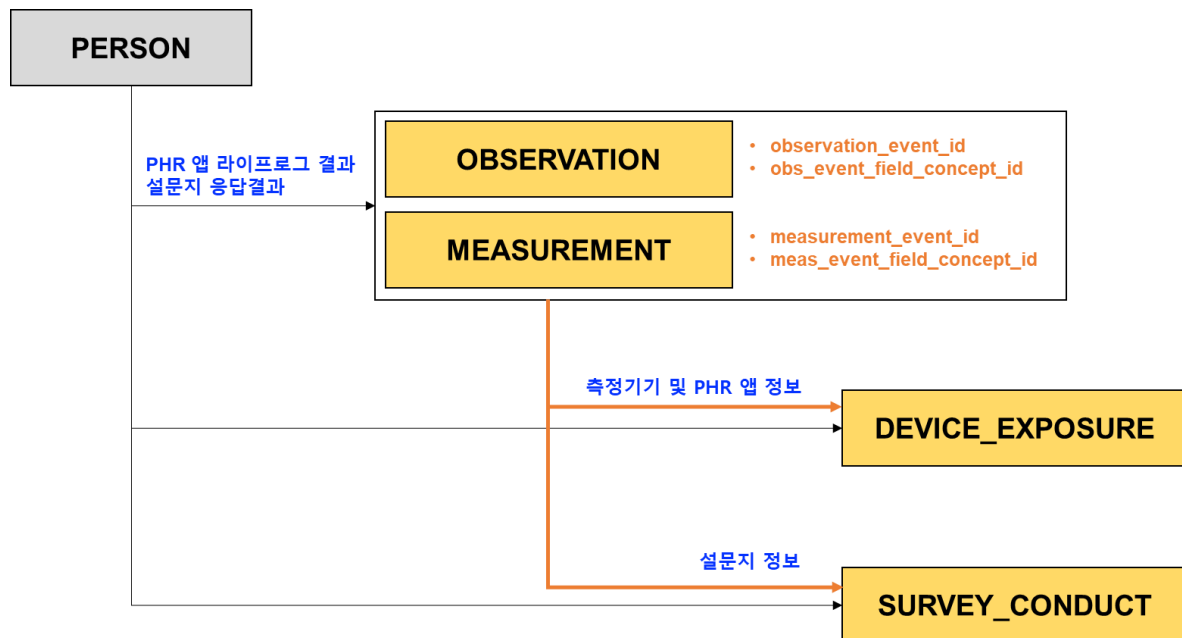


Figure 3. PGHD의 OMOP CDM 변환 개념도

II. PHR Application 라이프로그 데이터의 OMOP CDM 변환

1. 변환 대상 데이터 소개

PHR 앱으로부터 수집되는 라이프로그 데이터에 대한 CDM 변환 가능성을 검증하기 위해, 본 가이드에서는 분당서울대학교병원의 PHR 앱인 'Health4U'로부터 수집되는 건강정보를 대상으로 변환을 수행하였으며, 이를 통해 실제 의료 현장에서 수집 중인 PGHD 데이터로 변환 가능성을 보여주고자 하였다. Health4U에서 수집 중인 PGHD 데이터 종류는 아래 그림과 같다.



Figure 4. 분당서울대학교병원 Health4U (PHR App)의 수집 Data 종류

PHR 앱의 PGHD 변환 시 고려한 데이터 특성은 아래와 같다.

1) 다양한 데이터 수집 방법

- 최근 일상생활 중에 발생하는 다양한 라이프로그 데이터를 수집하기 위한 PHR Application 또는 Personal Device들이 다양하게 제품화 되어 있으며, 이를 사용하는 사람들도 지속적으로 증가하고 있다.
- 본 연구에 사용된 Health4U 앱도 특정 항목(예, 걸음수, 혈압, 혈당 등)에 대해 환자가 측정된 값을 보고 직접 수기로 입력할 수 있는 기능과 함께 개인 건강 기기를 연동할 수 있는 기능도 제공하고 있다.
- 이에 PHR 앱을 통한 수기 입력, 기기 연동 정보 등 데이터 수집 방법이 변환에 고려되었다.

2) 다양한 데이터 속성

- PGHD 데이터는 데이터 종류에 따라 다양한 속성 값을 가질 수 있다. 예를 들어, 본 가이드에서 사용한 Health4U 앱에서는 아래 예시와 같이 데이터 항목별로 속성에 해당하는 부가 정보들이 포함되어 있으며 이러한 부가 정보도 CDM 변환에 고려되었다.
 - ◆ 운동시간: 종류
 - ◆ 혈압 : 수축기, 이완기
 - ◆ 혈당 : 측정시점

3) 데이터 재가공을 통한 새로운 데이터 생성 및 변환

- PGHD 데이터를 분석에 활용하기 위해서는 수집된 raw 데이터뿐 아니라 원천 데이터를 재가공하여 새로운 데이터를 생성 및 추가할 수 있어야 한다.
- 본 가이드에서는 수집된 운동 정보를 기반으로 운동 강도에 해당하는 METs (metabolic equivalents)를 추가로 변환하는 방법을 고려하였다.

2. 표준 용어 매핑

1) 변환 대상 PGHD 종류

- 원 데이터 : 걸음수, 맥박, 수면 만족도, 수면 시간, 스트레스 정도, 운동 시간, 체중, 키, 혈당, 혈압(수축기/이완기)
- 재가공 데이터 : 운동 강도

2) PGHD 항목/속성/단위에 대한 표준 용어 매핑

➤ PGHD에서 수집되는 항목은 Observation 및 Measurement 테이블에 구성하게 되며 각 수집 항목들은 Domain에 따라 Observation_concept_id 및 measurement_concept_id 에 표현하며, 항목에 대한 단위 정보는 Unit_concept_id로 표현할 수 있다. 그리고 관련 source code에 대한 정보는 Observation_source_value/measurement_source_value와 Unit_source_value에 저장된다..

➤ 데이터 항목만 이용하여 변환 가능한 경우

source_value	concept_id	Concept_name	Vocabulary	Domain	Class	Unit
걸음수	40758552	Number of steps in unspecified time Pedometer	LOINC	Observation	Clinical Observation	44777557 (per hour)
맥박	3027018	Heart rate	LOINC	Measurement	Clinical Observation	8483 (counts per minute)
수면 만족도	2000000344	Sleep satisfaction score	SNUBH	Observation	Observable Entity	44777566 (score)
수면 시간	4086504	Duration of sleep	SNOMED	Observation	Observable Entity	8555 (second)
스트레스정도	4234506	Level of stress	SNOMED	Observation	Observable Entity	44777566 (score)
체중	3013762	Body weight Measured	LOINC	Measurement	Clinical Observation	9529 (kilogram)
키	3036277	Body height	LOINC	Measurement	Clinical Observation	8582 (centimeter)
수축기혈압	3004249	Systolic blood pressure	LOINC	Measurement	Clinical Observation	8876 (millimeter mercury)

						column)
이완기혈압	3012888	Diastolic blood pressure	LOINC	Measurement	Clinical Observation	8876 (millimeter mercury column)

➤ 데이터 항목과 속성 조합으로 변환하는 경우

Observation_ source_value	Observation_ concept_id	Qualifier_ concept_id	Concept_name	Vocabulary	Domain	Class
운동 한 시간	40758540		Exercise duration	LOINC	Observation	Clinical Observation
걷기(강함)		4044711	Walking	SNOMED	Observation	Observable Entity
걷기(약함)		4044711	Walking	SNOMED	Observation	Observable Entity
걷기(중간)		4044711	Walking	SNOMED	Observation	Observable Entity
검도		4032150	Sports activity	SNOMED	Observation	Qualifier Value
골프		4048266	Golfing	SNOMED	Observation	Qualifier Value
등산(강함)		4008063	Mountain climbing	SNOMED	Observation	Qualifier Value
등산(약함)		4008063	Mountain climbing	SNOMED	Observation	Qualifier Value
등산(중간)		4008063	Mountain climbing	SNOMED	Observation	Qualifier Value
베드민턴		4032150	Sports activity	SNOMED	Observation	Qualifier Value
볼룸댄스		4032150	Sports activity	SNOMED	Observation	Qualifier Value

Observation_ source_value	Observation_ concept_id	Qualifier_ concept_id	Concept_name	Vocabulary	Domain	Class
수영(배영)		4069419	Swimming	SNOMED	Observation	Qualifier Value
수영(자유형)		4069419	Swimming	SNOMED	Observation	Qualifier Value
수영(접영)		4069419	Swimming	SNOMED	Observation	Qualifier Value
수영(평영)		4069419	Swimming	SNOMED	Observation	Qualifier Value
스트레칭		4044212	Stretching action	SNOMED	Observation	Qualifier Value
아쿠아로빅		4032150	Sports activity	SNOMED	Observation	Qualifier Value
에어로빅		4032150	Sports activity	SNOMED	Observation	Qualifier Value
요가/필라테스		4032150	Sports activity	SNOMED	Observation	Qualifier Value
자전거(약함)		4114936	Does ride a bicycle	SNOMED	Observation	Clinical Finding
자전거(중간)		4114936	Does ride a bicycle	SNOMED	Observation	Clinical Finding
축구		4339223	Soccer	SNOMED	Observation	Qualifier Value
탁구		4032150	Sports activity	SNOMED	Observation	Qualifier Value
테니스		4131650	Tennis	SNOMED	Observation	Qualifier Value
헬스		4026921	Regular exercise	SNOMED	Observation	Qualifier Value
혈당	4017046		Blood sugar charts	SNOMED	Observation	Clinical Finding
아침식사 전		761108	Before breakfast	SNOMED	Observation	Qualifier Value
아침식사 후		764863	After breakfast	SNOMED	Observation	Qualifier Value
점심식사 전		46270469	Before lunch	SNOMED	Observation	Qualifier Value
점심식사 후		45757493	After lunch	SNOMED	Observation	Qualifier Value

Observation_ source_value	Observation_ concept_id	Qualifier_ concept_id	Concept_name	Vocabulary	Domain	Class
저녁식사 전		762254	Before dinner	SNOMED	Observation	Qualifier Value
저녁식사 후		45757489	After dinner	SNOMED	Observation	Qualifier Value
취침 전		4139386	Before sleeping	SNOMED	Observation	Qualifier Value
기타		9177	Other	SNOMED	Observation	Qualifier Value
식사 전		4145729	Before meal	SNOMED	Observation	Qualifier Value
식사 후		4088117	Postprandial	SNOMED	Observation	Qualifier Value
공복		37111339	After fasting	SNOMED	Observation	Qualifier Value
간식 후		45757496	With snack	SNOMED	Observation	Qualifier Value

- 기존 데이터를 재가공하여 추가 변환한 경우: 운동 종류 정보를 기반으로 운동 강도를 계산하여 아래와 같이 변환
- OBSERVATION.observation_concept_id: 운동 강도 컨셉
 - OBSERVATION.value_as_number: METs 수치 값
 - OBSERVATION.value_as_concept_id: 운동 강도에 대한 categorical value

Observation_ source_value	Observation_ concept_id	Value_as_ concept_id	Concept_name	Vocabulary	Domain	Class
운동강도	44786666		Exercise intensity	LOINC	Observation	Clinical Observation
		45877382	Light exercise intensity	LOINC	Meas Value	Answer
		45883346	Moderate exercise intensity	LOINC	Meas Value	Answer
		45878988	Vigorous exercise intensity	LOINC	Meas Value	Answer

3) CDM 변환 규칙

- CDMv6에 따라 PHR 앱의 CDM 변환 기준을 정의하였다

(<https://ohdsi.github.io/CommonDataModel/cdm60.html>)

- **Observation** : Observation 테이블에는 PGHD 데이터 중 OBSERVATION DOMAIN CONCEPT_ID로 매핑된 항목들이 구성되며, 수집에 사용된 개인건강기거나 응답한 설문지에 대한 정보가 있는 경우 observation_event_id와 obs_event_field_concept_id를 활용하여 연결한다.

컬럼명	ETL 기준
observation_concept_id	PGHD 항목의 concept_id
observation_date/datetime	PGHD 입력/등록 시점의 date 및 datetime
observation_type_concept_id	- App Data : 32865 (Patient self-report) 건강설문지 : 32862 (Patient filled survey)
value_as_number	수치형 값
value_as_concept_id	명목형 값의 concept_id
qualifier_concept_id	항목별 속성값의 concept_id
unit_concept_id	항목 단위의 concept_id
observation_source_value	항목에 대한 source code
observation_event_id	- App Data : Device_exposure.device_exposure_id - 건강설문지 : Survey_conduct.survey_conduct_id
obs_event_field_concept_id	- App Data : 1147693 (device_exposure.device_exposure_id) - 건강설문지 : (survey_conduct.survey_conduct_id)

- **Measurement** : Observation 테이블에는 PGHD 데이터 중 MEASUREMENT DOMAIN CONCEPT_ID로 매핑된 항목들이 구성되며, 수집에 사용된 개인건강기거나 응답한 설문지에 대한 정보가 있는 경우 measurement_event_id와 meas_event_field_concept_id를 활용하여 연결한다.

컬럼명	ETL 기준
measurement_concept_id	PGHD 항목의 concept_id
measurement_date/datetime	PGHD 입력/등록 시점의 date 및 datetime
measurement_type_concept_id	- App Data : 32865 (Patient self-report) 건강설문지 : 32862 (Patient filled survey)

value_as_number	수치형 값
value_as_concept_id	명목형 값의 concept_id
unit_concept_id	항목 단위의 concept_id
measurement_source_value	항목에 대한 source code
measurement_event_id	- App Data : Device_exposure.device_exposure_id - 건강설문지 : Survey_conduct.survey_conduct_id
meas_event_field_concept_id	- App Data : 1147693 (device_exposure.device_exposure_id) - 건강설문지 : (survey_conduct.survey_conduct_id)

- **Device_exposure** : PGHD 수집에 사용된 개인건강기기 및 앱에 대한 정보를 담는다. 본 가이드에서는 현재 Health4U 앱에서 수집되고 있는 기기 정보를 우선적으로 매핑하였다. 매핑이 불가능한 정보는 20억대의 custom concept_id로 임의 발행하였다.

컬럼명	ETL 기준
device_concept_id	DEVICE_INFO - 45759337 (Scale weight) : 개인용 체중계 - 4224372 (Blood glucose meters) : 개인용 혈당측정기 - 2614969 (Automatic blood pressure monitor) : 개인용 혈압측정기 - 2000000332 (Wearable device) : 개인용기기 - 45758275 (Patient health record information system application software) : PHR 앱 PLATFORM - 2000000011 : 삼성 S-Health - 2000000012 : APPLE Health-Kit
start/end_date	Device_concept_id 별 첫 기록 시점 ~ 마지막 기

start/end_datetime	록 시점
device_type_concept_id	Default 32865 (Patient self-report)
quantity	데이터 입력 건수
device_source_value	device_source_value platform_source_value

3. Data Quality Rules

1) Data Quality Rule 필요성

- PHR 앱을 통해 환자들이 수동으로 입력하는 데이터와 기기에서 측정된 연동 데이터를 수집할 수 있다. 자동으로 연동되어 전송된 기기 측정 데이터는 시스템 오류로 인한 극히 일부 상황을 제외하면 데이터에 오류가 있을 가능성이 낮으나, 수동 입력 데이터는 human error로 인한 오류가 있을 여지가 높다.
- 이에 PGHD 데이터 품질을 체크하기 위한 규칙이 필요하다.

2) Data Quality Rule 정의

본 가이드에서는 PGHD 앱 데이터에 대한 정해진 품질 기준이 마련되어 있지 않아, 논리적으로 성립할 수 없는 범위에 한정하여 Data Quality Rule을 개발하였다.

대상 항목	DQ Rule
걸음 수	20,000걸음(1시간당)보다 큰 경우
맥박	30보다 작은 경우
	250보다 큰 경우
수면 만족도	1~5점 척도 기준의 경우, 해당 범위를 벗어나는 경우
수면 시간	1초보다 작은 경우
	86400초(24시간)보다 큰 경우
스트레스 정도	1~5점 척도 기준의 경우, 해당 5점척도를 벗어나는 경우
운동 시간	1분보다 작은 경우
	600분(10시간)보다 큰 경우
체중	1kg보다 작은 경우
	250kg보다 큰 경우 ¹⁾
키	1cm보다 작은 경우
	만 18세 이상 성인 중 100cm보다 작은 경우
	250cm보다 큰 경우 ²⁾

대상 항목	DQ Rule
혈당	1보다 작은 경우
	600보다 큰 경우
혈압	DBP가 SBP보다 클 경우
	DBP와 SBP가 동일한 경우
	SBP와 DBP 중 어느 하나의 데이터가 없는 경우
	SBP가 1보다 작거나 300보다 큰 경우 ³⁾
	DBP가 30보다 작거나 200보다 큰 경우 ³⁾
기타	중복 레코드
	Observation 테이블의 obs_event_field_concept_id가 FK 오류인 경우

¹⁾ Hardy, R., J. Johnson, and A. Park. "CLOSER work package 1: Harmonised height, weight and BMI user guide. UK Data Service, 2016. Available from: http://doc.ukdataservice.ac.uk/doc/8207/mrdoc/pdf/closer_wp1_user_guide_v3_new_edition.pdf

²⁾ Muthalagu, A., et al. "A rigorous algorithm to detect and clean inaccurate adult height records within EHR systems." Applied clinical informatics 5.01 (2014): 118-126.

³⁾ OHDSI Data Quality Dashboard

III. 건강검진설문지 데이터의 CDM 변환

1. 변환 대상 데이터 소개

- 1) 분당서울대학교병원 건강증진센터에서 사용 중인 건강검진 설문지 서식 중 생활습관과 관련된 데이터를 우선적으로 선별하여 OMOP CDM 변환하였다.
- 2) 건강검진 설문지 데이터는 Observation 및 Survey_conduct 테이블에 적재된다. 설문지에 응답한 결과값은 Observation 테이블로 변환되고, 설문지의 응답이력 관리는 Survey_conduct 테이블에 저장되어, 각각의 두 테이블이 서로 observation.observation_event_id = survey_conduct.survey_conduct_id를 통해 연결된다.
- 3) CDM 변환 대상 데이터 종류는 아래 표에 제시된 바와 같다.

분류	세부항목
기본정보	결혼상태, 최종학력, 월 수입
건강정보	투약정보, 가족력, 과거력
생활습관	음주 정보: 음주 횟수, 음주량, 과음 횟수
	흡연 정보: 흡연력, 흡연량, 흡연시작 연령, 첫 흡연시간, 금연정보
	신체 활동 정보
	수면관련 정보: 수면시간, 졸림 정도, 수면중 무호흡증상
	식습관 정보: 규칙성, 과식여부, 식사 속도, 외식여부
정신건강	Psychosocial stress (BEPSI-K*) 설문지
	PHQ-9 quick depression assessment 설문지

-OBSERVATION Table-

CDM_Field	Type	Implement
observation_id	integer	Serial (unique) number as PK
person_id	integer	FK from Person table
observation_concept_id	integer	
observation_date	date	
observation_datetime	date	
observation_type_concept_id	integer	32862 (Patient filled survey) in general for observation_concept_id = 5609491 or 4269183, apply 32880 (Standard algorithm)
value_as_number	float	
value_as_string	varchar(60)	
value_as_concept_id	integer	
observation_event_id	integer	insert survey_concept_id
obs_event_field_concept_id	integer	1147832 (survey_conduct.survey_concept_id)

-SURVEY_CONDUCT Table-

CDM_Field	Type	Implement
survey_conduct_id	integer	Serial (unique) number as PK
person_id	integer	FK from Person table
survey_conduct_concept_id	integer	4159705 (General health questionnaire)
survey_start_date	date	
survey_start_datetime	date	
survey_end_date	date	
survey_end_datetime	date	
respondent_type_concept_id	integer	45884467: self 42530729: guardian
collection_method_concept_id	integer	42530834: Paper 42531021: Electronic
survey_source_value	varchar(100)	C_HPC_MEBI_NEW: 건강기록지 신문진 C_HPC_MEBI_OLD: 건강기록지 구문진

2. 표준용어 매핑

건강검진기록지 관련 concept_id는 분당서울대학교병원 외에도 삼성의료원과 협력하여 각기 매핑 시행한 결과를 토대로 화상회의를 통해 의견을 수렴한 결과, 각 병원의 concept_id를 통일하고, SNOMED-CT 및 LOINC 표준용어를 기반으로 채택하여 매핑 시행하기로 하였다.

지난 2021년 1월 OHDSI Forum을 통해 새로운 의료용어인 UK Biobank가 공표되어 검토하였으나, 검토 결과 가족력/과거력의 정보의 경우 UK Biobank의 데이터 원본이 Non-standard 컨셉을 가지고 Maps to relationship을 통해 SNOMED로 자동 변환되는 시스템을 구축한 것임을 확인하였다. 이에 분당서울대학교병원의 경우 건강검진설문지 원본이 UK Biobank와 일치하는 것이 아니기에, SNOMED로 직접 매핑하는 것을 선택하였다. 기타 흡연/음주 정보의 경우 UK Biobank의 컨셉이 Standard 형식을 취하고 있기는 하나, 분당서울대학교병원 건강검진설문지와 일치도를 감안하였을 때 LOINC가 더 적합하다고 판단되어 LOINC 기반의 매핑을 시행하였다.

설문지 질문문항은 observation_concept_id로 반영하고, 응답내용은 value_as_concept_id 혹은 value_as_number, 응답내용 원본은 value_as_string 필드에 저장하는 방식을 취하였다.

1) 환자 기본정보

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class
귀하의 결혼 상태를 표시해 주십시오	4076095		Marital or partnership status	SNOMED	Observation	Observable Entity
미혼		4053854	Single, never married	SNOMED	Observation	Clinical Finding
결혼		4338692	Married	SNOMED	Observation	Clinical Finding
별거 또는 이혼		4145800	Broken with partner	SNOMED	Observation	Clinical Finding
사별		4143188	Widowed	SNOMED	Observation	Clinical Finding
귀하의 최종 학력은 어떻습니까?	4076229		Education received in the past	SNOMED	Observation	Observable Entity

Observation_ source_value	Observation_ concept_id	Value_as_ concept_id	Concept_name	Vocabular y	Domain	Class
초등학교 졸업 이하		44800023	Educated to primary school level	SNOMED	Observation	Clinical Finding
중학교 졸업		4073580	Educated to junior school level	SNOMED	Observation	Clinical Finding
고등학교 졸업		43021808	Educated to high school level	SNOMED	Observation	Clinical Finding
대학교(전문대 포함) 졸업		4072735	Received university education	SNOMED	Observation	Clinical Finding
대학원 이상		4141366	Received graduate education	SNOMED	Observation	Clinical Finding
귀 가정의 평균 월 수입은 어떻 습니까?†	4076114		Household income	SNOMED	Observation	Observable Entity

† 월 수입 문항의 value 금액구간의 평균값을 value_as_number에 적재 (예: "500만원이상~800만원 미만"으
로 응답한 경우 value_as_concept_id = null, value_as_number = 650으로 적재)

2) 건강정보: 환자 투약 정보, 가족력 및 과거력

투약정보의 경우 응답에 해당하는 value값이 다양하여 LOINC 혹은 SNOMED-CT 단
일 표준용어로 매핑이 불가능하다. 이에 2가지 표준용어를 모두 사용하여 매핑 시행
하였고, 이 중 LOINC의 비중이 높아 질문에 해당하는 observation_concept_id를
LOINC 기반의 40768086(Are you currently taking any other medication on a daily
basis [PhenX]) 로 채택하였다.

건강정보 중 가족력 및 과거력과 관련한 information modeling은 THEMIS에서 논의
된 내용(<https://forums.ohdsi.org/t/how-to-represent-family-history/3386>)을 참조하여
이루어졌으며, UK Biobank의 precoordinated form에서 최종적으로
value_as_concept_id로 적재되는 형식을 따라서 고안하였다. Observation_concept_id
는 4083519(Family history of disorder; 가족력) 혹은 4214956(History of clinical
finding in subject; 과거력)의 Observation 도메인의 컨셉을 사용하고, 그에 대한
value_as_concept_id는 SNOMED-CT의 Condition 도메인에 해당하는 질환명을 매핑

하여 적용하였다.

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class
현재 복용하고 있는 약이 있으면 해당 항목에 표시하여 주십시오	40768086		Are you currently taking any other medication on a daily basis [PhenX]	LOINC	Observation	Clinical Observation
혈압약		45884606	Prescription medicine (other than water pill) for high blood pressure	LOINC	Meas Value	Answer
당뇨약		45882294	Prescription medicine for diabetes	LOINC	Meas Value	Answer
고지혈증약		40761394	Taken medication for high blood cholesterol	LOINC	Observation	Clinical Observation
아스피린		36308940	Aspirin	LOINC	Meas Value	Answer
항응고제(코마딘, 티클로피딘 등)		1035022	Anticoagulant	LOINC	Meas Value	Answer
갑상선약		2000000376	Thyroid hormone therapy	SNUBH generated		
호르몬제제(경구 피임약, 여성 호르몬제제 혹은 남성 호르몬제제 등)		45881892	Hormone Therapy Administered	LOINC	Meas Value	Answer
소염진통제 (관절염 약)		45876559	Naproxen, ketoprofen or other non-steroidal	LOINC	Meas Value	Answer
진정제 혹은 수면제		45883390	Sedatives, for example, sleeping pills, barbiturates, seconal, quaaludes, or chloral hydrate	LOINC	Meas Value	Answer
정신과 처방 약물		45882960	Prescription medicine for depression or anxiety	LOINC	Meas Value	Answer
한약 혹은 보약		2000000377	Oriental medicine	SNUBH generated		

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class
비타민제		45876566	Vitamins or minerals	LOINC	Meas Value	Answer
칼슘제		4047466	Calcium sulfate	SNOMED	Observation	Observable Entity
철분제		4113343	Iron intake	SNOMED	Observation	Observable Entity
기타 영양보조식품		45878128	Dietary supplements (other than vitamins or minerals)	LOINC	Meas Value	Answer
기타 약제		3024706	Other medications	LOINC	Observation	Clinical Observation
없음		37019178	Does not take medication	SNOMED	Observation	Clinical Finding
부모나 형제자매, 삼촌 이내의 가까운 친척 중 아래와 같은 질환을 진단받은 경우가 있습니까?	4083519		Family history of disorder	SNOMED	Observation	Context-dependent
뇌졸중		381316	Cerebrovascular accident	SNOMED	Condition	Clinical Finding
협심증, 심근경색		134057	Disorder of cardiovascular system	SNOMED	Condition	Clinical Finding
고혈압		316866	Hypertensive disorder	SNOMED	Condition	Clinical Finding
당뇨병		201820	Diabetes mellitus	SNOMED	Condition	Clinical Finding
고지혈증		4029305	Hypercholesterolemia	SNOMED	Condition	Clinical Finding
위암		443387	Malignant tumor of stomach	SNOMED	Condition	Clinical Finding
폐암		443388	Malignant tumor of lung	SNOMED	Condition	Clinical Finding
간암		194984	Disease of liver	SNOMED	Condition	Clinical Finding
대장암		443396	Malignant tumor of large intestine	SNOMED	Condition	Clinical Finding
유방암		4112853	Malignant tumor of breast	SNOMED	Condition	Clinical Finding
난소암		4181351	Malignant tumor of ovary	SNOMED	Condition	Clinical Finding
전립선암		4163261	Malignant tumor of prostate	SNOMED	Condition	Clinical Finding
갑상선암		4178976	Malignant tumor of thyroid gland	SNOMED	Condition	Clinical Finding

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class
췌장암		4180793	Malignant tumor of pancreas	SNOMED	Condition	Clinical Finding
기타		443392	Malignant neoplastic disease	SNOMED	Condition	Clinical Finding
치매		4182210	Dementia	SNOMED	Condition	Clinical Finding
귀하는 암으로 진단받으신 적이 있습니까?	4214956		History of clinical finding in subject	SNOMED	Observation	Context-dependent
위암		443387	Malignant tumor of stomach	SNOMED	Condition	Clinical Finding
폐암		443388	Malignant tumor of lung	SNOMED	Condition	Clinical Finding
간암		4246127	Malignant neoplasm of liver	SNOMED	Condition	Clinical Finding
대장암		443396	Malignant tumor of large intestine	SNOMED	Condition	Clinical Finding
갑상선암		4178976	Malignant tumor of thyroid gland	SNOMED	Condition	Clinical Finding
유방암		4112853	Malignant tumor of breast	SNOMED	Condition	Clinical Finding
자궁경부암		198984	Malignant tumor of cervix	SNOMED	Condition	Clinical Finding
전립선암		4163261	Malignant tumor of prostate	SNOMED	Condition	Clinical Finding
기타		443392	Malignant neoplastic disease	SNOMED	Condition	Clinical Finding
없음		45763843	No history of malignant neoplastic disease	SNOMED	Observation	Context-dependent
귀하는 다음의 질병으로 진단받은 적이 있습니까?	4214956		History of clinical finding in subject	SNOMED	Observation	Context-dependent
고혈압		316866	Hypertensive disorder	SNOMED	Condition	Clinical Finding
당뇨		201820	Diabetes mellitus	SNOMED	Condition	Clinical Finding
고지혈증		4029305	Hypercholesterolemia	SNOMED	Condition	Clinical Finding
지방간		4059290	Steatosis of liver	SNOMED	Condition	Clinical Finding
협심증, 심근경색		4329847	Myocardial infarction	SNOMED	Condition	Clinical Finding
뇌졸중		4310996	Ischemic stroke	SNOMED	Condition	Clinical Finding

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class
만성 간질환(간경화)		4064161	Cirrhosis of liver	SNOMED	Condition	Clinical Finding
담석, 담낭염		192353	Disorder of gallbladder	SNOMED	Condition	Clinical Finding
위궤양, 십이지장궤양		4265600	Gastric ulcer	SNOMED	Condition	Clinical Finding
갑상선질환		141253	Disorder of thyroid gland	SNOMED	Condition	Clinical Finding
결핵		434557	Tuberculosis	SNOMED	Condition	Clinical Finding
천식, 알레르기		317009	Asthma	SNOMED	Condition	Clinical Finding
전립선질환		196734	Disorder of prostate	SNOMED	Condition	Clinical Finding
녹내장		375252	Disorder of eye	SNOMED	Condition	Clinical Finding
신부전증		198124	Kidney disease	SNOMED	Condition	Clinical Finding
골다공증		80502	Osteoporosis	SNOMED	Condition	Clinical Finding
관절염		4291025	Arthritis	SNOMED	Condition	Clinical Finding
디스크		75344	Intervertebral disc disorder	SNOMED	Condition	Clinical Finding
척추 협착증		77079	Spinal stenosis	SNOMED	Condition	Clinical Finding
없음		40481925	No history of clinical finding in subject	SNOMED	Observation	Context-dependent
기타		45878545	Other/Unknown	LOINC	Meas Value	Answer

- 가족력 및 과거력의 경우 SNOMED-CT를 1순위 매핑하고, 약물의 경우 LOINC가 matched concept 비율이 높아 LOINC 1순위 채택함.
- 가족력의 경우 건강검진 설문지에는 가족구성원 중 누구의 가족력에 해당하는지 정보가 없어 CDM에 적용하지 않았으나, 초기간호정보기록지에서 추출되는 가족력 정보와 같이 가족력의 구성원 정보가 있는 경우, qualifier_concept_id에 해당 정보를 적재할 수 있다.

Observation_source_value	Observation_concept_id	Qualifier_concept_id	Concept_name	Vocabulary	Domain	Class
부모나 형제자매, 삼촌 이내의 가까운 친척 중 아래와 같은 질환을 진단받은 경우가 있습니까?	4083519		Family history of disorder	SNOMED	Observation	Context-dependent
부		4321888	Natural father	Relationship	SNOMED	Social Context
모		4277283	Natural mother"	Relationship	SNOMED	Social Context
형제		4218412	Natural sibling	Relationship	SNOMED	Social Context

3) 생활습관 설문지

① 음주 및 흡연 정보

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class	Value_as_number	Unit
▶ 음주 관련 문항								
귀하는 술을 얼마나 자주 마십니까?	40771103		How often do you have a drink containing alcohol	LOINC	Observation	Survey		
전혀 안 마심		45876662	Never	LOINC	Meas Value	Answer	0.00	44777559 (per week)
월 1회 이하		45879058	Monthly of less	LOINC	Meas Value	Answer	0.23	44777559 (per week)
월 2~4회		45885058	2-4 times a month	LOINC	Meas Value	Answer	0.75	44777559 (per week)
주 2~3회		45877711	2-3 times a week	LOINC	Meas Value	Answer	2.50	44777559

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class	Value_as_number	Unit
								(per week)
주4회 이상		45885059	4 or more times a week	LOINC	Meas Value	Answer	5.00	44777559 (per week)
거의 매일		45879676	Daily or almost daily	LOINC	Meas Value	Answer	7.00	44777559 (per week)
귀하는 한번에 술을 얼마나 마십니까? (술의 종류 구분 없이 각각의 술잔으로 계산합니다. 단, 맥주 1캔(355cc)은 1.4잔으로 계산합니다	40771104		How many standard drinks containing alcohol do you have on a typical day [SAMHSA]	LOINC	Observation	Survey		
1~2잔		45882591	1 or 2	LOINC	Meas Value	Answer	1.50	9287 (cup)
3~4잔		45883428	3 or 4	LOINC	Meas Value	Answer	3.50	9287 (cup)
5~6잔		45877926	5 or 6	LOINC	Meas Value	Answer	5.50	9287 (cup)
7~9잔		45877712	7 to 9	LOINC	Meas Value	Answer	8.00	9287 (cup)
10잔 이상		45882591	1 or 2	LOINC	Meas Value	Answer	10.00	9287 (cup)
한번의 술좌석에서 6잔 이상을 마시는 횟수는 어느 정도입니까?	40771105		How often do you have 6 or more drinks on 1 occasion [SAMHSA]	LOINC	Observation	Survey		
전혀 없다		45876662	Never	LOINC	Meas Value	Answer		
월 1회 미만		45885060	Less than monthly	LOINC	Meas Value	Answer		
월 1회		45879060	Monthly	LOINC	Meas Value	Answer		
주 1회		45877713	Weekly	LOINC	Meas Value	Answer		
거의 매일		45879676	Daily or almost daily	LOINC	Meas Value	Answer		
▶ 흡연 관련 문항								

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class	Value_as_number	Unit
귀하는 현재 담배를 피우십니까?	4041306		Tobacco use and exposure	SNOMED	Observation	Observation Entity		
전혀 피운 적이 없다. (평생 100개비 미만)		4222303	Non-smoker	SNOMED	Observation	Clinical Finding		
과거에는 피웠으나 지금은 피우지 않는다.		4310250	Ex-smoker	SNOMED	Observation	Clinical Finding		
현재도 피우고 있다.		4298794	Smoker	SNOMED	Observation	Clinical Finding		
담배를 끊으셨다면 언제 끊었습니까?	40766356		About how long has it been since you completely quit smoking cigarettes [TUS-CPS]	LOINC	Observation	Clinical Observation		9580 (Month)
끊기 전 2년 동안 하루 평균 몇 개비를 피웠습니까?	40766364		How many cigarettes per day do, or did, you smoke [FTND]	LOINC	Observation	Clinical Observation		45756923 (Cigarettes)
현재도 담배를 피우고 있다면 하루 몇 개비를 피웁니까?	40766929		How many cigarettes do you smoke per day now [PhenX]	LOINC	Observation	Clinical Observation		45756923 (Cigarettes)
처음 담배를 한대 피운 시기는 언제입니까?	40770348		At which age did you start smoking [PhenX]	LOINC	Observation	Clinical Observation		9448 (Year)
아침에 일어난 후 언제 첫 담배를 피웁니까?	40766360		How soon after you wake up do, or did, you smoke your first cigarette [FTND]	LOINC	Observation	Clinical Observation		
5분 이내		45882556	Within 5 minutes	LOINC	Meas Value	Answer		

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class	Value_as_number	Unit
			(3 points)					
6~30분		45884476	6-30 minutes (2 points)	LOINC	Meas Value	Answer		
31~60분		45882557	31-60 minutes (1 point)	LOINC	Meas Value	Answer		
1시간 이후		45881684	After 60 minutes (0 points)	LOINC	Meas Value	Answer		
최근 1년 동안 담배를 끊고 자 하루(24시간)이상 금연한 적이 있습니까?	40766643		In the past, have you ever made a serious attempt to quit smoking That is, have you stopped smoking for at least one day or longer because you were trying to quit [PLCO]	LOINC	Observation	Clinical Observation		
예		45877994	Yes	LOINC	Meas Value	Answer		
아니오		45878245	No	LOINC	Meas Value	Answer		
앞으로 담배를 끊을 계획이 있습니까?	40766642		Are you considering quitting smoking during the next 6 months [PLCO]	LOINC	Observation	Clinical Observation		
1개월 안에 금연할 계획이 있다		45876554	Yes, plan to stop within next 30 days	LOINC	Meas Value	Answer		
6개월 안에 금연할 계획이 있다.		45883499	Yes, plan to stop within next 6 months, but not within next 30 days	LOINC	Meas Value	Answer		
6개월 이내는 아니지만, 언젠가 금연할 계획이 있다.		45879997	Yes, sometimes	LOINC	Meas Value	Answer		

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class	Value_as_number	Unit
현재로서는 전혀 금연할 생각이 없다.		45881975	No, not thinking of quitting in next 6 months	LOINC	Meas Value	Answer		
최근 1년 동안 흡연 여부에 대하여 의사에게 질문을 받거나 금연권고를 받은 적이 있습니까?	40766647		During the past 12 months, did any doctor, dentist, nurse, or any other health professional advise you to quit smoking [PLCO]	LOINC	Observation	Clinical Observation		
아니오		45878245	No	LOINC	Meas Value	Answer		
흡연여부에 대한 질문을 받았으나, 금연권고를 받은 적은 없다.		45876555	No, was not advised to quit	LOINC	Meas Value	Answer		
흡연여부에 대한 질문을 받았으며, 금연권고를 간단히 받은 적이 있다.		45877994	Yes	LOINC	Meas Value	Answer		
흡연여부에 대한 질문 및 금연 권고 뿐 아니라, 구체적 금연방법에 대한 조언을 받은 적이 있다.		45877994	Yes	LOINC	Meas Value	Answer		

- 건강검진 문항에 따른 OMOP-CDM 의료표준용어 매핑은 위 도표와 같다.
- 흡연에 대한 문항은 SNOMED-CT 표준용어를 1순위로 두고, 흡연 외의 문항은 Survey Panel 내에 각각의 컨셉들 간의 Has-Answer Relationship 형성이 되어 있는 LOINC 표준용어를 1순위로 한다.
- 대다수의 문항은 원데이터를 표준용어 컨셉으로 매핑하여 적재하지만, 일부 문항의 경우 단순 문항의 대한 Mapping 뿐 아니라 연구 활용도를 고려하여 재가공 데이터를 생산한다. 예를 들면, 주간 음주량 및 흡연량(갑년)의 경우 아래 도표와 같이 Derived Value를 계산하여 CDM에 반영한다.
- Unit의 경우 주당 알코올 섭취량의 단위는 UK Biobank에서 제시한 32912(Generic unit for indivisible thing)로 반영하고, 흡연량의 단위는 45756923(cigarette; 개비)로 매핑한다. 흡연량의 단위가 non-standard이지만 OHDSI에서 제시한 가이드라인에 따르면 unit_concept_id의 standard 여부가 필수조건이 아니라 적용이 가능하다.

(<https://ohdsi.github.io/CommonDataModel/cdm60.html#OBSERVATION>)

Observation_concept_id	Concept_name	Vocabulary	Domain	Class	Unit	Type_concept_id
35609491	Alcohol units consumed per week	SNOMED	Observation	Observable Entity	32912 (Generic unit for indivisible thing)	32880 (Standard algorithm)
4269183	Cigarette pack-years	SNOMED	Observation	Observable Entity	45756923 (Cigarettes)	32880 (Standard algorithm)

② 신체활동 정보

신체활동 정보의 응답내용(일 or 시간)의 원본은 value_as_string 필드에 적재되어 있으며, 이를 각각의 unit에 맞게 환산한 값을 value_as_number 필드에 적재한다. 예를 들면, "3시간 40분"으로 응답한 경우 value_as_string 필드에는 "3시간 40분"으로 적재되지만 value_as_number 필드에는 $3*60+40 = 220$ 값이 그리고 unit_concept_id 필드에는 8550 (minute)으로 데이터가 적재된다.

Observation_source_value	Observation_concept_id	Concept_name	Vocabulary	Domain	Class	Unit
지난 7일간 무거운 물건 나르기, 달리기, 에어로빅, 빠른 속도로 자전거 타기 등과 같은 격렬한 신체 활동을 며칠간 하였습니다습니까?	36203099	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]	LOINC	Observation	Survey	8621 (day per week)
그런 날 중 하루에 격렬한 신체활동을 하면서 보낸 시간이 보통 얼마나 됩니까?	36203100	How much time did you usually spend doing vigorous physical activities on one of those days [IPAQ]	LOINC	Observation	Survey	8550 (minute)
지난 7일간, 가벼운 물건 나르기, 보통 속도로 자전거 타기, 복식 테니스 등과 같은 중간 정도 신체 활동을 며칠간 하였습니다습니까? (걷기는 포함시키지 마십시오.)	36203101	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	LOINC	Observation	Survey	8621 (day per week)

Observation_source_value	Observation_concept_id	Concept_name	Vocabulary	Domain	Class	Unit
		tennis for at least 10 minutes at a time [IPAQ]				
그런 날 중 하루에 중간정도 신체활동을 하면서 보낸 시간이 보통 얼마나 됩니까?	36203102	How much time did you usually spend doing moderate physical activities on one of those days [IPAQ]	LOINC	Observation	Survey	8550 (minute)
지난 7일간, 한 번에 적어도 10분 이상 걸은 날이 며칠입니까? (직장이나 집에서 걷기, 출퇴근이나 여가 시간에 걷기도 포함)	36203103	During the last 7 days, on how many days did you walk for at least 10 minutes at a time [IPAQ]	LOINC	Observation	Survey	8621 (day per week)
그런 날 중 하루에 걸으면서 보낸 시간이 보통 얼마나 됩니까?	36203104	How much time did you usually spend walking on one of those days [IPAQ]	LOINC	Observation	Survey	8550 (minute)
지난 7일간, 주중에 앉아서 보낸 시간이 보통 얼마나 됩니까? (학업이나 여가시간, 직장생활, 친구를 만나거나, 독서, 텔레비전 시청 시간도 포함)	36203105	During the last 7 days, how much time did you usually spend sitting on a weekday [IPAQ]	LOINC	Observation	Survey	8550 (minute)

③ 수면 정보

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class	
"낮 시간 동안 아래 증상이 얼마나 자주 나타났나요?" 텔레비전 시청할 때, 책을 읽을 때, 운전하거나 일상 활동을 하는 중 졸리다고 느꼈다.	40768306		How often do you feel so sleepy during the day that it interrupts your normal activities [PhenX]	LOINC	Observation	Clinical Observation	
1(지난 한 달 동안 없었다)		45882009	Never or nearly never	LOINC	Meas Value	Answer	
2(주 1회 미만)		45884563	1-2 times a month	LOINC	Meas Value	Answer	
3(주 1~2회)		45876595	1-2 times a week	LOINC	Meas Value	Answer	
4(주 3회 이상)		45880731	3-4 times a week	LOINC	Meas Value	Answer	
"낮 시간 동안 아래 증상이 얼마나 자주 나타났나요?" 텔레비전 시청할 때, 책을 읽을 때, 운전하거나 일상 활동을 하는 중 깜빡 졸았다	40767203		How often does this nodding off or falling asleep while driving occur [PhenX]	LOINC	Observation	Clinical Observation	
1(지난 한 달 동안 없었다)		45882009	Never or nearly never	LOINC	Meas Value	Answer	
2(주 1회 미만)		45884563	1-2 times a month	LOINC	Meas Value	Answer	
3(주 1~2회)		45876595	1-2 times a week	LOINC	Meas Value	Answer	
4(주 3회 이상)		45880731	3-4 times a week	LOINC	Meas Value	Answer	
"밤 시간 동안 아래 증상이 얼마나 자주 나타났나요?" 잠을 자려고 하는데 다리에 움직이고 싶거나 간지럽거나 불편한 느낌이 나타났다.	37021058		I felt as if I needed to move my legs a lot in past 7 days [PROMIS.PEDS]	LOINC	Observation	Clinical Observation	
1(지난 한 달 동안 없었다)		45882009	Never or nearly never	LOINC	Meas Value	Answer	
2(주 1회 미만)		45884563	1-2 times a month	LOINC	Meas Value	Answer	
3(주 1~2회)		45876595	1-2 times a week	LOINC	Meas Value	Answer	
4(주 3회 이상)		45880731	3-4 times a week	LOINC	Meas Value	Answer	

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class	
"밤 시간 동안 아래 증상이 얼마나 자주 나타났나요?" 잘 때 크게 코를 골았다. (또는 주위 사람이 그렇게 말했다.)	40767197		How often do you snore [PhenX]	LOINC	Observation	Clinical Observation	
1(지난 한 달 동안 없었다)		45882009	Never or nearly never	LOINC	Meas Value	Answer	
2(주 1회 미만)		45884563	1-2 times a month	LOINC	Meas Value	Answer	
3(주 1~2회)		45876595	1-2 times a week	LOINC	Meas Value	Answer	
4(주 3회 이상)		45880731	3-4 times a week	LOINC	Meas Value	Answer	
"밤 시간 동안 아래 증상이 얼마나 자주 나타났나요?" 잠을 자는 도중 숨을 멈추거나 몰아쉬었다.	40767199		Has anyone noticed that you quit breathing during your sleep [PhenX]	LOINC	Observation	Clinical Observation	
1(지난 한 달 동안 없었다)		45882009	Never or nearly never	LOINC	Meas Value	Answer	
2(주 1회 미만)		45884563	1-2 times a month	LOINC	Meas Value	Answer	
3(주 1~2회)		45876595	1-2 times a week	LOINC	Meas Value	Answer	
4(주 3회 이상)		45880731	3-4 times a week	LOINC	Meas Value	Answer	
보통 하룻밤에 몇 번쯤 깨어 나시나요?	42868707		Trouble falling or staying asleep in the last month [PCL-C]	LOINC	Observation	Survey	
1회		45882928	Once	LOINC	Meas Value	Answer	
2회		45876696	Twice	LOINC	Meas Value	Answer	
3회		46238127	3 times	LOINC	Meas Value	Answer	
4회		21498429	4 times	LOINC	Meas Value	Answer	
5회		21499051	5 or more times	LOINC	Meas Value	Answer	
6회		21499051	5 or more times	LOINC	Meas Value	Answer	
7회		21499051	5 or more times	LOINC	Meas Value	Answer	
8회 이상		21499051	5 or more times	LOINC	Meas Value	Answer	
지난 한 달 동안 실제 수면 시간은 얼마	40768255		How much sleep do you	LOINC	Observation	Clinical	8550

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class	
나 됩니까?			usually get on a typical night [PhenX]			Observation	(minute)

④ 식습관 정보

식습관 정보와 관련하여 SNOMED-CT 및 LOINC 표준용어체계에서 최대한 설문문항에 해당하는 observation_concept_id를 찾아 매핑하였으나, 일부 문항 및 응답의 경우 적절한 concept_id가 등록되어 있지 않아 20억대의 custom concept_id로 임의 할당하였다.

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class
식사를 규칙적으로 하십니까?	4266613		Regularity of eating	SNOMED	Observation	Observable Entity
거의 매일 3끼 먹는다.		2000000350	I eat three meals almost every day.	SNUBH generated	Meas Value	
끼니를 거르는 날이 일주일에 1-2번 있다.		2000000351	There are 1-2 days a week for skipping meals.	SNUBH generated	Meas Value	
끼니를 거르는 날이 일주일에 3번 이상 있다.		2000000352	There are more than three times a week for skipping meals.	SNUBH generated	Meas Value	
식사를 빨리 드십니까?	2000000345		How fast do you eat meals?	SNUBH generated	Observation	
매우 빠른편이다 (10분 이하)		2000000353	Very fast(10 minutes or less)	SNUBH generated	Meas Value	
빠른편이다. (15분 이하)		2000000354	Faily fast(15 minutes or less)	SNUBH generated	Meas Value	
보통이거나 느리다 (20분 이상)		2000000355	Moderate or slow(20 minutes or more)	SNUBH	Meas Value	

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class
				generated		
과식하는 경우가 있습니까?	4241588		Overeating	SNOMED	Observation	Clinical Finding
거의 과식하지 않는다. (주 0-1회)		36210209	Less than once a week	LOINC	Meas Value	Answer
가끔 과식하는 편이다. (주 2-3회)		45877711	2-3 times a week	LOINC	Meas Value	Answer
늘 과식하는 편이다. (주 4회 이상)		45885059	4 or more times a week	LOINC	Meas Value	Answer
아침식사를 하십니까?	4125259		Breakfast routine	SNOMED	Observation	Observable Entity
거의 먹지 않는다.(주 0-2회)		2000000356	Twice a week or less	SNUBH generated	Meas Value	
가끔 먹는다. (주 3-4회)		45880731	3-4 times a week	LOINC	Meas Value	Answer
자주 먹는다. (주 5회 이상)		37079490	5 or more times a week	LOINC	Meas Value	Answer
외식을 자주 하십니까? (급식은 제외됩니다)	46236008		Yesterday, how many times did the patient eat food from any type of restaurant? #	LOINC	Observation	Survey
거의 안 먹는다. (주 0-1회)		36210209	Less than once a week	LOINC	Meas Value	Answer
가끔 먹는다. (주 2-3회)		45877711	2-3 times a week	LOINC	Meas Value	Answer
자주 먹는다. (주 4회 이상)		45885059	4 or more times a week	LOINC	Meas Value	Answer

4) 정신건강 설문지

- 정신건강관련 설문지는 크게 Psychosocial stress (BEPSI-K*) 설문지와 PHQ-9 quick depression assessment 설문지 2가지가 있다. Psychosocial stress (BEPSI-K*) 설문지의 경우 OMOP-CDM 표준용어에 매칭되는 concept_id가 없어 20억대의 custom concept_id로 임의 할당하고, PHQ-9 quick depression assessment 설문지의 경우 LOINC 용어 내 Survey Panel을 활용하여 매핑하였다.

➤ Psychosocial stress (BEPSI-K*) 설문지

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class
(지난 한 달 동안) 정신적으로나 육체적으로 감당하기 힘든 어려움들을 느낀적이 있습니까?	2000000371		For the past 1 month, have you ever suffered from severe hardships that were emotionally and physically intolerable?	SNUBH generated		
전혀없다		45876662	Never	LOINC	Meas Value	Answer
간혹 있다		45876672	Rarely	LOINC	Meas Value	Answer
종종 여러번		45882528	Sometimes	LOINC	Meas Value	Answer
거의 매번		45884455	Often	LOINC	Meas Value	Answer
언제나 항상		45882536	Always	LOINC	Meas Value	Answer
(지난 한 달 동안) 자신이 생활신념에 따라 살아가려고 애쓰다가 좌절을 느낀적이 있습니까?	2000000372		For the past 1 month, have you ever felt frustrated trying to act up to your conviction or faith?	SNUBH generated		
전혀없다		45876662	Never	LOINC	Meas Value	Answer
간혹 있다		45876672	Rarely	LOINC	Meas Value	Answer
종종 여러번		45882528	Sometimes	LOINC	Meas Value	Answer
거의 매번		45884455	Often	LOINC	Meas Value	Answer
언제나 항상		45882536	Always	LOINC	Meas Value	Answer

Observation_source_value	Observation_concept_id	Value_as_concept_id	Concept_name	Vocabulary	Domain	Class
(지난 한 달 동안) 처한 환경이 인간답게 살아가는데 부족하다고 느낀적이 있습니까?	2000000373		For The Past 1 Month, Have You Ever Felt That Your Circumstances Were Insufficient To Live Humanly?	SNUBH generated		
전혀없다		45876662	Never	LOINC	Meas Value	Answer
간혹 있다		45876672	Rarely	LOINC	Meas Value	Answer
종종 여러번		45882528	Sometimes	LOINC	Meas Value	Answer
거의 매번		45884455	Often	LOINC	Meas Value	Answer
언제나 항상		45882536	Always	LOINC	Meas Value	Answer
(지난 한 달 동안) 미래에 대해 불확실하게 느끼거나 불안해진 적이 있습니까?	2000000374		For the past 1 month, have you ever felt uncertain or uneasy about your future?	SNUBH generated		
전혀없다		45876662	Never	LOINC	Meas Value	Answer
간혹 있다		45876672	Rarely	LOINC	Meas Value	Answer
종종 여러번		45882528	Sometimes	LOINC	Meas Value	Answer
거의 매번		45884455	Often	LOINC	Meas Value	Answer
언제나 항상		45882536	Always	LOINC	Meas Value	Answer
(지난 한 달 동안) 할 일들이 너무 많아 정말 중요한 일들을 잊은 적이 있습니까?	2000000375		For the past 1 month, have you ever had many things to do, so you forgot to do the things that were really important to you?	SNUBH generated		
전혀없다		45876662	Never	LOINC	Meas Value	Answer
간혹 있다		45876672	Rarely	LOINC	Meas Value	Answer
종종 여러번		45882528	Sometimes	LOINC	Meas Value	Answer
거의 매번		45884455	Often	LOINC	Meas Value	Answer
언제나 항상		45882536	Always	LOINC	Meas Value	Answer

➤ PHQ-9 quick depression assessment 설문지

Observation_source_value	observation_concept_id	value_as_concept_id	concept_name	vocabulary	domain	class
(지난 2 주 동안) 일을 하는 것에 대해 흥미나 재미가 없음	3042924		Little interest or pleasure in doing things in last 2 weeks [Reported.PHQ]	LOINC	Observation	Survey
전혀 아니다		45883172	Not at all	LOINC	Meas Value	Answer
여러 날 동안		45879886	Several days	LOINC	Meas Value	Answer
일주일 이상		45878994	More than half the days	LOINC	Meas Value	Answer
거의 매일		45882010	Nearly every day	LOINC	Meas Value	Answer
(지난 2 주 동안) 가라앉은 느낌, 우울감 혹은 절망감	3045858		Feeling down, depressed, or hopeless in last 2 weeks [Reported.PHQ]	LOINC	Observation	Survey
전혀 아니다		45883172	Not at all	LOINC	Meas Value	Answer
여러 날 동안		45879886	Several days	LOINC	Meas Value	Answer
일주일 이상		45878994	More than half the days	LOINC	Meas Value	Answer
거의 매일		45882010	Nearly every day	LOINC	Meas Value	Answer
(지난 2 주 동안) 잠들기 어렵거나 자꾸 깨어남, 혹은 너무 많이 잠	3045933		Trouble falling or staying asleep, or sleeping too much in last 2 weeks [Reported.PHQ]	LOINC	Observation	Survey
전혀 아니다		45883172	Not at all	LOINC	Meas Value	Answer
여러 날 동안		45879886	Several days	LOINC	Meas Value	Answer
일주일 이상		45878994	More than half the days	LOINC	Meas Value	Answer
거의 매일		45882010	Nearly every day	LOINC	Meas Value	Answer
(지난 2 주 동안) 피곤감, 기력이 저하됨	3044964		Feeling tired or having little energy in last 2 weeks [Reported.PHQ]	LOINC	Observation	Survey
전혀 아니다		45883172	Not at all	LOINC	Meas Value	Answer
여러 날 동안		45879886	Several days	LOINC	Meas Value	Answer

Observation_source_value	observation_concept_id	value_as_concept_id	concept_name	vocabulary	domain	class
일주일 이상		45878994	More than half the days	LOINC	Meas Value	Answer
거의 매일		45882010	Nearly every day	LOINC	Meas Value	Answer
(지난 2 주 동안) 식욕 저하 혹은 과식	3044098		Poor appetite or overeating in last 2 weeks [Reported.PHQ]	LOINC	Observation	Survey
전혀 아니다		45883172	Not at all	LOINC	Meas Value	Answer
여러 날 동안		45879886	Several days	LOINC	Meas Value	Answer
일주일 이상		45878994	More than half the days	LOINC	Meas Value	Answer
거의 매일		45882010	Nearly every day	LOINC	Meas Value	Answer
(지난 2 주 동안) 내 자신이 나쁜 사람이라는 느낌 혹은 내 자신을 실패자라고 느끼거나 나 때문에 나 자신이나 내 가족이 불행하게 되었다는 느낌	3043801		Feeling bad about yourself - or that you are a failure or have let yourself or your family down in last 2 weeks [Reported.PHQ]	LOINC	Observation	Survey
전혀 아니다		45883172	Not at all	LOINC	Meas Value	Answer
여러 날 동안		45879886	Several days	LOINC	Meas Value	Answer
일주일 이상		45878994	More than half the days	LOINC	Meas Value	Answer
거의 매일		45882010	Nearly every day	LOINC	Meas Value	Answer
(지난 2 주 동안) 신문을 읽거나 TV 를 볼 때 집중하기 어려움	3045019		Trouble concentrating on things, such as reading the newspaper or watching television in last 2 weeks [Reported.PHQ]	LOINC	Observation	Survey
전혀 아니다		45883172	Not at all	LOINC	Meas Value	Answer
여러 날 동안		45879886	Several days	LOINC	Meas Value	Answer
일주일 이상		45878994	More than half the days	LOINC	Meas Value	Answer
거의 매일		45882010	Nearly every day	LOINC	Meas Value	Answer

Observation_source_value	observation_concept_id	value_as_concept_id	concept_name	vocabulary	domain	class
(지난 2 주 동안) 남들이 알아챌 정도로 거동이나 말이 느림. 또는 반대로 너무 초조하고 너무 안절부절하지 못해서 평소보다 많이 돌아다니고 서성거림	3043785		Moving or speaking so slowly that other people could have noticed. Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual in last 2 weeks [Reported.PHQ]	LOINC	Observation	Survey
전혀 아니다		45883172	Not at all	LOINC	Meas Value	Answer
여러 날 동안		45879886	Several days	LOINC	Meas Value	Answer
일주일 이상		45878994	More than half the days	LOINC	Meas Value	Answer
거의 매일		45882010	Nearly every day	LOINC	Meas Value	Answer
(지난 2 주 동안) 나는 차라리 죽는 것이 낫겠다는 등의 생각 혹은 어떤 면에서건 당신 스스로에게 상처를 주는 생각들	3043462		Thoughts that you would be better off dead, or of hurting yourself in some way in last 2 weeks [Reported.PHQ]	LOINC	Observation	Survey
전혀 아니다		45883172	Not at all	LOINC	Meas Value	Answer
여러 날 동안		45879886	Several days	LOINC	Meas Value	Answer
일주일 이상		45878994	More than half the days	LOINC	Meas Value	Answer
거의 매일		45882010	Nearly every day	LOINC	Meas Value	Answer

3. Data Quality Rules

1) Data Quality Rule 필요성

- 건강검진 설문지 데이터의 경우 patient filled survey라는 특성 상 설문문항의 단위(unit)에 맞지 않는 대답을 하거나 논리적으로 성립될 수 없는 응답을 기재하는 human error가 종종 발생한다.
- 이에, 논리적으로 성립될 수 없는 데이터를 처리하지 않으면 데이터 분석에 bias가 발생할 수 있기에 이상치는 삭제하고, 문항의 단위 문제(연도, 나이 등)로 충분히 추론이 가능한 응답의 경우 재가공을 통해 데이터로 활용할 수 있게끔 한다.

2) Data Quality Rule 개발

- 건강검진설문지 항목들은 절대적인 이상치 범위는 있으나, 상식적인 범위에서의 허용치 범위(하한~상한)에 대해서는 명확한 기준은 확인하기 어렵다.
- 이에 본 연구에서는 논리적으로 성립할 수 없는 절대적 이상치 범위에 한정하여 Data Quality Rule을 개발하여 적용하였다.

분류	대상항목	DQ Rule
흡연정보	금연기간(개월)	금연기간은 연령보다 크거나 같을 수 없다.
	흡연량(개피)	흡연량은 1보다 작을 수 없다.
	첫 흡연연령(만 00세)	흡연연령은 연령보다 클 수 없다.
	갑년	0보다 크거나 같다.
신체활동	격렬한 신체활동 일수(일/주)	7일(1주일)보다 클 수 없다.
	격렬한 신체활동 시간(분/일)	활동시간은 1440분(하루)보다 클 수 없다.
	중간정도 신체활동 일수(일/주)	7일(1주일)보다 클 수 없다.
	중간정도 신체활동 시간(분/일)	활동시간은 1440분(하루)보다 클 수 없다.
	10분 이상 걷은 날(일/주)	7일(1주일)보다 클 수 없다.
	도보 시간(분/일)	활동시간은 1440분(하루)보다 클 수 없다.
	앉아서 보낸 시간(분/일)	활동시간은 1440분(하루)보다 클 수 없다.
수면	수면시간(분/일)	활동시간은 1440분(하루)보다 클 수 없다.

사사

이 연구는 2019년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임
(No.20004927, CDM 기반 바이오헬스데이터 통합 플랫폼 고도화 및 검증 기술개발)