

국내 의료 재료 용어에 대한 OMOP 표준용어로의 매핑 가이드라인

아주대학교 의료정보학과

**버전 1.0
04, 2021**

TABLE OF CONTENTS

1. 서론	1
2. 의학 용어 체계	3
2-1. SNOMED-CT	4
2-2. EDI 코드	4
3. 용어 매핑	6
3-1. OMOP 표준 용어	8
3-2. 매핑 프로세스	11
3-3. Usagi	13
4. 시술 용어 표준화 가이드라인	15
4-1. 시술 용어 표준화 결과	15
4-2. 3자 검토 사례 및 결과	16
4-3. 표준화된 시술 용어 적용방안	20
5. REFERENCES	22

1. 서론

Observational Health Data Sciences and Informatics (OHDSI, "Odyssey"라고 발음함) 연구 네트워크는 각 의료기관들이 보유한 서로 다른 형태로 존재하는 보건의료 데이터를 표준화하는 작업을 진행하고 있다. CDM 기반으로 데이터가 표준화되어 있는 기관의 경우, 동일한 분석 코드를 데이터 보유 기관에서 개별 실행하여 출력된 결과만을 통합하는 분산형 공동연구가 가능하다. 현재 해외 20 개국 이상 6 억명 이상의 환자 데이터가 CDM 기반으로 표준화되어 있으며, 국내 CDM 구축을 담당하는 분산형 바이오헬스 빅데이터 사업단 (FEEDER-NET)은 서울아산병원, 서울대학교병원, 성모병원 등의 국내 핵심 상급종합병원을 포함한 60 여 개 병원 및 20 여 개의 기업체, 컨소시엄들과 협업하여 분산 연구망 (Distributed Research Network; DRN)의 규모를 지속적으로 확대시키고 있다.

각 기관에 위치하는 보건의료데이터의 구조와 용어를 표준화하기 위해 공통 데이터 모델 (Common Data Model, CDM)의 동일한 데이터 구조뿐 아니라 Observational Medical Outcome Partnership (OMOP) Vocabulary 라는 공통 의료용어체계를 활용한다. 의학용어는 의료 정보를 표현하는데 가장 기본이 되는 요소로써, 의료 행위를 식별하고, 분류하고, 소통하기 위한 목적으로 합의된 어휘 및 전문용어들을 지칭한다. 의학 세부 분야, 혹은 기관, 국가에 따라 다양한 의학용어집들이 존재하는데, 각 기관 내에만 존재하는 기관코드부터 국제적으로 통용되는 용어체계(SNOMED, ICD, CPT4 등)까지 다양하게 존재한다. 따라서 동일한 의료행위일지라도 용어체계에 따라 다양한 방식으로 표현되고 있기 때문에 다양한 용어체계를 OMOP Vocabulary 아래 통합하는 작업이 필요하다.

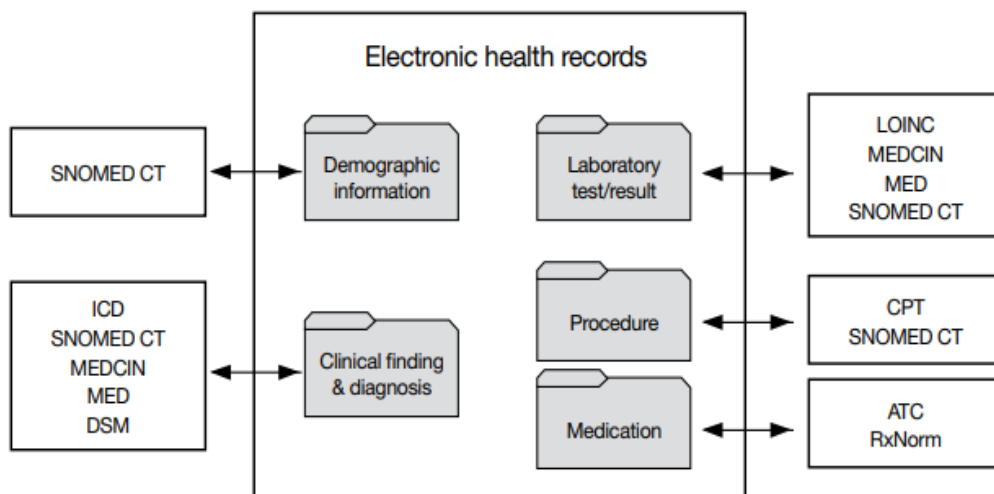


그림 1. 의료 데이터에서 널리 사용되는 용어체계

본 가이드라인에서는 의료 자료에 대해 한국의 사실상 표준(de facto standard)으로 사용되고 있는 Electronic Data Interchange(EDI)코드를 OHDSI 에서 정의하는 OMOP 표준 용어로의 매핑(Mapping) 방안을 제시함으로써, 국내 의료 자료 의학용어체계의 매핑 방법, 어려움, 해결 방안 등을 논의하고자 한다. 더 나아가, OMOP 표준 용어집으로의 매핑을 통해 국내 의료 자료 용어의 국제적 활용 가능성을 높이하고자 한다.

2. 의학 용어 체계

OHDSI 커뮤니티가 보건의료 데이터의 용어를 표준화하기 위하여 사용하는 용어집인 OMOP Vocabulary 는 서로 다른 의학 용어체계들의 집합이다. 따라서 동일한 의학적 개념들에게 하나의 표준화된 코드를 부여하기 위하여 용어체계들 간의 위계를 정하고 있다. 각 용어체계들 간의 상하위 관계뿐만 아니라 수평 관계도 정하며, 표준으로 사용되는 용어체계와 비표준으로 사용되는 용어체계도 정의해 두었다. OHDSI 커뮤니티의 Vocabulary 팀에서 용어체계들 간의 관계를 규명하는 일을 주도하고 있다. 개별 연구자도 매핑 작업을 진행한 후, OHDSI 측에 공식적으로 편입을 검토 및 요청할 수도 있다. OMOP vocabulary 는 ATHENA (<https://athena.ohdsi.org>)에서 조회 가능하다.

전 세계적으로 사용되고 있는 대표적인 의학용어 체계는 Current procedural terminology 4th edition(CPT4), dictionary for medicines and devices(dm+d), Healthcare common procedure coding system(HCPCS), International classification of disease, 10th version(ICD10), ICD10 procedure classification system(ICD10PCS), ICD10 clinical modification(ICD10CM), Japan medical data center(JMDC), Korean standard classification of disease(KCD7), EDI, National drug code(NDC), RadLex, Structured product labeling(SPL), Systematized nomenclature of medicine clinical terms(SNOMED CT) 등이 있다. 이들은 특정한 목적을 달성하기 위해 정부의 주도로 만들어지거나, 전자의무기록이나 의료보험 청구 등의 목적을 달성하기 위하여 민간 주도로 만들어지기도 한다. 현재 널리 사용되어 높은 영향력을 보유하고 있는 용어체계들이 OMOP 용어 사전집에 등재되어 있으며, 각 용어체계는 적게는 2 만개부터 많게는 100 만개 이상의 의학 용어들을 포함하고 있다. 특히, 의료 재료 부문의 의학 용어를 담은 대표적 용어 체계로는 SNOMED CT 가 있다.

표 1. OMOP 용어집(vocabulary) 내 포함 용어체계 및 용어의 수

	OMOP 내의 표준 용어 체계	용어의 수
1	RxNorm Extension	2,042,129
2	NDC	1,056,836
3	SNOMED	1,035,027
4	SPL	518,567
5	dm+d	383,469
	EDI	313,431
6	RxNorm	294,029
7	LOINC	250,287

8	OSM	203,339
9	ICD10PCS	194,981
10	DPD	193,647
11	Read	108,696
12	ICD10CM	96,924
13	AMT	76,706
14	ICDO3	64,304
15	VA Product	54,768
16	CIEL	50,881
17	BDPM	44,376
18	JMDC	38,798
19	NDFRT	37,486
20	NAACCR	34,473

표 2. OMOP CDM 기록에 사용되는 용어

구분	원천 자료	OMOP CDM 표준
Conditions	KCD	SNOMED CT
Drugs	EDI 약품코드	RxNorm
Procedures	EDI 수가코드	CPT4, SNOMED CT, ICD9PCS, HCPCS
Measurements	검사 코드, 검체 코드	LOINC, SNOMED CT
Devices	보험 EDI 치료 재료 코드	SNOMED CT

2-1. SNOMED-CT

영국 런던에 위치한 국제 비영리 표준개발기구인 SNOMED International 에서 관리하는 SNOMED Clinical Terms (SNOMED-CT) 는 각 의학 용어의 정의에 따라 계층, 속성, 타 용어와의 관계를 확립한 의학용어체계이다. 임상소견(clinical findings)과 증상(symptoms), 진단(diagnoses), 처치(procedures), 신체구조(body structures), 유기체(organisms)와 기타 병인(other etiologies), 물질(substances), 약재(pharmaceuticals), **기기(devices)**, 검체(specimens) 등 방대한 범위의 의학용어에 대한 구조를 체계화하고 있으며, 현재 약 30 만 개의 용어를 포함하는 가장 큰 생체 의학용어체계이다.

SNOMED-CT 의 용어들은 다른 용어집에 포함된 용어들로부터 여러 개의 유의어와 동의어 관계를 가질 수 있으며 다른 용어들과 속성 관계를 가질 수도 있다.

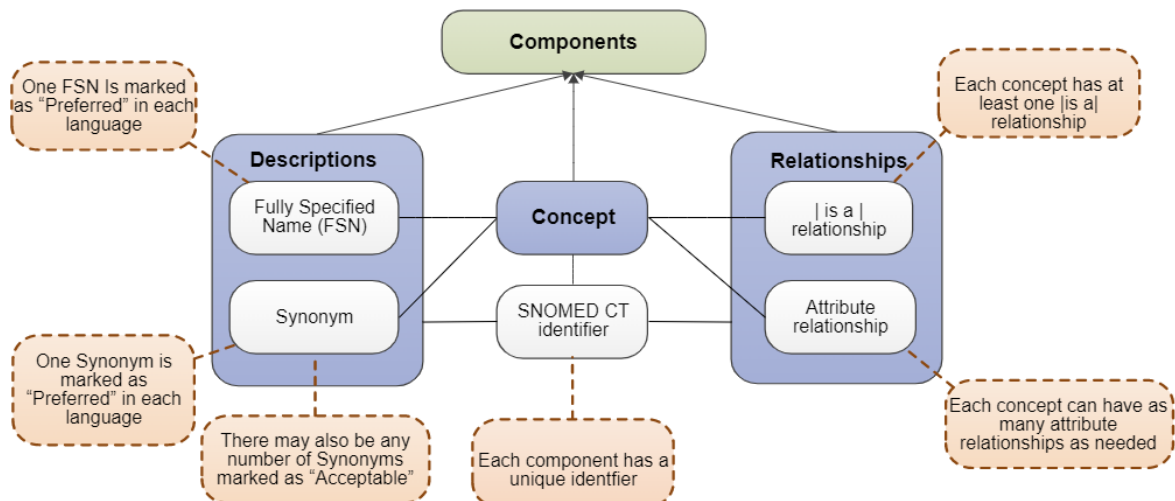


그림 2. SNOMED-CT 가 용어의 정의에 따라 계층, 속성, 타 용어와의 관계를 정의하는 모습

2-2. EDI 코드

EDI 코드란, 진료비 청구, 심사, 지급을 위해 국내에서 광범위하게 사용되는 의사소통 수단을 의미한다. 'SNOMED CT'는 체계적이고 구조화된 전산처리 가능한 국제 의료용어체계를 뜻한다.

EDI(Electronic Data Interchange)란 표준화된 상거래서식 또는 공공서식을 서로 합의된 표준에 따라 전자문서를 만들어 컴퓨터 및 통신을 매개로 상호 교환하는 것을 의미한다. 즉, EDI란 거래당사자가 인편이나 우편에 의존하는 종이서류 대신 컴퓨터로 행정서류 등 서로 합의한 표준화된 양식에 맞추어 상호 교환하여 재입력 과정 없이 직접 업무에 활용할 수 있도록 하는 새로운 정보전달 방식을 말한다.

EDI의 가장 기본 적인 구조는 아래 그림과 같다.



그림 3. EDI의 구조

건강보험심사평가원은 의약품 코드 일원화 방침에 따라 '약제 급여목록 및 급여상한금액표'가 전면 개정 고시됨에 따라 급여 EDI 코드를 표준코드(KD 코드)로 전면 변경했다. 표준코드는 2010년 1월 1일 진료분부터 적용되었으며, 자릿수는 9자리 기존 EDI 코드와 동일하나, 업체식별코드 4자리와 품목코드 5자리의 숫자로만 구성됐다. 즉, EDI 코드는 "제약사 코드(4) + 품목코드(4) + 포장단위(1)"를 뜻한다.

○ 변경 후(2010. 1. 1 진료분부터 적용)

- 의약품표준코드(13자리) 중 업체 식별코드 및 품목코드 9자리를 사용

자릿수(13)	3	4	5					1		
내 용	국가식별 코드	업체 식별 코드	품목코드					검증번호		
			한량포함한 품목코드		포장단위					
부여 예	880	6400-6999	0001-9999		0: 대표코드 1-9: 포장단위			0-9		
		①	②	③	④	⑤	⑥	⑦	⑧	⑨

그림 4. 표준코드로 변경된 급여 EDI 코드의 구조

국내에서는 의료용어 표준화를 위해 수많은 노력을 했지만, 실제로 의료 현장에 보급되어 사용되고 있는 코드체계는 드문 상황이다. 또한, 각 의료기관별로 OCS, EMR, PACS 등과 같이 의료정보를 수집하고 저장하는 환경이 상이하기 때문에 의료정보를 기록하는 그 식별자 대부분도 상이하다. 이러한 의료정보 환경을 고려했을 때, EDI 라는 사실상 표준으로 사용되는 코드 체계를 이용하는 것이 가장 효율적인 방법 중 하나라고 할 수 있다.

표 3. EDI 코드 상의 의료정보 구분 및 분류

EDI 구분	분류유형	내용
의·치과	1	수가(공상 포함)
	2	준용수가
	3	보험등재 의약품
	4	원료, 조제(제제) 의약품
	7	협약재료
	8	치료재료
한방	A	수가
	B	준용수가
	C	약가
	G	협약재료
	H	치료재료
	P	공상

EDI 는 국내 단일보험자체제의 특성상 전국민에 대한 건강보험 청구 및 상환을 위한 목적으로 개발되었으나, 결과적으로 한국 의료용어의 실질적인 표준으로 자리잡게 되었다. 또한, OMOP 용어집에 EDI 코드가 등재 되어있기 때문에, EDI-시술 용어와의 매핑 관계를 공식적으로 등재하여 관리할 수 있다. 사실상 표준으로 사용되는 EDI 코드와 OMOP 용어간의 매핑 관계를 정의하게 된다면, 국내 대부분의 의료기관 적용이 가능하기 때문에 재사용성의 효율이 높다.

EDI 코드(총 8 자리)는 **기본코드(5 자리)**와 **산정코드(3 자리)**로 구성되어 있는데, 이 중 기본 코드와 OMOP 표준 용어로의 매핑을 우선시한다.

예시) J2301062 Sclerosing needle, Sclerotherapy needle

J2301200 Sclerosing needle, Injection needle

J2301959 Sclerosing needle, BK-Injector

ID ▼	CODE ▼	NAME ▼	CLASS ▼	CONCEPT ▼	VALIDITY ▼	DOMAIN ▼	VOCAB ▼
42263933	HE538606	Abdominal MRA,Under 72 months radiologist reading	Procedure	Non-standard	Valid	Procedure	EDI
42365283	R4534	Abdominal Pregnancy	Proc Hierarchy	Non-standard	Valid	Procedure	EDI
42177139	R4534080	Abdominal Pregnancy,clinic, dental clinic(including Public Health and Medical care center) Nighttime, Saturday, holiday	Procedure	Non-standard	Valid	Procedure	EDI
42178947	R4534081	Abdominal Pregnancy,clinic, dental clinic(including Public Health and Medical care center) Nighttime, Saturday, holiday second surgery	Procedure	Non-standard	Valid	Procedure	EDI
42319213	R4534020	Abdominal Pregnancy,emergency	Procedure	Non-standard	Valid	Procedure	EDI
42303006	R4534040	Abdominal Pregnancy,Emergency holiday	Procedure	Non-standard	Valid	Procedure	EDI

그림 5. ATHENA 에서 확인한 EDI 코드의 OMOP vocabulary 등재 모습

3. MAPPING

3-1. OMOP 표준 용어

흔히 “용어 vocabulary” 라고 불리는 OMOP 표준 용어(Standardized Vocabularies)는 OHDSI 연구 네트워크의 기초적인 부분이자, 공통 데이터 모델(CDM)의 핵심 부분이다. OMOP 표준 용어는 데이터의 내용을 정의함으로써 분석 방법, 정의, 결과를 표준화하여 진정한 원격 네트워크 연구와 분석을 가능하게 한다.

의학 용어는 새로운 기술과 약물, 재료 등이 지속적으로 개발되고 사라지고 있음에 따라 규모의 범위와 복잡성이 확대되고 있다. 또한, 각 기관의 사용 목적에 따라 의학 용어의 분류나 체계 등이 상이한데, 이러한 용어집들의 상당수는 공공기관 혹은 정부 기관에서 장기적으로 관리하고 있다. OHDSI 는 관찰 연구를 위한 국제 표준인 OMOP CDM 을 개발했다. CDM 의 일부로, OMOP 표준 용어는 다음 두 가지 목적으로 사용할 수 있다.

- (1) 커뮤니티에서 사용되는 모든 용어의 공통 저장 자료
- (2) 연구에 사용하기 위한 표준화된 매핑

표준화된 용어는 커뮤니티에 무료로 제공되며, OMOP CDM 실제 사용 시마다 필수 참조 테이블로 사용되어야 한다. 이미 많은 용어집들이 관찰 데이터에서 사용되어 오고 있으며, 용어집의 복잡한 구성과 관리, 많은 이해관계자의 의견을 수렴해야 한다는 점 때문에 OHDSI 는 모든 용어를 처음부터 구축하는 것보다 기존에 널리 사용되던 용어집을 채택하는 것을 선호한다. 왜냐하면, 많은 용어집이 이미 공동체의 관찰 데이터에 사용되어 왔으며, 용어집의 구성 및 유지 관리가 매우 복잡하여 오랜 기간 동안 많은 이해관계자의 의견을 수렴해야 했기 때문이다.

이러한 이유로, OHDSI 는 몇 가지 특별한 내부 용어만 생성할 뿐, 이외의 용어들은 모두 기존에 존재하는 용어집으로부터 차용하여 쓰고 있다. 아래는 OMOP 표준 용어를 구성하는 용어집을 도식화한 그림이다.

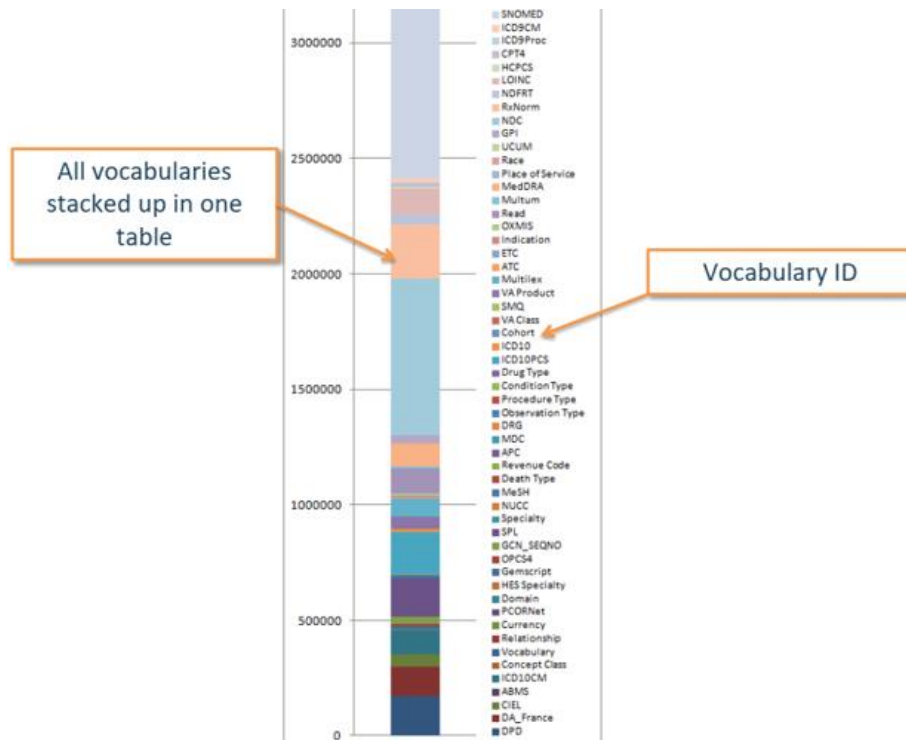


그림 6. OMOP 표준용어를 구성하는 용어집

OMOP 공통데이터모델 (Common Data Model, 이하 CDM)의 모든 임상 기록은 개념 concept 으로 표현되고 있으며, 이는 각 기록의 의미를 나타낸다.

CDM 에서 concept 은 데이터 기록의 기본적인 구성 요소로써, 모든 테이블을 거의 예외 없이 완전 정규화한다. CONCEPT 테이블은 CDM 내의 사전과 같으며, CDM 내부의 모든 concept 은 이 사전 안에 정의되어 있다. 달리 말해 CONCEPT 테이블에 없다면, 해당 용어는 CDM 내의 concept ID 를 갖지 못하고, 데이터를 정규화하여 기록하기 힘들어지기 때문에 시스템은 포괄적이어야 한다.

즉, 환자의 의료 경험(예를 들어, 진단명, 시술, 약물 노출 등) 및 의료시스템의 일부 관리 정보(예를 들어, 병원 방문, 관리 부위 등)과 관련된 모든 이벤트를 포괄할 만큼 충분히 많은 concept 이 있어야 한다.

CONCEPT_ID	313217	← Primary key
CONCEPT_NAME	Atrial fibrillation	← English description
DOMAIN_ID	Condition	← Domain
VOCABULARY_ID	SNOMED	← Vocabulary
CONCEPT_CLASS_ID	Clinical Finding	← Class in vocabulary
STANDARD_CONCEPT	S	← Standard, Source of Classification
CONCEPT_CODE	49436004	← Code in vocabulary
VALID_START_DATE	01-Jan-1970	← Valid during time interval
VALID_END_DATE	31-Dec-2099	
INVALID_REASON		

그림 7. OMOP CDM 에서 vocabulary 의 표준 표현.
(심방세동의 SNOMED 코드에 대한 CONCEPT 테이블 레코드)

각각의 concept 은 concept ID 를 기본 키 primary key 로 가진다. 무의미한 정수로 이루어진 ID 는, 해당 원본 코드를 대신하여 CDM 테이블에 데이터를 기록하는 데 사용된다. 각 concept 에는 하나의 이름이 있다. 그 이름은 항상 영어로 되어있고, 원천 source 으로부터 가져온다. 원천 용어에 둘 이상의 이름을 가지고 있으면, 가장 표현력이 높은 이름이 선택되고, 나머지 이름은 동일한 CONCEPT_KEY 로 CONCEPT_SYNONYM 테이블에 저장된다. 영어 이외의 이름은 CONCEPT_SYNONYM 에 기록되며, 적합한 language concept ID 가 LANGUAGE_CONCEPT_ID 필드에 기록된다. 이름은 255 자까지의 길이를 가지는데, 너무 긴 이름은 잘라내고 최대 1,000 자까지 저장 가능한 다른 이름의 동의어로 기록된다.

또한, 각 Concept 에는 DOMAIN_ID 필드에 도메인이 할당되는데, 정수로 구성된 Concept_ID 와는 달리, 대소문자를 구분하면서 길이가 짧은 고유한 문자 ID 이다. 이러한 각 도메인의 예로는 "Condition", "Drug", "Procedure", "Visit", "Device", "Specimen" 등이 있다. 모든 Concept 들은 각 주어진 도메인에서만 기록되어야 한다.

각 임상 사건의 의미를 나타내는 하나의 개념을 표준 개념 Standard concept 이라고 부른다. 심방 세동의 예를 들면, MeSH 코드 "D001281", CIEL 코드 "148203", SNOMED 코드 "49436004", KCD7 코드 "I48", ICD9CM 코드 "427.31" 및 Read 코드 "G573000"로 모두 condition 데이터에서는 SNOMED concept 만이 표준이고 질환을 의미할 수 있다. 나머지는 non-standard concept 또는 원천 개념인 source concept 으로, standard concept 에 매핑 되어야 한다. Standard concept 은 STANDARD_CONCEPT 필드에 "S"라고 표시한다. 그리고 이러한 standard concept 만이 "_CONCEPT_ID"로 끝나는

CDM 필드에 데이터를 기록하는데 사용된다. CONCEPT 테이블의 INVALID_REASON 은 NULL 이어야 한다.

분류 개념은 표준이 아니므로 데이터를 나타내는데 사용할 수 없다. 하지만 standard concept 과 어우러져 계층 구조를 나타냄으로써, 계층 쿼리 hierarchical queries 를 수행하는데 사용할 수 있다. 예를 들어, MedDRA 코드 10037908 의 모든 하위 항목에 대한 쿼리를 사용하면 (MedDRA license 를 받지 않는 사용자에게는 보이지 않음.)

심방세동 atrial Fibrillation 에 대한 표준 SNOMED concept 이 검색된다. 아래의 그림은 표준, 비표준 원천 및 분류개념과 condition 도메인에서의 계층적 관계를 나타낸다. SNOMED 는 대부분 standard condition concept (일부 ICDO3 에서 파생된 종양학 관련 concept)에 사용되고, MedDRA concept 은 계층 분류 concept 에 사용되며, 그 외 다른 모든 용어집은 non-standard concept 이나 혹은 계층 구조에 포함되지 않는 source concept 을 가지고 있다.

이렇게 표준으로 연결된 용어들 간의 관계를 매핑(Mapping) 관계라고 명시한다. 또한, 각 Concept 들은 서로 상·하위의 계층 구조로 구성되어 있고, 이러한 관계들은 CONCEPT_ANCESTOR 테이블에 저장되고 있다. 예를 들어, 심방 세동의 경우 가장 상위 Concept 으로 순환기 체계의 질병이라는 Concept 이 존재하고, 아래에는 만성 심방 세동 혹은 발작성 심방 세동 등의 Concept 이 존재할 수 있다. 하지만, 이러한 관계는 Drug 과 Condition 도메인에선 높은 품질을 보장하고 있지만, Procedure, Measurement, Observation 에서는 부분적으로만 구축되어 있다.

아래의 그림의 경우, "심방세동 atrial fibrillation" condition 계층으로 First degree ancestry 는 "Is a"와 "Subsumes" 관계를 통해 정의되며, 모든 higher degree 관계는 추론되고, CONCEPT_ANCESTOR 표에 저장된다. 각 개념은 분리 수준이 0 인 자기 자신의 하위 개념이라고 할 수 있다.

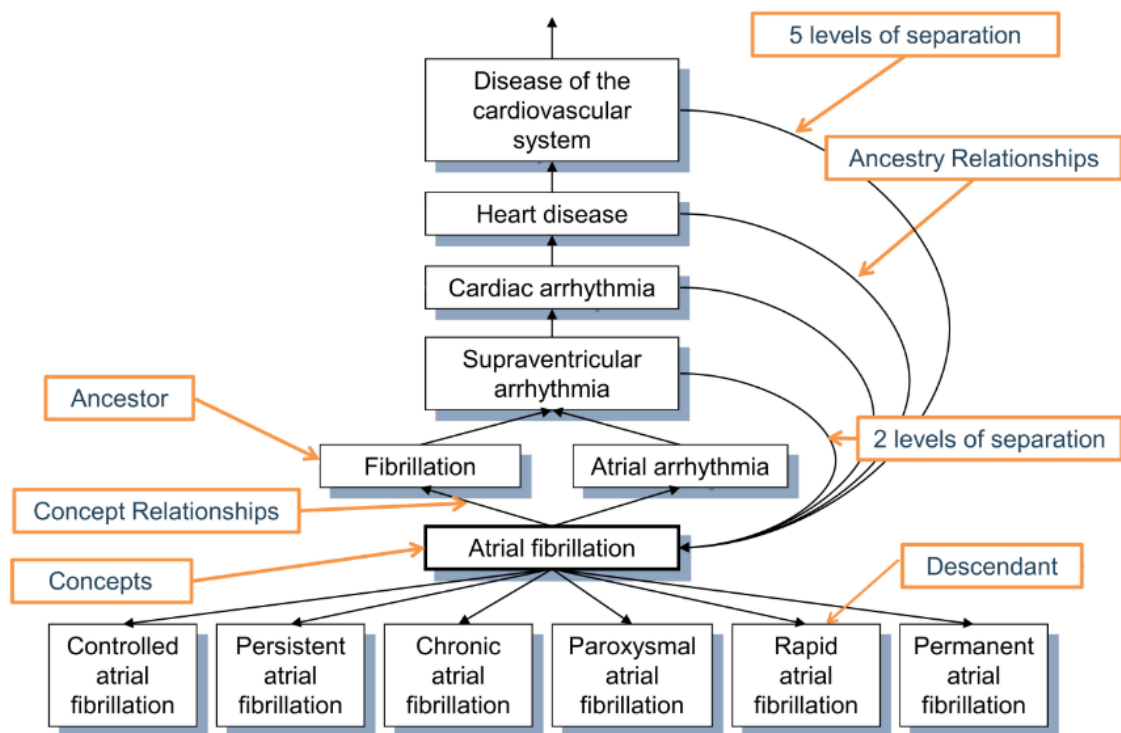


그림 8. OMOP 용어집 내 Concept 간 계층 구조

3-2. 매핑 프로세스(Mapping process)

용어 매핑을 통해 실제로 의미가 동일한 두 concept 은 도메인 혹은 원천 어휘집과 상관없이 특정 관계를 명시할 수 있다. 관계의 특성은 CONCEPT RELATIONSHIP 테이블의 RELATIONSHIP_ID 필드에서 대소문자를 구분하는 고유한 짧은 ID 로 표시를 하는데, 각 관계에 대해서 상호가 바뀐 대칭 관계가 존재하여, CONCEPT_ID_1 과 CONCEPT_ID_2 필드의 내용이 교환된다.

예를 들면, CONCEPT_ID_1-CONCEPT_ID_2 의 "Maps to" 관계는 CONCEPT_ID_2-CONCEPT_ID_1 의 "Mapped from"과 반대의 관계를 갖는다. 또한, 관계의 의미 중 "동등 개념 Equivalent concept"와 같이 동일한 개념을 나타낼 때도 있지만, 더 세밀한 병명을 나타내어 계층적으로 포함된 하위 개념을 의미하는 "하위 개념 Descendant concept"을 가질 때도 있다. 일부 매핑은 둘 이상의 표준 concept 에 연결된다. 예를 들어, ICD9CM 의 070.43 "Hepatitis E with hepatic coma"는 SNOMED 의 235867002 "Acute hepatitis E"뿐 아니라, SNOMED 의 72836002 "Hepatic coma"에도 매핑이 되어있다. 이는 원천 용어집에서 명시된 의미가 조합되어 (hepatitis + coma) 선 반영된 경우가 있기 때문이다. 이 외 매핑에 관련된 자세한 규칙은 Book of OHDSI 혹은 OHDSI wiki 에서 확인할 수 있다.

매핑 관계는 아래의 표와 같이 두 개의 relationship ID 쌍을 이용하여 Non-standard concept 를 standard concept 으로 변환하게 해준다.

표 4. 매핑 관계의 유형

Relationship ID 쌍	목적
“Maps to” and “Mapped from”	표준 개념에 매핑. standard concept은 자기 자신에게 매핑되고, non-standard concept은 standard concept으로 매핑된다. 대부분의 non-standard concept과 모든 standard concept은 한 standard concept과 이러한 관계를 맺는다. non-standard concept은 *_SOURCE_CONCEPT_ID에 저장되고 standard concept은*_CONCEPT_ID 필드에 저장된다. 분류 개념은 매핑되지 않는다.
“Maps to value” and “Value mapped from”	MEASUREMENT와 OBSERVATION 테이블의 VALUE_AS_CONCEPT_ID 필드에 배치할 값을 나타내는 개념에 매핑.

이러한 매핑 관계는 기존의 OHDSI 커뮤니티 혹은 기존에 존재하던 지식을 기반으로 명시하여 공개하고 있으나, 한국에서 고유하게 사용하고 있는 KCD 나 EDI 코드의 경우는 아직 매핑이 되어있지 않다. 물론 KCD 의 경우, 대부분 ICD 와 동일한 분류 체계 및 code 를 사용하고 있기 때문에, 일부 한국에서만 사용되고 있는 특이한 용어를 제외하면 KCD-ICD-SNOMED 순서를 통해 별도의 매핑 작업이 필요 없이 관계를 지을 수 있지만, EDI 의 경우 새롭게 그 관계를 명시해야 한다.

이처럼 기존 concept relationship 테이블에 의해 정의되지 않은 매핑 관계는, source to concept map 이라고 하는 테이블에서 그 관계를 표현할 수 있다. 예를 들어, 한국의 EDI 체계에서 “HE109”는 “기본자기공명영상진단-척추-경추-일반”을 의미하는데, 이 경우 SNOMED 의 “241646009”인 “MRI of cervical spine”와 매핑 될 수 있고, 각 각 SOURCE_CONCEPT_ID 필드와 TARGET_CONCEPT_ID 필드에 명시될 수 있다.

1단계. Source concept 살펴보기 (예시. Atrial fibrillation)

```
SELECT * FROM concept WHERE concept_code = '427.31';
```

CONCEPT_ID	CONCEPT_NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT	CONCEPT_CODE
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31

2단계. Source concept에 해당하는 표준 용어 검색하기

```
SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957  
AND relationship_id = 'Maps to';
```

CONCEPT_ID_1	CONCEPT_ID_2	RELATIONSHIP_ID	VALID_START_DATE	VALID_END_DATE	INVALID_REASON
44821957	313217	Maps to	01-Jan-1970	31-Dec-2099	

3단계. 표준 용어 살펴보기

```
SELECT * FROM concept WHERE concept_id = 313217;
```

그림 9. 매핑 관계 확인 방법

이러한 매핑 작업은 OMOP CDM의 품질 뿐 아니라 다기관 공동 연구의 질에 많은 영향을 줄 수 있기 때문에 정확한 코드로 매핑이 필요하다. 따라서, 일반적으로 매핑 작업은 의무기록사, 간호사, 의사 등과 같은 의학적 지식을 보유한 전문가에 의해 이루어져야 한다.

또한 용어 매핑의 품질을 높이기 위해서 2인 1조로 동일한 코드에 대해 독립적으로 매핑 작업을 진행하는 것이 권유된다. 이후 매핑이 완료된 결과에 대한 교차 검토를 통해, 불일치한 결과는 논의를 통해 합의하여 결정하거나, 합의가 이루어지지 않을 경우 3자의 검토를 통해 최종 매핑 코드를 승인하는 방식을 채택한다.

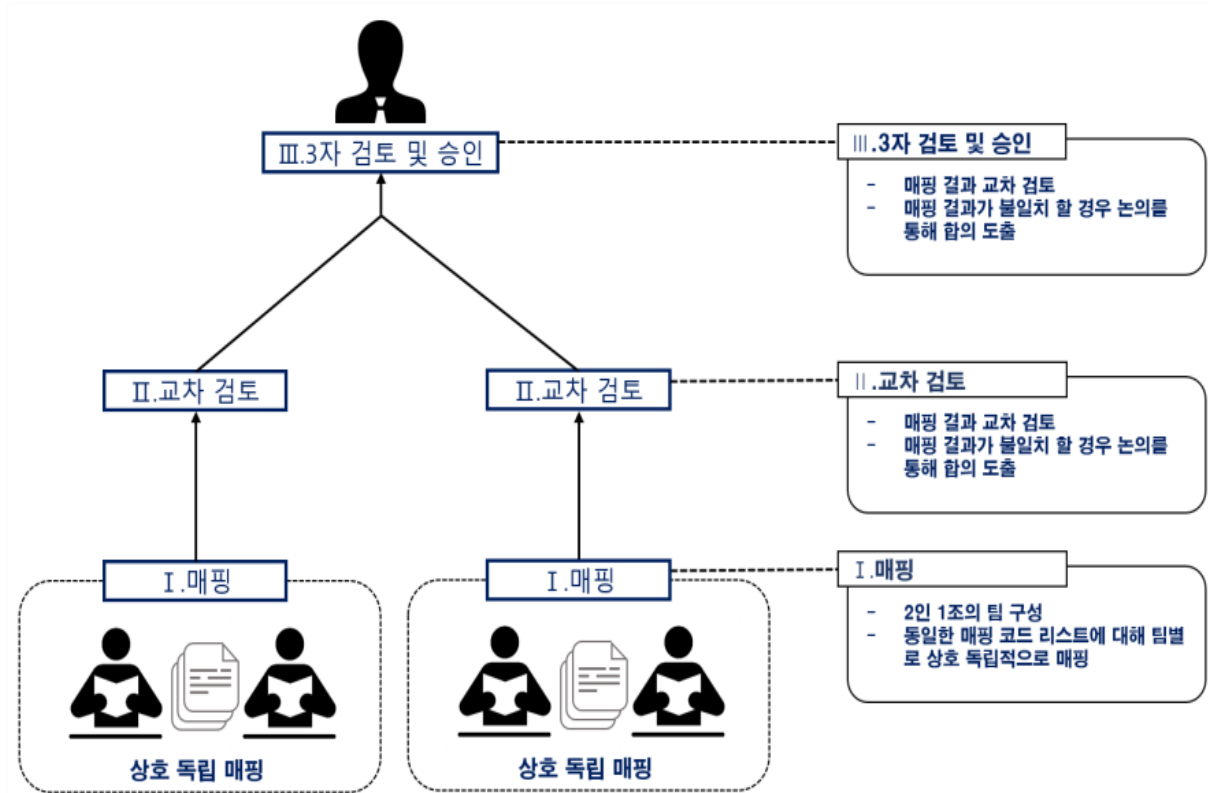


그림 10. 매핑 프로세스 모식도

3-3. Usagi

위에서 언급한 것과 같은 매핑 작업은 엑셀이나 R 과 같은 도구를 사용하여 수기로 매핑하는 방법도 있지만, OHDSI 컨소시움에서 개발한 Usagi 라는 도구를 사용할 수 있다. Usagi 는 용어 매핑 절차를 도와주는 도구로, 원천 코드 설명에 입력된 단어의 유사도에 기반하여 적절한 표준 용어를 추천해주는 도구이다. 만약 원천 코드의 설명이 영문이 아닐 경우엔, Google Translate 를 통해 해당하는 용어의 영어 번역을 확인할 수 있다. 또한, Usagi 의 용어 추천이 정확하지 않을 경우엔 사용자가 직접 적절한 표준 용어를 찾을 수 있다. Usagi 는 해당 GitHub repository (<https://github.com/OHDSI/Usagi>)를 통하여 사용할 수 있다. Usagi 소프트웨어를 사용하기 위한 순서는 다음과 같다:

1. 원천 시스템에서 사용하는 용어를 업로드
2. Usagi 의 단어 유사도 계산법을 이용하여 추천된 표준 용어로의 매핑
3. 매핑 결과를 용어의 SOURCE_TO_CONECEPT_MAP 으로 내보냄

Usagi - testUpCodesMapping.csv

File Edit View Help

Status	Source code	Source term	Frequency	CodeText	Match score	Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard con...	Parents	Children	Comment
Approved	K87.00	Hypertension...	694195	Hypertensie...	0.81	316866	Hypertensive...	Condition	Clinical Findi...	SNOMED	38341003	S	1	27	
Approved	L99.00	Other diseas...	680422	Andere ziekte...	0.47	0	Unmapped						0	0	Too generic
Approved	D01.00	Abdominal p...	678598	Gegeneralis...	0.61	197998	Generalized...	Condition	Clinical Findi...	SNOMED	102614006	S	1	0	
Unchecked	S99.00	Skin disease...	675817	Andere ziekte...	0.75	4317258	Disorder of s...	Condition	Clinical Findi...	SNOMED	95320005	S	2	193	
Unchecked	T88.00	Hunothuroidi...	667283	Hunothuroid...	1.00	4113642	Hunothuroidi...	Condition	Clinical Findi...	SNOMED	286910004	S	1	0	

Source code

Source code	Source term	Frequency	CodeText
S99.00	Skin disease, other	675817	Andere ziekte(n) huid/subcutis

Target concepts

Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard concept	Parents	Children
4317258	Disorder of skin	Condition	Clinical Finding	SNOMED	95320005	S	2	193

Remove concept

Search

Query

☒ Use source term as query

☐ Query:

Filters

☐ Filter by user selected concepts

☐ Filter by concept class: 2-dig nonbill code

☒ Filter standard concepts

☐ Filter by vocabulary: ABMS

☒ Include source terms

☒ Filter by domain: Condition

Results

Score	Term	Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard concept	Parents	Children
0.75	Skin disease	4317258	Disorder of skin	Condition	Clinical Finding	SNOMED	95320005	S	2	193
0.65	Skin Disease, Fungal	137213	Dermal mycosis	Condition	Clinical Finding	SNOMED	14560005	S	3	12
0.57	AIDS with skin dise...	4224566	Skin disorder associated with AIDS	Condition	Clinical Finding	SNOMED	421394009	S	2	2
0.56	Chronic skin disease	4134132	Chronic disease of skin	Condition	Clinical Finding	SNOMED	128236002	S	2	26
0.55	Disease, Otolitic	378161	Disorder of ear	Condition	Clinical Finding	SNOMED	25906001	S	4	43
0.55	Disease, Hers	4163346	Glycogen storage disease, type VI	Condition	Clinical Finding	SNOMED	29291001	S	2	0
0.55	Other peripheral va...	321052	Peripheral vascular disease	Condition	Clinical Finding	SNOMED	400047006	S	1	44
0.55	Other peripheral va...	4119612	Lower limb ischemia	Condition	Clinical Finding	SNOMED	233961000	S	2	3
0.55	Disease, Ormond	4176725	Retroperitoneal fibrosis	Condition	Clinical Finding	SNOMED	49120005	S	1	3
0.54	Pathological fractur...	73571	Pathological fracture	Condition	Clinical Finding	SNOMED	268029008	S	1	21
0.52	Disease, Tooth	4122115	Tooth disorder	Condition	Clinical Finding	SNOMED	234947003	S	3	58
0.52	Disease, Lip	135858	Disorder of lip	Condition	Clinical Finding	SNOMED	90678009	S	3	35
0.51	Disease, Ocular	4113600	Multiple congenital exostosis	Condition	Clinical Finding	SNOMED	264044004	S	6	0

Replace concept Add concept

Comment:

Approved / total: 5 / 1037 8.8% of total frequency

Approve

그림 11. Usagi 소프트웨어 작동 화면

4. 의료 재료 용어 표준화 가이드라인

4-1. 의료 재료 용어 표준화 가이드라인

국내 의료용어 표준화를 위한 수많은 노력이 있었지만, 실제 의료 현장에서 활용되는 표준화된 코드체계는 드물다. 각 의료기관별로 OCS, EMR, PACS 등과 같이 의료정보를 수집하고 저장하는 환경이 상이하여 의료정보를 기록하는 식별자 또한 다르게 쓰이고 있는 실정이다. 대한민국의 의료정보 환경을 고려하였을 때 요양급여비용 심사와 요양급여의 적정성 평가 업무를 수행하는 보건복지부 산하 준정부기관인 건강보험심사평가원이 관리하는 건강보험요양급여비용코드 (EDI 코드)가 국내 대표적인 표준화 코드체계로 활용되고 있다.

EDI 는 국내의 단일보험자체제의 특성상 전국민에 대한 건강보험 청구 및 상환을 위한 목적으로 개발되었으며, 국내 의료용어의 실질적인 표준으로 자리잡았다. 또한, OMOP-CDM 이 활용하는 표준화된 용어집인 OMOP vocabulary 에 EDI 코드가 등재되어있기 때문에, EDI 코드와 OMOP vocabulary 의 시술 용어와의 매핑 관계를 정의하여 관리할 수 있다. 사실상 국내 표준으로 사용되는 EDI 코드를 OMOP vocabulary 와 매핑시키는 작업이 OHDSI 커뮤니티 내에서 공식적으로 이루어진다면, 해당 작업은 국내 대부분의 의료기관으로의 적용이 가능할 것이다.

Device 도메인에서는 SNOMED, dm+d, Nebraska Lexicon 등의 용어집을 모두 표준으로 활용하고 있지만, SNOMED 가 보다 체계적인 구조로 광범위한 임상 용어를 분류한다. 따라서, 시술에 대한 EDI 코드를 OMOP 표준 용어로의 매핑 시 1 순위로 SNOMED 를 활용하고 있다. 또한, 각 나라별로의 의료 보험 체계 혹은 정책이 상이하기 때문에 정확히 일치하는 의료기기가 없을 경우나 다른 정보에서 의료기기를 식별할 수 있는 경우, 좀 더 상위 개념에 해당하는 용어로 매핑한다.

한국 OHDSI 의 공식 GitHub 인 OHDSI-Korea(<https://github.com/OHDSI-Korea>) 에서 위와 같은 가이드라인 하에 EDI-OMOP 표준 용어 간의 매핑 결과를 제시하고 있고, issue 란을 통하여 매핑 관계가 잘못되거나, 이 외 문제점들을 보고할 수 있다.

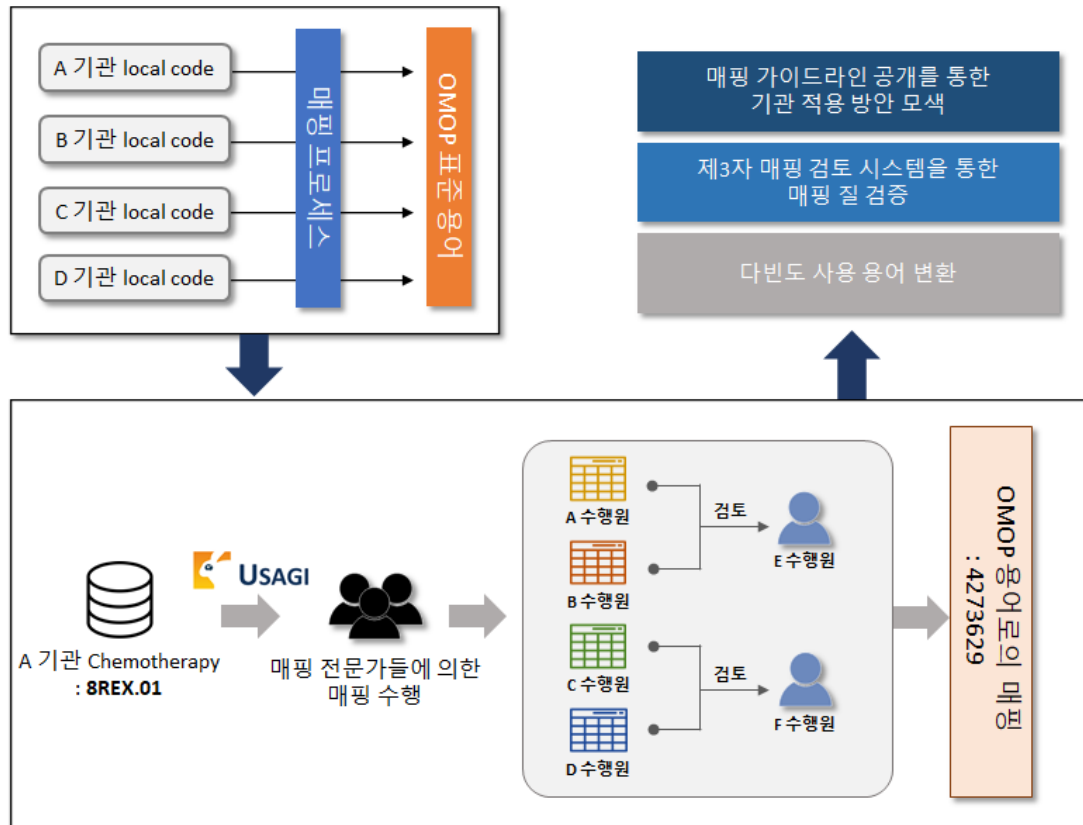


그림 12. 각 기관에서 고유하게 사용하는 코드를 표준용어로 매핑하기 위한 프로세스

4-2. 3 자 검토 사례

2 인 1 조로 이루어진 매핑팀에서 실시한 상호독립적인 매핑 결과가 불일치하고 합의가 도출되지 않거나, OMOP 표준 용어로의 매핑되어야 할 적절한 코드를 찾지 못한 경우, 의학적 지식을 가졌으면서 OMOP 표준 용어에 대한 전문가와 3 자 검토를 실시하여 적절한 코드를 도출해낼 수 있다. 다음 일부 사례는, 매핑 연구원끼리 합의가 되지 않거나, 혼동되는 일부 사례이다.

Case 1. B1011002: PROXIMATE

구분	내용
매핑 용어	연구원 A: 45765043 (Open-surgery manual linear cutting stapler, single-use) VS 연구원 B: 45764255 (Surgical staple, non-bioabsorbable)
3 자 검토 결과 및 사유	메스를 써서 피부나 근육 조직 등을 절개하여 근육, 뼈, 신경을 건드리는 관혈적 수술에 활용되는 의료 기기로서 보다 Open Surgery 라는 용어로 보다 정확한 의미를 담은 45765043 로 결정

Case 2. B2021001: ENDOCLIP

구분	내용
매핑 용어	연구원 A: 45758369 (Ligation clip, metallic) VS 연구원 B: 46273961 (Gastrointestinal endotherapy clip)
3 자 검토 결과 및 사유	해당 의료기기는 Gastrointestinal 분야의 수술에서 활용되는 내시경 클립으로 Gastrointestinal endotherapy 라는 용어로 보다 정확한 의미를 담은 46273961 로 결정

Case 3. F1401004: BLUE PECKER

구분	내용
매핑 용어	<p>연구원 A: 45765198 (Bone biopsy needle)</p> <p>VS</p> <p>연구원 B: 45761192(Bone trephine)</p>
3 자 검토 결과 및 사유	해당 의료기기는 bone biopsy 이외의 목적으로도 활용되는 의료기기로서 biopsy needle 이라는 단어로 사용목적은 한정짓지 않은 45761192 로 결정

Case 4. H0011006: LYOPLANT

구분	내용
매핑 용어	<p>연구원 A: 4104097 (Dural substitute)</p> <p>VS</p> <p>연구원 B: 37018156 (Dura mater graft, bovine)</p>
3 자 검토 결과 및 사유	해당 의료기기는 신경외과 수술에 활용되는 생체 흡수성 뇌막 류로 소의 심막에서 추출한 콜라겐을 활용하여 만들어졌기 때문에, 보다 정확한 의미를 담고 있는 37018156 로 결정

Case 5. B2021109: VISU-LOC SPRING CLIP

구분	내용
매핑 용어	<p>연구원 A: 46273961 (Gastrointestinal endotherapy clip)</p> <p>VS</p> <p>연구원 B: 45758369 (Ligation clip, metallic)</p>
3 자 검토 결과 및 사유	해당 의료기기는 내시경 수술 혹은 관혈적 수술에 활용되는 결찰재료로서 Gastrointestinal 분야 이외의 수술에도 활용될 수 있어 의미를 한정짓지 않은 45758369 로 결정

4-3. 표준화된 재료 용어 적용 방안

데이터의 용어와 구조가 표준화되어 있지 않은 각 의료기관의 원천 헬스케어 데이터베이스는 OMOP CDM 의 형태로 ETL(Extract Transform Load) 과정이 수행됨으로써 데이터의 표준화가 달성된다. OMOP CDM 은 보건의료 데이터를 효과적으로 표준화하기 위한 데이터 구조와 용어 시스템을 보유하고 있다. 원천 자료에 입력된 보건의료 데이터가 해당 기관에서만 고유하게 사용하는 로컬 코드인 경우, 로컬 코드와 표준 용어 사이의 관계를 정의하는 매핑 과정이 필요하다.

청구를 위해 보고되지 않는 비급여 약물이나 치료 같은 경우는 별도의 EDI 가 존재하지 않는다는 한계점을 보유하고 있음에도 불구하고, 보험 급여 청구가 되는 모든 보건의료 항목들에 대한 국내 표준으로 활용되고 있는 EDI 코드를 적극적으로 국제표준용어에 매핑하여 다국적 연구에 활용될 수 있는 기반을 마련하는 것이 바람직하다.

5. REFERENCES

- [1] LOINC database (Regenstrief Institute, Indiana University School of Medicine, Indianapolis, USA)
- [2] Study Plan ver.2.0. ICD-9-CM (International Classification Disease, 9th revision, Clinical Modification)
- [3] Guidelines for identifying non-personal information (2018.07, Korea).
- [4] 건강보험심사평가원 수가반영내역(19.07.01 기준).
- [5] 한국의료 QA 학회지 Volume 12, Number 1, 104-112.
- [6] 한국과학기술평가원(18.08)
- [7] LOINC user's guide. <http://www.regenstrief.org/loinc/>
- [8] NPU, LOINC, and SNOMED CT: a comparison of terminologies for laboratory results reveals individual advantages and a lack of possibilities to encode interpretive comments. *J Lab Med* 2018; 42(6):267-275. Andreas Bietenbeck et al.
- [9] Recent Developments in Clinical Terminologies - SNOMED CT, LOINC, and RxNorm. *Yearb Med Inform.* 2018 Aug;27(1):129-139. Bodenreider O, Cornet R, Vreeman DJ.
- [10] Issues in mapping LOINC laboratory tests to SNOMED CT. *AMIA Annu Symp Proc.* 2008 Nov 6:51-5. Bodenreider O.
- [11] Use of SNOMED CT® and LOINC® to standardize terminology for primary care asthma electronic health records. *J Asthma.* 2018 Jun;55(6):629-639. doi: 10.1080/02770903.2017.1362424. Epub 2017 Oct 9. Lougheed MD, Thomas NJ, Wasilewski NV, Morra AH, Minard JP.
- [12] Standardizing Germany's Electronic Disease Management Program for Bronchial Asthma. *Stud Health Technol Inform.* 2019 Sep 3;267:81-85. doi: 10.3233/SHTI190809. Sass J, Essenwanger A, Luijten S, Vom Felde Genannt Imbusch P, Thun S.
- [13] An update on the use of health information technology in newborn screening. *Semin Perinatol.* 2015 Apr;39(3):188-93. doi: 10.1053/j.semperi.2015.03.003. Epub 2015 Apr 29. Abhyankar S, Goodwin RM, Sontag M, Yusuf C, Ojodu J, McDonald CJ.
- [14] The LOINC RSNA radiology playbook - a unified terminology for radiology procedures. *J Am Med Inform Assoc.* 2018 Jul 1;25(7):885-893. doi: 10.1093/jamia/ocy053. Vreeman DJ, Abhyankar S, Wang KC, Carr C, Collins B, Rubin DL, Langlotz CP.
- [15] Single-Center Experience Implementing the LOINC-RSNA Radiology Playbook for Adult Abdomen/Pelvis CT and MR Procedures Using a Semi-Automated Method. *J Digit Imaging.* 2018 Feb;31(1):124-132. doi: 10.1007/s10278-017-0016-0. Sandhu RS1, Shin J, Wang KC, Shih G. Sandhu RS, Shin J, Wang KC, Shih G.

- [16] Observational Health Data Sciences and Informatics. The Book of OHDSI.
<https://ohdsi.github.io/TheBookOfOhdsi/>
- [17] Observational Health Data Sciences and Informatics. Standardized Vocabulary V5.0.
<https://www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary>