

## 보건의료서비스 연구를 위한 건강보험심사평가원 청구 데이터의 소개 및 활용

김지애 (건강보험심사평가원 부연구위원)

건강보험심사평가원(심평원)은 2009년도부터 연구용 청구데이터 개방을 준비해, 2012년도부터 점진적인 개방을 시작했다. 작년 2013년도는 “빅데이터” 활용이 화두가 되었던 해로, 빅데이터가 가진 잠재적 가치를 실제로 구현하기 위하여, 정부는 정부 3.0의 세부과제에 “공공기간 보유정보의 개방과 공유”를 추진함에 따라 청구데이터 공개는 더욱 가속화가 되었다. 현재 청구데이터를 이용한 다수의 연구들이 활발하게 진행되어 왔으며, 연구 결과는 국내 학술지에 게재되고, 보건의료서비스정책 결정과 수립에 과학적인 근거로 활용되고 있다. 저자는 본글을 통해 청구데이터를 이용한 연구에 관심 있는 보건의료 연구진들에게 우리나라 건강보험 청구데이터에 대한 간략한 소개와 사용 방법에 대한 안내를 하고자 한다.

**건강보험 청구데이터란?** “건강보험 청구데이터”란 요양기관이 의료서비스를 제공한 후 환자의 진료비용 중 ‘국민건강보험’이 부담하는 부분에 대해 지급의뢰를 하기 위하여 ‘건강보험심사평가원’에 보험급여 청구를 하면서 발생하는 데이터이다. 우리나라의 1년간 ‘건강보험 청구 환자수’는 주민등록인구의 90%에 해당하는 약 4천6백만 명으로, 전국의 8만여 개 요양기관으로부터의 방대한 청구건이 포함되어 있다. (2011년도 기준)

**연구용 청구데이터 구성 및 포함 변수** 연구자들에게 제공되는 청구데이터는 네개 테이블로 나눠져 있다. 첫번째는 명세서일반내역(Table 20)로 인구학적 특성변수(나이, 성별), 수진자 주민번호 대체키, 주상병과 부상병, 요양기관 대체키, 입원 및 외래 구분 변수(서식코드) 및 의료급여종별코드가 포함되어 있다. 이러한 변수들은 안내변수(instruction variables)라고 하며, 연구대상을 추출하는데 있어 활용도가 높은 변수들이다. 두번째 테이블은 진료내역(Table 30)으로 환자들이 입원하여 발생하는 모든 의료서비스 정보들을 포함하고 있다. 따라서 환자들에게 제공된 진료의 행위 및 약제 등(검사, 처치, 시술, 약제 등)에 대한 자세한 정보를 담은 진료내역 정보를 확인 할 수 있다. 세번째 테이블은 상병내역(Table 40)으로 주상병, 부상병을 포함하여 모든 동반상병정보를 담고 있어, 상병내역정보를 활용하는 경우는 임신과 같이 질병이 아닌 경우가 상병코드로 불는 경우 모든 동반질병내역을 확인하여 연구를 진행하는 경우를 예로 들 수 있다. 마지막으로 원외처방내역(Table 53)으로 환자가 외래처방으로 받은 약제에 대한 모든 정보를 포함하고 있다.

**청구데이터 사용 방법** 청구데이터를 이용하는데에는 세가지 방법이 있다. 첫번째는 심평원내에 위치한 자료처리실을 이용하는 방법이다. 자료처리실은 일종의 컴퓨터 랩(LAB)으로, 연구자들은 연구 주제에 맞게 맞춤형으로 데이터가 세팅된 컴퓨터를 할당 받아 이용하게 된다. 자료처리실은 주로 대용량의 자료 혹은 타기관의 자료와의 연계가 필요한 경우 이용된다. 외부보유자료 연계는 환자의 포괄적 사전 동의 획득이나, 건강보험청구자료와 연계에 대한 동의 획득 시 연계가 가능하다. 두번째 방식으로는 원격접속 서비스를 통한 데이터 이용이다. 2014년 4월 초부터 시행될 원격접속서비스는, 연구자들의 컴퓨터를 이용해서 심평원의 서버로 접속하여, 서버에 저장되어 있는 데이터를 접속하고 사용하게 되는 방법이다. 자료 처리실과 달리, 연구자들은 장소나 시간의 구애 없는 상시적인 데이터 이용이 가능하게 된다. 세번째로 무작위증화로 추출된 1년 단위의 표본데이터를 이용하는 방법이다. 표본자료는 2009-2011년도로 3년간 구축되어 있으며, 표본자료의 종류는 전체환자표본(NPS), 입원환자표본(NIS), 노인환자표본(APS), 소아·청

소년환자표본(PPS)으로 나누어 제공하고 있다. 표본데이터가 가지고 있는 제한성을 보완한 장기간 follow-up이 필요하고 보다 전문적이고 임상적인 영역 연구를 지원하기 위하여, 임상 코호트를 구축중으로 필요시 외부 자료와 연계를 지원하고 있다. 현재 산부인과와 COPD 질환의 코호트가 구축이 진행중이다.

**청구데이터 사용 절차** 자료처리실과 원격 접속을 통한 데이터 사용은 심의절차를 거쳐야 하며, 표본데이터의 경우는 별다른 심의과정 없이 소정의 데이터 사용 수수료를 지불한 뒤 연구자들이 직접 구입해 사용할 수 있다. 자세한 사항은 건강보험심사평가원 웹사이트에 안내되어 있다([건강보험심사평가원\\*정보\\*통계정보\\*진료정보이용안내](#)).

**청구 데이터의 유용성 및 고려할 점** 청구데이터는 전 국민의 의료서비스 내용을 대표하는 자료로써 일반화에 용이하며(대표성), 행위별수가제(fee-for-service)하의 청구데이터를 구축함으로써 처방약을 포함한 세부적인 의료 이용 내역을 포함하고 있다(포괄성). 또한 전국민의 의료서비스 사용을 포함하고 있으므로, 희소성(ex. 유병률 혹은 발병률이 낮은 희귀질병, 합병증, 약물 부작용, 희소하게 사용되는 진료행위)에 대한 확보가 용이하다. 청구데이터는, 엄격한 통제하에 수집된 일차자료와는 달리, 실 사회(real world)의 상태를 반영하고 있다. 즉 연구자들은 제한적이고 실험적 환경이 아닌 실제 의료보건 환경을 반영하여 현황 및 추세에 대한 관찰을 바탕으로 한 연구가 가능하다. 또한 가장 큰 장점 중 하나는 청구데이터는 이미 수집되고 구축된 데이터로 비용측면에서 효율적이며, 데이터 수집에 소요되는 시간을 단축할 수 있다. 이러한 유용성에도 불구하고 연구자들이 청구데이터가 가지고 있는 제한성을 인지하고 연구 설계시 고려해야한다. 우선 비급여 진료, 처방 없이 구입 가능한 약품(아스피린) 등에 대한 의료서비스 내역은 청구가 발생하지 않으므로 청구데이터에는 포함되어 있지 않다. 또한 의료급여 환자, 보건소진료환자, 포괄 수가제(Diagnostic related group, DRG) 대상 환자들의 경우, 명세서 일반 사항은 있으나 구체적인 진료 내역이 누락되는 경우가 많다. 또한 청구데이터에 포함된 상병 정보의 정확성에 대한 논란이 있다, 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 분석시 심각한 오류를 가져올 수 있다. 이러한 문제들은 비단 우리나라의 의료 청구데이터만 지니고 있는 제한점은 아니나, 심사와 평가라는 우리나라의 독특한 제도로 인하여 정확성 정도가 보다 심각할 수 있다는 우려가 제기된다. 청구데이터의 진단명 정확성 문제를 보완하기 위한 방법으로, 조작적 정의를 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. 마지막으로, 청구데이터에 포함되지 않은 중요한 보건의료 정보가 있을 수 있다. 예를 들면, 각종 임상 검사 및 신체기능에 대한 측정값에 대한 정보나 몸무게, 키, 혹은 질병력 및 가족력과 같은 수진자들에 대한 정보가 포함되어 있지 않으므로 청구데이터의 부족한 정보는 타기관이 보유하고 있는 데이터 (예. 병원 의무기록, 국민건강영양조사자료, 건강검진자료 등)와의 연계로 보완할 수 있는 방법이 있다. 심평원은 각 기관들이 수집하고 보유하고 있는 개별 데이터의 연계로부터 보건의료분야의 발전과 가치창출을 가져 올 수 있는 “보건의료 빅데이터”가 만들어질 수 있다는 판단하에 데이터 연계를 지원하고 있다.