

가이드라인 등록번호

[안내서-1128-01]

의료정보 데이터베이스 연구에 대한 가이드라인

(민원인 안내서)

2021. 6.



식품의약품안전처

지침·안내서 제·개정 점검표

명칭

의료정보 데이터베이스 연구에 대한 가이드라인(민원인 안내서)

아래에 해당하는 사항에 체크하여 주시기 바랍니다.

등록대상 여부	<input type="checkbox"/> 이미 등록된 지침·안내서 중 동일·유사한 내용의 지침·안내서가 있습니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	☞ 상기 질문에 '예'라고 답하신 경우 기존의 지침·안내서의 개정을 우선적으로 고려하시기 바랍니다. 그럼에도 불구하고 동 지침·안내서의 제정이 필요한 경우 그 사유를 아래에 기재해 주시기 바랍니다. (사유 :)	
	<input type="checkbox"/> 법령(법·시행령·시행규칙) 또는 행정규칙(고시·훈령·예규)의 내용을 단순 편집 또는 나열한 것입니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	<input type="checkbox"/> 단순한 사실을 대외적으로 알리는 공고의 내용입니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	<input type="checkbox"/> 1년 이내 한시적 적용 또는 일회성 지시·명령에 해당하는 내용입니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	<input type="checkbox"/> 외국 규정을 번역하거나 설명하는 내용입니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	<input type="checkbox"/> 정보제공 등 직원 교육용 자료입니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
☞ 상기 사항 중 어느 하나라도 '예'에 해당되는 경우에 지침·안내서 등록 대상이 아닙니다. 지침·안내서 제·개정 절차를 적용하실 필요는 없습니다.		
지침·안내서 구분	<input type="checkbox"/> 내부적으로 행정사무의 통일을 기하기 위하여 반복적으로 행정사무의 세부기준이나 절차를 제시하는 것입니까? (공무원용)	<input type="checkbox"/> 예(☞지침) <input checked="" type="checkbox"/> 아니오
	<input type="checkbox"/> 법령, 고시·훈령·예규와 같은 규정 또는 식약처장이 정한 특정한 사안에 대하여 그 절차 등의 내용을 알기 쉽게 풀어 설명하거나 식약처의 입장을 기술하는 것입니까? (민원인용)	<input checked="" type="checkbox"/> 예(☞안내서) <input type="checkbox"/> 아니오
기타 확인 사항	<input type="checkbox"/> 상위 법령을 일탈하여 새로운 규제를 신설·강화하거나 민원인을 구속하는 내용이 있습니까?	<input type="checkbox"/> 예 <input checked="" type="checkbox"/> 아니오
	☞ 상기 질문에 '예'라고 답하신 경우 상위법령 일탈 내용을 삭제하시고 지침·안내서 제·개정 절차를 진행하시기 바랍니다.	
상기 사항에 대하여 확인하였음. 2021 년 6 월 29 일 <div style="display: flex; justify-content: space-between;"> 담당자 확 인(부서장) 박예정 오정원 </div>		

이 안내서는 의약품 등의 의료정보 데이터베이스를 활용한 적절한 약물 역학 연구 실시에 참고를 위해 '의료정보 데이터베이스 연구'에 대해서 식품의약품안전처의 입장을 기술한 것입니다.

본 안내서는 대외적으로 법적 효력을 가지는 것이 아니므로 본문의 기술 방식('~하여야 한다' 등)에도 불구하고 민원인 여러분께서 반드시 준수하셔야 하는 사항이 아님을 알려드립니다. 또한, 본 안내서는 2021년 6월 현재의 과학적·기술적 사실 및 유효한 법규를 토대로 작성되었으므로 이후 최신 개정 법규 내용 및 구체적인 사실관계 등에 따라 달리 적용될 수 있음을 알려드립니다.

※ "민원인 안내서"란 대내외적으로 법령 또는 고시·훈령·예규 등을 알기 쉽게 풀어서 설명하거나 특정한 사안에 대하여 식품의약품안전처의 입장을 기술하는 것 (식품의약품안전처 지침서등의 관리에 관한 규정 제2조)

※ 본 안내서에 대한 의견이나 문의사항이 있을 경우,
식품의약품안전처 의약품안전국 의약품안전평가과에 문의하시기 바랍니다.
전화번호: 043-719-2711, 2715
팩스번호: 043-719-2700

제·개정 이력

연번	제·개정번호	승인일자	주요내용
1	안내서-1128-01	2021.6.29	제정

목 차

I. 서론	2
1. 배경 및 목적	2
2. 적용 범위	3
3. 용어의 정의	3
II. 데이터베이스의 종류 및 사용 시 고려사항	6
1. 국내 RWD 자료원 현황	6
2. 주요 데이터베이스 종류 및 특징	7
가. 건강보험 청구자료	7
나. 전자의무기록	8
다. 환자등록자료	9
3. 데이터베이스 사용 시 유의사항	9
가. 적절성	10
나. 신뢰성	10
III. 연구설계 및 선정 시 고려사항	12
1. 연구설계	12
가. 코호트 연구	12
나. 환자-대조군 연구	13
다. 코호트 내 환자-대조군 연구	14
라. 환자군 연구	15
2. 연구설계 선정 시 고려사항	18
가. 연구대상의 정의	18
나. 노출 및 결과의 정의	18
다. 비뚤림, 교란요인 및 효과변경인자	19
3. 데이터 관리	22
4. 분석	22
5. 개인정보보호 및 윤리	23
IV. 연구계획서 및 보고서 작성	24
1. 연구계획서 작성방법	24
2. 연구보고서 작성방법	29

I. 서론

1. 배경 및 목적

신약 등은 시판 전 임상시험을 통해 안전성 및 유효성을 평가받고 시판 승인 허가를 취득한다. 그러나, 임상시험은 한정된 시간 동안 소수의 제한된 대상자 집단에서 수행되기 때문에 다양한 특성을 가진 환자에게 오랜 기간 처방될 수 있는 실제 임상 환경을 반영하기 어렵다. 이러한 임상시험의 한계를 보완하기 위하여 실제 임상에서의 의약품 사용을 반영하는 실사용데이터(Real World Data, RWD)와 이 데이터를 가공·분석하여 생성한 실사용증거(Real World Evidence, RWE)를 활용하여 시판 후 의약품의 안전성·유효성을 평가하는 데 활용하려는 움직임이 전세계적으로 확산되고 있다.

국내에서는 건강보험 청구자료, 전자의무기록(Electronic Medical Records, EMR) 등의 RWD를 분석한 RWE를 시판 후 안전성 연구에 활용할 수 있으며 다양한 임상 환경을 반영한 RWE를 활용할 경우 임상시험보다 광범위한 대규모의 연구대상을 장기간 관찰하는 것이 가능하므로 시판 전 알 수 없었던 의약품의 안전성에 대한 보다 많은 정보를 얻을 수 있을 것으로 기대하고 있다.

우리나라는 국가 단일보험체계에 따라 전 국민을 포괄하는 건강보험 청구자료와 대다수의 병·의원이 EMR을 실시하여 활용 가능성이 높은 다양한 RWD를 보유하고 있다. 식약처는 최근 RWD를 의약품 안전성 평가에 활용할 수 있도록 ‘신약 등의 재심사 기준’을 개정하여 시판 후 조사의 한 방법으로 의료정보 데이터베이스 연구를 추가하였으며 의료 정보 데이터베이스를 활용한 조사의 실무에 대응하는 구체적인 가이드라인의 필요성이 대두되고 있다. 이에 따라 식약처는 ‘시판 후 약물감시 역량제고 방안 마련’ 및 ‘Real-World data 및 Evidence 기반으로 한 바이오의약품의 시장 출시 후 안전관리’와 같은 연구사업을 통해 데이터베이스를 활용한 시판 후 안전관리에 대한 참고자료를 마련하고자 하였다.

본 가이드라인은 이와 같은 연구사업을 통해 마련된 자료를 정리한 것으로 의료정보 데이터베이스를 활용한 적절한 약물역학 연구 실시를 위한 참고사항을 기술한 것이다.

2. 적용 범위

이 가이드라인은 의료정보 데이터베이스를 활용한 의약품 안전성 평가를 위한 약물역학 연구를 실시하는 경우 활용한다.

3. 용어의 정의

가. 데이터베이스

1) 실사용데이터(Real World Data, RWD)

실사용데이터는 기존의 중재적 방법에 의한 임상시험으로부터 수집된 데이터가 아닌 다양한 유형의 의료정보데이터를 포괄하는 용어이다. 실사용데이터는 의료제공자(의료기관), 보험자, 의료소비자(환자, 일반인) 등 다양한 출처로부터 생성될 수 있다.

2) 실사용증거(Real World Evidence, RWE)

실사용데이터를 분석하여 파생된 증거로써, 실사용데이터 분석을 통해 얻은 의약품의 사용 결과나 잠재적 이익 또는 위험성에 관한 임상적 근거를 말한다. 또한, 임상문헌을 분석하여 파생된 새로운 증거도 포함한다.

3) 의료정보 데이터베이스(database) 정의

의료정보 데이터베이스는 컴퓨터 시스템에 전자 방식으로 저장된 구조화된 정보 또는 데이터의 체계적인 집합을 의미하는 것으로 이 가이드라인에서는 일정기간 동안 수집된 환자의 의무기록, 건강보험 청구자료, 환자등록체계 기타 환자 건강 상태에 대한 데이터 등에 관한 정보를 체계적으로 전산화한 것을 의미한다.

4) 맞춤형연구 데이터베이스 DB (맞춤형자료)

맞춤형연구 DB데이터베이스 (<https://nhiss.nhis.or.kr/bd/ay/bdaya001iv.do>)는 국민건강보험공단이 수집, 보유, 관리하는 건강보험 청구자료를 정책 및 학술 연구목적으로 이용할 수 있도록 신청인이 제출한 연구계획에 맞게 맞춤형 자료로 정보주체를 알아볼 수 없도록 조치한 데이터 셋으로 가공하여 제공하는 종단형 데이터이다. 전 국민의 의료형태, 진단코드, 약제 코드 등이 포함되어 희귀질환, 희귀부작용 등에 관한 연구가 가능하다. 연구용 진료정보통계(<https://opendata.hira.or.kr/home.do>)는

맞춤형연구 데이터베이스와 유사하나 건강보험심사평가원이 제공하는 종단형 건강보험 청구자료이다.

5) 표본연구 DB 데이터베이스 (표본자료)

표본연구 데이터베이스는 국민건강보험공단 또는 건강보험공단이 수집, 보유, 관리하는 건강보험 청구자료 중 일부를 표본추출하여 제공하는 데이터이다. 건강보험공단에서는 표본코호트DB, 건강검진코호트DB, 노인코호트DB, 직장여성코호트DB, 영유아검진코호트DB 등을 제공하고 있다 (<https://nhiss.nhis.or.kr/bd/ab/bdaba010cv.do>). 건강보험심사평가원에서는 입원환자데이터셋, 전체환자데이터셋, 고령환자데이터셋, 소아청소년환자데이터셋을 제공하고 있다 (<https://opendata.hira.or.kr/op/opc/selectPatDataApplInfoView.do>).

6) 데이터 정제(data cleaning)

의료정보 데이터베이스 내 데이터를 삭제 또는 수정 등의 가공을 거쳐 분석 수행에 맞게 정비하는 작업이다.

나. 연구

1) 표본크기(sample size)

연구결과를 적용할 모집단으로부터 추출된 표본의 개체 수로 연구에 참여하는 대상자의 수를 의미한다.

2) 검정력(power)

통계학적으로 제2종 오류(대립가설이 참임에도 귀무가설을 채택할 확률)를 범하지 않을 확률이다. 연구 수행 시 통계적 유의성을 확보하기 위해서는 충분한 검정력을 가져야 한다. 검정력을 높게 설정할수록 통계적 유의성을 확보하기 위하여 많은 표본크기를 필요로 하기 때문에 검정력은 표본크기를 결정하는 기준이 된다.

3) 내적 타당성(internal validity)

관찰된 연구결과가 연구대상자의 특성을 얼마나 정확하게 나타내고 있는지 그 정도를 의미한다. 연구의 내적 타당성은 측정 오류나 연구대상자 선정 등에 의해 위협받을 수 있다.

4) 외적 타당성(external validity)

관찰된 연구결과가 본 연구의 연구대상자 외에 유사한 환자들에게도 적용되는지 그 정도를 의미한다. 주로 일반화 가능성이라고 하며, 연구의 내적 타당성이 확보되었다는 것을 확인한 후 외적 타당성을 판단한다.

다. 연구설계 및 수행

1) 노출(exposure)

환자에 대한 약물 또는 예방적, 치료적 개입이다.

2) 결과(outcome)

약물 노출 또는 예방적, 치료적 개입으로 일어날 수 있는 모든 건강상의 변화이다.

3) 공변량(covariate)

노출과 결과 이외에 추출되는 환자의 기본정보(예: 성별), 교란요인, 효과 변경 인자로서 분석에 포함되는 변수이다.

4) 비뚤림(bias)

연구설계, 수행, 분석 과정에서 발생하는 모든 계통적인 오류로 이는 노출이 결과에 미치는 영향을 잘못 추정하도록 한다.

5) 교란요인(confounder)

연구에서 설정한 노출과 결과를 중심으로 결과의 발생을 유발한다고 알려진 위험요인이자 노출과 연관성이 있는 요인을 의미한다. 비뚤림과 동일하게 교란요인의 작용으로 노출이 결과에 미치는 영향을 잘못 추정할 수 있으나, 하나의 오류가 아닌 요인과 결과 발생 위험 관계의 특성을 기술하는 현상이라는 점에서 비뚤림과 차이를 보인다.

6) 신환자(new-user)

연구대상 약물을 처음으로 처방받아 처방 시작 시점부터 추적관찰 되는 환자이다. 신환자가 연구에 포함될 경우 약물 처방 기간과 관련된 시간에 따른 위험을 평가할 수 있으며 약물 처방 전 연구대상자들의 일반적 특성을 포착하여 보정할 수 있다는 장점이 있다.

7) 활성 비교군(active comparator)

임상 현장에서 연구대상 약물과 유사한 적응증을 가지는 또 다른 약물을 처방받은

환자 집단이다. 연구대상 약물 노출군과 활성 비교군을 비교할 경우 적응증이 유사하므로 교란요인을 줄일 수 있다는 장점이 있다.

II. 데이터베이스의 종류 및 사용시 고려사항

1. 국내 RWD 자료원 현황

의약품 실사용데이터(RWD)는 실제 임상 환경에서 의약품 사용에 관한 임상 데이터를 의미하며, 주로 규모가 큰 집단으로부터 수집된다. 국내 활용가능한 RWD 자료원으로는 건강보험청구자료(건강보험공단, 건강보험심사평가원), 의료기관의 전자의무기록자료(Electric Medical Record), 환자·보건의료인 대상 설문조사, 환자등록정보 등이 있다. 아래는 의약품 데이터베이스연구에 활용 가능한 RWD 자료원의 목록이다.

표 1 국내 보유 데이터베이스의 특징 및 장단점

자료목록	정보원	장점	단점
건강보험 청구자료	기구축된 국내 건강보험가입자 청구자료	<ul style="list-style-type: none"> -보건의료 빅데이터 활용 으로 비용절감 -대규모 인구집단 -보험급여된 모든 정보 -환자 추적관찰 가능 -높은 일반화 가능성 	<ul style="list-style-type: none"> -청구과정과 코딩체계에 대한 사전지식 필요 -임상수치 접근의 어려움 -비급여 반영불가
전자의무 기록	의료제공자의 기록과 환자의 건강기록에 대한 구체적인 자료	<ul style="list-style-type: none"> -일상적 진료정보 파악, 청구자료에 비해 임상정보 풍부 -의학적/임상적 과거력을 포괄적으로 파악 가능 -추출된 자료 질이 우수함 (높은 진단명 타당도, 비 급여 기록) 	<ul style="list-style-type: none"> -의료기관에 따른 자료 질적/양적 차이 -과거력 파악 어려움(예: 신환자) -규모가 작아 검정력 확보 어려움 -연구결과 일반화 어려움
설문조사	환자의 경험 기반 직접적 자료	<ul style="list-style-type: none"> -비용 효과성 -설문지 항목에 연구자의 의도 반영 	<ul style="list-style-type: none"> -지속적인 환자 추적관찰이 어려움 -Recall bias 발생 가능
환자 등록자료	환자 등록을 통해 전향적으로 구축된 1차 자료원	<ul style="list-style-type: none"> -전향적으로 수집되므로, 특정 질병, 의약품에 관한 정보 파악 가능 	<ul style="list-style-type: none"> -자료구축에 많은 시간 소요 -표준화된 자료수집 도구

		-병원처방 외 의료정보 확보 가능	개발 필요
--	--	-----------------------	-------

2. 주요 데이터베이스 종류 및 고려사항

데이터베이스는 데이터 생성 배경에 따라 각각의 특징이 달라 데이터베이스 연구를 수행하는 경우 활용 가능한 여러 데이터베이스의 특징 및 장단점 등을 고려하여 연구 목적과 가설에 적합한 데이터베이스를 선택하여야 한다. 현재 의약품 데이터베이스 연구에 활용 가능한 국내 데이터베이스 종류 및 해당 데이터베이스를 선택하는 경우 고려해야 하는 사항은 다음과 같다.

가. 건강보험 청구자료

건강보험 청구자료란 요양기관이 의료서비스를 제공한 후 환자의 진료비용 중 일부를 보험자에게 청구하면서 발생하는 데이터로 전 국민의 인구학적 및 의료이용 자료가 포함된 대표적인 RWD이다. 건강보험공단 및 건강보험심사평가원의 건강보험 청구자료는 5천만 명이 넘는 인구 규모를 바탕으로 국내 자료원 중 의약품에 관하여 가장 많은 데이터를 얻을 수 있으며, 급여 청구된 항목에 대하여 임상정보를 얻을 수 있다. 주로 치료패턴의 확인, 치료효과의 비교, 복약 순응도를 확인하는 연구에 활용한다. 청구자료는 명세서, 원내진료내역, 상병내역, 원외처방내역으로 구성되어 있으며 아래의 인구학적 정보, 진단정보, 처방약물정보 변수 등이 있다(표 2).

표 2. 청구자료에서 이용 가능한 변수

	수진자 일반내역	명세서 내역	진료/상병내역	처방전교부 상세내역
변수	기준연월, 성별, 보험료분위, 출생년도, 사망일자	서식구분코드 (입원/외래 구분코드), 요양개시일자 (진료일자) 주·부상병 코드 입내원일수, 총처방일수	원내행위내역(처방, 수술, 재료), 1회투약량, 1일투약량, 1일투여량 또는 실시횟수, 총투여일수 또는 실시횟수, 주성분코드, 상병코드(모든 상병)	1회투약량 1회투여량 총 투여일수 일반명코드 (의약품 성분명코드) 약효분류번호

국민건강보험공단에서 제공하는 연구용 자료로는 맞춤형연구 데이터베이스와 표본연구 DB가 있다. 맞춤형연구 DB는 국민건강보험공단이 수집, 보유, 관리하는 건강보험 전수자료를 정책 및 학술 연구목적으로 이용할 수 있도록 가공하여 제공하는 전 국민의 의료형태, 진단코드, 약제코드 등이 포함된 데이터이다. 공공기관, 국가정책연구, 학술연구를 하는 기관 또는 사람, 공단과 협약을 체결한 기관 등을 대상으로 제공한다. 표본연구 DB는 전국민 2%코호트, 성인검진코호트, 고령환자코호트, 직장여성검진코호트, 영유아검진코호트 총 5가지 표본코호트데이터로 제공되어 시간에 따른 종단연구가 가능하다. 건강보험공단자료는 통계청과 연계한 사망자료를 이용할 수 있으며, 건강검진자료도 보유하고 있어 신청하는 경우 흡연, 음주 등의 생활습관 변수 및 임상 수치 변수를 신청할 수 있다. 단 표본코호트의 경우 민감상병 및 민감수가코드에 관한 정보를 제공하지 않는다.

건강보험심사평가원(Health Insurance Review and Assessment, HIRA) 청구자료도 전수자료 및 표본자료를 제공하고 있으며 이용할 수 있는 정보가 국민건강보험공단 자료와 유사하다. 표본자료는 입원환자데이터셋(National Inpatient Sample, HIRA-NIS), 전체환자데이터셋(National Patient Sample, HIRA-NPS), 고령환자데이터셋(Adult Patient Sample, HIRA-APS), 소아청소년환자데이터셋(Pediatric Patient Sample, HIRA-PPS) 총 4가지 표본데이터셋을 제공하고 있다.

청구자료는 비급여 정보(행위, 약제, 치료재료 등), 임상검사수치에 대한 자료, 의사의 처방 없이 약국에서 판매되는 약물의 정보가 포함되어 있지 않아 관련 연구를 수행하는데 제한사항이 있다. 질병의 중증도, 다른 질병의 이환 유무정보를 충분히 제공하지 않을 가능성이 있으며 흡연, 음주 등 생활습관에 대한 정보가 없어 해당 요인들은 교란요인에 반영하기 어렵다. 또한, 청구진단명이 최종 확진 진단명과 차이가 있을 가능성이 있으며 건강보험급여 정책에 따라 급여범위 변화 등 제도의 변화가 자료에 영향을 미칠 수 있기 때문에 보험급여정책에 대한 검토가 필요할 수 있다.

나. 전자의무기록

전자의무기록(이하 ‘EMR’)은 의료기관 내 진료과정에서 발생하는 입원 및 퇴원 기록, 진찰 및 수술기록, 투약기록, 임상검사기록, 영상검사기록 및 소견서, 간호기록 등을 전자화한 자료이다. EMR의 일차적인 목적은 환자 진료 및 의료행위에 대한 법적, 재정적, 행정적 기록이나 내원기간 동안 환자의 진료 및 관리에 대한 정보를 모두 포함하고 있기 때문에 주요 RWD로서 연구에 활용된다.

EMR은 청구자료와 달리 임상 및 영상정보 확인이 가능하기 때문에 진단명의 타당도가 높고 비급여 항목도 확인이 가능하다는 장점이 있다. 그러나 단일 의료기관의 내원 환자기록만 확인이 가능하여 자료에 포함된 환자의 수가 적어 연구 결과의 일반화가 어렵고, 환자가 타 의료기관을 이용한 경우 추적이 불가하다는 단점이 있다. 의료기관 간 EMR을 서로 연계하면 표본의 크기가 커지고 환자 이송 후에도 추적이 가능할 수 있으나 의료기관마다 의무기록의 정확도가 상이하고 표준화되어있지 않다는 점 등으로 인해 제한적이다. 국내에서는 의료기관 간의 자료를 연계하기 위해 각 병원의 데이터를 공통 형식으로 변환하는 공통데이터모델(Common Data Model, CDM)을 구축하였으며 이를 연구에 활용하는 방안도 검토해볼 수 있다.

다. 환자등록자료(Registry)

환자등록자료는 어느 특정 정보의 수집을 목적으로 작성되는 것이며, 수집하는 정보에 따라 다양한 종류의 레지스트리가 있다. 관심결과(질병)에 대한 연구대상을 직접 모집하여 추적관찰하는 ‘결과 레지스트리’, 관심노출(의약품)에 대한 연구대상을 직접 모집하여 추적관찰하는 ‘노출 레지스트리’, 특정 수술에 관련된 정보를 집적하여 추적관찰하는 진료 ‘행위 레지스트리’로 나눌 수 있다. 레지스트리는 각각의 목적에 따라 특정 데이터 항목이 수집되기 때문에 연구에 필요한 데이터 항목 중 수집 가능한 정보가 대부분 확보될 수 있다. 결과 레지스트리는 연구자의 설계에 따른 연구 변수 설정 및 타당도 높은 결과변수 이용이 가능하며 노출 레지스트리는 상대적으로 정확한 환자 식별이 가능하다.

레지스트리는 높은 구축 비용과 정보수집 간격 사이 정보 누락이 발생할 수 있으며, 노출 레지스트리의 경우 비교대상 의약품에 대한 정보수집이 없을 수 있다. 기존에 구축된 레지스트리를 다른 연구에 이용하고자 하는 경우 레지스트리 본래 구축목적과 연구자의 연구목적이 달라 필요한 데이터가 부족할 수 있으므로 추가 정보수집이나 다른 데이터로 연계하는 것에 대한 검토가 필요할 수도 있다.

3. 데이터베이스 사용 시 유의사항

적절한 데이터베이스를 선택하고자 하는 경우 연구목적 및 가설에 따라 데이터베이스의 적절성(Relevance) 및 데이터베이스의 신뢰성(Reliability)을 고려해야 한다.

가. 적절성(Relevance)

- 의약품의 사용, 노출 및 적절한 모집단에 대한 관심결과를 캡처하기에 충분한 세부 정보가 포함되는지
- 분석에 이용할 수 있는 데이터 요소가 타당하며 적절한 분석방법이 적용될 때 연구질문을 다룰 수 있는지
- RWD와 RWE는 임상 및 과학적 판단으로 해석이 가능한지
- 실제 모집단에서 의약품 사용의 대표성 여부와 평가 중인 관련 모집단에 일반화되는지
- 선택한 RWD가 지역, 국가 또는 국제적으로 사용되는지
- 선택한 RWD에서 캡처된 의약품에 대한 환자 노출의 전체 비율이 적절한지
- 선택한 RWD가 환자 집단의 경험을 얼마나 잘 반영하는지 평가하는 데 사용되는 검증 프로토콜 및 결과 데이터가 적절한지
- 연구설계, 연구 프로토콜, 분석계획이 규제 문제를 해결하기에 적절하고 충분히 시의적절하게 달성될 수 있는지
- 환자의 병력 및 기존 상태를 적절하게 파악하는지 여부 및 해결해야 할 문제를 평가하는 데 필요한 추적관찰(follow-up) 정보가 적절한지
- 노출 또는 관심 결과에 영향을 줄 수 있는 교란요인을 조정하기 위해 충분한 데이터 요소가 수집되는지
- 수행된 연계가 과학적으로 적절하고 출처에 따른 코딩 및 보고의 차이를 설명하는지
- RWD가 포함하는 기간과 현재 시점 사이의 간격 및 보고 기간을 포함한 RWD 소스 보고 일정이 적절한지
- 결과 기반 품질 평가, 검증된 예측 위험 모델링, 신호 탐지, 성능 개선, 벤치마킹 및 기타 임상적으로 의미있는 용도를 결정하기 위한 사전 문서화 된 RWD 소스의 사용이 적절한지 [예 : 동료 검토(peer-review) 간행물 또는 실무 지침]
- 수집된 데이터 요소가 결과를 평가하기에 충분한지

나. 신뢰성(Reliability)

1) 데이터 발생

아래 내용을 포함한 운영설명서 또는 기타 문서가 존재해야 한다.

- 사전 특정화된 수집 데이터 요소

- 데이터 요소 정의 (예: 공통 정의 프레임워크를 제공하기 위한 데이터 사전)
- 데이터 집계 방법, 문서 (예: 공통 사례 보고서 양식, 검증 가능한 출처의 요약) 및 데이터 요소 수집에 대한 관련 기간 (예: 일반적 시간 프레임워크)

2) 데이터 보증-품질 관리

- 데이터 요소 모집단의 품질 - 완전성·일관성을 위한 소스 검증 절차 및 데이터 수집 및 기록 절차 준수
- 교란 요인에 대한 조정을 포함한 특정 분석에 필요한 데이터의 완전성
- 장소 및 시간에 따른 데이터 일관성
- 참여 장소에서 데이터 수집 및 데이터 사전 사용을 위한 지속적인 교육 프로그램 평가
- 현장 및 데이터 모니터링 실무 평가나 데이터 품질 감사 프로그램 사용

III. 연구설계 및 선정 시 고려사항

1. 연구설계

임상연구는 중재 연구인 임상시험과 관찰연구로 구분할 수 있는데, RWD 자료원을 활용한 연구는 관찰연구에 해당한다. 임상연구의 목적은 가설제시와 가설검증으로 구분할 수 있다. 가설제시 단계에서는 노출과 질병 두 요인 간의 관련성의 가설을 제시하는 단계로 인과적 해석을 배제한 단계이며, 가설검증 단계에서는 노출과 질병의 관련성을 평가하며, 이 결과는 인과적 해석의 근거자료로 사용될 수 있다.

RWD 자료원을 활용한 가설 검증 연구로는 대표적으로 코호트 연구와 환자대조군 연구 설계가 있다. 코호트 연구는 질병발생의 원인으로 가정한 요인의 노출상태에 따라 인구집단을 구성하고, 일정기간 동안 추적관찰하여 특정 질병발생 여부를 확인하는 연구이다. 관찰연구 중 근거의 수준이 가장 높으며, 노출변수인 관심대상 약물이 환자수가 매우 드문 경우에 사용하기 적합하다. 환자대조군 연구는 관심 질병이 발생한 환자군과 대조군을 선정한 후, 두 집단에 속한 환자들의 과거 약물노출을 비교하여 위험요인을 파악하는 방법이다.

데이터베이스 연구를 수행하는 경우 비뚤림을 최소화하기 위해 연구목적 및 활용데이터의 특징, 연구설계 방법의 장단점을 고려하여 인과성을 판단할 수 있는 잘 설계된 연구설계 방법을 적용해야 한다.

가. 코호트 연구(cohort study)

코호트 연구는 부작용으로 의심되는 질병이 추적관찰 시작 시점 이전에 발생하지 않은 사람을 연구대상자로 선정하여 연구대상 약물의 노출 여부에 따라 대상자를 약물에 노출된 사람(노출군)과 약물에 노출되지 않은 사람(비노출군)으로 나눈 후 추적관찰하여 각 군에서의 질병의 발생률을 비교함으로써 약물 노출과 질병 발생 간 인과적 관련성을 평가하는 연구설계이다(그림 1).

코호트 연구에서는 연구자가 연구대상자에 대한 자료를 언제 수집하느냐에 따라 전향적 또는 후향적 코호트 연구로 구분하나 RWD를 활용할 경우 연구 시작 시점에 약물 노출과 질병 발생이 모두 일어났으므로 후향적 코호트 연구로 간주한다. 연구대상자는 모두 관찰을 시작하는 과거 시점에 해당 질병이 발병하지 않은 상태여야 한다.

코호트 연구에서 활용되는 핵심 통계 측도는 발생률(incidence rate)과 비노출군의 발생률 대비 노출군의 발생률을 비교한 발생률비(rate ratios, RR)이다. 발생률비는 상대발생률(relative rates), 위험비(risk ratio), 상대위험도(relative risk)라고 불리기도 한다. 발생률비는 '약물에 노출된 경우가 노출되지 않은 경우에 비해 질병 발생 위험이 RR배 높다'라고 해석한다. 발생률비 또한 오즈비와 마찬가지로 통계적 유의성을 파악하기 위해 95% 신뢰구간을 산출하며 신뢰구간이 RR=1을 포함하면 RR값은 통계적 유의성을 가지고 있지 않다고 간주한다.

코호트 연구는 약물 노출 여부에 따라 연구대상자를 나누므로 노출빈도가 높지 않은 경우에도 적용할 수 있으며, 하나의 노출에 대해 여러 건강 결과를 동시에 파악할 수 있다는 장점이 있다. 또한 관찰 시점에 연구대상 질병이 발생하지 않았기 때문에 노출과 결과 간 시간적 선후관계가 명확하여 인과관계를 밝히는 데 유리하다. 반면 환자수가 매우 드문 질병을 결과변수로 하는 경우 매우 큰 규모의 연구대상을 확보해야 하므로 시간과 비용이 많이 든다는 단점이 있다.

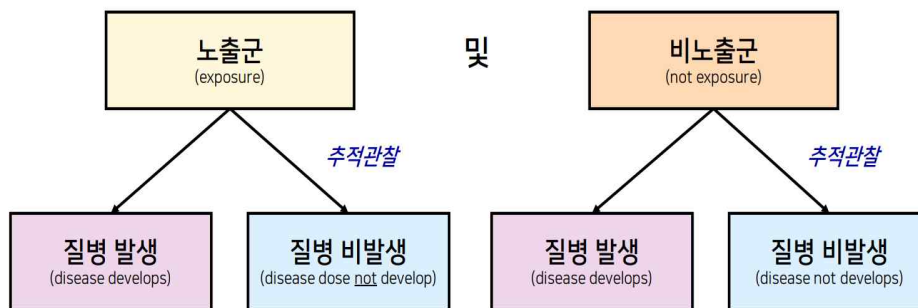


그림 1. 코호트 연구설계

나. 환자-대조군 연구(case-control study)

환자-대조군 연구는 약물 투여에 따른 부작용으로 의심되는 질병의 발병 여부에 따라 연구대상자를 질병을 진단받은 사람(환자군, case)과 진단받지 않은 사람(대조군, control)으로 나누어 과거에 해당 약물에 대한 노출력이 있는지 비교하는 연구설계이다(그림 2). 환자-대조군 연구의 목적은 과거 약물의 노출 유무와 질병 발생의 인과적 관련성을 파악하는 것이다.

연구대상자 선정 시 모든 환자군은 건강 관련 상태가 유사해야 하며, 대조군은 부작용으로 의심되는 질병을 가지고 있지 않다는 것을 제외하고 모든 면에서 환자군과 유사해야 한다. 이를 충족하기 위해 성별, 연령대, 사회경제적 특성 등에 따라 환자군과 대조군을 매칭(matching)할 수 있다.

환자-대조군 연구에서 활용되는 핵심 통계 측도는 오즈비(odds ratio, OR)이다. 오즈비는 ‘약물에 노출된 경우가 노출되지 않은 경우에 비해 질병 발생의 위험이 OR 배 높다’ 라고 해석한다. 오즈비의 통계적 유의성을 파악하기 위해 95% 신뢰구간(confidence interval, CI)을 산출하며 OR의 신뢰구간이 1을 포함할 경우 OR값은 통계적으로 유의하지 않다고 간주한다.

환자-대조군 연구는 질병 발생 여부에 따라 연구대상자를 나누기 때문에 발생률이 낮고 잠복기가 긴 질병을 연구하는 경우 적합하며, 하나의 질병에 대해 여러 위험요인의 연관성을 동시에 검정할 수 있다는 장점이 있다. 반면 이미 질병이 발생한 집단을 대상으로 하므로 질병의 발생률을 산출할 수 없으며 약물에 노출되는 경우가 드문 경우에는 연구 수행이 어렵다. 또한 대조군 선정의 오류로 인한 선택 비뚤림이 발생할 수 있으며 약물 노출과 질병 발생 간 시간적 선후관계가 불분명하여 인과성의 강도가 코호트 연구에 비해 낮다.

드물게 발생하는 관심결과에 코호트 연구설계를 이용할 경우에는 발생건수가 작아서 인과성을 파악하는 데 비뚤림이 생길 가능성이 크기 때문에 이런 경우 환자-대조군 연구설계를 고려할 수 있다. 코호트 연구에서 치료법과 결과변수 간의 연관성(association)을 나타내는 측도로 상대위험도를 사용할 수 있는 반면에, 환자-대조군 연구에서는 상대위험도를 계산할 수 없다. 따라서 환자-대조군 연구에서는 상대위험도 대신 오즈비를 이용하여 연관성을 나타낸다. 하지만 결과 발생이 드문 경우 오즈비는 상대위험도와 근사적으로 같아지므로, 관심 결과 변수가 드물게 일어나서 환자-대조군 연구설계를 이용하였을 경우 오즈비를 상대위험도와 같이 해석할 수 있다.

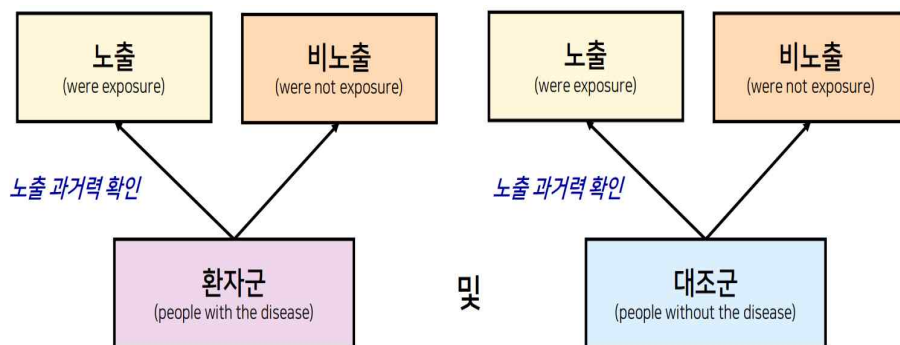


그림 2. 환자-대조군 연구설계.

다. 코호트 내 환자-대조군 연구(nested case-control study)

코호트 내 환자-대조군 연구는 코호트 연구에 환자-대조군 연구를 결합한 연구설계로

코호트 추적관찰 시작 시점 기준 과거기간에 관심 질병이 발생한 적이 없는 집단을 추적 관찰하여 질병이 발생한 대상자를 환자군으로 정의하고, 질병 발생 위험에 함께 놓여 있었으나 질병이 발생하지 않은 대상자 중 대조군을 선정하여 환자-대조군 연구와 동일하게 수행하는 연구설계이다(그림 3). 연구 시작 시점에 연구대상 질병에 이환되지 않은 사람을 연구대상자로 선정하기 때문에 원인과 결과의 시간적 선후관계가 불분명하다는 환자-대조군 연구의 한계점을 보완할 수 있다. 연구 초기에 선정된 대조군은 이후 관심 질병이 발생할 가능성이 있기 때문에 동일 연구에서 환자군이 될 수도 있다.

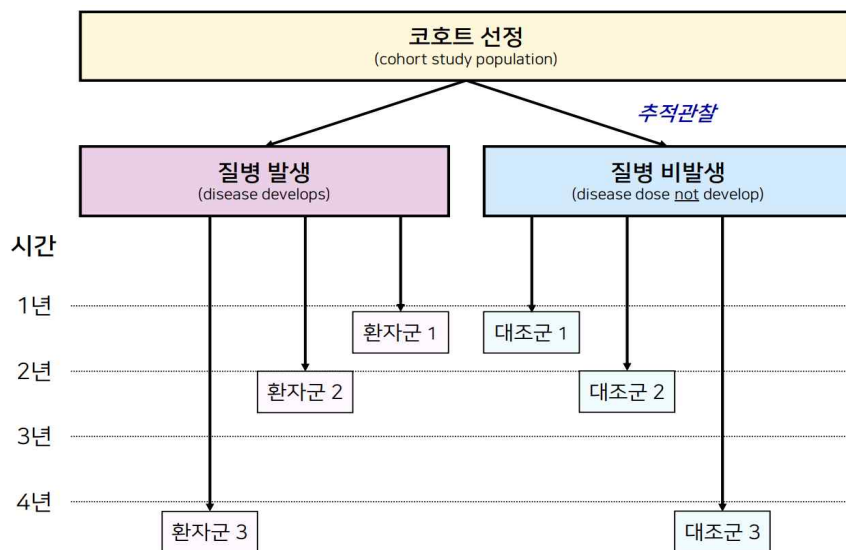


그림 3. 코호트 내 환자-대조군 연구설계.

라. 환자군 연구(case-only study)

보편적으로 활용되는 연구설계인 코호트 연구 및 환자-대조군 연구는 서로 다른 연구대상자들을 비교하기 때문에 대상자 간 특성의 차이에 따른 비뚤림을 배제할 수 없다. 이러한 한계를 보완하는 연구설계로 한 환자 안에서 대조군을 설정하는 환자군 연구가 있으며 환자-교차 연구(case-crossover study, CCO), 자신-대조 환자군 연구(self-controlled case series, SCCS)가 그 예이다. 환자군 연구는 관심 질병이 발생한 환자군만을 연구대상으로서 포함하여 서로 다른 연구대상자를 비교하는 것이 아닌 같은 연구대상자 안에서 시점에 따라 질병 발생의 위험을 비교한다. 즉, 한 환자에서 질병 발생 직전의 일정 기간을 위험기간(hazard period), 질병 발생 거리가 먼

일정 기간을 대조기간(control period)으로 설정하여 위험기간과 대조기간의 질병 발생 위험을 비교한다.

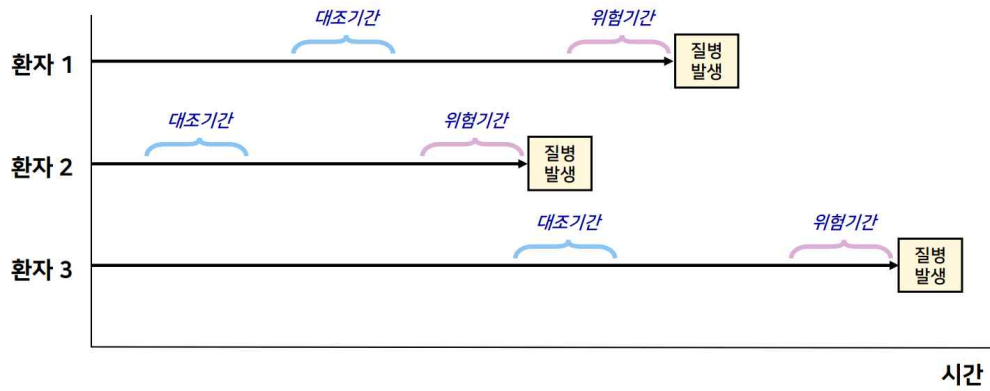


그림 4. 환자군 연구설계.

표 3. 연구설계의 장단점

	장점	단점
코호트 연구	<ul style="list-style-type: none"> • 약물 노출에 따른 여러 건강 결과에 대한 평가가 가능함. • 드문 노출(exposure)에 대한 연구에 적합함. • 질병 발생의 위험도와 발생률을 추정할 수 있음. • 약물 노출과 질병 발생 간 시간적 선후 관계가 명확함. 	<ul style="list-style-type: none"> • 드물게 나타나는 건강 결과(health outcome) 연구 시 상대적으로 많은 연구대상이 필요함. • 상대적으로 많은 연구대상이 필요함. • 연구대상 선정과 측정방법에 대한 통제가 비교적 어려움.
환자 대조군 연구	<ul style="list-style-type: none"> • 드물게 나타나는 건강 결과(health outcome) 연구에 적합함. • 비교적 적은 연구대상으로도 연구 수행이 가능함. • 연구대상 선정과 측정방법에 대한 통제가 비교적 쉬움. 	<ul style="list-style-type: none"> • 약물 노출에 따른 여러 건강 결과에 대한 평가가 불가능함. • 질병 발생의 위험도와 발생률을 추정하는 데 한계가 있음. • 드문 노출(exposure)에 대한 연구 시 상대적으로 많은 연구대상이 필요함. • 약물 노출과 질병 발생 간 시간적 선후 관계가 불명확함.
코호트 내 환자 대조군 연구	<ul style="list-style-type: none"> • 환자-대조군 연구의 단점인 시간적 선후 관계의 불명확성을 보완함. • 환자군과 대조군이 동일한 코호트에서 선정되므로 두 집단의 비교성이 높음. 	<ul style="list-style-type: none"> • 연구 초기에 선정된 대조군은 이후 건강 결과가 발생할 가능성이 있으며, 동일 연구에서 환자군이 될 수 있음. • 표본크기 감소로 검정력이 감소할 가능성이 있음.
환자군 연구	<ul style="list-style-type: none"> • 대조군을 선정하지 않아도 되기 때문에 환자군만으로 연구 수행이 가능함. • 환자 스스로가 대조군이 되기 때문에 기간의 영향을 받지 않는 교란요인을 보정할 수 있음. 	<ul style="list-style-type: none"> • 질병의 발생시점이 명확해야 기간을 설정할 수 있으므로 만성질환에 관한 연구가 어려움. • 위험기간과 대조기간 간 약물의 잔류효과가 없어야 함. • 약물의 노출이 장기간일 경우 다른 요인의 영향이 개입할 수 있음.

2. 연구설계 결정 시 고려사항

가. 연구대상의 정의

연구대상은 연구목적과 연구설계에 적절히 대응하는 집단으로 구성되도록 포함기준과 제외기준을 정의하여 선정해야 한다. 활용 예정인 데이터베이스 내 데이터 수집 시작일 및 종료일, 포함된 변수를 파악하여 정의해야 하며 관찰 단위는 인년(person-year), 인원수, 처방 건수 등에서 적절히 설정한다. 코호트 연구의 경우 연구대상이 선정되면 적절한 비교군을 선정해야 하는데 비노출군 또는 활성비교군(active comparator)로 정의할 수 있다. 비노출군으로 하는 경우 관심노출에 노출되지 않았다는 사실 이외에 모든 조건이 같아야 한다. 그러나 비노출군으로 정의하는 경우 치료의 시작이 데이터베이스에서 불완전하게 포착될 수 있어 비뚤림을 유발할 가능성이 크다. 따라서 일반적으로 데이터베이스에 유사하게 포착될 가능성이 큰 활성비교군을 사용한다.

나. 노출 및 결과의 정의

의약품 처방 정보는 데이터베이스마다 포함되는 의약품의 종류나 정확도가 다르기 때문에 이를 충분히 고려하여 노출을 정의한다. 즉 처방된 의약품의 코드 체계와 처방 년월일까지 모두 포함되어 있는지, 원외 처방 또는 원내 처방 의약품이 포함되어 있는지 등을 고려하여 정의해야 한다. 연구목적에 따른 주요 약물 노출 및 비교군에서의 비교 약물 노출에서 의약품 코드를 목록화하고 복수 제형 의약품일 경우 노출에 포함되는 제형을 명확히 한다. 노출 시작일이 처방일인지 조제일인지, 처방과 처방 간 간격을 노출이 계속된 것으로 간주할 것인지 여부와 노출 기간(시작 및 종료시점) 등을 설정할 필요가 있다. 처방 용량에 대한 정보가 필요할 경우 1일 투여용량 및 누적용량, 측정 기준 등의 산출 방법을 검토해야 한다. 또한 의약품을 새롭게 처방받은 환자(이하 ‘신환자’)만을 추출하여 노출 결과를 관찰할지 결정해야 하며, 필요한 경우 신환자를 선정하기 위한 조건(예: 처방 개시 전 6개월간 해당 의약품 처방이 없는 환자)을 설정한다.

데이터베이스를 활용하는 경우 데이터에 포함된 정보만으로 타당성 높은 결과를 정의하는 것이 필요하다. 연구목적에서 결과로 선정된 질병명뿐만 아니라 해당 질병의 진단 및 치료에 사용되는 의약품과 임상검사, 처치 등을 조합하여 정의함으로써 타당

성을 확보한다. 결과 정의를 위한 병명, 의약품명, 임상검사명, 또는 진료행위명 등의 코드를 목록화한다. 전자의무기록 자료처럼 임상검사 값이 포함되는 경우는 임상검사치의 이용을 검토하는 것이 권장된다. 복수 의료기관의 자료를 사용할 경우 의료기관마다 검사치의 기준 범위가 다른 점에 유의하여 결과를 정의한다.

다. 비뚤림, 교란요인 및 효과변경인자

약물 노출과 질병 발생 사이의 인과관계를 규명하고자 할 때 올바른 인과관계를 왜곡시키는 요인을 비뚤림(bias)이라고 한다. 주된 비뚤림의 종류로는 잘못된 연구대상 선정에 의한 선택 비뚤림, 연구대상으로부터 연구에 필요한 정보를 수집할 때 발생하는 정보 비뚤림이 있다. 데이터베이스를 활용한 연구에서의 비뚤림은 데이터베이스에 포함된 정보가 실제와 차이를 보일 때 발생할 수 있으며 비뚤림이 생길 것이라고 예상되는 경우, 분석의 강건성을 위하여 민감도 분석을 수행한다.

약물 노출과 질병 발생 간 아무런 인과성이 없음에도 제3의 변수로 인해 명백한 인과관계를 가지거나, 노출과 질병 발생 사이에 인과관계가 있음에도 제3의 요인을 통제하지 못하여 연관성이 없는 것처럼 나타날 경우 교란 비뚤림(confounding bias)이 발생했다고 하며, 이때 제3의 요인을 교란요인(confounder)이라고 한다. 교란요인은 연구 결과인 질병을 유발할 수 있는 위험요인이자 연구 약물과 상관관계를 가지고 있지만 약물 노출과 질병 발생의 중간 단계에 있지 않아야 한다. 데이터베이스에서 파악할 수 있는 관찰된 교란요인들을 통제하기 위해서는 적절한 통계분석 방법을 적용해야 한다. 교란요인은 데이터베이스에서 측정 가능한 것과 미측정인 것으로 구분되는데 측정 교란요인을 연구설계 단계에서 통제하는 방법은 연구대상을 제한하는 제한(restriction)방법, 교란요인이 동일 또는 유사한 대상자를 짝짓는 매칭(matching)방법이 있으며, 분석단계에서는 교란요인 별로 대상을 층화한 후 층별 분석을 하는 층화(stratification)분석, 교란요인을 다변수모형에 포함하여 보정하는 다변수회귀모형(multivariable regression model), 환자들의 다차원적인 공변량들을 일차원적인 성향점수로 변환하여 연구대상 그룹들의 공변량의 균형을 맞추는 성향점수(propensity score)가 있다.

교란요인과 달리 제3의 요인 수준에 따라 약물 노출과 질병 발생 간 연관성의 방향 또는 연관성의 정도가 달라지는 경우 제3의 요인을 효과변경인자(effect modifier)라고 한다. 예를 들어 약물 노출에 의한 부작용의 발현 정도가 남성과 여성에서 차이를

보인다면 성별은 효과변경인자로 간주할 수 있다.

1) 교란요인을 통제하는 분석방법

매칭(matching)은 환자-대조군 연구 시 환자군과 대조군 간 교란요인의 분포가 동일하도록 연구대상자를 선정함으로써 관찰된 교란요인의 효과를 통제하는 방법이다. 즉, 결과 발생에 따라 환자군을 선정한 후 환자군의 연령, 성별 등의 교란 특성과 맞추어 대조군을 선정하는 방법이다. 매칭의 종류로는 환자군의 교란요인 분포 분율에 맞추어 대조군을 선정(예: 환자군의 40%가 고혈압 환자라면 대조군의 40%가 고혈압 환자로 구성되도록 선정)하는 집단 매칭(group matching)과 각각의 환자군에 대해 교란요인이 유사하게 분포하도록 대조군을 선정(예: 첫 번째 환자가 20대 여성이라면 대조군도 20대 여성으로 선정)하는 개별 매칭(individual matching)이 있다. 매칭은 교란요인을 최소화할 수 있는 효과적인 방법이나 짝지을 교란요인이 많을수록 적절한 대조군을 찾기 어려우며, 짝지은 자료만으로 검정력을 확보할 수 있도록 표본의 수가 충분히 커야 한다.

층화(stratification)는 관찰된 교란요인을 범주별로 층을 나누어 층 내에서 분석하는 방법이다. 예를 들어 A 약물이 심근경색을 유발할 수 있다는 가설에 비만도가 교란요인으로 작용한다면 비만도를 정상, 경도 비만, 중도 비만, 고도 비만으로 층화하여 비만도가 정상인 노출군과 비노출군에서 A 약물이 심근경색에 미치는 영향을 파악하는 것이다. 층화는 적용이 간편하고 교란요인이 많지 않을 때 효과적이나 여러 개의 교란요인을 통제하기는 어렵다.

다중 회귀분석(multiple regression)은 관찰된 여러 교란요인을 동시에 통제하기 위하여 보편적으로 사용되는 통계분석방법으로 여러 개의 독립변수(약물 노출, 측정 가능한 교란요인 등)와 하나의 종속변수(결과 발생) 간 관련성을 구한다. 다중 회귀분석은 종속변수의 형태에 따라 활용 가능한 분석법이 다르다. 결과변수가 연속형 변수일 경우 선형 회귀분석, 이분형 변수일 경우 로지스틱 회귀분석, 결과 발생까지의 시간 인 경우 콕스(Cox) 비례위험모형, 비율(rate)일 경우 포아송 회귀분석을 활용한다. 결과변수의 형태에 따라 특정 다중 회귀분석을 수행할 경우, 연구의 핵심 통계 측도(예: OR)에 대한 보정된 추정치(예: adjusted OR)를 산출하여 교란요인의 효과를 보정한 노출과 결과 사이의 관련성을 파악할 수 있다.

성향점수(propensity score)는 연구대상자 간 관찰된 교란요인들의 균형을 맞추기 위해 사용되며 다차원 교란요인들을 성향점수라는 일차원 점수로 차원 축소하는 방법이다. 일차원으로 축소된 성향점수는 매칭, 층화, 공변량 보정 등에 활용할 수 있다. 그중 성

향점수 매칭(propensity score matching, PSM)는 짝지을 교란요인이 매우 많은 경우 일반 매칭 방법으로는 한계가 있기에 성향점수에 근거하여 짝짓는 방법으로 교란요인 통제를 위해 흔히 사용된다.

표 5. 교란요인을 통제하는 연구설계 및 분석방법

	연구설계	분석방법
측정 교란요인 (Measured confounder)	<ul style="list-style-type: none"> 제한(restriction): 특성이 동질한 연구대상으로 제한하여 효과를 추정 매칭(matching): 교란요인 분포가 동일 또는 유사한 대상자를 짝지어 효과를 추정 	<ul style="list-style-type: none"> 층화(Stratification): 교란요인 별로 층화 후 층별 분석 다변수회귀모형(Multivariable regression model): 교란요인을 다변수모형에 포함하여 보정 성향점수(Propensity score): 교란요인들을 점수화

2) 민감도 분석

데이터베이스를 활용한 관찰 연구는 변수의 정의부터 통계적 모델 선택까지 다양한 가정에 의존하며, 이 가정을 변형하여 시험했을 때에도 결과 값이 일관성 있게 도출되는 경우 분석 결과에 강건성(robustness)이 있다고 한다. 가정은 연구에서의 정의(노출, 결과, 교란요인의 정의 수정), 연구설계(데이터베이스 또는 연구대상의 변경 및 확장), 통계적 모델링(변수의 함수 형태 또는 정규성 검정 수정) 등에서 다양하게 변경될 수 있으며 변경된 가정을 검증하여 결과의 강건성을 확인하는 것을 민감도 분석(sensitivity analysis)이라고 한다. 민감도 분석은 여러 다른 비교그룹을 포함하거나 노출, 결과, 공변량의 정의를 변경하여 재분석하거나 선택 비뿔림이 우려되는 경우 외부데이터를 이용하여 수행할 수 있다(표 6).

표 6. 민감도 분석방법의 예시

비교군 (Comparison Groups)	여러 다른 비교군을 포함하고 결과의 견고성을 평가하는 방법으로 다른 비교군과 함께 잠재적인 비뚤림에서 관찰된 차이를 사용
노출 정의 (Exposure Definitions)	노출 정의를 변경하고 다른 정의에 따라 재분석
결과 정의 (Outcome Definitions)	결과에 대한 다른 정의를 사용하여 재분석
공변량 정의 (Covariate Definitions)	공변량 정의를 수정하여 교란요인을 얼마나 잘 처리하는지 평가
선택 비뚤림 (Selection Bias)	연구 중인 약물 중 하나에 노출된 사례의 과대표본 (또는 유사하게 과소표본)은 그럴듯하게 평가될 수 있는 범위에 걸쳐 효과 측정에 상당한 변화를 초래할 수 있다. 미측정 교란요인에 대해 식별할 수 있는 외부 검증 데이터를 사용하더라도 선택 비뚤림의 사소한 부분 이상을 설명하기는 어렵기 때문에 일반적으로 이용하는 데이터베이스에서 선택 비뚤림의 강력한 증거가 있는 경우 대체 데이터베이스를 찾는 것을 권장한다.

3. 데이터 관리

분석을 시작하기 전에 수행한 데이터 체크나 데이터 정제 및 분석용 데이터셋 추출 (선택한 변수 및 코딩 내용 등) 등의 데이터 관리에 대해 제3자가 재현 가능할 정도로 문서화 한다.

4. 분석

청구자료를 이용하여 관심있는 의약품, 치료법 등의 임상적 효과를 비교하기 위해 연구가설 및 연구설계에 적합한 방법론을 적용해야 하며 치료와 관련성이 있으면서 결과에 영향을 주는 교란요인을 통제할 수 있는 분석방법을 적용해야 한다. 사전 분석계획에는

노출과 결과에 대한 조작적 정의, 통계적 모델과 검정, 유의수준과 통계적 검정력, 결측값 처리 방법, 하위집단 분석, 교란요인 조정 방법 등에 대한 세부사항을 포함해야 한다. 사전 분석 계획과 계획 이후 추가한 분석을 정리하여 분석의 재현성을 위해 분석 과정을 알기 쉽게 문서화하고, 분석 프로그램은 제3자가 이해하기 쉽게 주석 등을 삽입하여 저장해두어야 한다.

5. 개인정보보호 및 윤리

이용하고자 하는 데이터베이스가 개인정보보호와 관련된 법률을 준수하였는지 확인해야 한다. 환자의 이름, 주민등록번호, 거주지 주소와 같은 직접 식별자 대신 환자식별키, 연령 및 추적관찰 시작 시점과 같은 간접 식별자를 이용하여 개인정보를 보호해야 한다. 또한 「생명윤리 및 안전에 관한 법률」(약칭: 생명윤리법)에 근거하여 인간대상 연구를 수행하고자 하는 연구자는 연구자가 소속된 연구기관, 교육기관 등에 소속된 기관생명윤리위원회(Institutional Review Board, IRB)로부터 연구계획서에 대한 심의를 받아야 한다.

IV. 연구계획서 및 보고서 작성

1. 연구계획서 작성방법

데이터베이스 조사 수행자는 연구수행 이전에 연구목적에 따라 연구설계, 분석방법 등 구체적인 연구계획서를 작성해야 한다. 연구계획서는 다음 요소를 포함하여 작성한다.

1) 연구계획서 제목

연구계획서에는 연구에 사용한 노출, 결과, 연구설계 등을 포함하여 연구내용을 간단히 파악할 수 있도록 기재한다.

2) 연구계획 요약

연구계획을 아래와 같은 항목을 중심으로 요약하여 기재한다.

- 연구대상 의약품의 품목 개요
: 승인일자, 승인번호, 승인요건, 유효성분, 함량 및 제형, 용법 및 용량, 효능 및 효과 등을 중심으로 기술
- 연구목적 및 가설
- 연구대상
- 조사에 활용되는 데이터베이스
- 연구설계
- 약물노출 및 안전성 결과
- 추적관찰기간
- 분석방법
- 비뚤림 및 교란요인 통제 방법
- 연구의 한계

3) 연구일정(최종 연구계획서 제출일, 중간보고일, 연구완료일, 최종보고일) 및 변경 이력(연구계획 변경사항, 연구결과 변경사항)

최종 연구계획서의 제출일, 중간보고일, 연구완료일, 최종보고일을 기재함과 동시에 확정 후 변경되는 경우에는 변경일, 변경 이유 및 내용을 기록하고 최종 결과보고서에 이를 반영하여 기술한다.

4) 연구자 및 기타책임자

책임연구자 및 연구에 참여한 연구자의 이름, 소속과 연락처(이메일 등)를 목록화하여 기재한다. 연구에 참여한 자문위원회, 검토위원회, 수행에 관여하였으나 연구자로 나열되지 않은 계약자 등이 있다면 그들의 이름, 소속, 연구에서의 역할을 목록화하여 기재한다. 또한 참여 연구자들의 데이터베이스 사용 경험, 전문성 및 자격 등 설명을 포함한다. 연구 수행 시 받은 연구비(funding)가 있다면 함께 기재한다.

5) 연구배경

연구 수행 배경이 된 사항을 기술하고, 왜 이 연구를 수행하는지를 명확히 한다. 또한, 참고한 선행 연구가 있으면 그 연구 결과와 한계 등 고려해야 할 사항을 기재한다. 주요한 의약품의 사용 상황, 같은 종류, 같은 효과가 있는 약의 유무 등에 대해서도 기재한다.

6) 연구목적

연구목적 및 가설을 기재한다. 연구목적은 아래 PICOT 요소를 포함하여 기재한다. 연구목적이 다수인 경우, 목적마다 연구계획서를 작성한다.

- 1) 조사 대상인 인구집단(Patient)
- 2) 해당 의약품 등 조사 대상인 노출(Intervention/Exposure)
- 3) 비교 의약품 등 비교 중재(Comparison)
- 4) 약물 노출 또는 예방적, 치료적 개입으로 일어날 수 있는 안전성 또는 유효성과 관련된 모든 건강상의 변화(Outcome)
- 5) 추적관찰 기간(Timing)

7) 데이터베이스 개요

연구에 사용되는 모든 데이터베이스 종류를 기재하고 연구목적이나 가설에 비추어 그 데이터 사용의 적절성을 설명한다. 기재항목은 아래 내용을 포함한다. 연구를 위해 기존의 데이터베이스에 추가해서 데이터를 수집한 경우는 그 수집 방법을 기재한다.

- 수집된 데이터의 본래 목적
- 데이터의 입수 방법
- 데이터의 신뢰성: 품질관리, 품질보증 등
- 데이터베이스 기간: 수집 또는 사용 가능한 데이터베이스의 데이터 기간
- 데이터베이스의 연계: 여러 데이터베이스를 연계해서 활용하는 경우 연계방법 기재
- 데이터 항목명: 수집 또는 사용 가능한 모든 데이터 항목명 기재하고 본 연구에서

사용할 데이터 항목명 또한 별도로 기술(부록으로 대체가능)

- 노출 및 결과 포착(capture)에 대한 데이터베이스의 적절성
- 연구의 한계로 이어질 수 있는 의료정보 데이터베이스의 특징
- 개인정보보호 및 윤리: 데이터베이스 활용 시 주의해야 할 개인정보 또는 윤리 관련 사항이 있는 경우 기재

8) 연구설계

연구설계의 선택 이유 및 한계에 대해서 기재한다. 연구설계는 의약품, 안전성 결과 및 연구가설의 특성에 따라 선택해야 한다. 비교군을 둔 디자인을 이용하는 경우가 많지만 활성 비교군을 설정하지 않는 경우(예: 적절한 활성 비교군이 없음)나 과거 비교군을 활성 비교군으로 바꿔 이용하는 경우에는 그 의미와 이유에 대해서도 기재한다.

9) 표본크기

연구에 필요한 표본크기, 표본크기 산출방법 및 근거, 검정력에 대해 기재한다. 표본크기 산출은 통계적 개념을 바탕으로 하며 고려해야 하는 사항으로는 통계적 유의수준, 검정력, 효과 크기(effect size)가 있다.

연구자는 일차적으로 연구에서 설정한 가설(A 약물은 B 질환을 유발한다)이 사실이 아니라고 가정하며 이러한 가정(A 약물과 B 질환은 관련이 없다)을 귀무가설, 연구에서 설정한 가설을 대립가설이라고 한다. 이때 연구 수행 결과에 따라 귀무가설이 참임에도 이를 기각하고 대립가설을 선정하는 오류를 제1종의 오류, 귀무가설이 거짓임에도 이를 선정하는 오류를 제2종의 오류라고 한다. 제1종의 오류와 제2종의 오류는 서로 반대되는 관계이며 제1종 오류가 일어날 확률을 0%로 둘 경우, 제2종 오류가 일어날 확률은 100%가 되므로 두 오류의 조화가 중시된다. 대개 제1종의 오류를 허용하는 유의수준(α)을 0.05로 두고 ‘1-제2종의 오류가 일어날 확률(β)’인 검정력(power)을 80%로 두어 표본의 크기를 산출한다. 마지막으로 효과크기는 연구에서 비교하는 두 집단 간 임상적 중요성이 있는 차이이다. 효과크기 산출 방법은 활용하는 통계 방법마다 다르며(예: 상관분석은 상관계수(r), 독립표본 t검정은 d 활용) 효과크기를 어느 정도로 사용할 것인지는 임상적 판단을 토대로 결정한다. 단, 모집단과 표본이 일치하는 전수대상 연구인 경우 표본크기 산출 근거를 제시하지 않을 수 있다.

10) 연구대상 정의

연구대상을 포함/제외기준으로 구분하여 정의하고 선정이유를 기재한다. 본 연구에서 사용하는 데이터 기간 범위를 기재하여 데이터베이스에서 연구대상을 얼마나 포함하는지 추정치를 제시한다. 연구질문과 관련이 있지만 데이터베이스에서 일부 연구대상을 완전히 파악할 수 없는 경우 이에 대한 해결방안을 기재한다. 관찰 시작과 종료, 중도절단(censoring) 등을 명확하게 표시하고, 이들의 정의나 조건이 적절한 이유를 기술한다. 가능한 흐름도(flow chart)로 설명한다.

11) 노출 정의

노출에 대한 정의는 활용할 데이터베이스의 코딩 시스템을 기반으로 하며 의약품 코드 목록을 연구수행 계획서에 첨부한다. 노출 기간(exposure risk window)의 정의에는 처방일 또는 조제일, 처방 기간, 다음 처방(조제)이 예상 일보다 지연됐을 때 이것을 중단으로 다룰 것인지에 관한 기준, 처방 기간이 중복됐을 때의 취급, 노출 기간 종료일 등이 포함된다. 대상자를 의약품의 신규 사용 환자(이하 ‘신환자’)에 한정하는 경우는 신환자의 정의도 기재한다. 또한, 노출 측정 기간에 영향을 줄 수 있는 관련 약물의 상호작용(예: 병용약물)도 정의한다. 각 정의에 대해서는 그 정의가 적절한 이유를 기술한다.

용량(dose)의 경우에는 환자가 처방을 준수하는지 여부를 알 수 없기 때문에 만성적 또는 고정된 용량 용법이 있는 의약품의 경우에는 데이터베이스에서 용량 정보를 유추하는 것이 적절할 수 있으며, 여러 투여 형태로 의약품이 활용되는 경우 이것이 투여량 계산에 어떻게 나타날 것인지 기재해야 한다. 필요한 경우 민감도 분석을 시행한다.

12) 결과 정의

결과의 지표가 되는 병명 코드나 기타 코드를 목록으로 계획서에 첨부한다. 병명과 치료약 또는 병명과 진료 행위 등을 조합해 결과를 정의하는 경우에는 그 조합 방식, 몇 가지 결과의 정의를 각각 기재한다. 첫 회 발생한 결과에 한정해서 측정할지, 여러 차례 발생을 측정할지에 대해서도 기재하며 어느 날짜를 결과 발생일로 할지에 대해서도 기재한다. 단, 노출이 결과의 정의에 영향을 미쳐서는 안 된다. 각 정의에 대해 그 정의가 적절한 이유를 기재한다. 연구목적으로 수집된 데이터베이스가 아닌 경우(예: 건강보험 청구자료) 결과에 대한 타당성 검증이 필요하며 타당성 검증방법에 대해 기재하도록 한다. 단, 현실적으로 타당성 검증이 불가능한 경우 그 사유를 제출하도록하고 결과가 이전에 검증된 연구의 경우 참고문헌을 인용한다.

13) 비뚤림 및 교란요인 통제 방법

데이터베이스 연구 수행자는 잠재적인 비뚤림을 식별하고 이를 처리하기 위한 프로세스를 설명해야 한다. 환자의 기본 정보, 교란요인 또는 효과 변경 인자 등 연구의 타당성에 영향을 미칠 가능성이 있는 공변량에 대해 정의하고 각 공변량의 측정 방법(노출이나 결과와 같이 코드로 정의하는 경우에는 코드 목록)을 기재한다. 교란요인을 식별하고 처리하기 위한 방법으로는 성향점수 매칭, 위험요인을 가진 환자를 배제하는 방법 등이 있다.

또한, 사용하는 데이터베이스에서 측정되지 않았지만 중요하다고 생각되는 공변량이 존재하는 경우, 대처 방법(민감도 분석의 수행 등)을 기재한다.

14) 데이터 관리방법

데이터 정제, 데이터 세트의 작성부터 최종적인 분석에 이르기까지 각 단계에서 수행하는 데이터의 취급 내용과 방법을 기재한다.

15) 통계분석

연구목적의 결과를 구하는 분석을 주요 분석으로 기재한다. 주요 분석을 보충하기 위해 수행하는 분석은 2차 분석으로 기재하며 세부사항 수준에 따라 부록에 포함한다. 사전 계획한 분석 내용은 모두 계획서에 기재하고, 연구 도중에 추가되는 분석과는 명확하게 구별해야 한다. 분석에 이용하는 모델, 표본 크기 추정, 유의수준 및 통계적 검정력, 결측값 처리, 교란요인 통제 방법 등을 상세하게 기술한다. 교란요인을 통제하기 위한 통계적 기법은 관련된 가정 및 한계에 대한 명확한 설명을 포함한다.

매칭 분석을 수행하는 경우는 매칭 요인과 방법 및 이들을 사용하는 이유를 기재한다. 분석 결과의 견고성을 확인하기 위해 민감도 분석을 수행하는 경우는 그 내용을 기재한다. 미리 계획한 민감도 분석은 모두 기재하고, 연구 도중에 추가한 민감도 분석과는 구별할 수 있게 해 둔다.

16) 연구의 한계

사용하는 데이터베이스에 포함된 정보량이 한정되어 있거나 정보의 타당성을 충분히 확인할 수 없는 등의 이유로 적절한 대처가 어려워 연구 결과의 타당성을 충분히 확보할 수 없을 때 이것을 예상되는 연구의 한계로 기재한다. 특히 연구의 한계가 결론의 내적 타당성 및 일반화 가능성(외적 타당성)에 어느 정도 영향을 미치는가에 대해 예상할 수 있는 범위 내에서 기재한다.

17) 결과공표방법

의약품 안전성 평가에 영향을 미치는 중요한 결과는 공표되어야 한다. 결과의 공표를 예정하는 경우에는 그 방법과 공표처를 기재한다.

18) 개인정보보호 및 윤리

개인정보의 이용 여부를 기재한다. 개인정보를 취급하는 경우에는 보안상의 대응 방법에 대한 개요를 기재한다. 개인정보에 관한 항목이 제거된 데이터 소스를 사용하는 경우에는 그 취지를, 연구 과정에서 데이터 소스로부터 개인정보에 관한 항목이 제외된 경우는 주된 처리 내용을 기재한다. 그 외에 윤리적으로 유의해야 할 사항이나 IRB, 대처해야 할 사항이 있으면 기재한다.

19) 이익상충관계

연구에 관여하는 모든 인원에게 대해, 연구의 타당성에 영향을 줄 수 있는 모든 요인에 대한 이익상충관계를 기재한다. 이익상충이 존재하는 경우에는 어떻게 연구의 투명성을 확보할 것인지 그 방법을 명시한다.

20) 참고문헌

연구계획서에 참고로 한 문헌을 기재한다.

2. 연구보고서 작성방법

가. 신청인(보고자)

- 1) 데이터베이스 연구대상 품목이 제조품목인 경우에는 업허가 번호를 기재하고, 수입 품목인 경우에는 수입업 신고 번호를 기재하되 수입업 신고 번호가 없는 경우에는 기재하지 아니한다.
- 2) 제조(영업)소 명칭, 제조(영업)소 소재지 및 성명은 식약처장으로부터 허가를 받은 명칭, 소재지 및 대표자 성명을 기재하고, 대표자의 생년월일을 기재한다.

나. 제조원

- 1) 제조품목의 경우에는 식약처장으로부터 허가받은 제조업소명과 소재지를 기재한다.

분공장인 경우에도 허가사항에 따라 기재한다(예 : ○○제약 서울 공장).

- 2) 수입품목인 경우에는 해당 품목허가증에 기재된 제조업소명, 소재지 및 제조국을 기재한다.

다. 데이터베이스 연구대상품목 정보

데이터베이스 연구 대상품목의 제품명, 성분명, 허가번호, 허가년월일, 효능·효과와 용법·용량을 기재한다.

라. 데이터베이스 연구결과

1) 연구결과 요약

모든 하위 내용을 포함하는 연구 결과보고서에 대한 요약을 기재한다. 제목 등의 키워드를 제외하고 500 단어 미만의 단어를 포함하도록 한다.

가) 제목

나) 주제어(keywords)

※ 연구의 주요 특징을 나타내는 5개 이하의 주제어를 제시한다.

다) 배경 및 목적

라) 연구설계 및 방법

마) 데이터베이스 및 주요 변수

바) 결과

사) 고찰

아) 신청인(보고자)

자) 책임연구자의 이름 및 소속

2) 약어 목록

연구 결과보고서에서 사용한 주요 약어를 정리하여 목록으로 작성한다.

3) 책임연구자, 연구자 및 참여인력

가) 책임연구자 및 연구에 참여한 연구자의 이름, 소속과 연락처(이메일 등)를 목록화하여 기재한다.

나) 연구에 참여한 자문위원회, 검토위원회, 수행에 관여하였으나 연구자로 나열되지 않은 계약자 등이 있다면 그들의 이름, 소속, 연구에서의 역할을 목록화하여

기재한다.

4) 진행단계

연구계획서 제출일부터 결과보고서 제출일까지의 날짜를 아래의 표와 같이 정리하여 기재한다. ‘계획일’은 연구계획서에 기재한 연구일정에 표시된 것과 동일하게 작성한다. ‘설명’은 계획일과 실제일 간 차이에 대한 간단한 설명을 작성한다.

마일스톤	계획일	실제일	설명
연구계획서 제출	[날짜]	[날짜]	[설명]
데이터 획득	[날짜]	[날짜]	[설명]
중간보고	[날짜]	[날짜]	[설명]
연구 완료	[날짜]	[날짜]	[설명]
결과보고서 제출	[날짜]	[날짜]	[설명]

5) 연구배경

연구 수행 배경이 된 사항을 기술하고, 왜 이 연구를 수행하는지를 명확히 한다. 또한, 참고한 선행 연구가 있으면 그 연구결과와 한계 등 고려해야 할 사항을 기재한다. 연구대상품목의 사용 상황, 같은 종류, 같은 효과가 있는 약의 유무 등에 대해서도 기재한다.

6) 연구목적

연구목적 및 가설을 기재한다. 연구목적은 PICOT 요소를 포함하여 기재한다.

7) 수정사항

연구계획서 제출일 이후 연구계획에서 수정 및 업데이트된 내용을 아래의 표와 같이 정리한다.

번호	변경일	연구계획서 내 목차	수정 전 내용	수정 후 내용	근거
1	[날짜]	[설명]	[설명]	[설명]	[설명]
2	[날짜]	[설명]	[설명]	[설명]	[설명]
...	[날짜]	[설명]	[설명]	[설명]	[설명]

8) 연구방법

다음의 내용을 포함하여 연구방법을 설명한다.

가) 연구설계

연구목적 규명을 위해 선정 및 수행한 연구설계에 대해서 기재한다. 연구목적에

답하기 위한 연구설계의 강점을 설명한다.

나) 설정

추적관찰 기간 등 연구에서 설정한 연구방법과 관련된 날짜를 기재한다.

다) 연구대상

연구대상의 포함 및 제외기준 정의를 기재하고 포함 및 제외기준을 구현하기 위한 조작적 정의(질병 진단 여부, 약물 처방 여부 등)를 서술한다. 관찰 시작과 종료, 중도절단(censoring) 등을 명확하게 표시하고, 이들의 정의나 조건이 적절한 이유를 기술한다. 표본이 되는 연구대상자의 수를 기재하고 표본 크기의 적절성을 판단하기 위하여 검정력을 기재한다. 단, 모집단과 표본이 일치하는 전수대상 연구인 경우 검정력을 기재하지 않을 수 있다.

라) 데이터베이스

연구에 사용한 데이터베이스의 종류와 설명(데이터 수집 배경 등)을 기재하고 연구목적에 비추어 해당 데이터베이스의 사용이 적절한지 설명한다.

연구 수행을 위한 데이터 코딩의 타당성(두 명 이상 연구자의 독립적인 검토 등)에 대하여 기재한다.

마) 주요 변수

대상품목에 의한 약물 노출, 결과, 교란요인 및 효과 변경 인자 등에 대하여 기재하고 연구를 수행하기 위해 해당 변수를 데이터베이스에서 어떻게 확인하였는지 각 변수별 조작적 정의(operational definition)를 서술한다.

바) 비뚤림

비뚤림의 잠재적인 원인을 평가하고 해결하고자 수행한 모든 처리 과정을 기재한다.

사) 통계분석

(1) 주요 분석

연구에서 적용한 통계적 방법 및 모델을 기재하고 이를 사용한 근거를 명시한다. 교란요인을 통제하기 위한 통계적 방법은 관련된 가정 및 한계에 대한 명확한 설명을 포함한다. 매칭 분석을 수행하는 경우는 매칭 요인과 방법 및 이들을 사용하는 이유를 기재한다.

(2) 결측값 처리

결측값을 처리하기 위해 수행한 방법을 기재한다.

(3) 민감도 분석

연구에서 통계적으로 수행된 모든 민감도 분석과 방법 및 이유를 기재한다.

(4) 사전 통계분석 계획의 수정

연구계획서 제출 시 기재한 통계분석 계획과 다르게 수행하였거나 추가적으로 수행한 분석에 대하여 기재한다.

아) 결과

수행된 연구결과를 토대로 표(table) 및 그림(figure) 등을 활용하여 시각화한다. 표 및 그림은 본문 또는 별첨에 기재한다.

(1) 연구대상

연구대상의 포함 및 제외기준에 따라 각 단계에서 포함 및 제외된 대상 수와 추적 및 평가 완료된 대상 수를 차례로 기재한다. 가능한 흐름도(flow chart)로 설명한다.

(2) 기술 통계량

연구대상의 주요 특징(예: 연령, 성별, 거주지역 등), 교란요인(예: 동반질환, 병용 약물 등) 및 연구가설과 관련된 기타 변수에 대한 양적 정보를 각 변수의 결측값을 포함하여 표로 작성한다.

(3) 주요 결과

연구에서 설정한 결과의 유무에 따른 연구대상의 수를 기재한다. 보정되지 않은 분석값과 보정된 분석값을 사전에 지정된 유의수준을 토대로 작성한다.

(4) 기타 결과

주요 결과에 포함되지 않은 하위그룹 분석, 민감도 분석 등을 기재한다.

자) 고찰

주된 연구결과를 간략히 언급하고 이를 뒷받침하거나 반박하는 선행 연구의 내용을 기재한다. 대상품목의 유익성-위해성 균형에 영향을 미칠 수 있는 결과에 대하여 기재한다. 데이터의 품질 또는 무결성에 영향을 주는 한계점, 비뚤림 등 연구의 한계가 될 수 있는 사항에 대하여 논의한다. 연구목적, 한계, 유사 선행 연구 및 기타 관련 내용으로부터 얻은 결과에 대한 전반적인 해석을 해야 하며, 해석은 연구의 배경이 되는 안전성 문제 등과 관련하여 설명한다. 연구대상의 특성과 포함 및 제외 기준, 데이터베이스의 특성 등을 고려하여 연구 결과의 일반화 가능성을 기재한다.

차) 기타 정보

연구계획서 혹은 정기보고서에서 기술되지 않은 사항에 대한 추가 또는 보완된 정보를 기재한다.

카) 결론

연구를 통해 도출된 연구의 주요 결론을 별첨으로 하여 기재한다.

타) 참고 문헌

참고문헌

- 김호. (2008). 튜토리얼: R과 SAS를 이용한 표본수 및 검정력 계산. 한국보건정보통계학회지, 33(1), 1-10.
- 대한의료정보학회. (2014). 보건의료정보학 제3판. 서울: 현문사.
- 이수현. (2018). 의료 빅데이터 분석을 위한 공통 데이터 모델. 정보통신기획평가원 주간기술동향. 1878. 14-24.
- 박병주. (2016). 약물역학 제2판 (대한약물역학위해관리학회 역). 서울대학교출판문화원. 302-320.
- 손원. (2018). 보건의료 빅데이터 종류와 내 연구에 맞는 자료 선택하기. 대한간학회, 191-194.
- 장은진. (2017). 보건의료빅데이터 활용방안 - 성과연구 중심으로. HIRA 정책동향, 11(1). 8-16.
- 최남경. (2015). 건강보험심사평가원 자료를 활용한 보건사회약학연구. 한국보건사회약료경영학회지, 4(1), 42-49.
- 신주영(2020), 'Real-world data 및 evidence 기반으로 한 바이오의약품의 시장 출시 후 안전관리'
- ABPI. (2011). Demonstrating value with Real World Data: A practice guide.
- Corrao, G. (2013). Building reliable evidence from real-world data: methods, cautiousness and recommendations. Epidemiology, Biostatistics and Public Health, 10(3), 1-40.
- EMA. (2011). Guidelines for Good Database Selection and use in Pharmacoepidemiology Research.
- EMA. (2012). Guidance for the format and content of the protocol of non-interventional postauthorisation safety studies.
- ENCePP. (2018). The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) - Guide on Methodological Standards in Pharmacoepidemiology.

FDA. (2013). Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets.

FDA. (2017). Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices Guidance for Industry and Food and Drug Administration Staff.

FDA. (2018). Framework for FDA: Real-world evidence program.

FDCA(Federal Food, Drug & Cosmetic Act) section 505(o)(3).

Gordis, L. (2009). 역학 (한국역학회 역). 서울: EPUBLIC.

ISPE, & ISPOR. (2017). Good practices for Real-World Data studies of treatment and/or comparative effectiveness: Recommendations from the Joint ISPOR-ISPE special task force on real-world evidence in health care decision making.

ISPE. (2020). ISPE' s position on real-world evidence (RWE).

PMDA. (2014). 의료정보 데이터베이스 등을 이용한 의약품 안전성 평가에 있어서 약물역학 연구 실시에 관한 지침(医療情報のデータベース等を用いた医薬品の安全性評価における薬剤疫学研究の実施に関するガイドライン).

PMDA. (2017). 제조판매 후 의약품 안전성 감시의 의료정보 데이터베이스 이용에 관한 기본적 생각에 대해(製造販売後の医薬品安全性監視における医療情報データベースの利用に関する基本的考え方について).

PMDA. (2017). 제조판매 후 의약품 안전성 감시에서 의료정보 데이터베이스의 이용에 관한 기본적인 생각에 대해(製造販売後の医薬品安全性監視における医療情報データベースの利用に関する基本的考え方について).

PMDA. (2018). 의약품 제조판매 후 데이터베이스 조사의 신뢰성 담보에 관한 유의사항(医薬品の製造販売後データベース調査における信頼性担保に関する留意点について).

PMDA. (2018). 제조판매 후 데이터베이스 조사 실시 계획서의 기재 요령(製造販売後データベース調査実施計画書の記載要領).

Velentgas, P., Dreyer, N.A., Nourjah, P., et al. (2013) Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. US: Agency for Healthcare Research and Quality.