

약물부작용 감시를 위한 공통데이터모델 기반 임상데이터웨어하우스 구축

노미정*

*단국대학교 공공·보건과학대학

〈Abstract〉

Development and Lessons Learned of Clinical Data Warehouse based on Common Data Model for Drug Surveillance

*Mi Jung Rho

**College of Health Science, Dankook University, Republic of Korea*

Purposes: It is very important to establish a clinical data warehouse based on a common data model to offset the different data characteristics of each medical institution and for drug surveillance. This study attempted to establish a clinical data warehouse for Dankook university hospital for drug surveillance, and to derive the main items necessary for development.

Methodology/Approach: This study extracted the electronic medical record data of Dankook university hospital tracked for 9 years from 2013 (2013.01.01. to 2021.12.31) to build a clinical data warehouse. The extracted data was converted into the Observational Medical Outcomes Partnership Common Data Model (Version 5.4). Data term mapping was performed using the electronic medical record data of Dankook university hospital and the standard term mapping guide. To verify the clinical data warehouse, the use of angiotensin receptor blockers and the incidence of liver toxicity were analyzed, and the results were compared with the analysis of hospital raw data.

Findings: This study used a total of 670,933 data from electronic medical records for the Dankook university clinical data warehouse. Excluding the number of overlapping cases among the total number of cases, the target data was mapped into standard terms. Diagnosis (100% of total cases), drug (92.1%), and measurement (94.5%) were standardized. For treatment and surgery, the insurance EDI (electronic data interchange) code was used as it is. Extraction, conversion and loading were completed. R language-based conversion and loading software for the process was developed, and clinical data warehouse construction was completed through data verification.

Practical Implications: In this study, a clinical data warehouse for Dankook university hospitals based on a common data model supporting drug surveillance research was established and verified. The results of this study provide guidelines for institutions that want to build a clinical data warehouse in the future by deriving key points necessary for building a clinical data warehouse.

Key Words: Common Data Model, Data Warehouse, Distributed Research Networks, Electronic Health Record, Multi-institutional Research, MOA CDM

* 투고일자 : 2023년 01월 25일, 수정일자 : 2023년 08월 04일, 게재확정일자 : 2023년 08월 05일

** 본 연구는 2022년도 한국약품안전관리원 용역비 지원에 의해 수행되었음

‡ 교신저자 : 노미정, 단국대학교 공공·보건과학대학, 041-550-1474, rhomijung@dankook.ac.kr

I. 서론

병원에서 생성되는 의료 데이터의 경우, 물리적으로 데이터를 외부로 이관할 수 없는 제약으로 인해 다기관 연구의 한계를 지닌다. 그러나 코로나19와 같은 감염병 연구에서 볼 수 있듯이 의료 분야의 연구는 한 국가에 한정하거나, 한 기관에 한정하여 진행하는 것은 유의미한 결과를 도출하고 연구의 질을 높이는데 걸림돌이 된다.

의료분야에서 다기관 연구는 매우 중요한 분야로, 이를 위해서 분산형연구네트워크(Distributed research networks, DRNs) 개발이 중요하다. 관련된 해외 연구 인프라로는 National Patient-Centered Clinical Research Network (PCORnet)[1-3], Biobanking and BioMolecular Resources Research Infrastructure-European Research Infrastructure Consortium (BBMRI-ERIC)[4-6], ELIXIR[7-9] 등이 있다. DRNs을 구축하여 활용하기 위해서는 네트워크에 참여한 기관들의 데이터가 공통데이터모델(Common Data Model, CDM)로 구축되어야 한다. CDM 사례로는 Observational Medical Outcomes Partnership (OMOP) CDM, Sentinel CDM[10, 11], PCORnet CDM[12] 등을 들 수 있다. OMOP CDM은 Systematized Nomenclature of Medicine(SNOMED), RxNorm, Logical Observation Identifier Name Code(LOINC) 등 각종 국제표준에 대한 OMOP CDM의 Concept ID를 전체 매핑하여 다기관 연구에 활용 가능한 데이터베이스 모델을 제시하고 있다. 2021년 5월 British Medical Association BMJ 저널에는 OMOP CDM 이용한 다국적 네트워크 코호트 연구를 구성하여, 코로나19 환자의 약물용도 변경 및 보조제 사용 연구를 발표하였다[13]. 한국을 비롯한 전 세계 국가에서 OMOP CDM을 약물부작용을 포함한 다양한 주제로 다기관 연구 프로젝트에 활용하고 있다[14]. 또한 미국 Food and Drug Administration(FDA)의 Sentinel은 Active Risk Identification and Analysis(ARIA) 시스템을 통해 Electronic Health Record(EHR) 데이터와 연결된 청구데이터를 기반으로 4단계 분석을 통해 약물 안전성 등을 모니터링하고 있다[15]. 다기관 연구를 하고자 하는 의료기관들은 CDM을 기반으로 하여 데이터의 민감 정보 유출에 자유로우며, 분석의 유연성과 투명성, 재현성 등이 가능하다. 많은 국내·외 의료기관들은 기관별 상이한 데이터 특징을 상쇄하고 다기관 연구를 위해서 노

력하고 있으며, CDM을 구축하여 다양한 연구 활동을 지원하고 있다.

본 연구는 CDM 기반 약물부작용 감시를 위한 단국대학교병원 임상데이터웨어하우스(Clinical Data Warehouse, CDW)를 구축하고, 필요한 주요 사항을 도출하였다.

II. 이론적 배경

1. 공통데이터모델(Common Data Model, CDM)

Europe-Adverse Drug Reaction(EU-ADR), Sentinel, OMOP 등은 약물 및 백신 안전 영역에서 능동적 감시 연구 수행을 위하여 여러 의료 데이터베이스(Database, DB)를 결합하여 사용하는 것을 시작했다[16]. 각 지역의 거버넌스 규칙과 국가별 관련 법률이 환자 데이터 공유를 제한하기 때문에 가장 실용적인 협업을 위한 방법은 각 DB 사이트에서 로컬로 분석을 수행한 뒤, DB 간에 결과를 병합하는 것이었다. 또한 동일한 프로토콜을 공유하는 것뿐 아니라 분석 코드를 공유하여 어느 기관에서나 동일하게 분석이 수행되도록 해야 했다. 이외에 각 DB가 CDM을 도입하여 유사한 형식으로 데이터를 표준화하는 과정이 필요했으며, 모든 DB 네트워크는 자체 CDM을 개발하였다[16]. CDM을 사용하여 데이터 형식을 표준화하면, 서로 다른 기관의 다른 데이터를 일관된 연구 방법을 적용할 수 있고, 그에 따라 비교 가능한 결과를 효과적으로 얻을 수 있는 장점을 지닌다[17].

현재 전세계적으로 가장 많은 협업을 이루며 많은 연구가 진행되고 있는 사례는 OMOP CDM이다. 2023년 1월 기준 [*_datetime] 필드를 필수 필드로 전환한 OMOP CDM(6.0)이 있다. Sentinel CDM은 2023년 1월 현재 8.1.0 버전으로 COVID-19 항체 검사 결과 등의 항목을 추가했다[11]. PCORnet CDM은 미국 FDA 감시 플랫폼인 Sentinel CDM을 기반으로 개발되었으며, 2023년 1월 현재 PCORnet CDM 6.0 버전이다[18].

본 연구는 약물부작용 감시를 위한 단국대학교병원 CDW를 구축하기 위해 OMOP CDM을 활용하였다.

2. 분산형연구네트워크(Distributed research networks, DRNs)

CDM을 기반으로 다기관 연구를 위한 DRNs가 존재한다. PCORnet은 여러 도메인의 전자 건강 기록 데이터를 국가 규모로 통합한 최초의 DRNs 중 하나이다. DRNs의 경우, 데이터 기반 임상 연구뿐만 아니라, 거버넌스, 데이터 품질, 정책 관련 연구, 개발 방향 등 다양한 연구가 진행되고 있다[19–23]. 특히 코로나19와 관련하여 전 세계적인 다기관 연구가 필요함에 따라 더욱더 CDM 기반의 DRNs의 중요성이 강조되고 있다.

한국은 보건복지부의 DRNs와 한국의약품안전관리원의 Medical record Observation & Assessment for drug safety(MOA) CDM과 병원자료분석네트워크 등이 있다. 보건복지부 DRNs는 연구자가 보건복지부 소속 산하 기관들인 국민건강보험공단, 건강보험심사평가원, 국립암센터, 질병관리본부가 보유한 CDM을 활용하여 연구를 수행할 수 있도록 지원한다. 보건복지부 DRNs는 2010.01.01.~2020.12.31일 사이의 약 101억건 명세서 총 5,648 환자의 Health Insurance Review and Assessment Service(HIRA) CDM 1.0을 구축하여 운영하고 있다. 코로나19의 신속한 대응을 위해 HIRA CDM을 활용한 국제협력 연구지원이 2020년 3월부터 시작되었다[24]. 2023년 1월 기준 JAMA, Lancet Psychiatry 등에 HIRA CDM을 활용하여 다양한 연구가 보고되고 있다[25–28].

한국의약품안전관리원 MOA 병원자료 분석네트워크의 경우 2018년부터 총 27개(2022년 기준) 대형종합병원과 국립병원 및 대학병원 참여를 통해 약 3,300만명의 환자 데이터의 환자 정보 표준화를 완료하여 병원 EHR 기반 다기관 확장을 꾀하고 있다. 현재 MOA CDM을 기반으로 한 논문이 지속적으로 발간되고 있으며, MOA CDM을 활용한 다양한 연구를 기반으로 한 의약품 부작용 모니터링 및 안전관리 대응 체계를 마련하고 있다. 2019년에는 ‘EHR-CDM 기반 한국 약물감시 시스템’ 논문이 발간되었으며[29], 2021년에는 알로푸리놀 사용과 갑상선 자극 호르몬 수치 증가 위험 사이의 연관성 연구가 발표되었다[30].

III. 연구방법

1. 대상 자료

연구에 사용된 데이터는 단국대학교병원에서 추출된 2013.01.01.~2021.12.31. 기간의 데이터이다. 추출된 원자료는 단국대학교병원 전산팀의 익명화 과정을 거쳐 전달받았다. 전달받은 데이터는 OMOP CDM의 표준화된 임상 데이터 17개의 테이블 중에서 단국대학교병원 CDW 구축이 가능한 최종 테이블을 선정하였다. 단국대학교병원 EHR 데이터 중에서 실제 수집된 항목과 매칭이 되는 최종 8개 테이블을 선정하였다(그림 1).

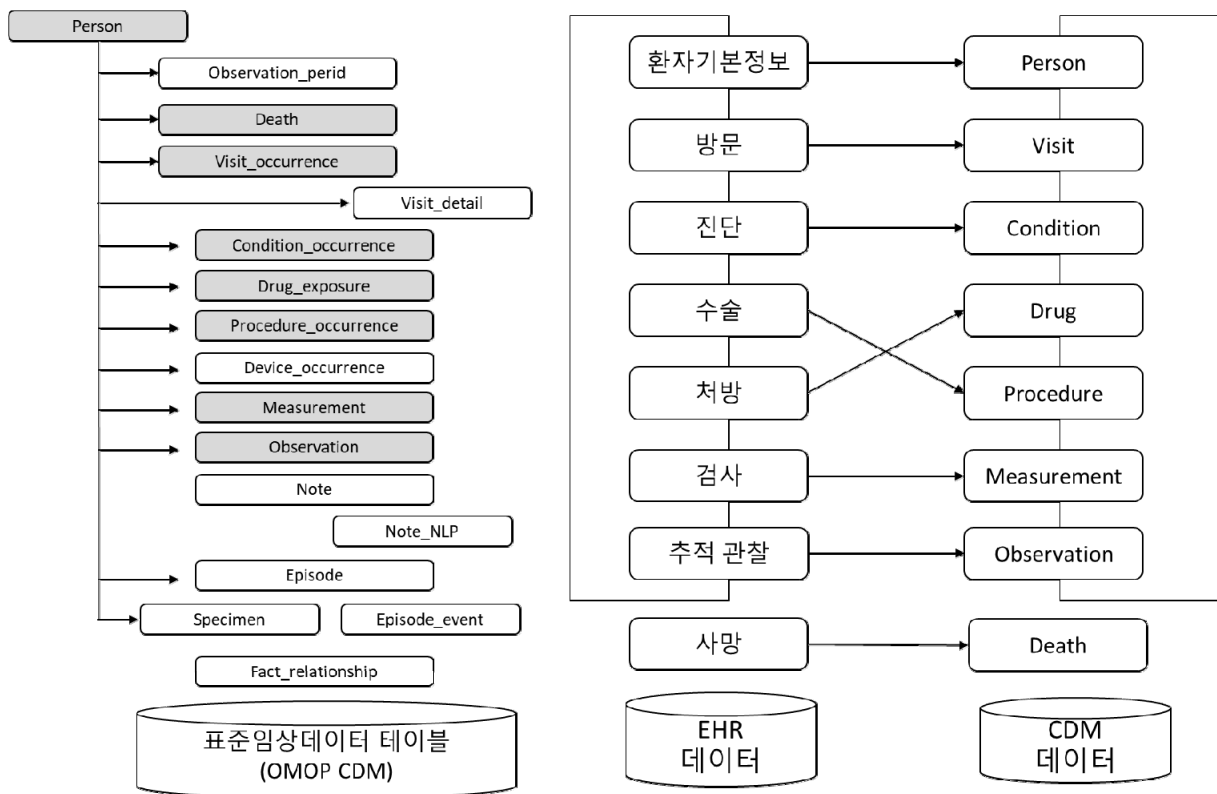
각 선정된 테이블(Person, Visit, Condition, Drug, Procedure, Measurement, Observation, and Death)에 해당하는 항목(Birthdate, Sex, Visit source, Discharge, Condition, Drug source, Dose, Unit, Route, Vaule, Cause 등)을 추출하였다. 또한 용어 매핑 시 사용된 테이블은 OMOP CDM의 표준화된 사전정보(Standardized vocabularies)을 사용하였다.

2. CDW 구축 과정

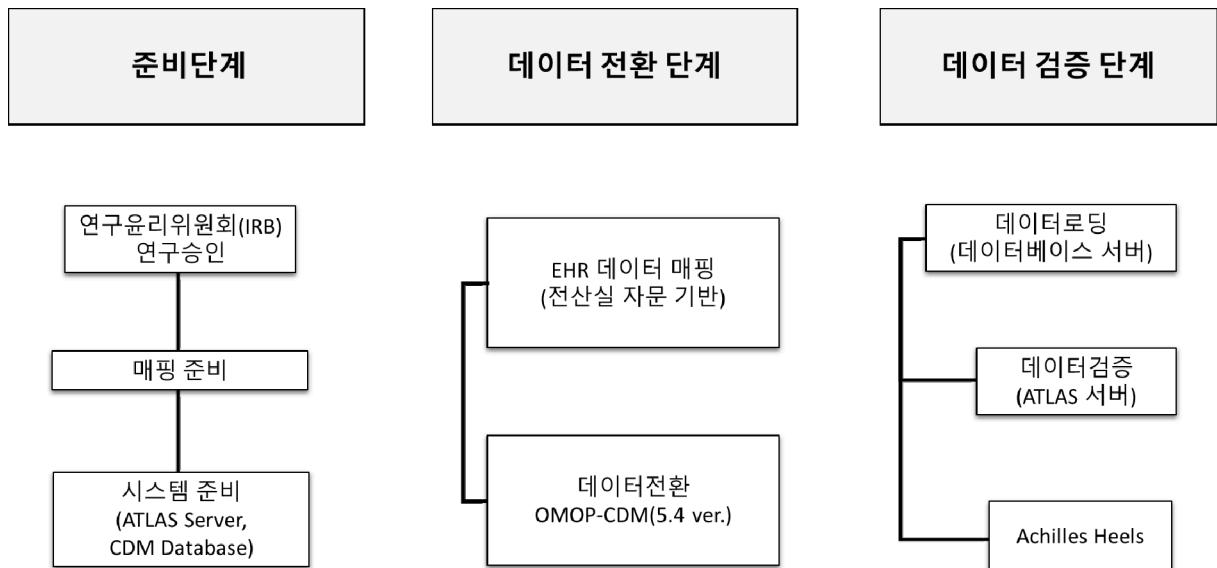
본 연구의 CDW 구축 과정은 (그림 2)와 같다.

본 연구로 구축된 CDW는 추출, 변환, 로딩(Extraction, Transformation, and Loading, ETL)과정을 거쳤다(그림 3). 해당 과정을 위한 R 언어 기반의 변환 및 로딩 소프트웨어를 개발했다. 개발된 소프트웨어는 EHR 추출 자료를 원본 데이터로 사용한다.

단국대학교 병원에서 추출된 데이터는 용어 매핑 과정을 거쳤다. 데이터 매핑은 OMOP CDM(버전5.4)을 기준으로 실시하였다. 용어 매핑은 국제표준용어체계 기준으로 진행되었다. 진단용어 매핑은 Korean Standard Classification of Disease and Cause of Death(KCD)와 SNOMED 간 매핑을 원칙으로 진행하였다. KCD의 경우 연도별 버전이 달라, 이를 고려하여 2020년까지의 데이터는 KCD-7로 진행하고, 2021년 데이터는 KCD-8로 구분하여 매핑하였다. 약물 용어 매핑은 보험 EDI 코드와 RxNorm 코드 간의 매핑을 실시하였다. 검사 용어는 보험 EDI 약품 코드와 LOINC 간 매핑을 하였다. 처치 및



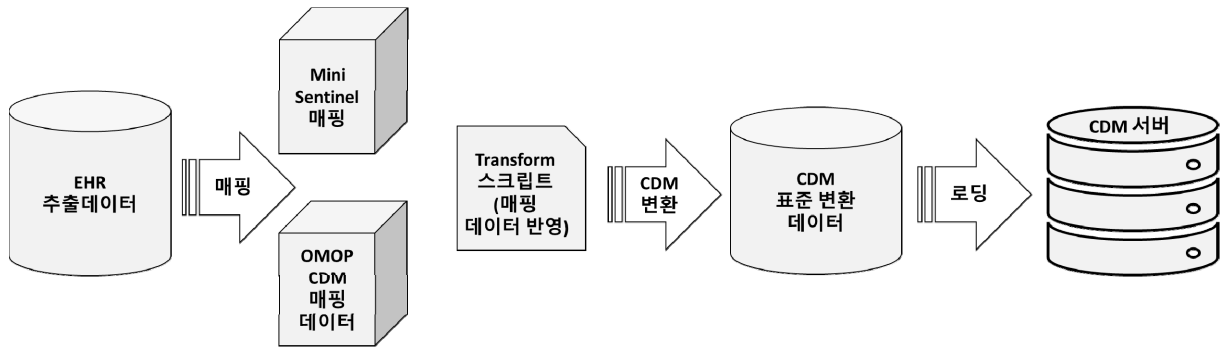
<그림 1> 단국대학교 CDW 구축을 위한 CDM 테이블



<그림 2> CDM 기반 단국대학교병원 DW 구축 과정

수술의 경우 보험 EDI 코드를 그대로 사용하였다. 매핑을 위한 가이드라인을 제작하여 매핑을 실시하였다. 사망 용어에 대한 매핑은 병원에서 제공받은 질병사인분류체계인 KCD 코드와 SNOMED-CT 간 매핑을 원칙으로 하였다.

예를 들어, 아스피린(Aspirin) 100mg의 약물 정보를 641100060으로 원본에서 추출했다면, 매핑 데이터는 641100060이라는 약물에 대해 1113143이라는 공통의 코드로 매핑된다고 저장하고 있다. 단, 여기서 사용되는



<그림 3> 추출, 변환, 로딩 과정

매핑 데이터는 국제표준 또는 사실상의 표준(Standard de facto)으로써, 약물의 경우는 의료영역에서 사용하는 RxNorm 용어 표준을 근간으로 한다. 따라서, 641100060은 RxNorm의 32929로 매핑이 먼저되고, 해당 매핑은 1113143이라는 공통 코드로 변환된다. 32929에서 1113143으로의 매핑은 이미 사실상의 표준으로 정의되어 있기 때문에, 현재 641100060이라고 쓰는 약물을 또 다른 코드로 쓰는 기관이 있다고 하더라도 똑같은 32929표준으로 매핑하게 되면, 전세계 어디에서도 같은 코드인 641100060를 표준으로 데이터를 다루고 분석할 수 있게 된다. 정리하면 각 기관이 사용하는 약물코드 정보를 RxNorm 코드로 매핑하고 해당 매핑된 정보는 똑같은 공통의 코드로 나타내게 되므로 CDM으로 변환된 데이터의 다기관 분석이 가능하다.

변환 과정에서는 OMOP CDM(버전5.4)을 활용하여 원본의 메타정보를 CDM의 메타정보로 변환했다[31]. 변환에는 두 가지 요소, 즉, 메타정보의 매핑 데이터와 원본 데이터가 필요하다. 원본 데이터의 코드는 메타정보를 통해 다른 기관에서 보는 것과 동일한 형태의 새로운 메타정보로 매핑된다. 매핑 데이터 자체는 연결정보만을 갖는 것이기 때문에 실제데이터는 CDW로 로딩되는 두 번째 단계를 거쳐야 한다<그림 3>.

매핑 작업을 위해 원본 데이터 파일의 언어 포맷을 Universal Coded Character Set+Transformation Format-8-bit(UTF-8)로 변환하는 과정을 거쳤다. 해당 변환은 리눅스 서버에서 R 스크립트를 사용하여 작업되었다. UTF-8 포맷으로 변환되어진 원본 데이터는 Comma Separated Values(CSV)포맷으로 저장되어있다. CSV원본데이터로부터 R 스크립트를 이용하여 병원에서 사용 중인 진단코드, 약품코드, 검사코드 등을 추출

한다. 추출된 검사코드는 유일한 값을 갖도록 전처리하여 매핑 정보를 생성한다. 생성된 매핑정보는 원본코드, 표준코드정보, 그리고 매핑된 결과값 코드로 이루어진다. 해당 매핑정보는 Postgresql DB로 업로드된다.

로딩 작업은 매핑 정보를 활용하여 입력된다. 예를 들어, 10명의 환자가 같은 질환으로 진단받은 경우 해당 질환 정보의 매핑 데이터가 CDW의 질환 정보를 나타내게 되므로, 10명의 환자 수에 매핑 정보를 곱한 수가 CDW의 실제로 저장된 숫자가 된다.

데이터 로딩 작업은 매핑 정보와 함께 원본 데이터를 필요로 한다. 원본 데이터는 R 스크립트를 통해 Postgresql DB로 먼저 로딩된다. 원본 데이터는 아직 표준화된 코드와 연결되지 않았기 때문에 매핑 정보를 담은 DB의 테이블 정보와 원본 데이터를 담은 DB의 테이블 정보를 조합하여 데이터를 표준 정보를 담은 DB의 테이블로 로딩하게 된다.

데이터의 조합과정은 DB 쿼리를 위한 4세대 언어인 Structured Query Language(SQL)을 사용한다. 직접적인 SQL 쿼리의 실행은 DB 접속 소프트웨어를 사용해서 가능하나, 데이터 변환 및 로딩 과정의 자동화를 위해서 R 스크립트를 통해 SQL을 실행하는 방식으로 데이터 조합과정이 진행되었다. 조합하는 SQL은 매핑 정보가 갖는 원본 데이터의 코드값과 실제데이터를 담은 원본테이블의 코드값이 일치하는 경우를 조건으로 조합되며, 조합된 결과는 매핑 정보가 갖는 표준코드를 목표 테이블에 저장하게 된다.

변환 및 로딩 작업에서 사용된 R 스크립트는 자동화가 가능하도록 작성되었으며, EHR 데이터로부터 CDM 데이터를 구성하는 마지막 단계까지의 전 과정을 모두 R 언어로 작성한 소프트웨어로 실행할 수 있다.

3. 자료 품질 및 활용성 검증

단국대학교병원 CDW으로 구축된 데이터의 활용성 검증은 안지오텐신 수용체 차단제 사용과 간손상 발생 현황을 분석하고, 병원 원자료와 구축된 단국대학교병원 CDW 데이터 간 결과를 비교하였다. 일반적으로 안지오텐신 수용체 차단제에 의한 간손상은 잘 알려져 있지 않으나, 성분별 개별 사례가 보고된 바 있다. 2021년 한국의약품안전관리원에 따르면 아질사르탄 복용 후 간손상이 발생한 환자의 피해 구제 신청 건 검토 결과, 약물 사용과 이상사례 발생간의 인과 관계가 인정되어 진료비를 지급한 사례가 존재한다. 이에 단국대학교병원 CDW으로 구축된 데이터의 활용성 검증 대상으로 안지오텐신 수용체 차단제 사용과 간손상 발생으로 한정하였으며, 이에 대해 한국의약품안전관리원의 협의하에 진행되었다.

본 연구에서 간독성은 약물로 인한 간 장애를 의미하며, 약물에 의해 간이 손상되거나(Drug-induced liver injury), 약물에 의한 간독성(Drug hepatotoxicity)으로 정의된다[32]. 본 연구에서 간독성은 KCD(7차, 8차) K71 독성간질환(Toxic liver disease)으로 한정하며, 하위코드 포함한다. 대상약물은 별첨1을 참조할 수 있다.

간독성 발생 및 미발생은 안지오텐신 수용체 차단제 첫 처방 후 6개월 동안을 기준으로 산정한다. 추출에 사용된 데이터 기간은 2018.01.01.부터 2021.06.30.까지이다. 연구대상자는 해당 기간 중 단일 성분 안지오텐신 수용체 차단제를 1회 이상 사용한 모든 환자를 대상으로 하며, 2종 이상의 안지오텐신 수용체 차단제 성분 병용 시 제외하였다. 병원 방문유형은 입원, 응급, 외래 등의 구분없이 모두 포함하였다. 분석 시 연령은 안지오텐신 수용체 차단제 첫 처방일 기준으로 계산하였다.

4. 구축 인력

전체 CDW 전체 구축 기간은 2022.03.10.~2022.11.30.로 약 9개월이었다. 데이터 추출 및 전처리 기간은 3개월이었으며, 표준용어 매핑은 4개월이 소요되었다. CDM

변환 및 자료 적재에는 1개월이 필요하였으며, 자료 품질 및 활용성 검증에 1개월이 소요되었다.

전체 9개월 동안 CDW를 구축하기 위해서 총 9(박사급 4명, 석사급 3명, 학사급 2명)명의 인력이 투입되었다. CDM에 대한 정확한 이해를 가지고 있는 의료정보 전문가 1명, 용어 매핑 전문가인 보건정보 전문가 1명, 관련 연구원 3명이 투입되었다. 의료정보 전문가의 경우 CDM 구축 과정에 대한 정확한 이해와 병원 EHR의 구조와 특징 등을 정확히 이해하고 있는 경력 11년 이상의 전문가로 구성하였다. 또한 CDW 구축 환경 세팅 및 데이터 검증을 담당한 소프트웨어 전문가 1명이 투입되었다. 소프트웨어 전문가는 DB 서버 설치 및 구축하여 매핑 된 정보와 원본 데이터의 전처리, 전처리 된 정보로부터 정보의 변환, 변환된 정보를 목표 CDM DB로 로딩하는 작업을 수행했다. 소프트웨어 담당자는 13년 경력을 보유하고 있으며, CDM 및 EHR에 대한 전문성을 가지고 있다. 용어 매핑의 경우 보건의료정보관리사 자격증을 보유한 36년 경력의 보건정보 전문가로 구성하였다. 마지막으로 단국대학교 병원 EHR 추출, 매핑 데이터 정확성 검증 및 보안 점검을 위해 전산팀 3명이 이를 수행하였다.

5. 구축 환경

CDM 기반 단국대학교병원 CDW 구축에 사용된 하드웨어 및 소프트웨어는 다음과 같다. 하드웨어의 경우 DB 서버 1대, PC 5대(분석용, 매핑용)로 구성된다. <표 1>의 하드웨어에 사용된 서버는 DB를 위한 4TB 저장소를 갖는 서버급 장비를 사용했다.

운영체제 소프트웨어는 리눅스 중 안정성을 위하여 Ubuntu 리눅스의 16.0.4 버전을 사용했다. 데이터 저장을 위한 Database Management System(DBMS)는 Postgresql 11.16 버전을 사용했다. 운영체제 및 DB 관리 소프트웨어는 모두 오픈소스를 사용했다. CDW 표준화 유지를 위한 품질관리 소프트웨어로써 오픈소스인 Atlas¹⁾와 Achilles²⁾가 설치되었다. CDW 데이터의 파이프라인 자동화를 위한 소프트웨어로 R언어 3.6.3 버전이

1) ATLAS는 CDM 형식의 표준화된 환자 데이터에 대한 분석의 설계 및 실행을 용이하게 하는 도구임. [cited 2023 July 1]. Available from <https://www.ohdsi.org/software-tools/>
2) ACHILLES는 CDM을 준수하는 데이터베이스의 특성화 및 시각화를 제공하는 소프트웨어 도구임. [cited 2023 July 3]. Available from <https://www.ohdsi.org/software-tools/>

<표 1> 단국대학교병원 CDW 구축 환경

하드웨어	소프트웨어
1. DB 서버	· DB 서버 운영체제: Ubuntu 16.0.4 LTS
- PowerEdge T140 타워타입	· DB 서버: Postgresql 11.16
- CPU: 인텔 제온 E-2124 3.3GHz, 8M 캐시, 4C/4T(71W)	· DB 서버 설치 소프트웨어: - Atlas / Achilles / R 3.6.3
- 메모리: 8GB 2666MT/s DDR4 ECC UDIMM	· 분석용 PC 설치 소프트웨어: - 윈도우즈 10
- 저장소: 4TB (2TB hdd x 2ea)	- IBM SPSS Statistics 26
2. 컴퓨터 총 5대	- R, RStudio
- 분석용 데스크탑 PC 2대	
- 매펀용 전용 데스크탑 PC 3대	

설치되었다. 데이터 분석을 위한 IBM SPSS Statistics (버전26)과 오픈소스인 R 언어 및 RStudio를 설치하였다.

IV. 연구결과

6. 연구윤리

본 연구는 단국대학교 IRB 승인하에 이루어졌다 (IRB00U -2022-06-008). 모든 데이터는 익명 처리되어 개인정보를 포함하고 있지 않다.

본 연구로 구축된 단국대학교병원 CDW는 총 670,933 건(2013.01.01.~2021.12.31.)의 데이터로 구성된다. 구축된 CDW에는 방문, 진단, 처방, 처치, 검사, 그리고 사망 정보에 대한 데이터를 포함하고 있다

1. 단국대학교병원 원자료 분석 결과

안지오텐신 수용체 차단제 성분별 처방 현황에 대해 단

<표 2> 단국대학교병원 안지오텐신 수용체 차단제 성분별 처방 현황

구분	2018~2121.06		2018		2019		2020		2021(~6월)	
	N	%	N	%	N	%	N	%	N	%
환자수(전체)	18,048	100	10,223	100	10,396	100	10657	100	9618	100
아질사르탄	217	1.20	47	0.46	106	1.02	113	1.06	113	1.17
칸데사르탄	1,802	9.98	1,145	11.20	1,148	11.04	1,096	10.28	977	10.16
에프로사르탄	593	3.29	473	4.63	474	4.56	447	4.19	401	4.17
피마사르탄	978	5.42	664	6.50	685	6.59	684	6.42	649	6.75
이르베사르탄	1,614	8.94	1,265	12.37	1,225	11.78	1,189	11.16	1,125	11.70
로사르탄	2,909	16.12	1,534	15.01	1,558	14.99	1,586	14.88	1,455	15.13
올메사르탄	667	3.70	472	4.62	455	4.38	446	4.19	410	4.26
텔미사르탄	4,144	22.96	2,053	20.08	2,061	19.82	2,267	21.27	1,923	19.99
발사르탄	5,124	28.39	2,570	25.14	2,684	25.82	2,829	26.55	2,565	26.67
처방건수(전체)	177,247	100	50,864	100	49500	100	50,374	100	26,509	100
아질사르탄	1,459	0.82	84	0.17	329	0.66	522	1.04	524	1.98
칸데사르탄	19,091	10.77	5,938	11.67	5,748	11.61	5,102	10.13	2,303	8.69
에프로사르탄	5,523	3.12	1,566	3.08	1,735	3.51	1,499	2.98	723	2.73
피마사르탄	5,413	3.05	1,521	2.99	1,562	3.16	1,494	2.97	836	3.15
이르베사르탄	17,712	9.99	5,168	10.16	5,008	10.12	4,890	9.71	2,646	9.98
로사르탄	27,040	15.26	7,616	14.97	7,564	15.28	7,718	15.32	4,142	15.62
올메사르탄	5,204	2.94	1,580	3.11	1,500	3.03	1,438	2.85	686	2.59
텔미사르탄	45,545	25.70	13,075	25.71	12,212	24.67	13,255	26.31	7,003	26.42
발사르탄	50,260	28.36	14,316	28.15	13,842	27.96	14,456	28.70	7,646	28.84

국대학교병원 원자료 분석을 수행한 결과는 다음과 같다 <표 2>.

원자료 분석에 사용된 환자수는 18,048명이다(2018.01.01.~2021.06.30.). 성분에 따른 처방환자수는 발사르탄 5,124명(28.39%), 텔미사르탄 4,144명(22.96%), 로사르탄 2,909명(16.12%), 칸데사르탄 1,802명(9.98%) 순이었다. 4개의 안지오텐신 수용체 차단제 성분이 전체의 77.45%를 차지했다. 처방건수도 발사르탄 50,260건(28.36%), 텔미사르탄 45,545건(25.70%), 로사르탄 27,040건(15.26%), 칸데사르탄 19,091건(10.77%) 순으로 많았다. 전체 처방건수인 177,247건 중 상위 4개의 성분이 80.08%를 차지하였다.

단국대학교병원 원자료 내 간독성 발생 현황을 살펴보면, 간독성 발생은 0건이었다<표 3>.

2. 구축된 단국대학교병원 CDW 자료 분석 결과

안지오텐신 수용체 차단제 성분별 처방 현황에 대해 구축된 단국대학교병원 CDW 데이터를 분석한 결과는 다음과 같다<표 4>.

분석에 사용된 환자수는 18,111명이다(2018.01.01.~2021.06.30.). 성분에 따른 처방환자수는 발사르탄 5,144명(28.40%), 텔미사르탄 4,375명(24.16%), 로사르탄 2,672명(14.75%), 칸데사르탄 1,828명(10.09%) 순이었다. 4개의 안지오텐신 수용체 차단제 성분이 전체의 77.41%를 차지했다. 처방건수도 발사르탄 54,071건(29.14%), 텔미사르탄 49,821건(26.85%), 로사르탄

25,310건(13.64%), 칸데사르탄 20,206건(10.89%)순으로 많았다. 전체 처방건수인 185,583건 중 상위 4개의 성분이 80.51%를 차지하였다.

구축된 단국대학교병원 CDW 내 간독성 발생 현황을 살펴보면 간독성 발생은 0건으로 나타났다<표 5>.

VI. 고찰 및 결론

본 연구는 CDM 기반 약물 부작용 감시를 위한 단국대학교병원 CDW를 구축하였고, 병원 CDW 구축에 따른 주요 결과와 의미는 다음과 같다.

CDM 기반 다기관 연구를 위한 CDW를 구축하기 위해서는 크게 의료정보 및 보건학 분야의 전문가를 포함하고 있어야 한다.

먼저, CDW를 구축하기 위한 의료정보 전문가의 역할이 매우 중요하다. 의료정보 전문가는 병원의 특징, CDM의 특징과 구축과정에 대한 정확한 이해와 관련 전문성을 가지고 있어야 한다. CDM은 OMOP CDM, Sentinel CDM, PCORnet CDM 등 다양한 종류가 있으며, DRNs를 구성하는 기관에 따라 추구하는 방향이 모두 다르기 때문에, 병원 규모와 목적에 맞는 CDM 구축하고, 다기관 대규모 연구가 가능하도록 준비하여야 한다. 또한 최근에는 다른 CDM 간 데이터 변환을 용이하게 하고, 데이터 일관성을 평가하기 위한 연구들이 진행되고 있다[33]. 따라서 CDW를 구축하려는 기관의 특성과 CDM을 정확히 이해하는 의료정보 분야 전문가가 CDM 구축에 참여하

<표 3> 단국대학교병원 간독성 발생 현황

성분명	안지오텐신 수용체 차단제 첫 처방 후 6개월 이내	
	간독성 발생	간독성 미발생
전체	0	18,048
아질사르탄	0	217
칸데사르탄	0	1,802
에프로사르탄	0	593
피마사르탄	0	978
이르베사르탄	0	1,614
로사르탄	0	2,909
올메사르탄	0	667
텔미사르탄	0	4,144
발사르탄	0	5,124

<표 4> 구축된 단국대학교병원 CDW 내 안지오텐신 수용체 차단제 성분별 처방 현황

구분	2018~2121.06		2018		2019		2020		2021(~6월)	
	N	%	N	%	N	%	N	%	N	%
환자수(전체)	18,111	100	10,250	100	10,529	100	10,787	100	9,788	100
아질사르탄	209	1.15	48	0.47	100	0.95	108	1.00	107	1.09
칸데사르탄	1,828	10.09	1,159	11.31	1,177	11.18	1,127	10.45	1,005	10.27
에프로사르탄	593	3.27	473	4.61	476	4.52	449	4.16	403	4.12
피마사르탄	977	5.39	663	6.47	688	6.53	684	6.34	649	6.63
이르베사르탄	1,660	9.17	1,299	12.67	1,266	12.02	1,220	11.31	1,158	11.83
로사르탄	2,672	14.75	1,373	13.40	1,422	13.51	1,435	13.30	1,340	13.69
올메사르탄	653	3.61	458	4.47	450	4.27	441	4.09	406	4.15
텔미사르탄	4,375	24.16	2,178	21.25	2,237	21.25	2,461	22.81	2,120	21.66
발사르탄	5,144	28.40	2,599	25.36	2,713	25.77	2,862	26.53	2,600	26.56
처방건수(전체)	185,583	100	52,981	100	51,871	100	52,879	100	27,852	100
아질사르탄	1,409	0.76	85	0.16	321	0.62	473	0.89	530	1.90
칸데사르탄	20,206	10.89	6,256	11.81	6,062	11.69	5,443	10.29	2,445	8.78
에프로사르탄	5,572	3.00	1,574	2.97	1,759	3.39	1,510	2.86	729	2.62
피마사르탄	5,463	2.94	1,531	2.89	1,580	3.05	1,511	2.86	841	3.02
이르베사르탄	18,563	10.00	5,452	10.29	5,258	10.14	5,099	9.64	2,754	9.89
로사르탄	25,310	13.64	7,054	13.31	7,041	13.57	7,211	13.64	4,004	14.38
올메사르탄	5,168	2.78	1,551	2.93	1,501	2.89	1,433	2.71	683	2.45
텔미사르탄	49,821	26.85	14,146	26.70	13,344	25.73	14,661	27.73	7,670	27.54
발사르탄	54,071	29.14	15,332	28.94	15,005	28.93	15,538	29.38	8,196	29.43

<표 5> 구축된 단국대학교병원 CDW 내 간독성 발생 현황

성분명	안지오텐신 수용체 차단제 첫 처방 후 6개월 이내	
	간독성 발생	간독성 미발생
전체	0	18,111
아질사르탄	0	209
칸데사르탄	0	1,828
에프로사르탄	0	593
피마사르탄	0	977
이르베사르탄	0	1,660
로사르탄	0	2,672
올메사르탄	0	653
텔미사르탄	0	4,375
발사르탄	0	5,144

고, 향후 운영에 따른 고도화 과정까지 전체적으로 참여를 이끌어야 한다.

다음으로, 다기관 용어를 현재 버전으로 매핑하고 이를 검증할 수 있는 의료 용어 분야의 전문가가 필요하다. 특히 의료기관 간의 용어는 매우 다양하여 다기관 의료 데이터를 이용하여 연구하는 것은 매우 어려운 일이다. CDM이 이러한 이슈를 해결할 수 있지만, 기관에 맞는 매

핑은 전문인력과 매핑의 정확성과 버전 업데이트 등의 통일성이 요구된다. 또한 데이터 추출, 변환, 로딩 뿐만 아니라 DB를 세팅할 수 있는 DB 전문가가 필요하다. 마지막으로 구축된 DW의 데이터를 활용하기 위해서는 정확한 분석 코딩을 작성하고 활용할 수 있는 데이터 분석 전문가도 포함되어야 한다. 모든 인력을 갖춘 대형병원의 경우 인력 확보가 가능하나, 그렇지 않은 경우에는 한 인

력이 여러 역할 수행이 불가피하다. 기관 규모에 맞는 인력 투입이 되는 것이 현실이나, 적절한 인력배치와 역할을 통해 정확한 CDW가 구축되도록 해야 할 것이다.

본 연구는 단국대학교병원 EHR 중 2013.01.01.~2021.12.31 기간의 환자 데이터를 표준화하고 총 670,933건 데이터를 기반으로 CDW를 구축 및 검증하였다. 안지오텐신 수용체 차단제 첫 처방 후 6개월 이내 간독성 발생 건수의 경우, 단국대학교병원 CDW 자료의 경우 발생 건수와 원자료 분석 결과 모두 0건으로 나타났다.

간독성 미발생 건수에서 주요 성분명 상위 분포의 경우, 상위 분포는 발사르탄, 텔미사르탄, 로사르탄, 칸데사르탄 순으로 나타났다. 발사르탄, 텔미사르탄, 로사르탄, 칸데사르탄 순의 상위 안지오텐신 수용체 차단제 성분의 분포가 병원 원자료의 경우 77.45%, 구축된 단국대학교병원 CDW의 경우 77.41%로 확인되었다. 단국대학교병원 CDW의 간독성 미발생 건수는 18,111건, 병원 원자료의 경우 18,048건으로 병원원자료 전체 건수가 구축된 단국대학교병원 CDW 건수보다 63건 적은 것이 확인되었다. 단국대학교병원 CDW 데이터와 병원 원자료 분석을 통해 구축된 데이터 검증을 실시한 결과, 구축된 CDW를 연구용으로 사용하는데 문제가 없는 것으로 나타났다.

한국의약품안전원의 경우 2018년부터 2022년까지 총 27개 병원의 연구네트워크를 이루어 데이터를 표준화하고 이를 활용한 다양한 연구를 진행하고 있다. 이외에도 국내외로 CDM을 기반으로 한 다양한 연구들이 배출되고 있으며, 한국의 경우 오딧세이 컨소시엄 등의 연구사업이 진행된 바 있다[34]. 이처럼 CDM 기반 CDW는 의료 분야의 약물감시 뿐 아니라 정신건강의학, 감염병까지 다양한 연구의 근간이 되고 있다[35, 36]. 이번 연구로 개발된 CDM 기반 단국대학교병원 CDW도 다기관 연구를 위한 여러 기관과 연구진들의 협조가 필요하다. 또한 정확하고 다양한 형태의 데이터 확보 및 꾸준한 데이터 업데이트 등이 반드시 요구되어 진다.

의료기관에서 구축된 CDM 기반 CDW 활용을 위해서 DRNs와 같은 연구자가 데이터 접근 및 활용 가능한 플랫폼이 필요하다. DRNs의 경우, 데이터 기반 임상 연구뿐만 아니라, 거버넌스, 데이터 품질, 정책 관련 연구, 보안 기술, DRNs 구축 오픈소스 및 개발 방향 등 다양한 연구가 진행되고 있다[37-41]. 기존에 CDM 기반의 CDW 구축이 이루어진 대형병원의 경우 활용할 수 있는 플랫폼

구축이 구축되었거나, 구축이 진행되고 있다. 이처럼 CDM 기반 CDW 구축되더라도 내부 연구진이 이를 활용하기 위해서는 다양한 내부 활용 촉진 방안이 필요하다. 또한 구축 기관들의 협업과 다기관 연구를 지원할 수 있는 촉진 요인이 필요할 것이다. 또한 CDM 기반 CDW가 여러 기관이 함께 공유할 수 있는 기관 간 협력이 절실히 필요하다.

마지막으로 병원 EHR 기반 CDW 구축에 있어 가장 많은 시간을 소요해야 하는 부분은 데이터 표준과 매핑 및 검증 과정이다. 상대적으로 병원 CDW 구축 과정 중 매핑 관련 부분이 정확성 등을 요구하기 때문에 많은 시간과 전문성이 요구되어 진다. 타 기관들의 구축 수준과 현재 용어 버전 등과 그 특징을 정확히 반영하여 데이터 매핑을 수행해야 할 것이다. 이를 인공지능과 같은 기술을 이용하여 매핑을 자동화하기 위한 연구들이 나오고 있으나 [38], 수작업이 필요한 부분이 존재하기 때문에 이에 대한 충분한 시간과 노력이 필요하다.

본 연구는 다음과 같은 한계를 지니고 있다. 첫 번째, CDM 기반 다기관 연구를 위한 단국대학교병원 CDW 구축에 관한 연구로, 향후 이에 대한 운영에 대한 부분은 본 논문에서 다루고 있지 않다. 병원의 DW 구축 이후 얼마나 잘 활용될 수 있는지는 매우 중요한 이슈이며, 현실적으로 활발한 활용이 제한적이다. 향후 효율적으로 운영하는데 필요한 연구가 발표되는 것이 필요할 것이다. 두 번째, 이 연구는 텍스트 기반 단국대학교병원 EHR 데이터를 기반으로 CDW를 구축하였다. 현재 CDW에는 이미지 등의 데이터를 포함하고 있지 않아 이를 확장 할 필요가 있다. 세 번째, 본 연구를 통해 구축된 단국대학교 CDW는 단국대학병원의 EHR 데이터를 CDW로 구축하였다. 그러나 암 환자의 시계열 데이터를 활용한 연구를 위해서는 보완이 필요하다. 암 환자의 치료 및 예후 관련 연구를 위해서 CDM을 시계열 기반 에피소드 도메인 데이터로의 업그레이드가 필요하다. OMOP CDM의 경우 암환자 연구를 위해 2020년에 계층 어휘를 확장한 바 있다[39]. 암 연구를 위해서는 관련 어휘 확장 및 에피소드 기반의 암환자 도메인 데이터와 시계열 데이터 기반 연구를 위한 변환이 필요하다. 향후 암 연구를 위한 이러한 작업을 추가하는 것이 필요하다. 네 번째, 본 연구는 CDW로 구축된 데이터의 활용성 검증을 위해 안지오텐신 수용체 차단제 사용과 간손상 발생에 대해서만 진행하였다. 연구비

와 연구 기간의 제한이 있어, 다수의 약물과 약물 부작용에 대해 다루지 못하였다. 향후 연구는 다수의 약물과 약물 부작용에 대해서 다루는 것이 바람직 할 것이다. 다섯 번째, 본 연구는 약물부작용을 위한 CDM CDW에 대하여 참고할 만한 국내외 선행연구가 많지 않아, 이를 결론에 충분히 제시하지 못하였다. 추후 관련한 연구가 이어지길 바란다. 마지막으로 구축 CDM이 잘 구축되었는지를 병원 원자료와 살펴볼 수 있는 통계학적 접근은 아직까지 부족하여, 추후 관련 연구가 진행된다면 구축된 CDW의 정확도 등을 확인할 수 있을 것이다.

병원의 규모나 상황에 따라 병원이 EHR 데이터를 표준화하고, CDM 기반 CDW를 구축하는 것은 비용과 인력 면에서 의료 기관에게 쉬운 일이 아니며, 반드시 구축해야 하는 사항이 아닐 수 있다. 그러나 다양한 CDM을 기반으로 임상 의사결정지원시스템을 구축하고자 하는 노력들이 지속적으로 이루어지고 있다[40, 41]. 또한 이러한 CDM은 의료기관 간 상이한 데이터를 활용하여 의미 있는 연구를 가능하게 하는 효율적인 방법을 제공함으로써 다기관 연구가 가능하게 한다[17]. 이에 본 연구는 CDM 기반 약물부작용 감시를 위한 단국대학교병원 CDW를 구축하고, CDW 구축에 필요한 주요 사항을 도출하여 향후 CDW를 구축하고자 하는 기관이 소속 기관에 맞는 CDW를 구축하는데 유용한 참고 자료가 될 수 있을 것이다.

Reference

- [1] The National Patient-Centered Clinical Research Network [cited 2023 July 31]. Available from <https://pcor-net.org/>
- [2] Fleurence R, Curtis L, Califf R, Platt R, Selby J, Brown J, Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc* 2014; 21(4) 578-582.
- [3] McTigue K, Wellman R, Nauman E, Anau J, Coley R, Odor A, Comparing the 5-year diabetes outcomes of sleeve gastrectomy and gastric bypass: the National Patient-Centered Clinical Research Network (PCORNet) Bariatric Study. *JAMA surg* 2020; 155(5): e200087-e200087.
- [4] The Biobanking and BioMolecular Resources Research Infrastructure [cited 2023 July 31]. Available from <https://www.bbmri-eric.eu/>
- [5] Litton J. Launch of an infrastructure for health research: BBMRI-ERIC. *Biopreserv Biobank* 2018; 16(3) : 233-241.
- [6] Vuorio E. Networking biobanks throughout Europe: the development of BBMRI-ERIC. *Biobanking of Human Biospecimens*. Springer, Cham, 2017. pp. 137-153.
- [7] The European infrastructure for life science data [cited 2023 July 31]. Available from <https://elixir-europe.org/>
- [8] Harrow J, Drysdale R, Smith A, Repo S, Lanfear J, Blomberg N. ELIXIR: providing a sustainable infrastructure for life science data at European Scale. *Bioinformatics* 2021; 37(16): 2506-2511.
- [9] Crosswell LC, Thornton JM. ELIXIR: a distributed infrastructure for European biological data. *Trends Biotechnol* 2012; 30(5): 241-2.
- [10] Curtis LH, Weiner MG, Boudreau DM, Cooper WO, Daniel GW, Nair VP. et al. Design considerations, architecture, and use of the MiniSentinel distributed data system. *Pharmacoepidemiol Drug Saf* 2012; 21: 23-31.
- [11] Sentinel Common Data Model, [cited 2023 January 22]. Available from <https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model>
- [12] PCORnet Common Data Model, [cited 2023 January 22]. Available from <https://pcor-net.org/data/>
- [13] Prats-Urbe A, Sena AG, Lai LYH, Alghoul H, Alser O, Alshammari TM, et al. Use of repurposed and adjuvant drugs in hospital patients with covid-19: multinational network cohort study. *Bmj* 2021; 373.
- [14] Reisinger SJ, Ryan PB, O'Hara DJ, Powell GE, Painter JL, Pattishall EN, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. *J Am Med Inform Assoc* 2010; 17(6): 652-662.

- [15] Drug Studies [cited 2023 January 22]. Available from <https://www.sentinelinitiative.org/assessments/drugs>
- [16] Trifirò G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how?. *J Intern Med* 2014; 275(6): 551–561.
- [17] Reisinger S, Powell G, Dreyfus B, Schneider G. Use of common data model to meaningfully compare treatment patterns for depression among disparate databases. *Value Health* 2013; 16(3): A24.
- [18] PCORnet CDM 6.0 [cited 2023 January 22]. Available from https://pcornet.org/wp-content/uploads/2022/01/PCORnet-Common-Data-Model-v60-2020_10_221.pdf
- [19] Juárez D, Schmidt EE, Stahl-Toyota S, Ückert F, Lablans M. A generic method and implementation to evaluate and improve data quality in distributed research networks. *Methods Inf Med* 2019; 58(02/03): 086–093.
- [20] Qualls LG, Phillips TA, Hammill BG, Topping J, Louzao DM, Brown JS, et al. Evaluating foundational data quality in the national patient-centered clinical research network (PCORnet®). *Egems* 2018;6(1).
- [21] Marsolo KA, Brown JS, Hernandez AF, Hammill BG, Raman SR, Syat B, et al. Considerations for using distributed research networks to conduct aspects of randomized trials. *Contemp. Clin. Trials Commun* 2020; 17: 100515.
- [22] Panozzo C. Distributed research networks and applications to pragmatic randomized trials. *Pragmatic Randomized Clinical Trials*. Academic Press, 2021. pp. 237–256.
- [23] Oh S, Sung M, Rhee Y, Hong N, Park YR. Evaluation of the Privacy Risks of Personal Health Identifiers and Quasi-Identifiers in a Distributed Research Network: Development and Validation Study. *JMIR Med Inform* 2021;9(5): e24940.
- [24] 조도연. HIRA ISSUE: 코로나 19의 세계적 유행에 대응하는 국제협력 연구플랫폼. 건강보험심사평가원 심사평가연구소: .14 2020.
- [25] You S, Rho Y, Bikdeli B, Kim J, Siapos A, Weaver J, et al. Association of ticagrelor vs clopidogrel with net adverse clinical events in patients with acute coronary syndrome undergoing percutaneous coronary intervention. *Jama* 2020; 324(16): 1640–1650.
- [26] Williams R, Markus A, Yang C, Salles T, Falconer T, Jonnagaddala J, et al. Seek COVER: Development and validation of a personalized risk calculator for COVID-19 outcomes in an international network. *MedRxiv* 2020: 05.
- [27] Kim H, Yoon J, Kwak M, Cha J. Gastrointestinal and nongastrointestinal complications of esophagogastroduodenoscopy and colonoscopy in the real world: A nationwide standard cohort using the common data model database. *Gut Liver* 2021; 15(4): 569.
- [28] Rho Y, Son Y, Lee Y, Kim J, Lee H, You S, et al. COVID-19 international collaborative research by the health insurance review and assessment service using its nationwide real-world data: database, outcomes, and implications. *J Prev Med Public Health* 2021; 54(1): 8.
- [29] Son N, Kim B, Chung S, Han S. Korean Pharmacovigilance System Based on EHR-CDM. *MedInfo* 2019; 264:1592–3.
- [30] Choi W, Yang YS, Chang DJ, Chung YW, Kim H, Ko SJ, et al. Association between the use of allopurinol and risk of increased thyroid-stimulating hormone level. *Sci Rep* 2021;11:1–10.
- [31] OMOP CDM 5.4 [cited 2023 January 22]. Available from <http://ohdsi.github.io/Common DataModel/>
- [32] Núñez M. Clinical syndromes and consequences of antiretroviral-related hepatotoxicity. *Hepatology* 2010; 52(3): 1143–1155.
- [33] Yu Y, Zong N, Wen A, Liu S, Stone DJ, Knaack D, et al. Developing an ETL tool for converting the PCORnet CDM into the OMOP CDM to facili-

- tate the COVID-19 data integration. *J Biomed Inform* 2022; 127: 104002.
- [34] Park R. The Distributed Research Network, Observational Health Data Sciences and Informatics, and the South Korean Research Network. *Korean J Med* 2019; 94(4) : 309-314.
- [35] Byeon SJ, Park HA, Kang DY, Choi IY, Choi JH. Surveillance of Adverse Drug Reactions to First-line Antituberculosis Drugs Using Common Data Model. *Journal of Pharmacoepidemiology and Risk Management* 2021; 13(2): 78-86.
- [36] 하재호, 이은영, 이동윤, 조용혁, 이혜림, 박범희, 외. 우울증 환자에서 항우울제 단독요법과 항정신병제와 항우울제의 병용요법의 자살위험비교: 정신건강의학 공통 데이터 모델을 활용한 파일럿 연구. *J Korean Neuropsychiatr Assoc* 2020; 59(3) : 243-249.
- [37] Juárez D, Schmidt EE, Stahl-Toyota S, Ückert F, Lablans M. A generic method and implementation to evaluate and improve data quality in distributed research networks. *Methods Inf Med* 2019; 58(02/03): 086-093.
- [38] Kang B, Yoon J, Kim HY, Jo SJ, Lee Y, Kam HJ. Deep-learning-based automated terminology mapping in OMOP-CDM. *J Am Med Inform Assoc* 2021; 28(7): 1489-1496.
- [39] Ahmadi N, Peng Y, Wolfien M, Zoch M, Sedlmayr M. OMOP CDM Can Facilitate Data-Driven Studies for Cancer Prediction: A Systematic Review. *Int J Mol Sci* 2022; 23(19): 11834.
- [40] 안윤애, 조한진. 공통데이터모델 기반의 임상 의사결정 지원시스템에 관한 연구. *한국융합학회논문지* 2019; 10(11): 117-124.
- [41] Rho MJ, Kim SR, Park SH, Jang KS, Park BJ, Hong JY, et al. Common data model for decision support system of adverse drug reaction to extract knowledge from multi-center database. *Inf. Technol. Manag* 2016; 17(1): 57-66.

<별첨 1> 안지오텐신 수용체 차단제 ATC 코드

성분명	ATC
azilsartan	C09CA09, C09DA09
candesartan	C09CA06, C09DB07, C09DA06, C09DX06
eprosartan	C09CA02, C09DA02
fimasartan	C09CA10, C09DB09, C09DA10, C10BX16
irbesartan	C09CA04, C09DB05, C09DA04, C09DX07
losartan	C09CA01, C09DB06, C09DA01
olmesartan	C09CA08, C09DB02, C09DA08, C09DX03
telmisartan	C09CA07, C09DB04, C09DA07
valsartan	C10BX10, C09CA03, C09DX02, C09DB01, C09DA03, C09DB08, C09DX05, C09DX04, C09DX01