

Correspondence to:

Nam-Kyong Choi
Department of Health Convergence,
Ewha Womans University,
52 Ewhayeodae-gil, Seodaemun-gu,
Seoul 03760, Korea
Tel: +82-2-3277-6585
Fax: +82-2-3277-2867
E-mail: nchoi@ewha.ac.kr

Received: February 26, 2024

Revised: March 12, 2024

Accepted: March 13, 2024

Current Issues and Improvement Strategies for Promoting the Utilization of National Health Insurance Claims Data

Hee-Jin Kim¹, Dong Wook Kim², Myeong Gyu Kim^{3,4}, Mi-Sook Kim⁵,
Ye-Jee Kim⁶, Jeong Ah Kim¹, Suvin Park¹, SeungJin Bae^{3,4},
Aesun Shin⁷, Ju-Young Shin⁸, Song Vogue Ahn¹, Jeonghoon Ahn¹,
Bo Ram Yang⁹, Heehyun Won¹, Seung-Mi Lee¹⁰, Joongyub Lee⁷,
Hui-Eon Lee⁴, Sun-Young Jung¹¹, Haerin Cho¹ and Nam-Kyong Choi^{1,4}

¹Department of Health Convergence, College of Science and Industry Convergence, Ewha Womans University, Seoul, Korea

²Department of Information and Statistics, Research Institute of Natural Science, Gyeongsang National University, Jinju, Korea

³College of Pharmacy, Ewha Womans University, Seoul, Korea

⁴Graduate School of Industrial Pharmaceutical Science, College of Pharmacy, Ewha Womans University, Seoul, Korea

⁵Medical Research Collaborating Center, Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea

⁶Department of Clinical Epidemiology and Biostatistics, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

⁷Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea

⁸School of Pharmacy, Sungkyunkwan University, Suwon, Korea

⁹College of Pharmacy, Chungnam National University, Daejeon, Korea

¹⁰College of Pharmacy, Daegu Catholic University, Gyeongsan, Korea

¹¹College of Pharmacy, Chung-Ang University, Seoul, Korea

국내 건강보험 청구자료의 활용 촉진을 위한 현안 및 개선방안

김희진¹, 김동욱², 김명규^{3,4}, 김미숙⁵, 김예지⁶, 김정아¹, 박수빈¹, 배승진^{3,4}, 신애선⁷, 신주영⁸, 안성복¹,
안정훈¹, 양보람⁹, 원희현¹, 이승미¹⁰, 이종엽⁷, 이희연⁴, 정선영¹¹, 조해린¹, 최남경^{1,4}

¹이화여자대학교 신산업대학교 융합보건학과, ²경상국립대학교 자연과학대학 정보통계학과, ³이화여자대학교 약학대학, ⁴이화여자대학교 약학대학 제약산업학과, ⁵서울대학교병원 의생명연구원 의학연구협력센터, ⁶서울아산병원/울산대학교 의과대학 의학통계학과, ⁷서울대학교 의과대학 예방의학교실, ⁸성균관대학교 약학대학, ⁹충남대학교 약학대학, ¹⁰대구가톨릭대학교 약학대학, ¹¹중앙대학교 약학대학

Abstract

While health insurance claims data in Korea have the potential as evidence for drug regulatory decision-making, its effective utilization remains limited. This study aims to identify the challenges encountered by researchers in utilizing claims data and discuss improvement strategies. We summarized practical difficulties encountered by researchers who have experience using claims data from the National Health Insurance Service and the Health Insurance Review and Assessment Service. Challenges encountered by researchers included difficulties in acquiring drug data, delays in data provision, limited provision of information on specific conditions, drugs, and treatments, small data capacity, short data usage periods, and spatiotemporal considerations when accessing data. To maximize the utility of claims data in drug regulatory decision-making, continuous communication between researchers and data providers is necessary for addressing these challenges. (PeRM 2024;16:1-10)

Key Words: Healthcare administrative claims; Real-world data; Real-world evidence; Pharmaceutical research; Regulatory science

Copyright © Korean Society for
Pharmacoepidemiology and Risk
Management



This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

서론

식품의약품안전처 가이드라인에 따르면 Real-World Data (RWD)란 기존 중재적 임상시험으로부터 수집된 데이터가 아닌 다양한 유형의 의료정보데이터를 포괄하는 용어로 의료전문가, 보험자 또는 의료소비자가 직접 수집하거나 체계적으로 수집되는 데이터이며 대표적으로 건강보험 청구자료, 의무기록자료, 공통데이터모델(Common Data Model, CDM), 레지스트리(registry), 의약품이상사례보고자료가 있다.¹⁾ 이 중에서도 건강보험 청구자료는 전국민 97%의 인구학적 및 급여의료이용에 대한 정보를 포함하고 있으며 뛰어난 접근성, 타당성 및 대표성으로 인해 국내에서 가장 활용성이 높은 RWD이다. 국민건강보험공단(이하 '공단') 및 건강보험심사평가원(이하 '심평원')은 의료기관에게 의료서비스에 대한 진료비용 일부를 지불하는 과정에서 발생한 청구자료를 학계 및 제약업계가 연구용으로 활용할 수 있도록 제공하고 있다.²⁾ 심평원은 2009년부터 연구용 청구데이터 개방을 준비하여 2015년부터는 본격적으로 일반 연구자들이 활용할 수 있도록 지원하였으며 2015년부터 2020년까지 1,110건의 연구과제에 데이터 제공을 승인하였고 현재도 활발히 제공 중이다.³⁾ 공단 또한 2014년부터 연구자들이 이용할 수 있도록 개방하여 2014년부터 2023년까지 9,870건의 연구과제에 데이터 제공을 승인하였으며 제공 건수가 해가 지날수록 증가하고 있어 국내 청구자료를 활용한 연구는 현재까지도 활발히 진행되고 있는 것으로 보인다.⁴⁾

이렇게 청구자료의 활용이 고도화되는 것에 비해 국내 의약품 개발 전주기에서 규제적인 목적으로는 아직까지 적극적인 활용이 되지 않고 있으며 여기에는 관련 법/제도의 미비, 이해관계자들의 이견, 개인정보보호문제 등 복합적인 문제가 존재한다.⁵⁾ 제약업계의 경우 영리적 사용을 우려해 청구자료 이용에 대한 심의승인이 저조할 뿐 아니라 학계를 통해서 또는 센터 방문을 통해서만 자료에 접근이 가능하고 제공받는 정보 또한 제한되는 경우가 많다.⁶⁾ 학계는 자료접근에 대한 장벽이 제약업계에 비해 낮지만 자료를 활용하여 규제사결정에 적합한 고품질 연구결과를 도출하기까지는 여러 가지 어려움을 겪고 있다.

따라서 본 연구는 공단과 심평원의 자료 신청과정과 신청 및 활용 시 유의사항을 간단히 소개하고 자료의 활용성을 촉진하기 위해 연구자들이 자료를 신청 및 활용하는 과정에서 겪는 현실적인 문제점과 개선방안을 파악하고자 수행되었다.

청구자료 신청방법과 신청 및 활용 시 유의사항

공단 및 심평원이 보유하고 있는 자료의 정보와 신청방법은 거의 유사하나 가용범위 및 이용규정 등에 약간의 차이가 있다(Table 1). 공단이 제공하고 있는 자료는 표본자료로 제공되는 표본코HORT(National Sample Cohort, NHIS-NSC), 건강검진코HORT, 노인코HORT와 공단이 보유한 전수자료를 연구목적에 따라 가공하여 제공하는 맞춤형DB (NHIS customized DB)가 있다. 심평원이 제공하고 있는 자료는 표본자료인 입원환자데이터셋(National Inpatient Sample, HIRA-NIS), 전체환자데이터셋(National Patient Sample, HIRA-NPS), 고령환자데이터셋(Adult Patient Sample, HIRA-APS), 소아청소년환자데이터(Pediatric Patient Sample, HIRA-PPS)와 심평원이 보유한 전수자료를 연구목적에 따라 가공하여 제공하는 맞춤형 연구자료(HIRA customized DB)가 있다. 본 연구는 공단의 맞춤형DB, 심평원의 맞춤형 연구자료의 신청방법 및 유의사항을 중심으로 소개하였다.

1. 국민건강보험공단 자료

1) 공단 자료 신청방법

전국민의 2002년부터 2022년까지의 자료를 연구목적에 맞게 추출하여 제공하는 맞춤형 자료 신청은 건강보험 자료 공유서비스(nhiss.nhis.or.kr)에서 가능하며 홈페이지 내에 기본정보 및 연구과제요약서를 기

Table 1 Data availability and rules of usage for National Health Insurance Service and Health Insurance Review and Assessment Service

	National Health Insurance Service (NHIS)		Health Insurance Review & Assessment Service (HIRA)	
Database	NHIS-NSC 2.2 DB	NHIS customized DB	HIRA-NPS	HIRA customized DB
Population/ sample cohort	Sample cohort (about 1,000,000; 2% of population)	Population	Sample of total patients 2009-2018 (about 1,400,000/year; 3% of population); Sample of total patients 2019-2020 (about 1,000,000/year; 2% of population)	Population
Data period	2002-2019	2002-2022	2009-2020	Between 2007 and 8 months ago from the present
Components	Qualification DB, Death DB, Treatment statement DB (T20), Details of treatment DB (T30), Type of disease DB (T40), Details of prescription DB (T60), Medical institution DB, Long term care DB, Health screening DB		Treatment statement DB (T200), Details of treatment DB (T300), Type of disease DB (T400), Details of prescription DB (T530), Medical institution DB	
Access	Remote	Center visit	Remote	Remote or center visit
Available DB usage period	- Virtual room: maximum of 6 months - Storage: maximum of 1 year or approved end date for the study - Reuse: until the end date of storage or study period	- Virtual room: maximum of 1 year (extension of up to 3 months is possible) - Storage: maximum of 1 year or until approved end date for the study - Reuse: until the end date of storage or study period	- Virtual room: maximum of 2 months - Storage: maximum of 1 year or until approved end date for the study - Reuse: until the end date of storage or study period	- Virtual room: maximum of 90 days (extension of up to 180 days is possible) - Storage: maximum of 1 year or until approved end date for the study - Reuse: until the end date of storage or study period
Available DB capacity	421.1 GB	Around 700 GB	34 GB	Around 300 GB
Available virtual room capacity	Around 1.5 TB	Around 1.5 TB	Around 1 TB	Around 1 TB
Access authority	All pre-approved research participants (simultaneous access not possible)		Up to 2 pre-approved analysts (simultaneous access not possible)	
The average duration from application to DB acquirement	About 3-4 months (varies by study)	About 6-7 months (varies by study)	About 2 months (varies by study)	About 3-5 months (varies by study)
Characteristics of DB	- Random sampling based on eligible individuals as of 2006 - Some masked information (e.g., gender-specific conditions, psychological disorders, etc.)	- Request for single-brand drug is impossible - Medicine and treatment information (T30, T60) not provided* - Only the first 4 digits of the principal components code are provided†	- Annual medical service usage history of sample patients by subject (data sets by year cannot be linked to each other) - Cross-sectional analysis is preferred - Some masked information (e.g., psychological disorders, rare diseases, abortion, etc.)	- Request for single-brand drug is impossible - Lack of qualification information (e.g., insurance premium level, area of residence, etc.) and date of death - Some masked information (e.g., psychological disorders, rare diseases, sexually transmitted infections, AIDS, abortion, COVID-19 information, etc.)* - Only the first 4 digits of the principal components code are provided†

Table 1 Continued

	National Health Insurance Service (NHIS)		Health Insurance Review & Assessment Service (HIRA)	
Statistical analysis tool	SAS, R, SPSS, STATA	SAS, R, SPSS, STATA, Python	SAS, R	SAS, R
Data export	Maximum of 25 MB (spending at least 3 days)		Maximum of 300 MB (limited to 10 MB per file)	
Potential linkage with other DB	Not applicable	Vaccination DB in KDCA, National Cancer Center, Public hospitals, Statistics Korea [†]	Not applicable	Vaccination DB in KDCA, National Cancer Center, Public hospitals [†]

NHIS, National Health Insurance Service; NSC, National Sample Cohort; HIRA, Health Insurance Review & Assessment Service; NPS, National Patient Sample; DB, database; KDCA, Korea Disease Control and Prevention Agency; AIDS, Acquired Immune Deficiency Syndrome; COVID-19, Coronavirus disease.

*A separate protocol or application is required if you need it.

[†]A separate application is required if you want to receive more than 4 digits.

[‡]A separate permission is required.

입하고 사전 승인된 IRB계획서, IRB결과통지서, 소속 증빙서류와 공단 서식에 맞춘 모든 연구자의 동의서 및 서약서를 제출해야 한다. 공단은 연구 목적에 따라 정책연구, 학술연구, 공동연구 등으로 구분하며 연구 유형에 따라 자료제공범위, 필요서류 및 수수료가 다르기 때문에 하고자 하는 신청 시 연구가 공단규정에서 어느 유형에 해당하는지 사전에 확인해야 한다. 정책연구는 정책연구보고서를 제출하기 위한 연구를 의미하며 연구용역계획서 등의 정책연구임을 입증할 수 있는 서류와 주관부서 발행 공문을 함께 첨부해야 한다. 자료는 자격 및 보험료DB, 사망DB, 명세서DB (T20), 진료내역DB (T30), 상병내역DB (T40), 처방교부상세DB (T60), 요양기관DB, 장기요양DB, 건강검진DB 등 보유하고 있는 모든 자료에 대해 신청이 가능하지만 신청 시 각각의 자료가 연구목적에 적합한지에 대한 확인이 필요하다. 예를 들어, 건강검진 자료의 경우 일반 청구자료에서 파악하기 어려운 검사수치, 생활습관 등의 유용한 정보가 포함되어 있지만, 검진 간격이 2년이기 때문에 변수에 따라 일부 결측치가 존재할 수 있다. 신청 마지막 단계로 변수명, 대상자 정의 등 최종 자료추출에 대한 상세조건을 홈페이지 내에 설정한다. 기존에는 추출조건을 공단의 서식(엑셀)에 기재하여 제출했으나 최근 자동추출 방식으로 변경되었다. 자료를 이용하고자 하는 모든 연구자들의 등록 및 본인 인증을 거치면 신청이 최종적으로 완료된다.

2) 약제자료 신청

공단이 보유하고 있는 모든 자료에 대한 신청이 가능하나 약제 및 치료재료 정보(T30 및 T60)를 이용하고자 하는 경우 연구에 필요한 약제정보를 공단 약제활용계획서 서식에 맞춰 작성하여야 한다. 또한 요청하고자 하는 약제의 주성분코드, 제품코드, 제품명, 업체명, 제품 수 등의 정보를 포함한 전체약품목록을 공단에서 제공하는 엑셀서식에 약효분류 별로 그룹화하여 작성 및 제출해야 한다. 단, 약제의 제조사가 식별될 수 있는 단독등재 제품의 경우 제공이 되지 않고 기본적으로 주성분코드 9자리 중 4자리만 제공하고 있어 5자리 이상(즉, 용량 및 제형)의 정보를 제공받고자 하는 경우 별도의 사유서를 작성하여 제출해야 한다.

3) 자료심의절차

제출한 신청서에 대한 심의는 매월 둘째 넷째 주 화요일에 개최되고 심의일 2주 전에 신청서 제출이 마감되므로 홈페이지에 공표된 마감일 및 심의일을 반드시 확인해야 한다. 또한 심의일 2주 전부터는 신청서 접수가 마감되기 때문에 제출한 신청서 및 추출조건에 대한 변경이 불가하다. 연구과제요약서 및 IRB계획서가 심의에서 안건으로 상정되기 때문에 작성 시 각별히 주의해야 하며 연구과제요약서, IRB계획서 및 홈페이지 자동추출 시스템에 기입한 자료추출 조건이 상세하고 모두 일치하게 작성되어야 한다. 신청일부터 심의결과 통지까지는 약 1~2개월이 소요되며 심의가 결정되고 나면 자료구축이 완료될 때까지 약 3~4개월정도 대

기가 필요하다. 자료 구축이 완료되고 나면 원주에 소재한 공단 본원에 방문하여 자료가 제대로 추출되었는지 확인하는 절차를 거친다. 자료확인 후에는 분석가능한 가상화 룸을 배정받아야 하는데 사용하고자 하는 분석도구를 SPSS, STATA, SAS, R, Python 중 선택할 수 있으며 Python은 특정 가상화 룸에서만 이용할 수 있어 사전에 별도로 요청해야 한다. 또한 이미 가상화 룸 이용자 및 대기자가 많다면 보편적으로 룸 배정을 위한 1-2개월 정도의 대기기간이 필요하다. 요약하면 자료신청부터 자료를 제공받기까지 짧게는 3-4개월에서 평균 6-7개월 정도가 소요될 수 있는데 만약 자료 심의 이전에 보완요청이 발생하거나 추출된 자료 확인 시 원하는 대로 자료가 추출되지 않았을 경우에는 재추출이 필요하여 대기기간이 더 많이 지연될 수 있기 때문에 최초 신청 시 꼼꼼한 서류제출 및 연구기간에 여유를 두고 자료를 신청할 것을 권장한다.

4) 공단 자료 신청 이후 이용 시 유의사항

공단의 표본자료는 어디에서나 원격으로 분석이 가능하지만 맞춤형 자료의 경우 9시-18시 사이에 공단의 지역 분석센터를 방문해야만 접근이 가능하기 때문에 약간의 시공간적 제약이 존재한다. 공단 가상화 룸의 가용 용량은 연구 당 1.5 TB 정도이지만 충분한 분석 및 저장공간을 확보하고자 공단은 700 GB 내외의 자료를 신청할 것을 권고하고 있다. 제공받은 자료이용기간은 최대 1년, 연장 3개월, 자료보관은 3개월 단위로 최대 1년까지 신청 및 재사용이 가능하고 필요 시 보관연장 신청도 가능하나 자료이용, 자료보관, 재사용은 모두 승인된 연구기간 내에 이루어져야 한다.

2. 건강보험심사평가원 자료

1) 심평원 자료 신청방법

전국민의 2007년부터 현재날짜 기준 8개월 이전까지의 자료를 연구목적에 맞춰 추출하여 제공하는 심평원 맞춤형 자료 신청은 보건 의료 빅데이터 개방시스템(opendata.hira.or.kr)에서 가능하다. 먼저 자료추출 조건을 홈페이지에 가입해야 하는데 필요한 상병코드, 약물코드 등의 정보를 코드명, 영문명, 국문명 등으로 검색할 수 있고 입력과정의 대부분이 체크박스 등의 폼 형태로 시스템화 되어있어 매우 간편하다. 심평원은 명세서DB (T200), 진료내역DB (T300), 상병내역DB (T400), 원외처방내역DB (T530), 요양기관DB를 제공하나 자격 및 보험료, 사망정보를 보유하고 있지 않아 제공되지 않는다. 또한 정신질환, 희귀질환, 성매개 감염병, 후천성면역결핍증, 유산, 학대, 코로나19 상병정보를 특정 상병정보로 분류하여 마스킹 후 제공하는데 만약 이를 이용해야 하는 연구의 경우에는 특정 상병정보 활용 신청서를 제출해야 한다. 또한 특정 상병코드의 가장 적은 연평균 환자수가 일정 수 미만일 경우에는 연령은 5세 단위로 범주화 되어 제공되며 의료기관 소재지 정보는 범주화 되거나 제공이 불가하다. 2024년도 부터는 진료기간 5년 이상의 자료를 요청하는 경우 연구내용에 해당 기간의 자료가 필요한 사유를 작성하여 제출해야 한다.

위와 같은 자료추출 조건을 홈페이지에 입력하여 신청하면 담당자가 1:1로 배치되어 해당 추출조건을 기반으로 자료의 물리적 제공 가능성을 검토하는 상담과정이 먼저 진행된다. 담당자가 1차적으로 자료를 추출할 때까지 약 2개월정도가 소요되며 용량이 300 GB를 초과하거나 제공되기 어려운 정보가 발견되거나 자료 추출 조건에 모순이 발견되는 등 수정사항이 발생하면 이에 대한 수정을 요청한다. 신청자가 담당자 안내에 따라 수정하여 재신청 하면 이에 대한 2차 상담이 진행되고 상담자와 신청자간 합의에 도달할 때까지 상담을 반복한다. 최종적으로 상담 완료되면 더 이상 자료추출에 대한 수정이 불가하며 사전 준비한 승인된 IRB계획서, IRB결과통지서, 소속 증빙서류와 서식에 맞춘 이용개요서, 동의서 및 서약서를 첨부하고 이용자를 등록한다. 공단과 유사하게 연구 유형이 일반연구와 정책연구로 분류되어 있어 수수료 및 제출서류에 약간의 차이가 있는데 정책연구의 경우 연구용역계획서 등 정책관련 연구임을 입증할 수 있는 서류와 주관부서 발행 공문을 제출해야 한다. 이용자 등록은 인원 제한이 없지만 데이터에 접근하고 분석할 수 있는 분석자는 최대 2명만 지정할 수 있다.

2) 약제자료 신청

공단 대비 제출서류가 간소하지만 기본적으로 주성분코드 9자리 중 4자리만 제공하는 것으로 2024년 변경되어 그 이상의 정보를 제공받고자 하는 경우 약제활용신청서를 작성하여 제출해야 하고 심의를 통해 약제의 제조사가 식별될 수 있는 단독등재 제품의 경우 제공이 되지 않는다.

3) 자료심의절차

관련 서류제출 및 이용자 등록이 완료되면 심의가 진행되는데 매월 둘째 넷째 주에 개최된다. 결과적으로 신청일부터 상담을 거쳐 심의결과 통지까지는 약 3-5개월 정도의 기간이 소요되나 상담이 반복되거나 대기자가 많을 경우 더 장기화될 가능성이 있다. 심의가 결정되고 수수료 납부 및 이용승인요청을 하면 추출된 자료를 즉시 이용할 수 있지만 자료를 물리적으로 확인할 수 있는 과정이 없어 상담 이후에는 재추출이 불가하다.

4) 심평원 자료 신청 이후 이용 시 유의사항

심평원 맞춤형 자료는 직접 방문 없이 원격으로 분석이 가능하지만 이용승인요청을 통해 사전 승인된 고정 IP에서만 접근할 수 있다. 분석도구는 SAS와 R만 제공되고 가상화 룬의 가용 용량이 연구 당 1 TB 정도이지만 충분한 분석 및 저장공간 확보를 위해 심평원은 300 GB 내외의 자료를 신청할 것을 권고한다. 자료이용기간은 최초 90일 및 연장 180일, 자료보관은 이용기간 종료 후 최대 1년까지며 재사용은 자료보관 기간 내 1개월 단위로 최대 1년까지 신청이 가능하지만 자료이용, 자료보관, 재사용이 모두 승인된 연구기간 내에 이루어져야 하고 총 이용기간이 최초 자료신청 이후 2년을 초과할 수 없다.

실제 청구자료 활용과정에서 직면하는 애로사항

청구자료 신청 및 활용 시 문제점 및 개선방안은 약물역학, 임상역학, 임상약학, 보건경제학 및 보건통계학 등 보험청구자료를 활용한 연구 경험이 매우 풍부한 전문가 13명을 대상으로 개별 서면 자문을 통하여 수집되었다. 서면 자문 문항은 연구 시 실제 활용한 자료명칭과 해당 자료 활용 과정에서 겪은 애로사항 및 해결 방안을 자유롭게 기술(narrative description)할 수 있도록 구성되었다. 자료 활용 과정은 4단계인 자료신청 및 획득단계, 제공된 자료내용 및 범위 및 용량 측면, 자료분석 및 활용단계, 자료보관 및 그 이후 활용단계와 기타 추가 의견으로 구분하여 각 단계에 따라 직면하는 어려움과 개선방안을 최대한 상세히 기술할 수 있도록 하였다. 각 전문가들은 독립적으로 의견을 제시하였으며, 모든 의견이 취합 정리된 내용을 확인하고 수정하였다. 본 연구에서 자문을 통해 수집한 정보는 인간대상에 대한 정보가 아닌 전문가로서의 제도 및 활용 경험에 대한 주관적 의견이기에 IRB 생명윤리법의 '연구대상자'가 의미하는 인간대상연구의 대상이 되는 사람에 해당하지 않는다고 판단하여 별도의 IRB 심의를 받지 아니하였다.

1. 자료신청 및 획득단계

1) 약제자료 신청의 어려움

약제 및 치료재료 정보는 관찰연구에서 주요 노출 또는 공변량으로 간주되기 때문에 대부분의 보건의료 연구 시 활용이 필수적이다. 청구자료는 원시자료 그대로 제공되지 않아 연구에 필요한 약제 정보만 요청해야 하는데 이를 준비하는 과정에서 상당한 인력과 시간이 소요된다. 그래도 최근에는 약가 전체 마스터파일이 제공되고 있고 심평원은 품 형태로 시스템화 되어 비교적 약제정보의 신청이 간편하나 공단은 필요한 주성분코드 및 관련 약물을 모두 정리하고 제조사가 드러나지 않도록 계열별로 그룹화 하는 과정을 거쳐야 하며 아직은 최근 도입된 자동추출방식과 기존 엑셀파일로 제출하는 재래식 형태가 혼합되어 있어 신청에 어려움이 있다. 또한 단독등재 제품인지 여부를 식별하기 위해 제공받고자 하는 연구 약물을 주성분코드 4자리 기준으로 제품명, 업체명, 업체 수를 정리하여 제출하도록 되어있어 연구와 별개로 추가 행정적인 노동이 발생한다.

약제정보를 제공받고자 하는 경우 약제 연구활용계획서를 제출해야 하는데 구체적인 기술항목이 제시되어 있지 않아 연구 내용에 따라 10-20장 내외 분량의 계획서를 제출하는 경우가 발생할 수 있고 작성하여 제출한다고 하더라도 심의에 따라 계획서가 불충분하다고 판정이 되어 자료제공이 반려된다면 이의신청을 통해 다음 심의를 대기해야 한다. 그러나 이의신청도 반려되거나 대부분 연구기간이 길지 않기 때문에 다음 심의를 대기할 수 없는 경우 최종적으로 약제정보 제공받는 것을 포기해야 하는 상황이 발생하기도 한다.

2) 자료 제공의 지연

자료신청 후 실제 연구자가 자료를 받기까지 약 6-7개월이라는 짧지 않은 기간이 소요되게 되는데 만약 심의에서 승인이 거부되어 재신청 또는 이의제기가 필요하거나, 이용자가 많아 가상화 공간 배정이 지연되거나, 타기관 연계자료를 이용하고자 하는 경우에는 7개월에서 1년 이상의 대기기간이 발생할 수도 있다. 이런 일은 예상보다 빈번하게 발생하기 때문에 단기간 수행되는 연구의 경우 연구결과 도출에 어려움을 겪을 수 있다. 자료 대기기간이 길어져서 지연된 기간 동안 축적된 최신 자료를 추가적으로 업데이트 하여 제공하고 싶어도 자료 신청 당시를 기준으로 가능했던 기간의 데이터만 제공하고 있기 때문에 제공받은 데이터 기간과 실제 연구 기간의 간극이 발생하여 시의성이 중요한 연구에서는 효용성이 떨어지는 경우가 있다.

2. 제공된 자료의 내용, 범위 및 용량 측면

1) 일부 상병, 약제 및 치료재료 정보제공의 제한

공단 및 심평원의 표본자료는 맞춤형 자료에 비해 연령, 민감상병 등에 추가적인 비식별화 및 그룹화 조치가 이루어져 맞춤형 자료보다 제한적인 정보가 제공된다. 희귀질환 등 대상환자수가 적은 자료의 경우 이러한 제한적인 조치가 개인정보 보호를 위한 것으로 합리적이거나 대상자가 많은 상병(예: 정신과 관련 상병)의 경우 맞춤형 자료 대비 개인정보 보호 관련 조치가 더 많이 되어 있는 표본자료를 이용하여 연구를 진행하는 것이 더 합당하다고 사료되는데 불구하고 상병 정보가 마스킹 되어있어 이용하는 것이 불가능하다. 따라서 표본자료를 통해 충분히 신속하고 의미 있는 연구를 진행할 수 있음에도 맞춤형 자료를 신청해야 하는 경우가 발생한다. 또한 현재 제공되는 공단 표본코호트 2.2 자료는 2019년 12월 31일까지만 자료가 제공되고 있어 시의성 측면에 한계가 있다.

공단의 맞춤형 자료는 기본 인구학적 정보, 건강검진결과 등 심평원 보다 다양한 정보를 제공한다는 장점이 있으나 약제에 대해서는 제한적인 정보를 제공한다. 의약품 연구의 경우 성분명에 따른 분석 및 동일 성분 내에서도 함량에 따른 용량-반응(dose-response) 평가가 매우 중요하지만 정책연구 외에는 약제코드 9자리 중 주성분 일련번호인 앞의 4자리에 대한 정보만 제공하는 것을 원칙으로 하여 용량, 제형, 투여경로에 대한 정보를 이용할 수 없고 앞의 4자리가 같지만 서로 다른 성분명을 가진 약제들도 있어 주의가 필요하다. 또한 연구대상자에 대해 신청 시 요청한 약물 정보만 제공하고 있기 때문에 연구 대상자의 전체 약물사용현황을 파악하거나 이를 공변량으로 활용하는 것이 제한적이다.

심평원의 맞춤형 자료는 약제정보 제공이 공단보다 비교적 유연하다는 장점이 있으나 기관 특성 상 사망(보험 자격), 보험료, 실거주지와 같은 자격정보를 제공하는 것이 불가능하며 대상자 수가 일정 규모 미만인 경우 연령이 5세 단위로 제공되는 등의 제한점이 있다.

2) 자료 용량의 제한

공단과 심평원 맞춤형 자료는 용량을 각각 700 GB, 300 GB 내외로 제한하여 제공하는데 코호트 연구와 같이 비교군이 필요한 연구의 경우 용량이 부족하여 적절한 비교군을 확보하기 어려운 경우가 발생한다. 자료의 크기를 줄이기 위하여 심평원은 규정 용량 이내로 자료의 규모를 줄이도록 안내하며 공단은 대상자 일부

를 랜덤 추출(sampling)할 것을 권고하고 있지만 이 경우 전수자료를 활용하는 맞춤형 자료의 의미가 축소될 수 있다. 시간이 지남에 따라 원시자료가 점점 축적되면서 가용한 자료의 기간과 범위가 넓어지지만 제공되는 자료의 용량은 변화가 없어 대규모 연구가 어렵다.

3. 자료분석, 활용과정 및 활용기간 단계

1) 자료 활용 중 갱신의 불가

심평원은 맞춤형 자료를 실제로 제공받기 전까지 상담자가 추출한 자료에 대한 물리적인 확인이 불가하며 공단은 자료 제공받기 전 제대로 추출이 되었는지 연구자가 직접 확인하는 과정을 거치지만 확인 당시에는 추출된 자료에 오류가 있는지 또는 연구 의도에 따라 추출되었는지 여부를 위주로 확인하기 때문에 실제 연구과정에서 보정변수 또는 층화 등을 위해 자료의 추가, 변경이 필요한 경우가 발생할 수 있다. 그러나 공단 및 심평원은 원칙적으로 심의 계획서에 명시한 테이블 및 변수에 대해서만 제공하기 때문에 자료가 한번 제공되고 나면 업데이트가 불가하며 이를 추가하고자 하는 경우에는 자료를 새로 신청해야만 한다.

2) 자료접근에 대한 시공간적 제약

공단 및 심평원의 맞춤형 자료는 사전 승인된 IP에서 원격으로 접속하거나 분석센터에 직접 방문해야만 접근이 가능하기 때문에 기본적으로 시공간적 제약이 따른다. 공단의 맞춤형 자료는 공단 지정 분석센터에서 오전 9시부터 오후 6시 사이에 방문을 통해서만 분석이 가능하며 지리적 제약이 있으며 사전에 자리를 예약해야 하는데 이용자가 많은 경우 공간이 부족한 경우가 있다. 자료 이용기간 또한 맞춤형 자료의 경우 공단은 최대 1년, 심평원 최대 270일(최초 90일 + 연장 180일)으로 제한되어 있고 표본자료의 경우 공단 최대 6개월, 심평원 최대 2개월(최초 1개월 + 연장 1개월)로 제한되어 있다. 여기에는 반출이나 반입 등 행정처리에 소요되는 기간도 포함되어 있기 때문에 중간 결과에 대해 연구자 간 논의를 거쳐 진행방향을 결정하는 연구의 경우 물리적인 시간이 부족하다. 또한 연구자마다 할당되는 가상화 분석공간의 용량이 심평원은 1 TB, 공단은 1.5 TB이지만 제공되는 자료의 용량을 제외하면 각각 약 700 GB 정도만 분석공간으로 활용이 가능하기 때문에 딥러닝 등 분석자원을 많이 필요로 하는 연구의 경우는 진행하기 어렵다.

3) 분석도구의 제한

심평원은 분석도구로 SAS 및 R을 지원하고 있으며 공단은 SAS, R, STATA, SPSS, Python을 지원한다. 대부분의 연구자들은 SAS를 활용하지만 분석적 구현의 한계가 있어 최근에는 자유도가 높은 R과 Python의 활용도가 높아지는 추세이다. 그러나 R을 이용하기 위해서는 별도의 패키지 반입승인 절차를 거쳐야 하며 SAS와 R에서 이용되는 편리한 명령어 및 패키지가 빠르게 개발되고 있는 것에 비해 가상화 공간에 설치된 프로그램의 버전이 낮아 적용이 되지 않는 경우가 많다.

4) 자료 반출 제한

공단 맞춤형 자료의 경우 반출까지 최소 3일(공휴일 제외)이 소요되고 분석센터 내 전자기기 및 사진촬영이 금지되어 있어 실시간으로 분석결과를 공유하거나 제출해야 하는 경우 어려움이 많다. 또한 일일 반출 가능한 용량이 최대 25 MB로 제한되고 있어 최소 300 dpi가 요구되는 논문에 삽입될 고화질 그래프를 반출하기에 제약이 있다.

4. 자료보관 및 추후 활용 단계

자료활용 기간이 종료되면 신청했던 분석 데이터는 원칙적으로 삭제되며, 추가적으로 자료를 이용하고자 하는 경우 별도의 자료보관 신청을 해야 한다. 자료보관은 심평원의 경우 최대 1년 동안 10 GB 이내, 공단의 경우 최대 1년 동안 20 GB 이내만 보관할 수 있기 때문에 분석에 활용했던 데이터의 일부만 남겨놓을 수

있으며 보관기간이 종료되면 최초 투고 이후 논문 심사가 지연되거나, 심사에서 결과보완을 요청하거나, 게재절로 인한 재투고가 필요한 경우 추가 분석이 필요하더라도 진행하기 어렵다.

청구자료 활용성 촉진을 위한 개선방안

국내 청구자료는 전국민 의료이용에 대한 전체 정보를 활용할 수 있어 국제적으로도 인정받는 매우 유용한 RWD임이 분명하다. 따라서 본 연구는 궁극적으로 규제의사결정에서 청구자료의 활용을 극대화하기 위해 개선이 필요한 몇 가지 논의사항들을 확인하였으며 이에 대한 5가지 해결방안을 아래와 같이 제언하고자 한다.

먼저 자료준비기간 포함 자료신청부터 실제 연구자가 자료를 획득하기까지의 기간이 7개월 이상 소요되는데 연구의 시의성을 높이기 위해서라도 이 기간이 축소되어야 할 것이다. 현재는 단독등재 품목 여부를 확인하기 위해 연구에 필요한 약제 관련 정보를 업체별, 품목별로 정리하여 제출해야 하는데 이를 단독등재 유무 여부만 확인하여 표기하는 방식으로 간소화하여 자료신청 준비기간을 단축시킬 수 있는 방안이 있다. 또한 신청과정에서 보완요청이 빈번하게 발생하게 되면 그만큼 심의와 자료 수급이 지연되기 때문에 신규 신청자를 위한 신청서, 약제활용계획서 등 자유롭게 작성하여 제출해야 하는 서류의 사례(샘플) 등이 제공되면 도움이 될 것이다. 기관 내 데이터베이스를 다루는 인력을 충원하여 자료추출기간을 단축시키거나 데이터를 빨리 받을 수 있는 '신속제공(비용 추가)' 옵션을 지원하는 것도 해결방안 중 하나로 고려해볼 수 있을 것이다.

둘째, 제3자의 이익침해가 발생하지 않는 경우 연구의 공익성을 판단하여 약제나 치료재료 정보의 제공을 유연하게 할 필요가 있다. 현재 약제 및 치료정보가 포함된 T30, T60 제공에 제한이 있고 주성분코드를 앞 4자리까지만 제공하며, 특정 성분 또는 단독 성분 의약품에 대해서는 자료요청이 불가하기 때문에 국내 제약산업에 필요한 신약에 대한 안전성 및 유효성 연구, 오리지널-제네릭 연구가 실질적으로 불가능한 상황이다. 특히 정책연구 외 연구에는 보다 보수적으로 자료가 제공되는 경향이 있어 발주처가 공공기관이나 정부 인지 여부로 공익성을 결정하기 보다는 연구결과의 파급효과 등을 고려하여 판단해 약제와 의료가기 등에 대한 제공범위를 지금보다 전향적으로 확대할 필요가 있을 것으로 사료된다.

셋째, 대체적으로 연구기간이 길지 않기 때문에 자료이용에 대한 시공간적 제약을 개선할 수 있는 방안이 마련되어야 한다. 모든 연구가 자료를 원격으로 이용하는 것이 불가능하다면 일부 학술 연구 등은 원격으로 제공받을 수 있는 옵션을 제공하거나 분석센터 예약을 일단위가 아닌 시간단위로 변경하여 효율성을 개선할 수 있다. 국외 다수 자료원들은 분석이 필요한 연구자들에게 원자료를 분양하고, 정보유출 등이 발생할 경우 강력히 처벌하도록 하고 있는 반면 국내 체계는 자료가 클라우드 방식으로 되어있고 보안 우려로 인해 센터에 방문해야만 분석이 가능하도록 운영되고 있는데 국제적인 운영방식을 검토하여 보다 효율적으로 운영될 수 있도록 개선하는 것이 필요해 보인다.

넷째, 제공되는 자료의 용량, 이용 및 보관기간이 연구의 성격에 따라 유연하게 조정될 수 있어야 한다. 현재는 300 GB, 700 GB와 같이 용량이 제한되며 1년 내외의 짧은 이용기간으로 인해 대규모 연구가 불가능하기 때문에 장기간 큰 용량의 데이터를 지원할 수 있는 추가비용지불 등의 옵션을 마련해볼 수 있다. 또한 자료 보관기간이 종료되면 분석 데이터가 지속적으로 보관되지 않는다는 점은 데이터 과학의 재현성 측면에서도 큰 문제가 있어 사유서 제출 및 심의를 거친다면 보관종료 이후에도 논문 게시 전까지는 자료보관을 할 수 있는 방안이 마련되어야 할 것이다. 만약 공단 및 심평원이 자체적으로 장기간 자료활용이 필요한 과제들에 대해서 직접 내규를 수립하고 시행하는 것이 어렵다면 정책연구가 근거하는 특별법으로 운영지침을 우

회하는 방법도 고려해볼 수 있을 것이다.

다섯째, 데이터 활용 도중에 필요한 정보를 갱신하여 활용할 수 있는 방안을 고려해보아야 한다. 현재 최초 신청서 심의 이후에는 자료를 변경할 수 없고 변경을 원하는 경우 재신청을 해야 하는데 대체적으로 연구기간이 길지 않아 현실적으로 어렵기 때문에 갱신하고자 하는 정보가 최초 연구계획에서 크게 벗어나지 않고 개인정보 보호 등을 추가로 위반할 가능성이 없다면 심의를 거쳐 기존 자료에 추가로 자료를 제공할 수 있는 방안이 마련되었으면 한다.

결론적으로 지금보다 의약품 규제의사결정에서 청구자료의 활용성을 극대화하기 위해서는 개인정보가 침해되지 않는 범위 내에서 연구의 성격에 따라 자료, 용량 및 기간 등이 유연하게 조정될 수 있는 방안이 마련되어야 하며 이를 위한 자료 보유기관과 연구자의 지속적인 의사소통이 필요할 것이다.

감사의 말씀

본 연구는 2023년도 식품의약품안전처의 출연과제예산으로 수행되었으며 이에 감사의 말씀을 드립니다 (과제 고유번호: 21183실임상543).

이해상충

저자들은 본 논문의 내용과 관련하여 그 어떠한 재정적인 상충이 없다.

References

1. Ministry of Food and Drug Safety. Guidelines for research using medical information databases. June 29, 2021. Available from: https://www.mfds.go.kr/brd/m_1060/view.do?seq=14861&srchFr=&srchTo=&srchWord=&srchTp=&itm_seq_1=0&itm_seq_2=0&multi_itm_seq=0&company_cd=&company_nm=&page=1. Accessed January 02, 2024.
2. Ministry of Food and Drug Safety. Considerations for using medical big data in Korea. April 20, 2022. Available from: <https://www.medic.or.kr/Uploads/BLibrary/%EA%B5%AD%EB%82%B4%20%EC%9D%98%EB%A3%8C%EB%B9%85%EB%8D%B0%EC%9D%B4%ED%84%B0%20%ED%99%9C%EC%9A%A9%EC%9D%84%20%EC%9C%84%ED%95%9C%20%EA%B3%A0%EB%A0%A4%EC%82%AC%ED%95%AD.pdf>. Accessed January 2, 2024.
3. Lee SW. Opening and provision of health insurance data from the Health Insurance Review and Assessment Service. Korea Institute of Public Administration 2021;37:8-15.
4. Song HY. National Health Insurance “Efforts to create data value by combining de-identification information”. Etnews. December 4, 2023. Available from: <https://www.etnews.com/20231204000135>. Accessed January 2, 2024.
5. Ha MY, Lee JY, Kwak MS, et al. Recommendations on laws and systems to increase RWD/RWE utilization in regulatory decision-making for pharmaceuticals. Regulatory Research on Food, Drug & Cosmetic 2022;17:163-71.
6. Cho H, Lim E, Park S, et al. The current status and perspective of real-world data use of pharmaceutical industries in Korea: A survey study. Journal of Korean Academy of Social & Managed Care Pharmacy 2023;11:9-16.