

II 의료보건 연구를 위한 건강보험심사평가원 청구 데이터의 소개 및 활용

김지애, 김록영

건강보험심사평가원 통계센터 통계분석팀

중심 단어: 건강보험심사평가원, 청구데이터, 의료보건 연구, 표본 데이터

‘건강보험 청구데이터’란 요양기관이 의료서비스를 제공한 후 환자의 진료비용 중 ‘국민건강보험’이 부담하는 부분에 대해 지급의뢰를 하기 위하여 ‘건강보험심사평가원(이하 심평원)’에 보험급여 청구를 하면서 발생하는 자료이다. 청구데이터는 진료 내역(처치, 시술, 검사, 처방전 약 등), 진단명, 보험자 지급 비용, 환자 부담금, 환자 인구 특성, 요양기관 정보 등과 같은 다양하고 방대한 정보를 포함하고 있으며, 이러한 정보는 의료보건 연구에 활용 가치가 높다. 심평원은 2009년도부터 연구용 청구데이터를 개방하기 시작하여 현재 자료처리실, 직접 제공 및 원격접속을 통해 연구자들의 데이터 이용을 지원하고 있다. 본 논문은 청구데이터를 이용한 의료보건연구 수행에 관심 있는 연구자들에게 청구데이터에 대한 자세한 소개와 활용에 대한 안내를 목적으로 한다.

1. 도입

보건의료 연구를 계획하고 설계할 때 가장 우선적으로 고려해야 할 사항 중 하나는 연구를 통해 분석할 수 있는 데이터의 존재 파악 및 습득 여부이다. 연구자는 크게 두 가지 방식으로 데이터를 확보할 수 있다. 첫 번째는 연구 목적에 맞게 데이터를 직접 수집하는 방식(일차 데이터: primary data)이며, 두 번째는 기존에 다른 목적 혹은 용도로 이미 수집되어 있는 데이터를 활용하는 방법이다(이차 데이터: secondary data). 의료보건분야에서 대표적인 이차 데이터로 의료청구데이터가 있다.

건강보험심사평가원(이하 심평원)은 2009년도부터 연구용 데이터 개방을 준비하여, 2012년도부터 점진적인 개방을 확대해왔다. 올 2013년도에는 정부 3.0의 세부과제인 “공공기관 보유정보의 개방과 공유”의 시행 방안으로 연구용 청구데이터 공개는 더욱 가속화되고 있는 추세다.

본 논문은 의료보건 분야 연구자들에게 우리나라 건강보험 청구데이터에 대해서 비교적 자세한 정보를 제공함으로써 청구데이터를 이용한 의료보건 연구 활성화를 도모하고자 한다. 이를 위하여, 본 논문의 주제는 세 가지로 구성된다. 첫 번째는, “청구데이터의 성격과 이를 활용한 연구 분야에 대한 소개”를 통해 활용방안을 소개하고, 두 번째는 “우리나라 건강보험 청구데이터 생성, 구성 및 내용”에 대한 세부적인 설명을 통해 청구데이터에 대한 이해도를 높이도록 한다. 세 번째는 “심평원 정보 개방 다양화”를 통한 데이터 이용 방식과 제공되는 데이터 종류에 대해서 알아본다. 마지막으로 “청구데이터의 유용성과 제약성 등과 같은 연구 시 고려해야 할 사항”으로 구성된다.

1) 청구데이터의 성격과 이를 활용한 연구 분야에 대한 소개

(1) 의료 청구데이터를 이용한 연구 분야 및 해외 사례: 의료청구 데이터는 의료서비스 제공업체(병원, 의원, 치과, 보건서 등)가 환자에게 제공한 진료 서비스를 보험자에게 청구하면서 발생하게 되는 데이터이다. 청구데이터는 진료 내역(처치, 시술, 검사, 처방전 약 등), 진단명, 보험자 지급 비용, 환자 부담금, 환자 인구 특성, 요양기관 정보 등과 같은 다양하고 방대한 정보를 포함하고 있으며, 이러한 정보는 의료보건 연구에 활용 가치가 높다.

의료청구데이터는 성과연구(outcome research), 비교효과연구(comparative effectiveness studies), 의료의 질(quality of care), 의료정책 및 평가 연구(health policy and evaluation)와 같은 의료보건분야 연구에 활발히 활용되고

있다. 또한 고비용의 질병 발생의 위험부담이 큰 환자들을 예측하여 예방적 의료를 제공하기 위한 predictive modeling 개발에도 활용될 수 있다.

최근에는 특히 근거 중심의 의료(evidence-based medicine) 제공 필요성에 대한 사회적 요구가 증가함에 따라 청구데이터를 활용한 비교효과연구를 포함한 성과연구는 꾸준히 증가세를 보이고 있다. 미국의 경우, 오바마 정부는 근거 중심의 의료서비스 시술에 대한 상환(reimbursement) 정책 추진을 배경으로, 2009년도 경기부양 패키지(Obama's stimulus package) 일환으로 11억 USD라는 막대한 자원을 비교효과연구(comparative effectiveness research)지원을 해 오고 있으며, 청구데이터는 이러한 연구 수행을 위한 주요한 데이터 소스가 되고 있다.

2009년도에 연구용으로 청구데이터가 개방된 우리나라에 비해, 미국은 청구데이터의 연구 이용에 오랜 역사를 지닌다. 사보험 중심의 의료보험 시장 중심의 미국은, 공공보험의 경우 제한적으로 특정한 인구에게 제공되어 보완재의 성격을 띠고 있다. 연방정부가 운영하는 65세 이상 고령층 대상의 메디케어(Medicare)와 주정부가 운영하는 저소득층 인구를 대상의 메디케이드(Medicaid; Medicaid는 주정부에서 운영하나, 재원의 일부분은 연방정부에서 지원을 한다)가 대표적인 공공보험이라 할 수 있다.

이러한 공공보험은 연방정부 기관인 Center for Medicare and Medicaid Services (CMS)에서 주관하고 있다. CMS는 두 곳의 연구 센터와의 협력을 통하여 연구자들에게 CMS 데이터(Medicare 의료청구데이터와 Medicaid 의료청구데이터) 이용을 적극 지원하고 있다. CMS는 효율적인 데이터 활용을 위해 수진자(beneficiaries) 중심으로 데이터를 재구축하여 제공하고 있으며, 자료연계가 필요한 연구자들에게 데이터 연계를 지원하고 있다. 주요 지원책 중 하나로 연구자들에게 연간 약 4회 이상의 무상 데이터 교육 프로그램을 제공하여 CMS 데이터의 연구 이용을 정책적으로 활성화하고 있다. CMS는 또한 메디케어 인구의 의료서비스 사용 및 성과에 대한 다양한 파일럿 스터디(pilot studies) 수행을 통해 선제적이면서 상시적으로 데이터의 질을 점검(quality check)하고 관리한다.

2) 우리나라 건강보험 청구데이터 생성, 구성 및 내용

‘건강보험 의료청구데이터’란 요양기관이 의료서비스를 제공한 후 환자의 진료비용 중 ‘국민건강보험’이 부담하는 부분에 대해 지급의뢰를 하기 위하여 ‘심평원’에 보험급여 청구를 하면서 발생하는 자료이다. 우리나라의 1년간

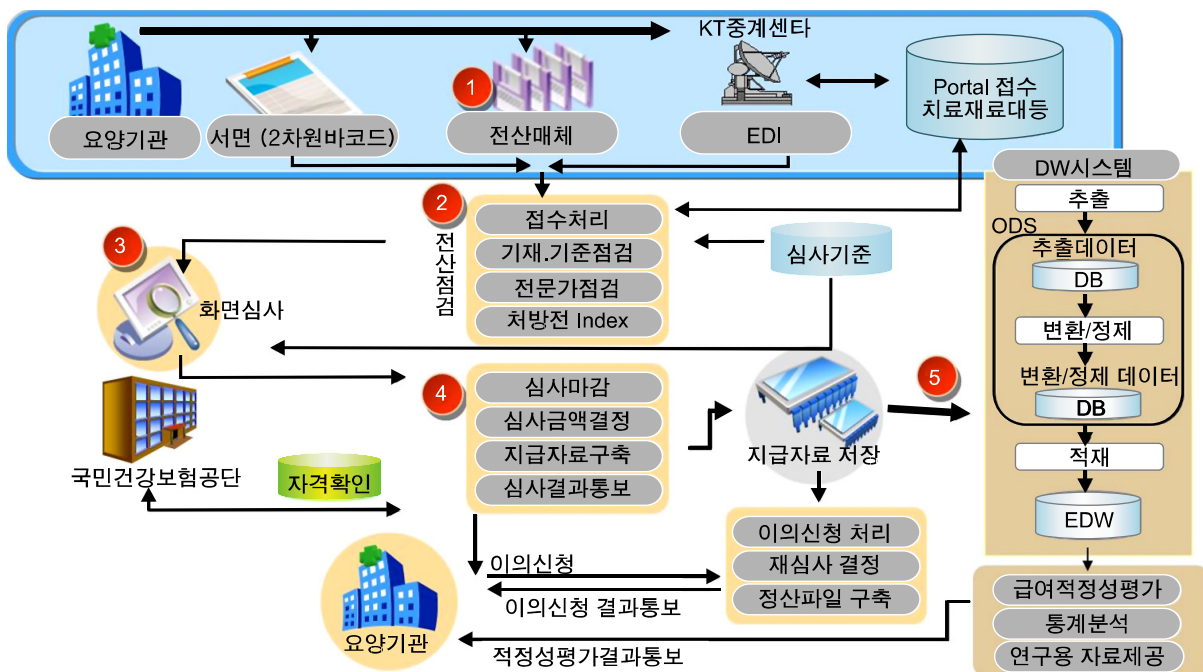


Figure 1. 청구데이터 흐름도, EDI: Electronic Data Interchange, EDW: Electronic Data Warehouse.

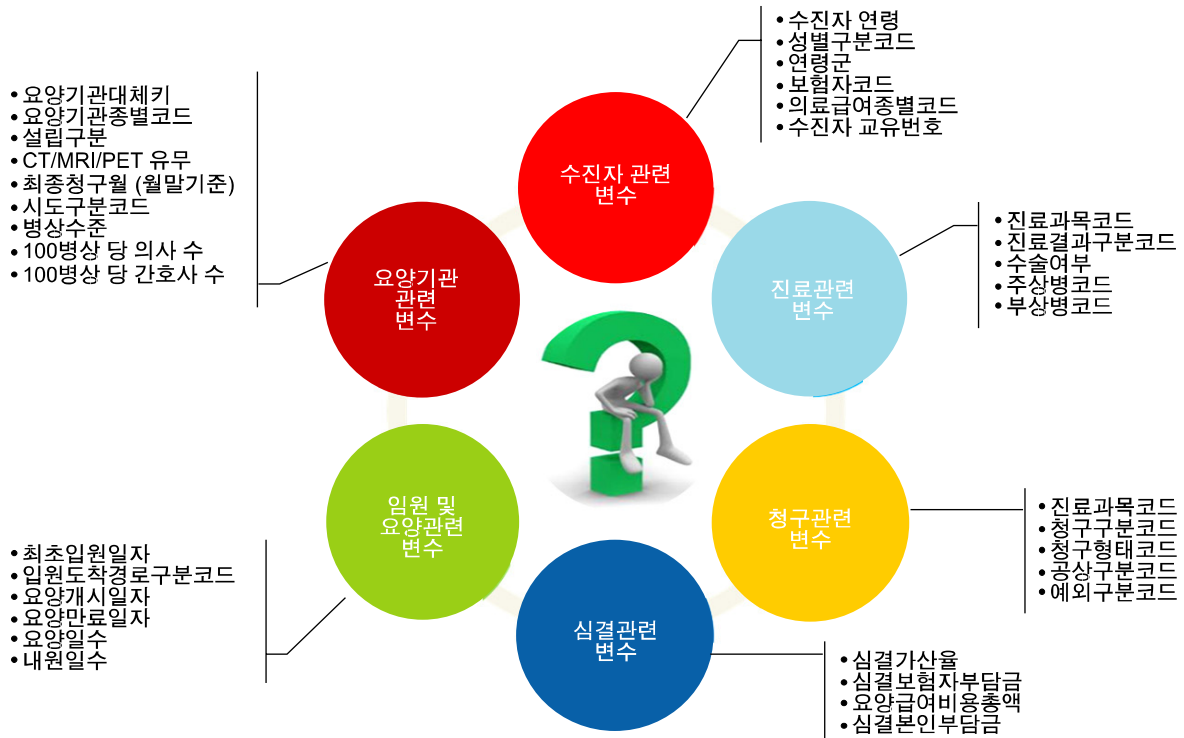


Figure 2. 청구데이터 보유 대표변수.

‘건강보험 청구 환자수’는 주민등록인구의 90%에 해당하는 약 4천6백만 명으로, 전국의 8만여 개 영양기관으로부터의 청구 건이 포함되어 있다(2011년도 기준).

요양기관에서는 환자 진료 후 급여 상황을 위하여 심평원에게 Figure 1과 같은 절차에 따라, “요양급여비용명세서”를 이용하여 청구하게 된다. 요양기관으로부터 진료비 청구 명세서가 접수되면, 전산 점검 및 심사 단계를 거쳐, 데이터로 누적어 되고, 누적된 데이터는 Data Warehouse (DW) 시스템에 저장된 후 급여적정성평가와 통계생성에 활용되거나, 연구용 데이터로 제공된다.

청구데이터는 크게 두 가지 소스—요양급여비용 청구명세서와 요양기관현황—로부터 생성된다. 청구명세서는 수진자 정보를 담은 일반사항이 기재된 일반내역정보와 상병, 진료과목, 요양급여비용, 항목별 진료비(행위, 의약품 등)와 같은 세부사항이 기재된 진료 세부정보로 구분된다. 이러한 내용들은 네 개의 테이블로 나뉘어져 연구용 데이터로 제공되며 ‘연구용 데이터 테이블 구성’에서 보다 자세한 사항을 언급하기로 한다. 요양기관현황 데이터에는 요양기관의 시설, 인력, 보유 장비 등과 같은 요양기관 정보들이 포함되어 있다. 이러한 두 가지의 자료원으로부터 청구데이터의 주요 변수들을 추출하거나 생성한다(Figure 2).

(1) 연구용 의료청구데이터 구성 및 포함 변수: 연구자들에게 제공되는 의료청구데이터는 당해년도를 기준으로 최근 5년간의 청구된 정보가 포함되어 있다. 이러한 데이터는 네 개의 테이블로 나뉘어져 있다(Table 1). 첫 번째는 명세서 일반내역 테이블이다. 명세서일반내역(Table20)에는 인구학적 특성변수(나이, 성별), 수진자 주민번호 대체키, 주상병과 부상병(상병코드는 한국표준질병사인분류표의 상병분류기호를 사용하고 있다), 요양기관 대체키, 입원 및 외래 구분 변수(서식코드) 및 의료급여종별코드가 포함되어 있다. 이러한 변수들은 안내변수(instruction variables)라고 하며, 연구대상을 추출하는 데 유용하게 사용하는 변수들이다.

두 번째 테이블은 진료내역(Table30)으로 환자들이 외래 혹은 입원하여 발생하는 모든 정보들을 포함하고 있다. 따라서 환자들에게 제공된 진료의 행위 및 약제 등(검사, 처치, 시술, 약 등)에 대한 자세한 정보를 담은 진료내역 정보를 확인할 수 있다. 진료내역 테이블의 약제 정보는 원내 처방이 이루어진 약제만 포함되어 있다.

Table 1. 청구데이터 테이블 구성

| | |
|---------------------|--|
| Table20 (명세서일반내역) | 명세서조인키, 보험자코드, 수진자개인식별대체키, 성별구분, 수진자연령, 의료급여종별, 요양기관식별대체키, 요양종별코드, 지역코드, 서식구분코드, 주상병코드, 부상병코드, 진료과목코드, 당원요양개시일자, 요양종료일자, 최초입원일자, 입내원일수, 요양일수, 원외처방일수, 원외처방약제비, 원외처방건수, 초진회수, 청구요양급여비용총액, 청구분인부담금, 수술여부, 공상구분, 상해외인구분코드, 특정기호구분, 진료결과구분, 입원도착경로구분, 청구형태코드, 청구구분코드, 심사년월 |
| Table30 (진료내역) | 명세서조인키, 항목코드, 분류코드구분, 분류코드, 일반명코드, 1회 투약량, 1일 투약량, 일일투여량 또는 실시횟수, 총투여일수 또는 실시횟수, 총사용량 또는 실시횟수, 단가, 금액, 가산적용금액, 진료예외구분코드, 1_2구분 |
| Table40 (수진자상병내역) | 명세서조인키, 일련번호, 상병코드, 상병분류코드, 청구진료과목코드, 청구요양개시일자, 내과세부전문과목 |
| Table53 (원외처방전상세내역) | 명세서조인키, 처방전교부번호, 분류코드구분, 분류코드, 일반명코드, 1회투약량, 1일투여횟수, 총투여일수, 총사용량, 단가, 금액 |

세 번째 테이블은 상병내역(Table40)으로 주상병, 부상병을 포함한 수진자가 가진 모든 동반상병정보를 담고 있다. 수진자가 가진 모든 상병내역 정보가 필요한 경우에 사용해야 하는 테이블이다. 예를 들어, 임신과 같이 질병이 아닌 경우가 상병코드로 붙는 경우 모든 동반질병내역을 확인하여 연구를 진행하는 경우를 예로 들 수 있다. 건강상태를 통제하기 위한 변수 생성(ex. Charlson Score Index 혹은 Prescription-Risk Adjustment)을 위해서는 상병내역 테이블을 이용하게 된다.

마지막으로 원외처방내역(Table53)은 수진자에게 원외 처방으로 발생된 모든 약제에 대한 정보를 포함하고 있다. 심평원은 청구가 이루어지는 약에 일반명코드(혹은 주성분 코드)를 부여하여 개별약을 식별하도록 하고 있다. 개별 약제에 대한 일반명은 있으나, 약물치료군(therapeutic drug class)에 관한 정보는 없기 때문에 연구자들은 개별적으로 매칭을 해야 한다.

3) 건강보험심사평가원 정보 개방 다양화

(1) 데이터 이용 방식: 심평원은 공공정보의 적극 개방 및 공개요구에 따라, 다양한 형태를 통한 데이터 공개를 통하여 국민의 알 권리 충족 및 공공데이터 활용을 위해 보유정보를 단계적으로 개방하고 있다. 심평원은 2009년도부터 청구데이터를 연구용으로 제공하기 위한 프로젝트를 시작하면서, 연구자들에게 데이터를 개방해 오고 있으며, 현재 세 가지 방식으로 정보를 이용할 수 있다.

첫 번째는 심평원 내에 위치한 자료처리실을 이용하는 방법이다(Figure 3). 자료처리실은 일종의 컴퓨터 랩(LAB)으로, 연구자들은 연구 주제에 맞게 데이터가 세팅된 컴퓨터를 할당받아 이용하게 된다. 자료처리실은 주로 대용량의 자료 혹은 타기관의 자료 연계가 필요한 경우 주로 사용된다. 자료처리실을 이용한 데이터 이용은 전국민을 대상으로 한 데이터이므로 발생이 적은 희귀질환 분석이 가능하나, 이용자적은 국책사업 및 공공연구나 국가·행정기관에서 업무수행과 관련하여 요청하는 경우로 제한된다. 외부보유자료 연계는 환자의 포괄적 사전 동의 획득이나, 건강보험 청구자료와 연계에 대한 동의 획득 시 연계가 가능하다. 자료처리실을 이용할 경우, 심의과정을 거쳐야 한다.

두 번째는 데이터의 직접 제공이다. 심평원은 연구자들의 데이터 접근성과 편의성을 위하여, 무작위추출로 추출된 1년 단위의 표본 자료를 2012년도부터 제공해오고 있다(Table 2). 표본자료는 2009~2011년도로 3년간 구축되어 있으며, 표본자료의 종류는 전체환자표본(National Patient Sample, NPS), 입원환자표본(National Inpatient Sample, NIS), 노인환자표본(Aged Patient Sample, APS), 소아·청소년환자표본(Pediatric Patient Sample, PPS)으로 나누어 제공하고 있다. 환자의 특성과 대표성을 반영하도록 특정 계층에 대해 별도로 표본을 추출함으로써, 그 특정 계층만이 지니고 있는 환자의 대표성을 높여 연구에 대한 활용도를 높이도록 하였다.

환자표본자료의 제한점은 모든 표본 자료 공통의 한계점으로서 표본자료 내의 관측치는 확률에 의해 추출되는 자료이기 때문에 적정수준 이상의 표본수를 확보해야 대표성, 유의성을 보장되므로, 특정 연령대의 희귀질환 발생빈도의 경우 표본추출 빈도가 너무 적어 대표성과 설명력이 떨어질 수 있다. 따라서 표본자료의 설명력은 다빈도 상병일수록 커지며, 상병의 발생 빈도가 떨어지면 감소하게 된다. 또한 환자표본 자료는 일 년짜리 단면 데이터

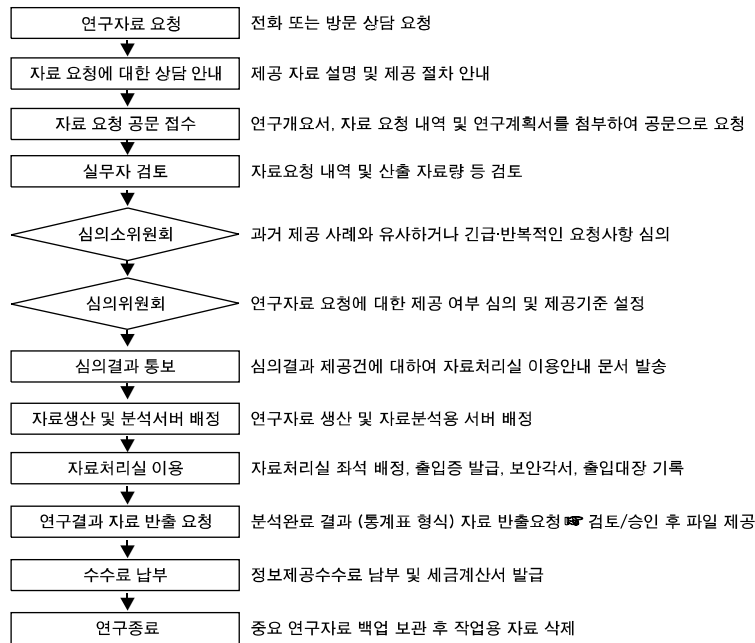


Figure 3. 자료처리실 이용 절차.

Table 2. 표본데이터 종류 및 산출기준

| 표본자료 종류 | | 산출 기준 |
|-----------------|----------|--|
| 2012년도 제공 시작 | HIRA-NIS | 1년 단위 입원환자 약70만 명(13%), 외래환자 약40만명(1%) |
| | HIRA-NPS | 1년 단위 전체 환자 약140만 명(3%) |
| 2013년도 추가 제공 시작 | HIRA-APS | 1년 단위 65세 이상 환자 약100만 명(20%) |
| | HIRA-PPS | 1년 단위 20세 미만 환자 약110만 명(10%) |

NIS: National Inpatient Sample, NPS: National Patient Sample, APS: Aged Patient Sample, PPS: Pediatric Patient Sample.
 ※각 환자표본자료의 표본 한계치는 환자수 150만 명 또는 영역별 20% 이내를 기준.

(crosssectional data)이므로 각 년도마다 연계가 되지 않는다. 따라서 장기간의 추적 관찰을 하여 인과 관계를 도출해 내야 하는 연구 수행은 불가능하다. 연구자들은 소정의 수수료를 지불하고 표본데이터를 구입할 수 있으며, 심의과정 없이 모든 연구자들에게 제공된다.

세 번째 방식으로는 원격접속 서비스를 통한 데이터 이용이다. 2013년 12월 중순부터 시행하게 되는 원격접속서비스는 연구자들이 본인들의 컴퓨터를 이용해서 심평원의 서버로 접속하여, 서버에 저장되어 있는 데이터를 사용하게 되는 방식이다. 자료처리실과 달리, 연구자들은 장소나 시간의 구애 없는 상시적인 데이터 이용이 가능하게 된다. 특히 서울 이외 지역에 있는 연구자들은 자료처리실을 이용한 데이터 사용에, 서울지역 연구자들보다 더 큰 물리적 시간적 장벽을 경험하고 있어 청구데이터가 연구에 활발하게 활용되지 못하고 있다. 원격접속 서비스는 이러한 서울 수도권 지역 이외의 의료보건 연구자들의 데이터 사용 진입 장벽을 낮출 수 있을 거라고 예상되며 의료보건 연구자 저변층을 확대하는 데 일익을 담당할 것으로 예상된다.

(2) 임상 표본 코호트 구축 계획: 심평원은 임상분야별 임상학회와의 워킹그룹을 통하여 개인정보가 보호되는 수준에서 다년간의 표본추적 코호트 데이터를 구축할 계획이다. 이는 1년 단위 환자표본자료로는 분석이 불가능한 연구-회귀 질병, 노출과 결과가 장시간에 걸쳐 발생하는 경우, 약물 부작용-에 제한이 되는 문제점을 보완하기 위하여 환자들의 장기간 follow-up이 필요한 보다 전문적이고 임상적인 영역을 대상으로 구축하고자 한다. 반면, 환자들을 장기간 추적에 따른 개인정보유출 문제가 도래할 위험 부담이 크므로 개인정보 보호를 위해 환자식별 대체기를

Table 3. 임상 표본 코호트 종류

| 코호트 명칭 | 내용 |
|-------------|---|
| 만성기도질환 산부인과 | <ul style="list-style-type: none"> • 2008년도 만성기도질환 환자 대상으로 5년간의 추적 코호트 자료 구축(전 수 또는 약 20% 추출) • 2008년도에 출산한 산모 49만 명에 대한 표본 산모 10만 명(전체 산모의 약 20%) 추출 |

연구자마다 다르게 부여하여 자료의 이동 경로를 추적할 수 있고, 연구자들 간의 데이터 연계가 불가능하도록 개인정보 보호에 초점을 두고 개발 중에 있다.

Table 3은 만성기도질환 환자를 대상으로 한 코호트 자료와 산모의 출산일을 기준으로 과거 1년부터 출산 이후 5년간 추적한 산부인과 코호트 자료에 대한 설명을 보여준다.

산부인과 코호트를 예로 보자면, 2008년도에 출산한 산모 49만 명에 대해 표본 산모 10만 명을 추출하여, 2008년도에 출산한 산모의 출산일을 기준으로 과거 1년간의 진료내역(후향적 코호트)과 산모의 출산일 기준으로 산모뿐만 아니라 신생아의 5년간 진료내역(전향적 코호트)을 추가할 계획이며, 이는 전향적·후향적 추적조사를 통해 산모의 과거 건강상태를 고려한 질적인 연구가 활발히 진행될 것으로 기대한다.

이러한 코호트를 통해, 산모가 보유하고 있는 질환(고령산모, 임신성당뇨 등)으로 인하여 신생아 주별 사망률 등과 같은 신생아에게 미치는 영향을 파악할 수 있으며, 희귀난치성 신생아가 가져오는 가계부담을 파악이 가능해진다. 또한 장기간의 데이터를 구축하게 되면 소아질환이 성인질환까지의 지속성 파악이 가능해질 것으로 판단된다.

향후, 심평원은 다양한 보건의료분야 연구의 지속적인 활성화를 위하여 만성기도질환 코호트와 산부인과 코호트 자료에서 더 나아가 의약품분야와 내과분야 코호트 자료 등으로 다양화할 계획이다.

4) 청구 데이터의 연구 이용 시 고려 사항

(1) 청구데이터를 이용한 관찰연구: 청구데이터로부터 연구 인구 대상을 추출하게 되는 경우는 대부분이 관찰 연구(observational studies)가 된다. 관찰 연구는 연구대상자를 자연 환경(natural setting)에서 직접관찰하여 노출(exposure)과 결과(outcome)와의 관계를 도출해내는 연구이다. 따라서 무작위 배정 임상시험(randomized controlled trial)과 달리 노출은 개인이 가지고 있는 특성, 의료 구조 및 의료 정책 등과 같은 사항에 의해 결정되는 경우가 많으므로, 연구로부터 도출해내는 노출과 결과의 인과관계 설명력인 내적 타당도에 위협을 초래한다. 선택 바이어스(selection bias)와 혼란 요인(confounding factors)이 두 가지 주요 바이어스이다. 따라서 연구자들은 청구데이터를 이용한 관찰연구 시 이러한 바이어스 여부를 점검하고, 발견 시 바이어스가 통제되도록 연구를 설계해야 한다. 바이어스 제거 방법으로는 층화(stratification), 매칭(matching), 제약(restriction), 다변량 분석(multivariate analysis) 혹은 도구 변수(instrumental variable) 등이 사용되고 있다.

(2) 청구데이터의 유용성: 의료청구데이터는 연구 이용에 여러 가지 장점을 지니고 있다. 전 국민의 의료서비스 내용을 대표하는 데이터로서 일반화에 용이하다(representative). 두 번째로, 행위별수가제(fee-for-service)하의 의료 청구자료를 구축함으로써 처방약을 포함한 세부적인 의료 이용 내역을 포함하고 있다(complete). 또한 전국민의 의료서비스 사용을 포함하고 있으므로, 희소하게 발생하는 사건(ex. 유병률 혹은 발병률이 낮은 희귀질환, 합병증, 약물 부작용, 희소하게 사용되는 진료행위)에 대한 통계 파워를 확보할 수 있는 샘플 사이즈가 가능하다.

의료청구자료는 엄격한 통제하에 시행된 무작위 배정 임상시험하에서 수집된 데이터와는 달리, 실제 현실(real world)의 상태를 반영하고 있는 데이터이다. 즉 연구자들은 제한적이고 실험적 환경이 아닌 실제 의료보건 환경을 반영하여 현황 및 추세에 대한 관찰을 바탕으로 한 연구가 가능하다. 따라서 청구데이터가 주요 데이터 소스가 되는 연구로부터 도출되는 결과는 효능(efficacy)이 아닌 효과(effectiveness)를 볼 수 있다.

청구데이터가 가지고 있는 또 하나의 큰 장점은 비용과 시간에 있어서 경제적이라는 점이다. 청구데이터는 이미 수집되고 구축된 데이터이기 때문에 데이터 수집을 위한 비용과 시간이 들지 않는다. 따라서 연구실적이 필요한 젊은 연구자들에게 유용한 데이터 소스가 될 수 있다.

(3) 청구데이터의 제한점: 청구데이터의 의료보건의료를 위한 많은 유용성에서 불구하고, 연구자들이 청구데이터 활용 시 고려해야 할 제약성에 대해서는 인지할 필요가 있다. 의료서비스는 비급여 진료, 처방 없이 구입 가능한 약품(아스피린) 등에 대한 의료서비스 내역은 청구가 발생하지 않기 때문에 청구데이터에는 포함되지 않는 정보들이다. 또한 의료급여 환자, 보건소진료환자, 포괄수가제(Diagnostic Related Group) 대상 환자들의 경우, 명세서 일반 사항은 있으나 구체적인 진료 내역이 누락되는 경우가 많다. 따라서 이러한 연구대상 혹은 관련 질환 연구 시에는 주의를 할 필요가 있다.

두 번째로 고려해야 할 사항은, 의료청구데이터에 포함된 상병 정보의 정확성에 대한 논란이다. 의료청구데이터의 상병 정보에 오류를 일으키는 요인으로는 의도적 오류와 비의도적 오류가 있다. 진료비 삭감이나 불리한 평가를 피하기 위한 상병 코드 허위 추가나 업코딩시 발생하게 되는 ‘의도적 오류’와 정보 부족, 부실한 정보관리 및 상병 코딩 지침 미숙지에 따른 비의도적 오류로 구분할 수 있다.

의료청구데이터상의 진단명 정확성은 질병, 요양기관 종별, 입원 및 외래별로 정도의 차이를 보이고 있다. 2003년도에 실시된 ‘건강보험 질병코드의 타당도 제고방안 및 자료의 활용 방안 수립 연구’에 따르면, 주상병의 경우, 입원은 70%, 외래는 56% 일치도를 보였으며, 부상병의 경우, 49%, 38%의 일치도를 보였다. 또한 중증 질환이 경미한 질환보다, 병원이 의원보다 더 높은 정확성을 보였다¹.

이러한 진단명 정확성의 논란은 연구 대상(study population)이 해당 질환 유무 기준으로 추출 시 연구대상의 정확성에 대한 신뢰도가 저하되게 된다. 즉 해당 질환의 진단명이 존재한다고 하여 수진자가 반드시 그 질환을 가졌다고 볼 수 없으며, 연구 결과 분석 시 심각한 오류를 가져올 수 있다. 이러한 문제들은 비단 우리나라의 의료청구데이터만 지니고 있는 제한점은 아니나, 심사와 평가라는 우리나라의 독특한 제도로 인하여 진단명 타당성에 대한 문제 정도가 보다 심각할 수 있다는 우려가 제기된다.

진단명의 부정확성을 해결하기 위한 가장 이상적인 방법으로는 오류를 발생시키는 원인을 규명하고 시정하면 될 것이다. 하지만 심사와 평가라는 현 시스템에 제도적 수정을 가하는 데에는 사회적 합의와 법규정의 개정이 필요하기 때문에 장기간의 접근을 필요로 한다. 따라서 연구자들은 현재 주어진 데이터를 최대한 잘 활용하여 진단명 타당성에 대한 문제를 접근할 필요가 있다.

의료청구데이터의 진단명 정확성 문제를 보완하기 위한 보다 현실적인 방안으로, 조작적 정의(working definition)를 이용한 질환 보유 유무를 추려낼 수 있다. 조작적 정의를 진단명과 아울러 관련 질환을 치료하기 위한 의료서비스 내역(시술, 처치 혹은 처방전 약) 사용이 동시에 발생한 수진자들만이 실제로 해당질환을 가졌다고 정의하는 것이다. 이와 같은 조작적 정의와 더불어 수행되어야 할 사항은 개발된 조작적 정의가 어느 정도의 정확성을 가지는가를 검증하는 후속 연구(validation)가 요구된다. 질환별 조작적 정의 개발 및 검증은 연구자들에게 의료청구데이터를 활용한 좋은 연구 주제가 되기도 한다.

마지막으로, 의료청구데이터에 포함되지 않은 중요한 보건의료 정보가 있을 수 있다. 각종 임상검사(lab data) 및 신체기능에 대한 측정값, 몸무게, 키, 흡연여부, 소득, 거주지와 같은 수진자에 대한 정보가 부족하다. 또한 성과연구에서 중요한 결과 중 하나인 사망 정보 역시 포함되어 있지 않다. 이렇게 부족한 정보들은 타기관이 보유하고 있는 데이터와의 연계(병원 의무기록, 국민건강영양조사자료, 건강검진자료 등)로 보완할 수 있다. 따라서 청구데이터의 활용도를 제고하는 데 있어서는 타기관과의 연계 여부가 의료보전 관련 데이터를 보유한 기관들이 연구 활성화를 위해 풀어야 할 과제라고 할 수 있다.

참 고 문 헌

1. Park BJ, Jung JH, Park KD, Soe SW, Kim SW. Study on Diagnosis Accuracy for Health Insurance Claims data in Korea, report to Health Insurance Review and Assessment Services, Seoul: Medical School of Seoul National University; 2003.