

서울성모병원 CDM 소개 자료

CDM이란

■ CDM (Common Data Model)

- 여러 소스에서 수집된 임상 데이터는 기관별로 구성, 형식, 용어에 차이로 공동연구가 어려움¹⁾
- 이를 극복할 방법으로 제시된 하나가 Common Data Model(CDM)이며, 데이터를 공통된 형태로 만들 수 있도록 해주는 데이터 모델을 의미.
- CDM을 구성하는 주요 항목은 국제 표준 용어를 사용
- CDM은 다른 목적에 따라 생성된 다른 형식의 데이터를 수용할 수 있어 국내외의 데이터 소스를 활용한 공동연구도 지원이 가능²⁾

1)S.J. Reisinger, P.B. Ryan, D.J. O'Hara et al., Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare data-bases, J Am Med Inform Assoc 17 (2010), 652-662

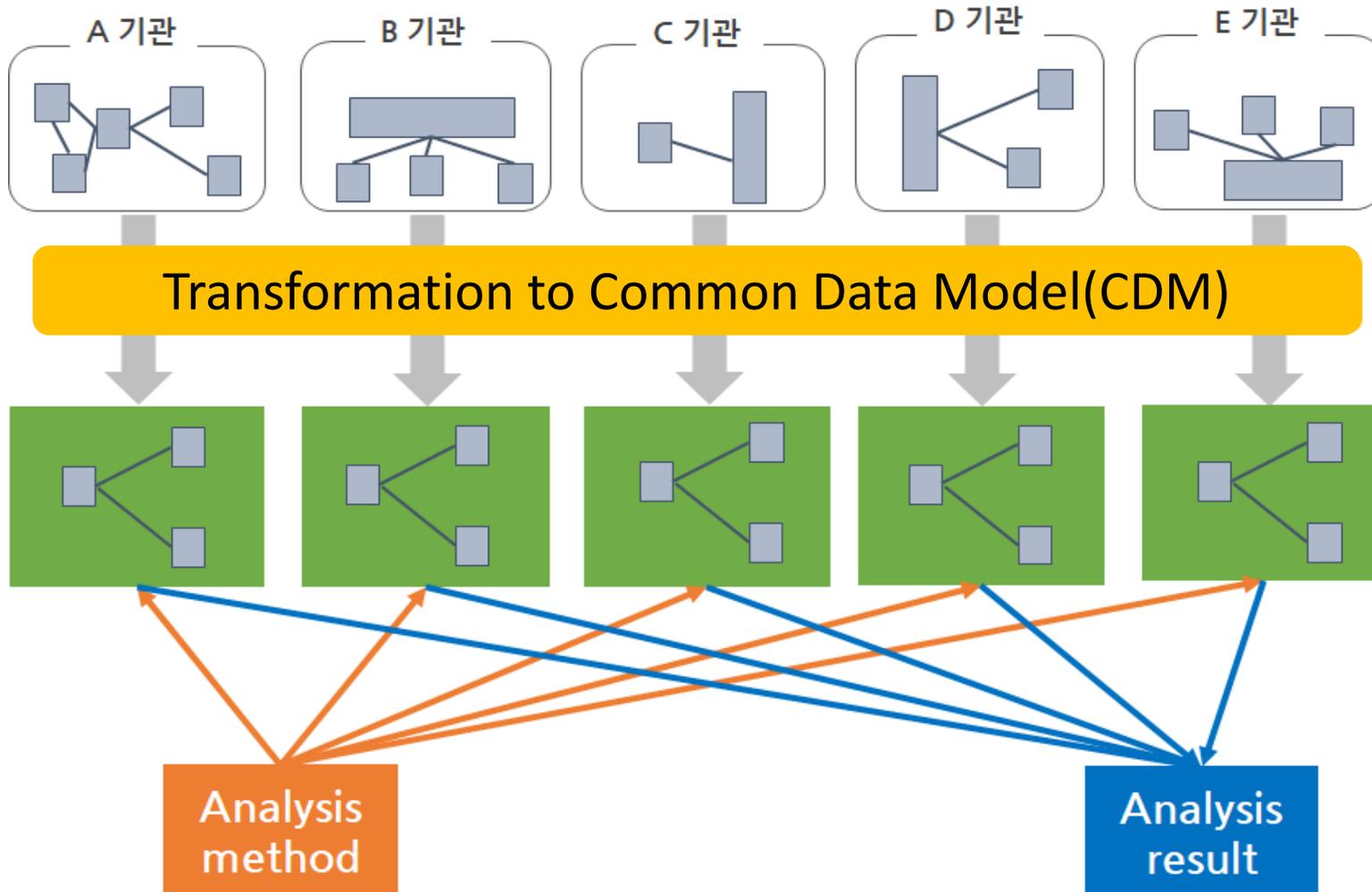
2)<https://www.ohdsi.org/data-standardization/>

OMOP CDM

- The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)
- By OHDSI (Observational Health Data Sciences and Informatics)

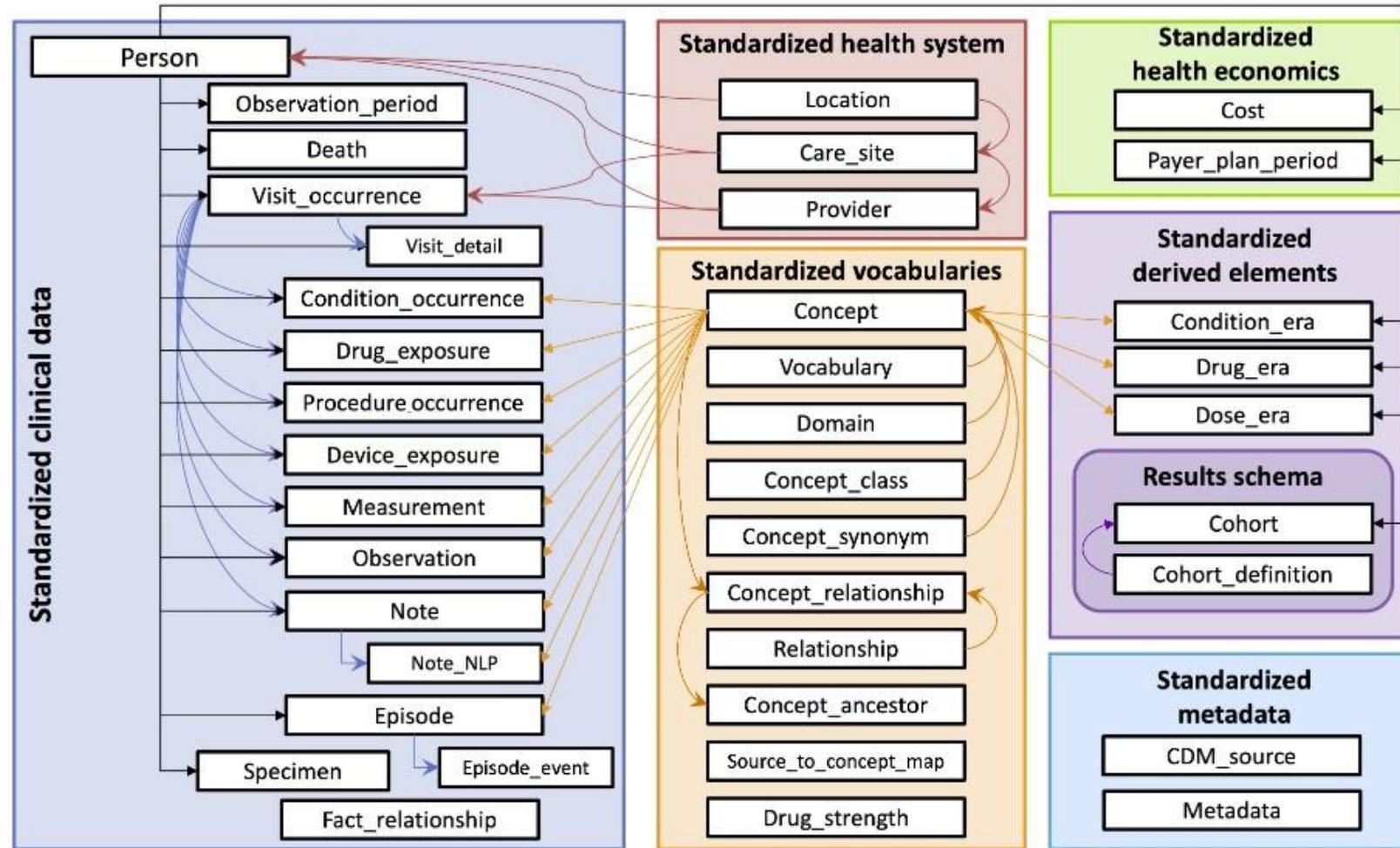
- 연구 방법 : population-level estimation and patient-level predictions
- 목적 : 신뢰할 수 있는 과학적 증거를 생성하여 의료 의사 결정

CDM 활용



OMOP¹⁾ CDM V5.3.1

1) OMOP; Observational Medical Outcomes Partnership



CDM 주요 데이터 테이블

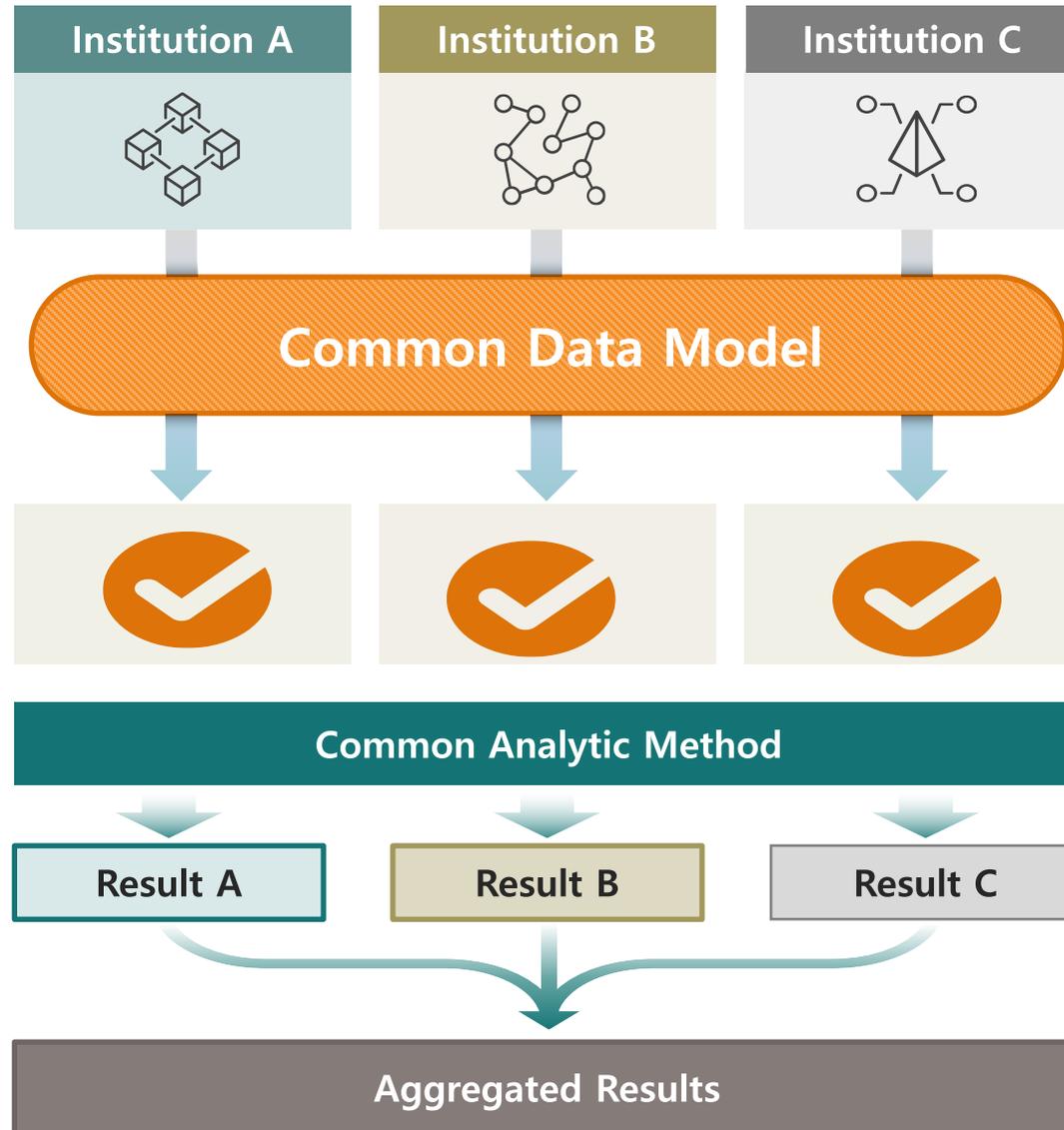
구분	테이블명(EN)	테이블명(KR)	테이블 내 컬럼항목 정보
1	Person	환자기본정보	성별, 생년월일, 출생일시, 사망일시, 인종, 민족성, 환자 주소 ID, 의료진 ID, 병원지역 ID, 환자의 주소 정보
2	Observation_period	환자관찰기간 (최초방문일, 최종방문일)	의료기관 최초 방문일, 의료기관 최종 방문일, 방문기간 타입, 급여대상자 여부, 투약에 대한 보험 여부, 차트열람 가능 여부
3	Visit_occurrence	수진정보 (입원, 외래, 응급)	방문 유형, 방문 시작일, 방문 종료일, 방문 기록 유형, 의료진 ID, 병원 지역 ID, 원자료의 방문 기록, 입원 기록, 퇴원 기록, 개인의 이전 방문 ID, 병원 이름, 환자의 퇴원 시 상태, 퇴원 후 환자의 위치, 환자의 입원 전 위치
4	Visit_detail	수진정보 (전과,전동,전실 포함)	방문 상세 시작일, 방문 상세 종료일, 방문 상세 기록 유형, 의료진 ID, 병원 지역 ID, 원자료의 방문 상세 기록, 입원 기록, 퇴원 기록, 개인의 이전 방문 상세 ID, 개인의 고유 방문 ID
5	Condition_occurrence	진단정보 (진단, 주호소)	진단 표준용어 concept id, 진단 시작일, 진단 기록 유형, 진단 상세 기록, 개인의 고유 방문 ID, 방문일, 병원 이름, 방문 유형, 진단코드 값, 진단코드 종류 및 버전, 원자료의 진단코드 값, 주진단/부진단 여부
6	Death	사망정보	사망일자, 사망정보 원자료, 데이터 신뢰도
7	Drug_exposure	약물 처방 및 투약기록	약물의 표준용어 concept id, 현재 약물 노출 시작일/ 종료일, 처방 기록 유형, 약 총 처방량, 처방기간, 용량용법, 약물 노출경로 정보, 의료진 ID, 개인의 고유 방문 ID, 방문 상세 ID, 약물의 EDI 코드, 원자료의 약물코드 concept id, 원자료의 약물 투여경로 정보, 약물 노출 용량 concept id, 원자료의 약물 처방 용량 단위, 환자 ID, 약물 처방일, 총 처방 기간(일 단위), 총 처방량
8	Procedure_occurrence	수술, 처치, 검사처방	처치·수술 표준용어 concept id, 처치·수술 일자, 처치·수술 기록 유형, 시술 수식어 concept id, 처치·수술 시행 수, 의료진 ID, 개인의 고유 방문 ID, 방문 상세 ID, 원자료의 처치·수술명, 원자료의 처치·수술 코드 값, 처치코드, 방문일, 병원 이름, 방문 유형

CDM 주요 데이터 테이블

구분	테이블명(EN)	테이블명(KR)	테이블내 컬럼항목 정보
9	Device_exposure	기기 및 재료	원자료의 기기,치료재료 노출 고유 ID, 기기/치료재료 수식어 concept id, 노출 시작일, 노출 종료일, 기기/치료재료 유형, 수량, 의료진 ID, 개인의 고유 방문 ID, 방문 상세 ID, 기기/치료재료 기록, 원자료의 기기
10	Measurement	검사결과, 임상관찰기록, 병리보고서	개인의 고유 검사 ID, 검사 표준용어 concept id, 검사 일자, 검사 시간, 검사 유형, 검사 결과 숫자형 값/단위/하한값/상한값, 의료진 ID, 방문 상세 ID, 원자료의 검사 코드, 원자료의 검사 결과 값, 검사명 약어, 검사 분류, 공복 검사 여부, 검사 시료의 종류, 검사의 긴급성, 검사 시료 추출 장소, 검사 결과 위치, 검사의 EDI코드, 검사 코드 종류, 검체채취 날짜/시간, 검사처방 날짜, 결과보고 날짜/시간, 검사 결과 concept id, 검사 결과 값 단위 concept id, 검사결과의 정상범위 하위 문자형, 검사결과의 정상범위 상한값, 검사결과의 정상범위, 상위 문자형, 검사결과와 정상범위 비교결과, 검사결과값(문자형), 검사결과값(숫자형), 원자료의 검사결과 단위, LOINC 검사 코드 값, 검사처방한 과, 의료기관식별부호
11	Note	기록지 (입퇴원기록, 수술기록, 병리기록, 영상기록)	노트 이벤트 ID, 노트 일자, 노트 일시, 노트 유형 concept id, 노트 범주 concept id, 노트 제목, 노트 내용, 문자 인코딩, 언어 concept id, 의료진 ID, 개인의 고유 방문 ID, 방문 상세 ID, 노트의 자료원
12	Specimen	검체 정보	검체 고유 ID, 검체 concept ID, 검체 유형 concept ID, 관찰일, 해부학 부위 기록, 질병 상태 기록
13	Observation	관찰 결과 정보	관찰의 concept ID, 관찰일, 관찰 유형 concept ID, 관찰 결과 숫자화, 관찰 결과 문자화, 관찰 결과 concept id, 검사 유닛 concept ID, 의료진 ID, 개인의 고유 방문 ID, 방문 상세 ID, 관찰 기록 concept ID
14	Location	위치 및 주소 정보	위치 및 주소정보(시군구)
15	Care_site	의료기관 정보	의료기관 ID, 의료기관명, 병원지역ID, 의료기관 영문명
16	Provider	의료진 정보	진료과 concept ID, care_site ID, specialty_source_value(진료과 한글명)
17	Drug_Era	약물 노출 정보	(Drug_exposure 데이터에 대한) 특정약성분에 노출된 것으로 예상되는 시간, 길이
18	Condition_Era	진단 기간 정보	(Condition_occurrence 데이터에 대한) 처방된 진단상태를 갖는 것으로 예상되는시간, 길이

서울아산병원 CDM 소개 자료

CDM (Common Data Model) for Collaborative Research & Real World Evidence



✔ **Integrated data analysis across institutions**

✔ **Easily expandable based on standards**

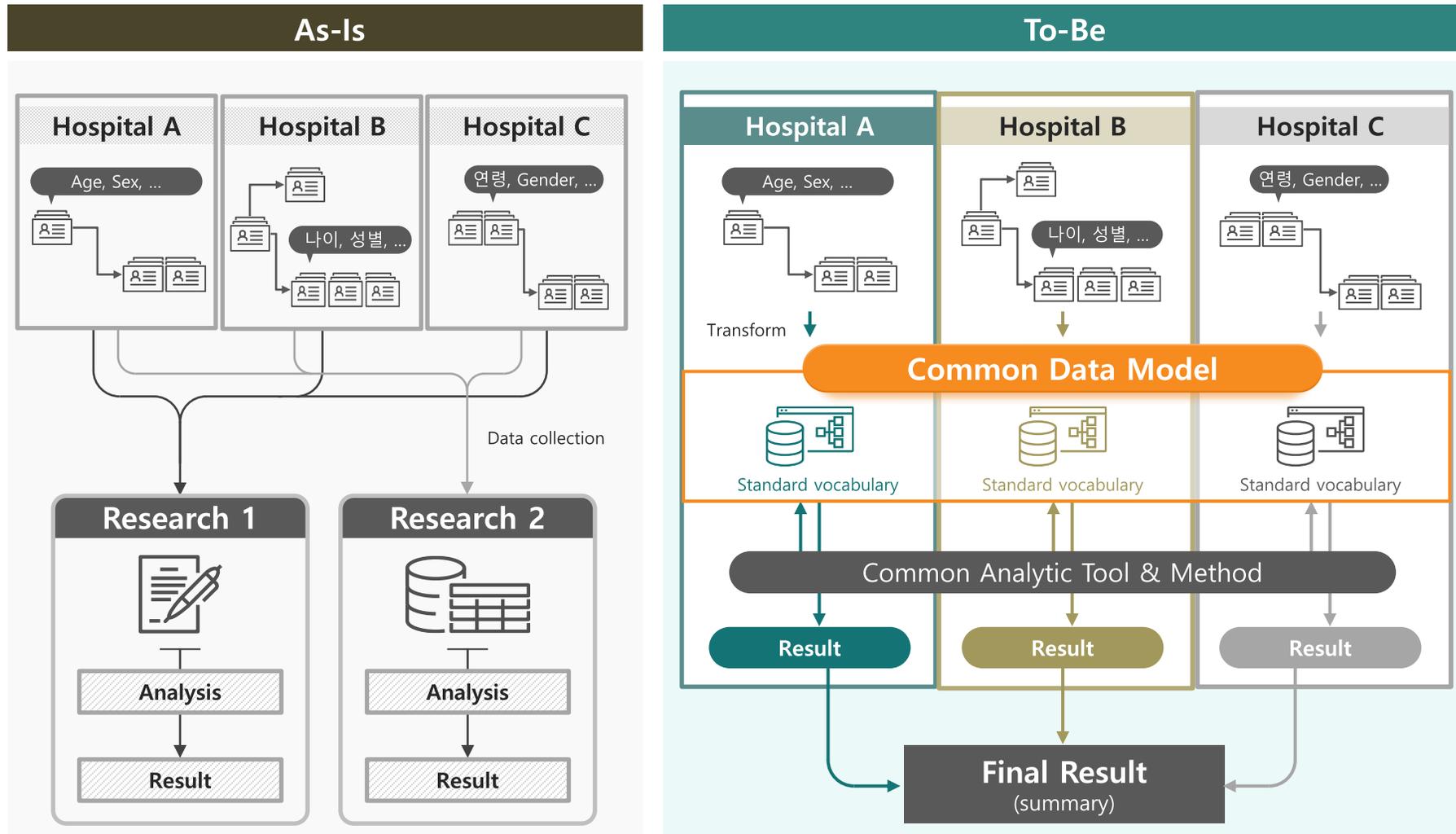
✔ **Minimize sensitive personal information exposure**

CDM data standard team

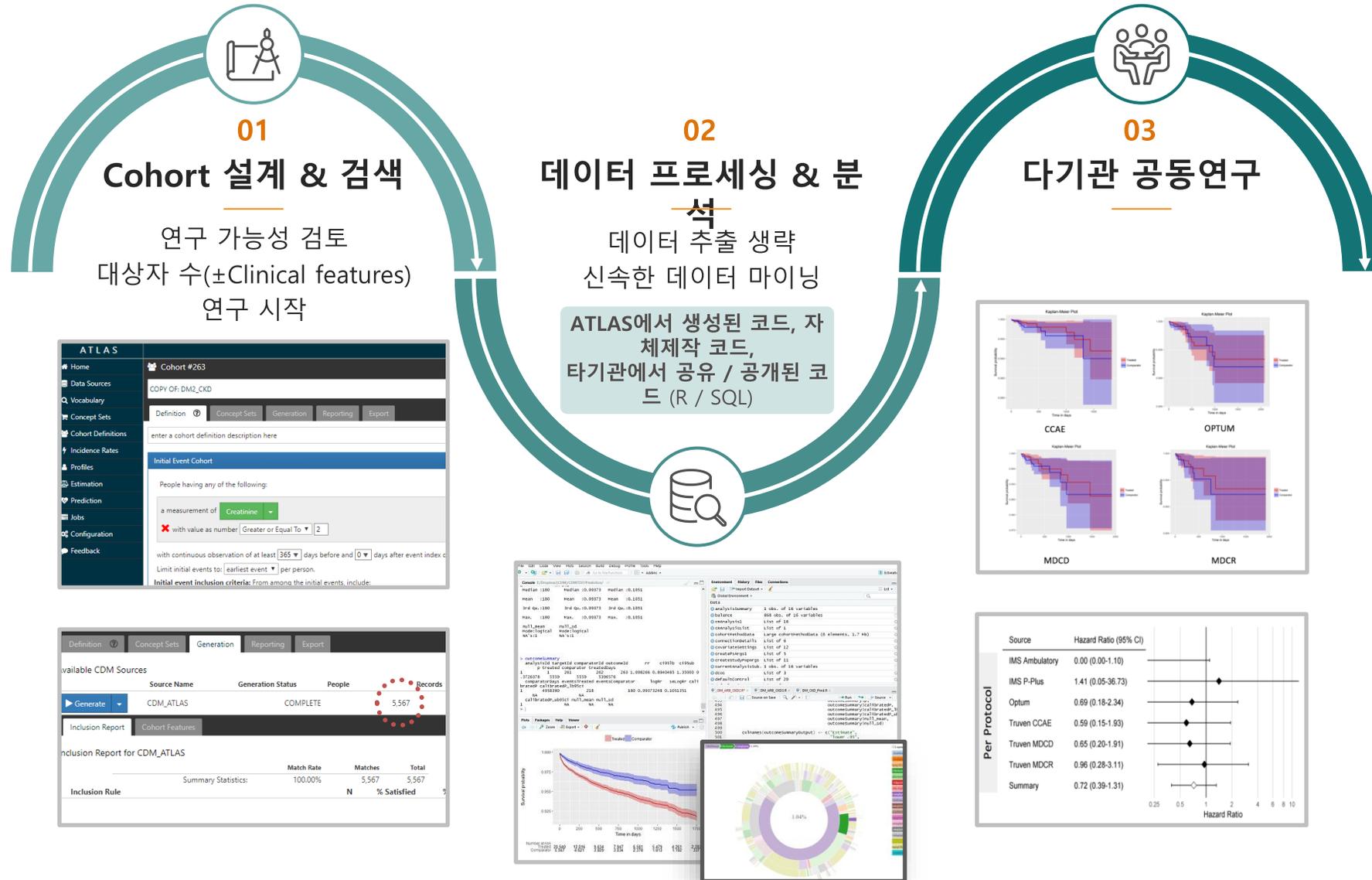
- **Unifying/managing medical terminology**
- **Vocabulary Mapping to OMAP standards**
(SNOMED-CT, LOINC, RxNorm)
 - Condition
 - Drug
 - Procedure
 - Measurement
 - Device...

CDM-based Multi-Center Study

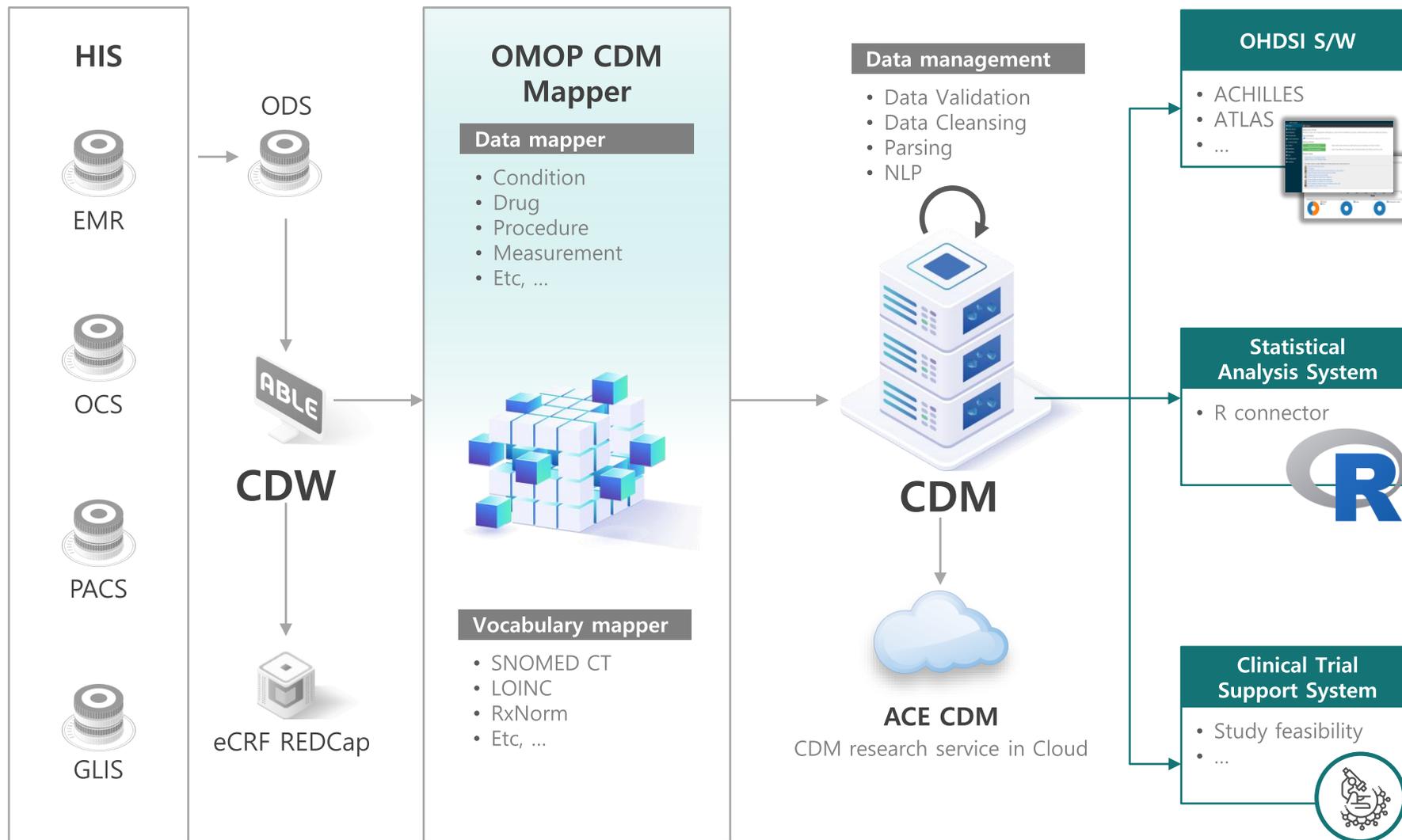
의료 데이터를 이용하여 **질병관리 및 건강증진**에 필요한 과학적 증거를 생성하기 위한 **다기관 공동 연구**를 효율적이고 효과적으로 수행하기 위한 표준화된 **데이터 모델**



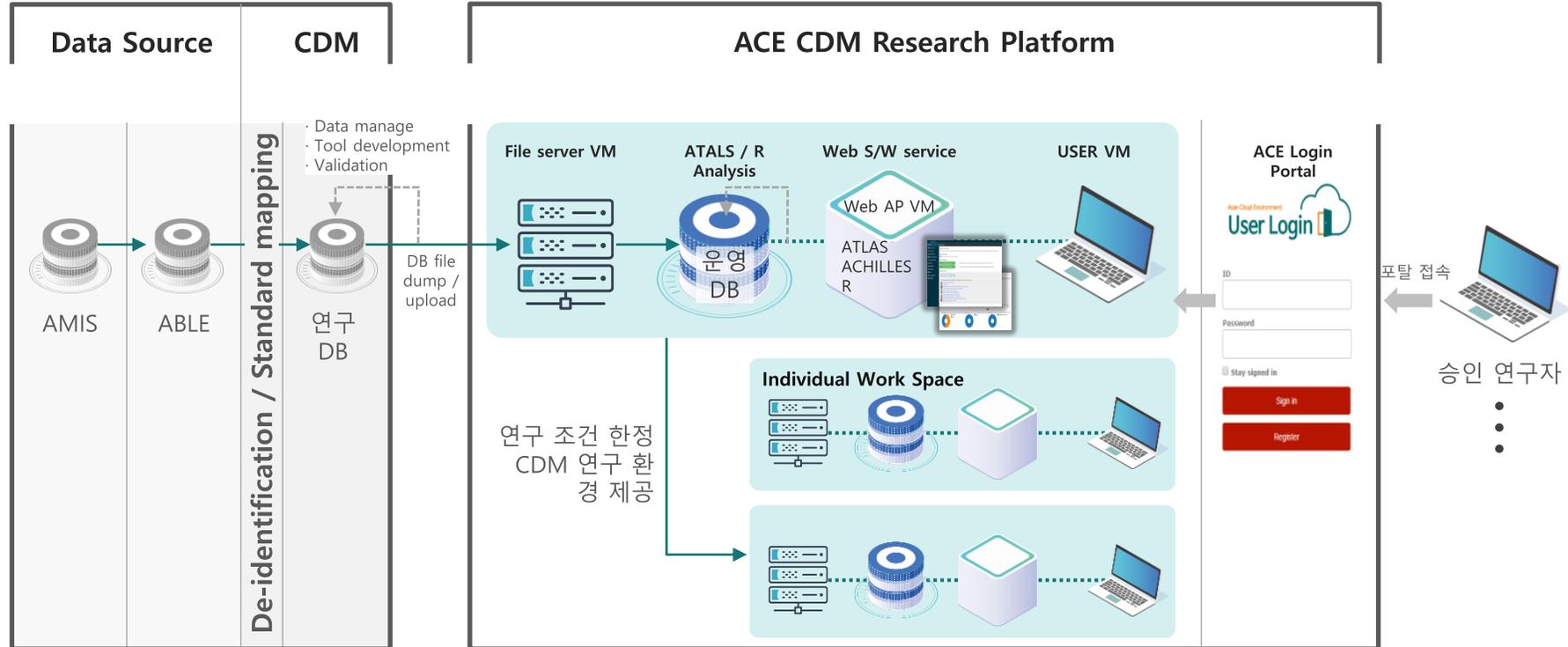
CDM-based Multi-Center Study



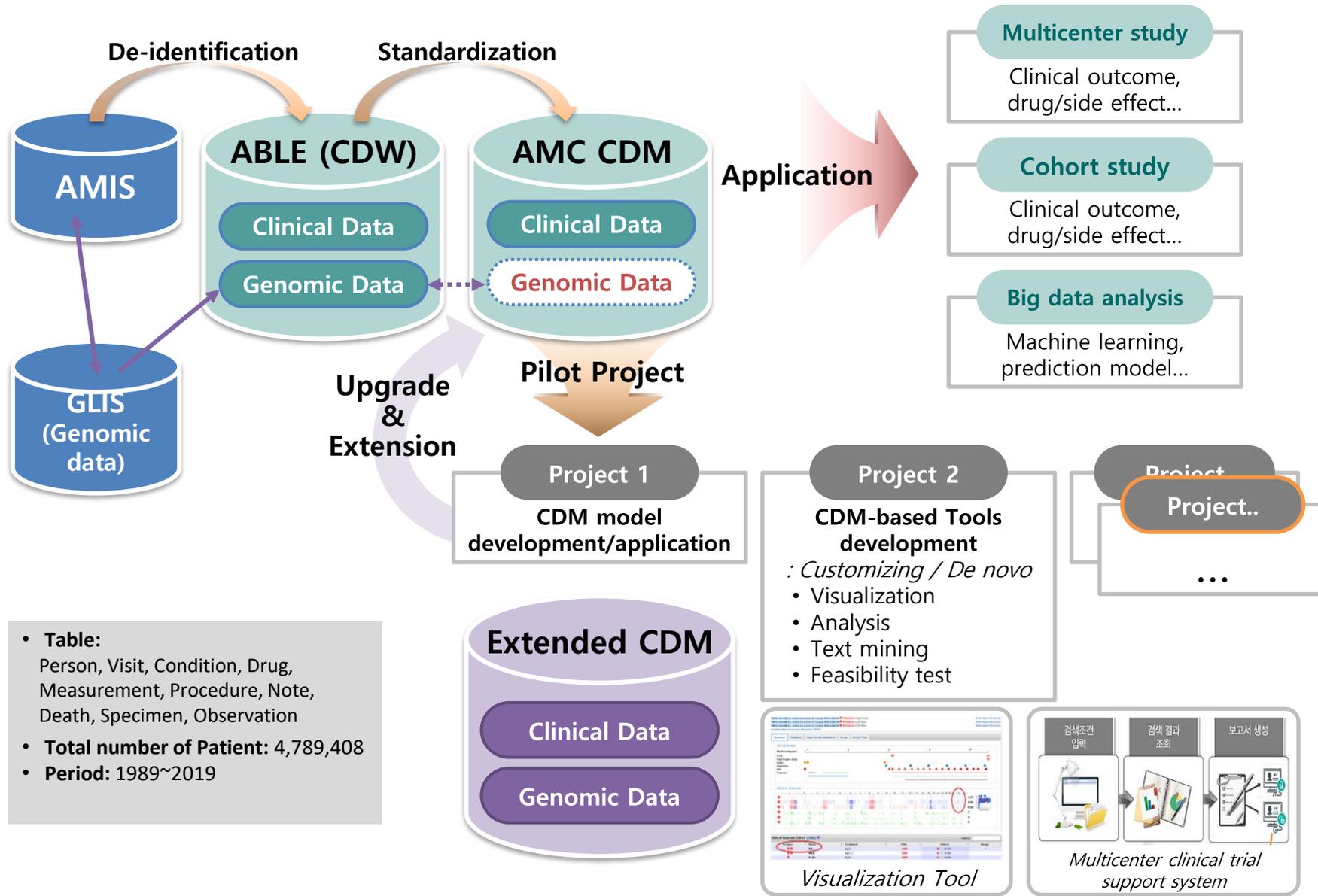
Clinical Data & Research Flow in AMC (2)



Cloud-based CDM Research Platform (ACE: Asan Cloud Environment)



AMC CDM(Common Data Model) Researches for Precision Medicine



- **Table:** Person, Visit, Condition, Drug, Measurement, Procedure, Note, Death, Specimen, Observation
- **Total number of Patient:** 4,789,408
- **Period:** 1989~2019

국립암센터 CDM 소개 자료

CDM이란

■ CDM (Common Data Model)

- 여러 소스에서 수집된 임상 데이터는 기관별로 구성, 형식, 용어에 차이로 공동연구가 어려움¹⁾
- 이를 극복할 방법으로 제시된 하나가 Common Data Model(CDM)이며, 데이터를 공통된 형태로 만들 수 있도록 해주는 데이터 모델을 의미
- CDM을 구성하는 주요 항목은 국제 표준 용어를 사용
- CDM은 병원에서 발생한 주요한 임상 정보를 표준화 된 데이터 규격에 맞춰 적재되도록 설계되어 국내외의 데이터 소스를 활용한 공동연구 지원이 가능²⁾

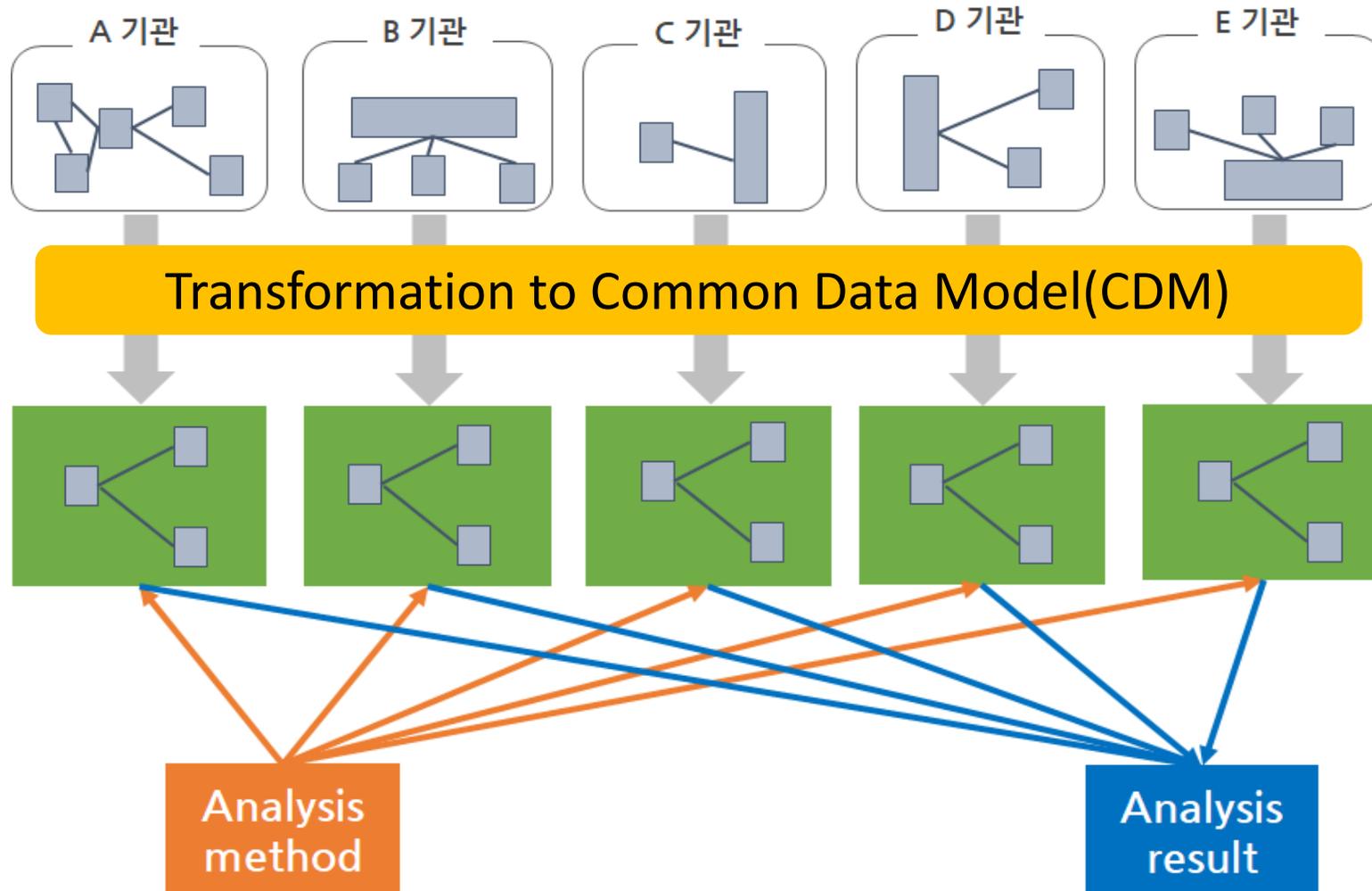
1) S.J. Reisinger, P.B. Ryan, D.J. O'Hara et al., Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare data-bases, J Am Med Inform Assoc 17 (2010), 652-662

2) <https://www.ohdsi.org/data-standardization/>

OMOP CDM

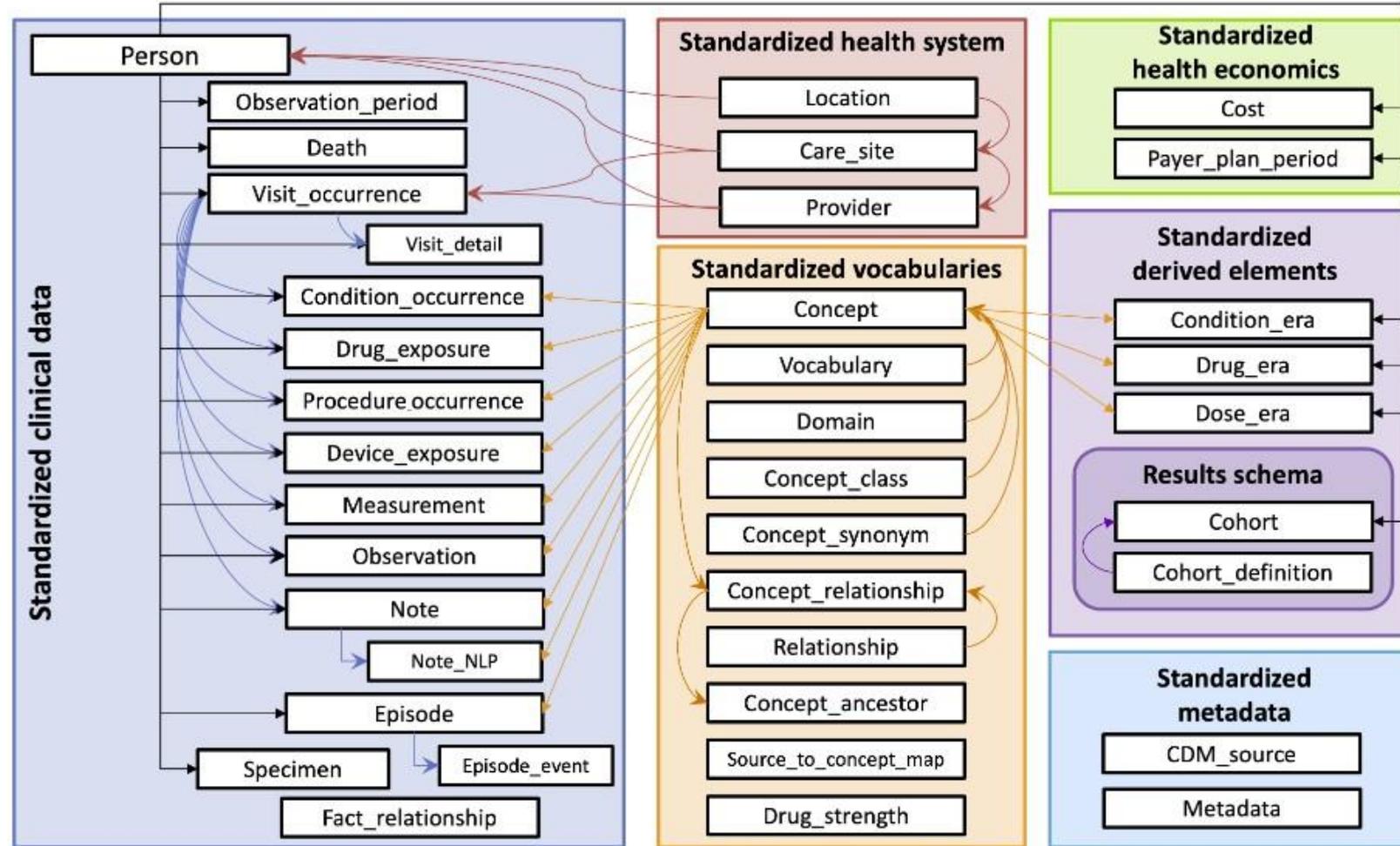
- The Observational Medical Outcomes Partnership(OMOP) Common Data Model(CDM)
- By OHDSI(Observational Health Data Sciences and Informatics)
- 연구 방법 : population-level estimation and patient-level predictions
- 목적: 신뢰할 수 있는 과학적 증거를 생성하여 의료 의사 결정 지원

CDM 활용



OMOP¹⁾ CDM V5.3.1

1) OMOP; Observational Medical Outcomes Partnership



CDM 주요 데이터 테이블

구분	테이블명(EN)	테이블명(KR)	테이블 내 컬럼항목 정보
1	Person	환자기본정보	CDM 환자ID (연계키), 성별_concept ID, 출생 년도, 출생 월, 민족 concept ID, 성별_원내기록_concept id
2	Observation_period	환자관찰기간 (최초방문일, 최종방문일)	관찰 기록 ID, CDM 환자ID (연계키), 최초 관찰 기록일자, 최종 관찰 기록일자, 관찰기록 출처
3	Visit_occurrence	수진정보 (입원, 외래, 응급)	내원 유형 ID, CDM 환자ID (연계키), 내원 유형_concept ID, 내원 시작일, 내원 종료일, 내원 출처정보, 내원 유형_원내기록_concept id
4	Condition occurrence	진단정보 (진단, 주소소)	상병 기록 ID, CDM 환자ID (연계키), 상병 기록_concept ID, 상병기록 시작일자, 상병기록 유형_주/부상병, 내원 기록 ID
5	Drug_exposure	약물 처방 및 투약기록	투약 기록 번호, CDM 환자ID (연계키), 투약 약품정보 concept ID, 투약 시작일자, 투약 종료일자, 투약 출처정보, 1일당 투약 약품량, 투약 일수, 약품 투여방법_concept ID, 내원 유형 ID, 약품 투여방법_원내기록값, 약품 용량 단위의 원내기록값
6	Procedure_occurrence	수술, 처치, 검사처방	행위처방 기록ID, CDM 환자ID (연계키), 행위처방 concept ID, 행위처방 일자, 시행 횟수, 내원 유형 ID
7	Measurement	검사결과, 임상관찰기록, 병리보고서	검사 및 측정 식별 ID, CDM 환자ID (연계키), 검사 및 측정 concept ID, 검사 및 측정 일자, 검사 및 측정 시간, 검사 및 측정 정보 출처 concept ID, 검사 및 측정 결과값, 내원 유형 ID, 검사 및 측정 결과치에 대한 단위의 원내 기록값, 검사 및 측정 결과_원내기록값
8	Observation	관찰 결과 정보	관찰 기록 ID, CDM 환자ID (연계키), 관찰 정보 concept ID, 관찰 일자, 관찰 정보 출처 concept ID, 관찰 정보 결과_원내기록값, 내원 유형 ID
9	Death	사망정보	CDM 환자 ID(연계키), 사망 일자, 사망정보 출처_concept ID, 사망 원인_concept ID

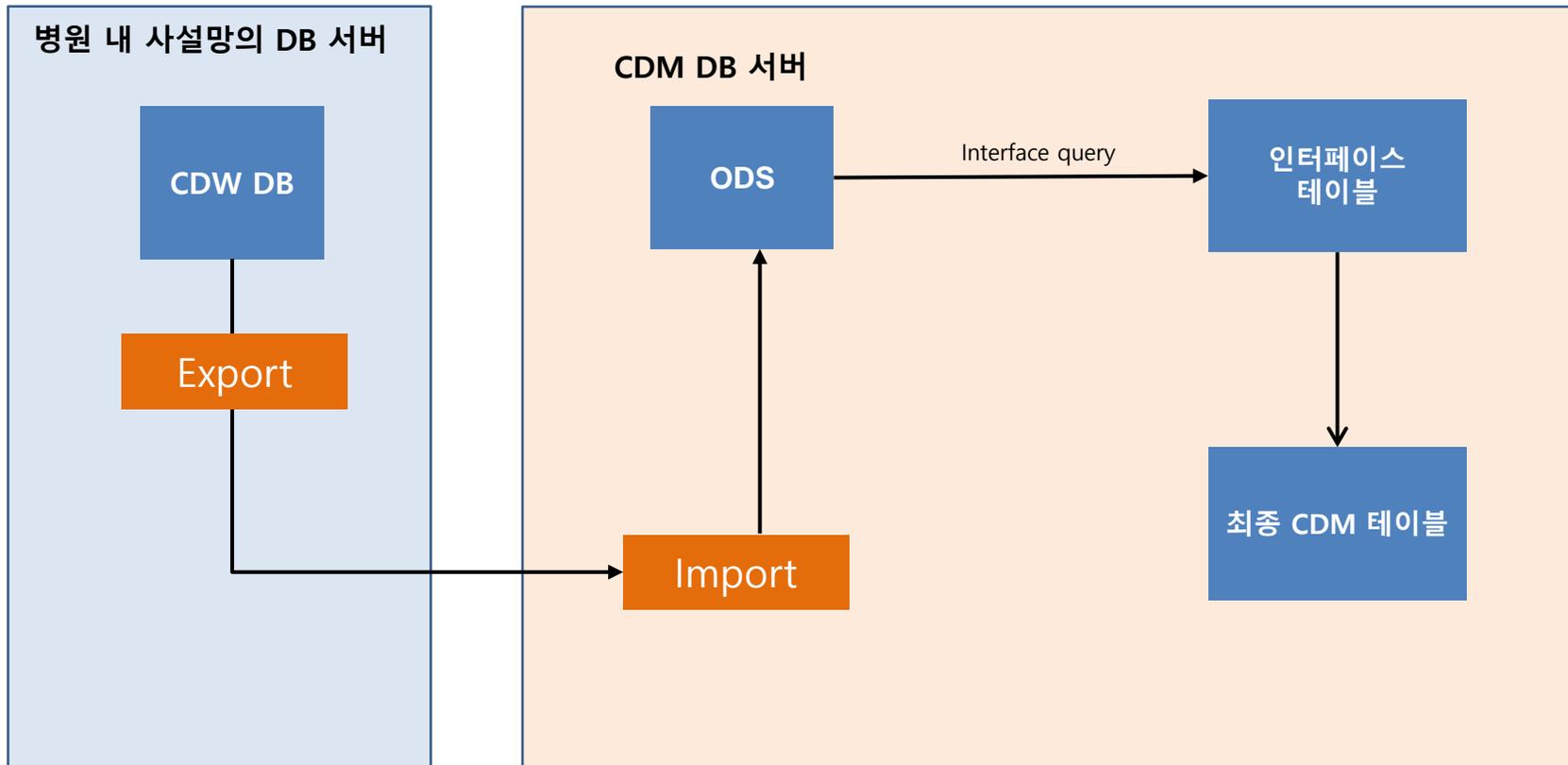
중앙대학교병원 CDM 소개 자료

1. CDM 소개

OMOP- CDM Version	CDM v5.3.1 (See full CDM specification file on the OHDSI Web)
데이터 기간	2022년 06월 ~ 2025년 01월(데이터 현행화 중)
환자수	전체 약 35만 명
CDM 변환	CDW to CDM (Oracle to PostgreSQL)
주기 변환	Monthly Update

2. CDM 변환 과정

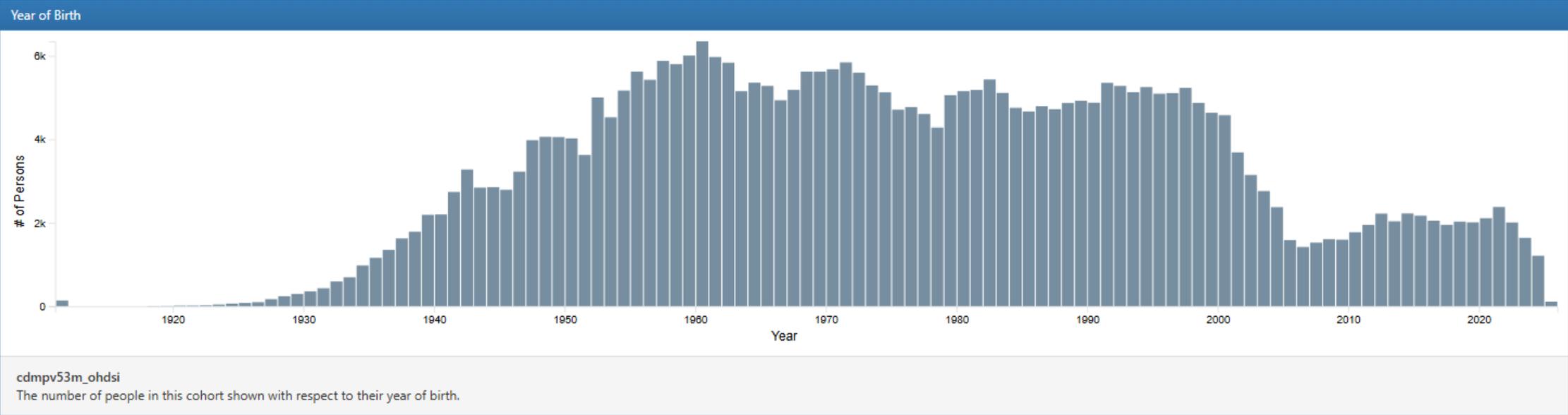
- ▶ 필요한 원천 테이블을 Export/Import하여 CDM DB 서버에서 변환 작업 진행.



3. CDM 주요 데이터 현황 (1/2)

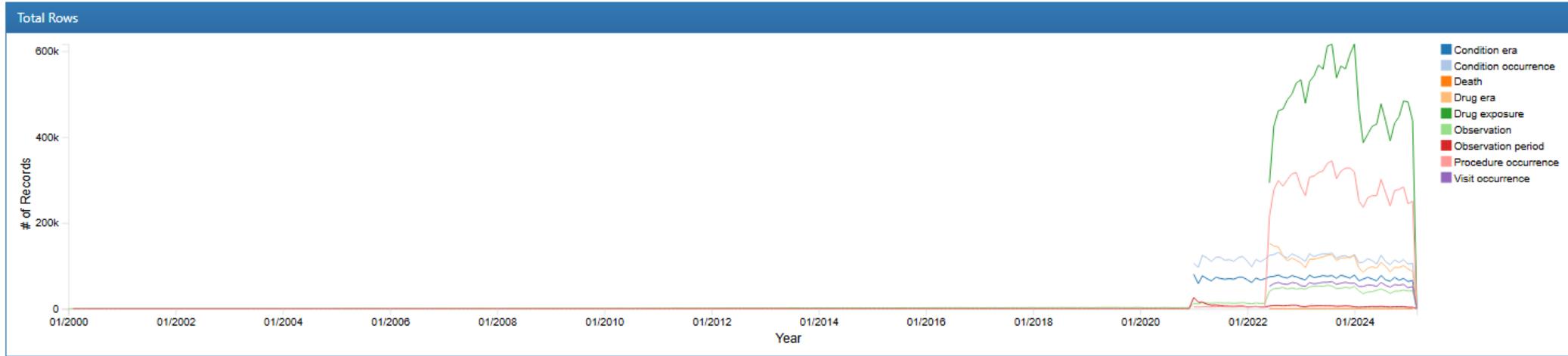
환자 연령대 별 현황

cdmpv53m_ohdsi Person Report

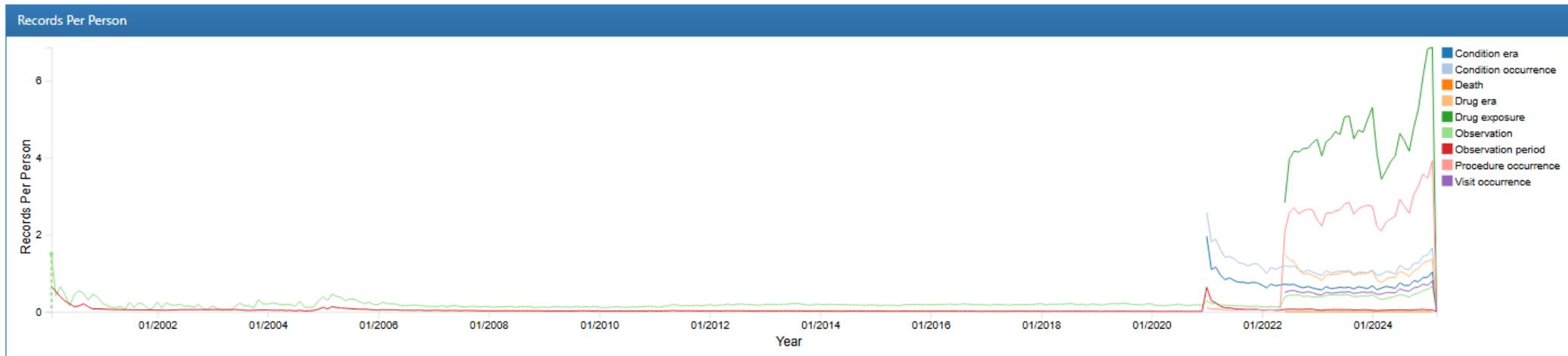


3. CDM 주요 데이터 현황 (2/2)

Total Rows



Records Per Person



4. 주요 CDM 테이블 정보

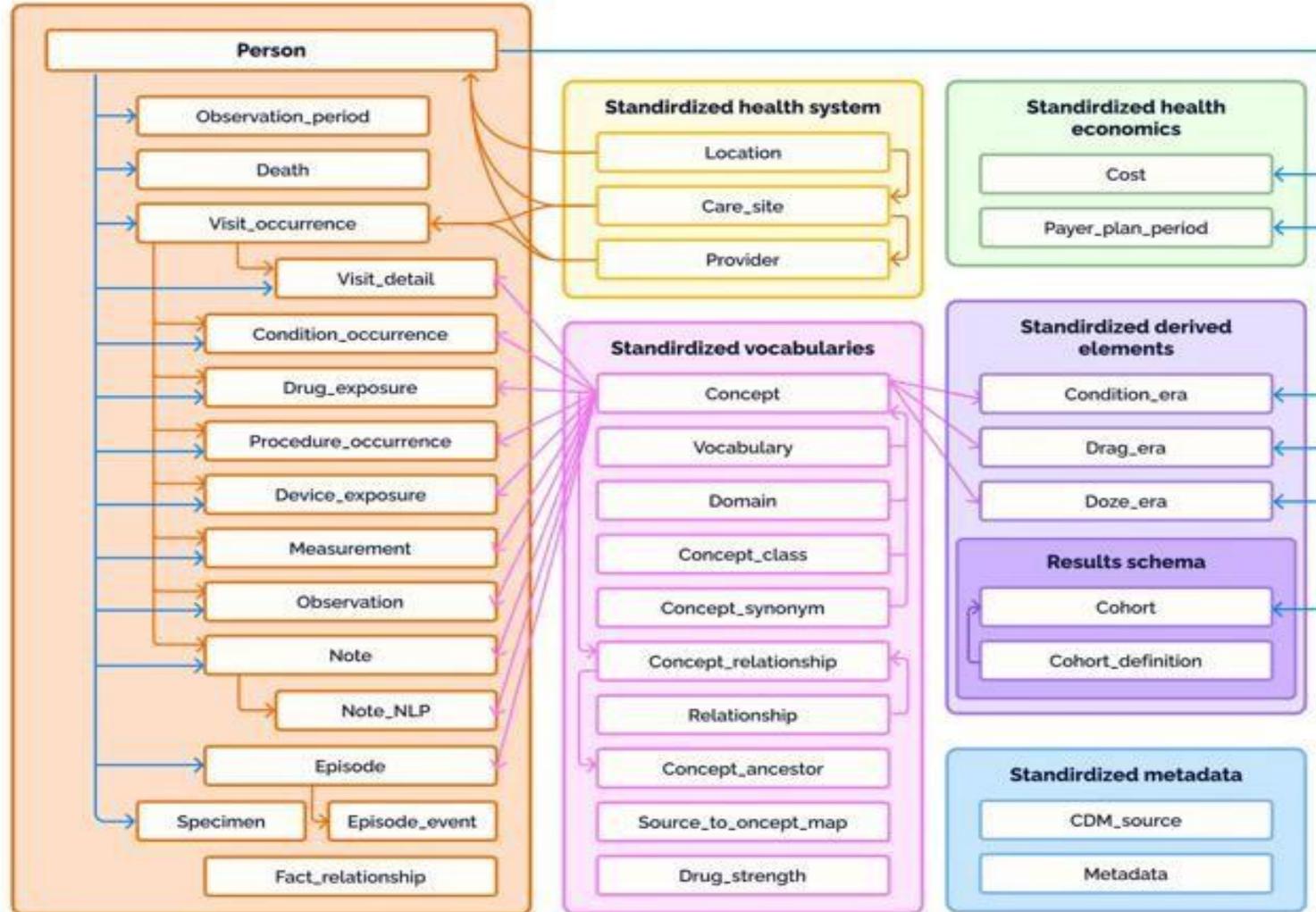
NO	테이블명	원천 테이블(EMR, CDW 등)	표준코드 (사용코드)	비고
1	PERSON	환자마스터테이블	없음	-
2	VISIT_OCCURRENCE	내원정보테이블	없음	-
3	CONDITION_OCCURRENCE	진단정보테이블	SNOMED CT	-
4	DRUG_EXPOSURE	투약정보테이블	RXnorm	-
5	PROCEDURE_OCCURRENCE	수술/처치 처방테이블	EDI, SNOMED CT	-
6	MEASUREMENT	검사(Lab, 영상, 병리) 테이블	LOINC	진단검사 신체측정
7	DEATH	사망정보 테이블(원내사망만)	없음	-
8	NOTE	의무기록 서식 및 내용 테이블	없음	-
9	SPECIMEN	병리테이블	SNOMED CT	-

건국대학교병원 CDM 소개 자료

CDM

- CDM
 - Common Data Model
 - 다양한 데이터 형식을 표준화하여 동일한 구조로 변환하는 데이터 모델
 - 데이터의 일관성 유지, 여러 기관에서 공통된 방식으로 데이터를 활용할 수 있음
- OMOP CDM
 - 여러 CDM 중 하나로, Observational Medical Outcomes Partnership(OMOP)에서 만든 모델
 - 특히 의료 데이터를 연구하는 데 최적화된 모델
 - SNOMED-CT, LOINC, RxNorm 같은 표준화된 용어 체계 사용하여 데이터의 일관성과 신뢰성이 높음

OMOP CDM V5.3.1



CDM 현황

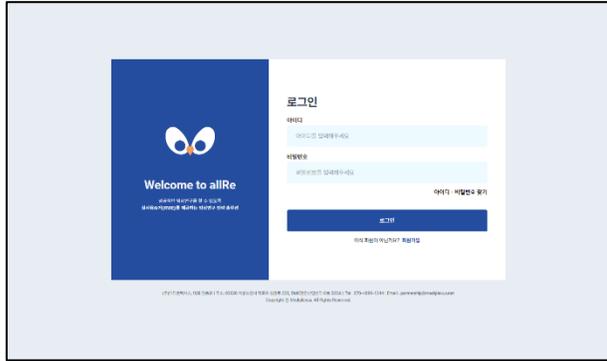
NO	테이블명	원천 테이블(EMR, CDW 등)	표준코드 (사용코드)	비고
1	Person	환자기본정보	없음	-
2	Observation_period	환자관찰기간	없음	최초방문일 최종방문일
3	Visit_occurrence	수진정보	없음	입원 외래 응급
4	Visit_detail	수진정보	없음	전과 전동 전실
5	Condition_occurrence	진단정보	SNOMED-CT, ICD,Cancer Modifier, OMOP Extension	진단
6	Death	사망정보	없음	-
7	Drug_exposure	약물 처방 및 투약기록	RxNorm, RxNorm Extension, KCD	-
8	Procedure_occurrence	수술, 처치, 검사처방	ICD9Proc,OMOP Extension,RxNorm,OMOP Genomic,OPCS4,CPT4,SNOMED,NAACCR,HCPCS,ICD10PCS	-
9	Device_exposure	기기 및 재료	AMT,SNOMED,HCPCS	-
10	Measurement	검사결과, 임상관찰기록, 병리보고서	LOINC	-

CDM 현황

NO	테이블명	원천 테이블(EMR, CDW 등)	표준코드 (사용코드)	비고
11	Note	기록지	없음	병리결과, 영 상판독, 입퇴 원기록지
12	Specimen	검체정보	SNOMED	-
13	Observation	가족력/과거력, 음주, 흡연, 문진결과 등	SNOMED	-
14	Location	위치 및 주소정보	없음	-
15	Care_site	의료기관정보	없음	-
16	Provider	의료진정보	없음	-
17	Drug_Era	(Drug_exposure 데이터에 대한) 특정약성분에 노출된 것으로 예상되는 시간 길이	없음	-
18	Condition_Era	(Condition_occurrence 데이터에 대한) 처방된 진단상태를 갖는 것으로 예상되는 시간 길이	없음	-

CDM 학습용 시스템

- 주요 화면



Original dataset	Deleted	Used in PSM
360	0	360

RAAS_분류	RAAS_분류	Count
1	1	475
2	2	14
3	1	95
4	2	2

과거사_고혈압	과거사_고혈압	Count
1	1	340
2	2	149
3	1	46
4	2	42

eGFR_1년변화량 (N: 0 - 326, 1 - 60)

* Mean: red point, Median: black bar

data	diff	p-value
eGFR_1년변화량	-0.32	117
eGFR_1년변화량	0.977	6.792

subset	N	Mean	Std	Missing
0	61	-0.977	10.867	32
1	58	-1.48	5.559	25