공통데이터모델(CDM) 분석 이용 가이드

<건강보험심사평가원 빅데이터전략부>

1. 이용절차

	주요절차	주체	내용	일정
1	**연구과제 선정**	심사평가원	▪ 선정 결과 통보 *E-mail 개별 통보 ▪ 분석코드 제출 요청	**8.17.(수)~**
▼				
2	**분석 신청**	신청자	▪ 분석코드는 한 개의 zip파일로 제출	**상시**
▼				
3	**분석코드 수정**	심사평가원 ↔ 신청자	▪ 실행오류 발생 시 분석코드 수정 제출 - 필요시 직접 방문 가능하되, 방문일자는 반드시 사전협의	**요청 후 5일 이내**
▼				
4	**결과값 제공**	심사평가원	▪ 결과(통계)값 반출 ※ 연구과제 선정 결과에 따라 순차적으로 제공하되, 분석코드 수정 등의 상황에 따라 반출 순서는 달라질 수 있음	**상시**
▼				
5	**연구 종료**	신청자	▪ 이용 종료 확인서 제출 ▪ 결과물(연구 산출물)에 출처 표기 ▪ 결과 활용 통보서 제출	**상시**

* 일정은 변경·조정될 수 있음

2. 세부절차

○ **(연구과제 선정)** 접수된 총 42개 과제 대상 선정결과 통보

○ **(분석 신청)** 연구자의 분석환경에서 작성한 R코드를 E-mail로 제출

- 예상 결과값과 예상 결과 데이터 구성형태 간략한 설명(ex. 연령별 질병 발생률 테이블 및 도표) 또는 분석자 환경에서 도출한 결과값 캡처

* 자료 제출 후 추가 보완자료를 요청할 수 있음

○ **(분석코드 수정)** 실행오류 발생 시 분석코드 수정 요청 후 5일 이내에 수정하여 제출해야 하며, 제출기한을 넘길 경우 후순위로 제공되거나 제공 불가할 수도 있음

○ **(결과값 제공)** 결과값은 한 번 반출 가능하며 여러 번 제공 지원 불가

- **반출대상:** 쿼리, 그림, 그래프, 통계분석표, 집계표 등 분석 결과값
- **파일형식:** CSV, TXT, R, PDF, 이미지 파일(JPEG, GIF, TIFF, PNG, BMP)

※ sas7dbat, html, css, rmd, srx, Atlas를 통한 View 캡처, 기타 형식 반출 불가

- **반출기준:**

① 통계분석표나 집계표 반출 시 변수명에 대한 설명 기재 필수

② 수치형변수(연령, 내원일수, 총사용량 등)를 기준으로 집계 시 그룹화 필수

③ 식별가능정보(연령, 지역, 종별 등)를 조합하여 집계 시 적절한 그룹화 필요

- **반출불가 기준:**

① 식별자(수진자/명세서/요양기관)가 포함된 경우

② 수진자, 명세서, 요양기관별 통계

③ 의약품성분코드, 치료재료코드 단위 통계

④ 그룹별 집계표 결과값이 2 이하인 건이 30%를 초과할 경우

⑤ 줄단위 데이터, 중간 산출물 등 재가공이 가능한 데이터 형태 자료

⑥ 그 밖에 개인 식별 우려가 높다고 판단되는 경우

반출 검토 기본 원칙

① **(연구·반출 목적 부합성)** 연구목적과 반출자료 활용 목적이 부합하는지 검토
② **(가명처리의 적정성)** 반출 자료가 k-3 익명성을 만족하는지 여부
③ **(정보침해 가능성)** 법인·단체의 경영·영업상 정보의 침해가 있는지 여부
④ **(재식별 가능성)** 추가 정보 또는 다른 정보와 결합하여 재식별 가능성이 있는지 여부

○ **(연구 종료)** 연구 종료 시 [붙임1] 이용 종료 확인서를 E-mail로 제출

- 우수 연구자*는 향후 심사평가원 CDM 개방(2단계) 이용 신청 시 우선 순위 제공 및 수수료 면제

  * 연구 활용물 3건 이상, 국제학술대회 발표 1건 이상 등

- 미제출자는 향후 심사평가원 CDM 이용 제한 등 패널티 부여

3. 표준용어 사전

○ 심평원 CDM 매핑용어사전 및 **Athena**(https://athena.ohdsi.org) 참고

○ **(코로나19 정보)** 확정 진단 여부와 확정 진단 일자 제공, 관련 매핑 코드는 아래와 같음

- Observation_concept_id: '704996'
- Condition_concept_id: '37311061', '439676', '4100065'

4. 분석코드 작성 및 제출 가이드

○ **(분석환경)** CDM 공통도커를 통해 R Session 생성 후 해당 세션에서 프로젝트 빌드 및 설치를 독립적으로 지원하며 과제별로 라이브러리 추가 가능

※ R 버전 변경 불가(V. 3.5.1)

○ **(필수선언)** DB연결부 및 결과 저장파일 생성

- *ATLAS 활 용: 연구과제 디자인 후 출력한 프로젝트 폴더 전체를 제출 (SQL Oracle로 Export)
- ATLAS 미활용: DB연결부 작성 시 *DatabaseConnector* 패키지를 사용하고 최종 결과저장은 파일형태로 이뤄지도록 작성 요망

*ATLAS(아틀라스) - OHDSI 제공 환자 수준의 CDM데이터 분석 응용 프로그램

○ **(제출방식)** 작성된 R소스코드를 압축파일 형태로 제출

- R소스코드 제출 전 연구자별 바이러스 검사 및 검사결과 캡처 이미지 첨부

○ (추가 안내 및 유의사항)

구 분	내 용	비 고
Docker	- **CDM 공통도커 운영** ※ 연구과제별(=연구자별) 도커 이미지 파일 생성 불가	R Ver 3.5.1
Library	- 프로젝트별 독립적으로 패키지 추가 설치 가능 ※ CRAN 제공 패키지만 추가 가능	R Ver 3.5.1
R 프로젝트	- 코호트 작성 및 통계처리 부분의 로직을 분리하여 작성 요망 - 빌드하지 않은 상태로 압축만 하여 제출 ※ 심평원에서 .tar 파일로 빌드 후 설치하여 코드 실행 예정	**압축 전 바이러스 검사 必**

DB연결	- **오라클(Oracle) 데이터베이스 연결 파라미터 변수 5종 관련**
	· dbms, user, password, server, pathToDriver 변수명 사용
	· DB변수 정보 값은 심평원에서 입력하여 코드 실행
	`# Details for connecting to the server:`
	`connectionDetails <- DatabaseConnector::createConnectionDetails(dbms = dbms,`
	`user = user,`
	`password = password,`
	`server = dbServer,`
	`pathToDriver = pathToDriver`
	`)`
	[참고코드] *DatabaseConnector.createConnectionDetails* 함수 사용예시	

임시 데이터 관련	- 각 연구과제별 전용 테이블스키마 및 테이블스페이스(30GB) 할당 - 최종결과가 나온 뒤, 코호트 등 임시 정보는 모두 삭제하는 코드 작성 요망	
디버그	- 에러 발생에 따른 코드 디버깅 지원 불가	
결과 저장	- 최종결과 저장은 *파일 형태로 이뤄지도록 작성	*최종결과 파일만 반출

5. 문의처

소 속	담당업무	전화번호	메일주소
빅데이터실 빅데이터전략부	CDM 전산 등	033-739-1006	cdm@hira.or.kr
	CDM 운영 등	033-739-1088	

별첨1. ‘CDM ETL 정의서’

별첨2. ‘CDM 매핑용어사전’

별첨3. ‘CDM 임상테이블15종 데이터샘플’

[붙임1]

이용 종료 확인서

연구명	국문		
	영문		
연구책임자		소속/직위	성명
분석결과 활용 목적		□정책/연구보고서 □학술지논문 □학술대회 발표 □학위논문 □기타( )	

위와 같이 CDM 연구분석 이용을 종료하며, 향후 논문, 보고서, 발표자료, 보도자료 등의 결과물이 발생하는 경우 다음 사항을 준수할 것을 확약합니다.

① 결과물(산출물)에 자료 출처 명시

- (출처표기방법)
  - 국문: 본 자료는 건강보험심사평가원 OMOP-CDM자료를 활용한 것이며, 연구결과는 건강보험심사평가원 및 보건복지부와 관련이 없음을 밝힙니다.
  - 영문: This study used HIRA OMOP-CDM data made by Health Insurance Review & Assessment Service(HIRA). The views expressed are those of the author(s) and not necessarily those of the HIRA and the MOHW.

② 발생일로부터 30일 이내에 결과 활용 통보서 제출

- 제출방법: 전자우편(cdm@hira.or.kr)

년 월 일

연구책임자 (서명)

건 강 보 험 심 사 평 가 원 장 귀 하

[붙임2]

결과 활용 통보서

연구명	국문		
	영문		
연구책임자		소속/직위	성명

구분	□정책/연구보고서 □학술지논문 □학술대회 발표 □학위논문 □기타( )
발표연월	
제목	
저자	
공개여부	
원문URL	
발표지	
내용	

건 강 보 험 심 사 평 가 원 장 귀 하