

# 의약품 안전관리를 위한 빅데이터의 활용

정선영<sup>1</sup> · 최남경<sup>2</sup> · 이종엽<sup>2</sup> · 박병주<sup>1,3</sup> | <sup>1</sup>한국약품안전관리원 <sup>2</sup>서울대학교병원 의학연구협력센터, <sup>3</sup>서울대학교 의과대학 예방의학교실

## Use of big data for drug safety monitoring and decision making

Sun-Young Jung, PhD<sup>1</sup> · Nam-Kyong Choi, PhD<sup>2</sup> · Joongyub Lee, MD<sup>2</sup> · Byung-Joo Park, MD<sup>1,3</sup>

<sup>1</sup>Korea Institute of Drug Safety and Risk Management, Seoul, <sup>2</sup>Medical Research Collaborating Center, Seoul National University Hospital, <sup>3</sup>Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea

The development of information technologies has led to the era of big data; such enormous collections of information on drugs and adverse drug reactions are stored in either a structured, a semistructured, or an unstructured form. Because of the nature of the emerging issue of drug safety, it is common for policy makers and healthcare professionals to make decisions without sufficient evidence. Big data may be used as an efficient pharmacovigilance tool, which enables us to recognize adverse drug reactions that may not have been identified in pre-marketing clinical trials, in order to capture the patterns of drug utilization and adverse events, and to predict the occurrence of adverse drug reactions. National surveillance systems using electronic health databases have been established successfully in the US and Europe. The Korea Institute of Drug Safety and Risk Management (KIDS) plans to establish a big data platform for pharmacovigilance in Korea. The big data may be effectively used for evidence-based regulatory and clinical decision making in the field of drug safety and risk management.

**Key Words:** Big data; Pharmacovigilance; Pharmacoepidemiology

### 서론

빅데이터는 여러 가지 형태로 정의되고 있으나 대표적으로 미국 IT리서치기업인 가트너는, 빅데이터는 "의사결정의 기반을 강화하고 새로운 통찰이나 가치를 제공할 수 있도록, 비용-효과적이고 혁신적인 처리과정을 필요로 하는 큰 규모,

빠른 속도 및 높은 다양성을 가지는 정보"를 의미한다고 정의한 바 있다[1]. 최근 빅데이터를 분석하여 유용한 정보를 생성할 수 있는 기술이 발전하면서 마케팅, 금융, 교육, 치안, 환경, 교통, 농업, 보건의료 등 다양한 분야에 활용되고 있으며, 가트너는 빅데이터를 '21세기의 새로운 원유'라고 지칭하기도 하였다. 미국에서는 빅데이터 진흥정책으로서 2012년 3월 19일 '빅데이터 R&D 이니셔티브'를 발표한 이후 대규모 데이터를 수집, 저장, 보관, 관리, 분석, 공유하는데 필요한 핵심기술을 개발하고, 전문인력을 양성하고 있다. 우리나라에서도 역시 정보화진흥원 산하에 '빅데이터분석활용센터'를 설립, 운영하는 등 미래창조과학부와 정보화진흥원을 주축으로 빅데이터를 효과적으로 활용하기 위한 방안을 마련하기 위하여 국가차원에서 적극적으로 노력하기 시작하였다.

Received: April 6, 2014 Accepted: April 20, 2014

Corresponding author: Byung-Joo Park  
E-mail: bjpark@snu.ac.kr

© Korean Medical Association

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

대용량 자료로부터 일정 패턴을 파악하여 새로운 통찰을 제공하고 예측에 필요한 근거를 제공함으로써 의사결정에 도움을 줄 수 있다는 유용성이 밝혀지면서, 보건 의료 분야에서 빅데이터 기술을 적용하여 질병발생양상을 파악하여 확산을 방지하고, 개인 특성에 맞는 맞춤형 치료, 예측, 위험요인 규명 등에 필요한 근거기반을 제공할 수 있으리라는 기대로 이 분야에 대한 관심이 크게 증가하고 있다[2]. 보건 의료 분야에서의 대표적인 빅데이터 활용사례로 언급되는 Google flu trend는 구글 검색키리를 이용하여 인플루엔자 의사환자(influenza-like illness)를 파악하는 알고리즘으로, 이것이 해당 질환으로 의료기관을 방문하는 빈도의 추이와 높은 상관성을 보인다는 것이 확인되어 인플루엔자 발생동향의 실시간 감시에 활용되고 있으며, 향후 인플루엔자 발생의 조기예측에도 활용될 것으로 기대된다[3,4].

의약품은 임상시험을 통하여 유효성과 안전성을 평가한 후 임상적으로 그 가치가 인정될 때 시판이 허가되지만, 시판 이후에는 남녀 구분 없이 모든 연령층, 다양한 질환을 가진 많은 수의 환자들에게 임상시험에서 확인되지 않은 용량과 용법으로도 의약품을 사용하게 되므로 임상시험 단계에서는 미처 발견하지 못하였던 드물게 발생하는 중대한 약물유해반응이 나타날 가능성이 항상 존재한다. 따라서 시판 후에 발견되는 안전성 문제로부터 환자를 보호하여 안전하고 합리적이며 효과적인 의약품 사용이 가능하도록 지속적인 약물감시(pharmacovigilance)가 필수적이다. 약물감시의 근간이 되는 방법은 자발적부작용보고제도로써, 식품의약품안전처 산하의 공공기관인 한국의약품안전관리원에서는 시판 후에 의약품으로 인하여 발생하였다고 의심되는 유해사례를 제약업체, 의사와 약사 및 간호사 등 환자진료에 직접 참여하는 의료전문가들 또는 소비자로부터 자발적으로 보고받고 있다. 자발적 부작용보고자료는 부작용 문제를 파악하는데 중추적인 역할을 담당하고 있으며, 약물의 위험을 조기에 파악하여 시기적으로 빠른 결정을 내릴 수 있도록 실마리정보(Signal)를 감지해내는 데이터마이닝 기법이 개발되어, 국내에서도 자발적 부작용보고자료를 대상으로 데이터마이닝기법을 적용하여 안전성정보를 생성하고 있다. 그러나 자발적으로 보고되는 정보는 실제 발생한

유해사례 중 일부만 보고되기 때문에 과소보고의 문제가 있고, 치료기간과 같은 임상정보가 불충분한 경우가 많으며, 약물을 투여받은 분모집단에 대한 정보가 없기 때문에 실제 환자들에서 발생한 위험수준을 정확하게 평가하기 어려운 한계점이 있다[5]. 따라서 자발적 부작용보고의 한계점을 보완하면서 의미 있는 안전성 문제를 파악하기 위한 능동적 약물감시방법(active surveillance)으로서 빅데이터가 또 하나의 중요한 자료원으로 활용될 수 있다.

이와 같이 전통적으로 이루어지고 있던 수동적 약물감시체계의 미비점을 보완하기 위하여 의료보험청구자료, 병원의무기록자료, 사망자료, 인터넷 상의 정보 등을 다각도로 활용하여 이전에 알려지지 않았던 안전성 문제를 능동적으로 파악해내고자, 이 분야에도 빅데이터의 활용에 대한 관심이 증대하고 있다. 본 고에서는 의약품안전관리 분야에서 빅데이터 활용의 유용성과 활용동향을 살펴보고자 한다.

### 근거중심 의약품안전성 의사결정에 빅데이터 활용의 유용성

실제로 약물부작용으로 시판이 철회된 경우를 살펴보면 이전에 밝혀지지 않았던 중대한 부작용이 발견되어 시장에서 철회되기까지 시판 후 수년에 걸친 긴 시간이 소요된다. 임신초기 입덧에 효과적인 항구토제로 폭발적으로 사용된 이후 태아에서 사지결손증을 유발한 탈리도마이드는 1957년 독일에서 허가 받은 이후 유럽, 아프리카, 일본을 포함하여 전 세계 40여 개국에서 널리 사용되었으나 결국 1만 명 이상의 기형아 출산을 야기하고 1961년에 시판중지 결정이 내려졌다. 위식도 역류성질환의 대표적 치료제로 세계적으로 엄청난 매출을 기록하였던 시사프라이드의 경우 미국 내 1993년 허가된 이후 부정맥 등 심장부작용 사례 341건이 자발적으로 보고됨에 따라서 2000년에 시판중지 되었으며[6], COX-2선택적 억제제로 개발된 블록버스터 약물이었던 로페록시브는 1999년 시판된 이후 시판 후 임상시험과 약물역학연구에서 심혈관계 부작용의 발생위험이 높다는 문제

가 제기되어 2004년에 시판을 중단하게 되었다. 최근 호주 연구에 따르면, 로페콕시브가 1999년 시판되기 시작한 이후, 2001년 말까지 호주 식약청(Australian Therapeutic Goods Administration)으로 자발적으로 보고된 심근경색 증 사례는 단 3건이었지만, 보험청구자료를 이용한 분석에서는 2001년에 이미 심근경색증 발생위험이 의미있게 증가한다는 실마리정보를 찾아낼 수 있었고, 2002년에는 자발적 부작용보고자료를 이용한 분석에서 실마리정보를 찾아내어, 데이터마이닝기법을 이용한 실마리정보 검색을 통하여 2004년 시판철회에 앞서 미리 의약품안전성 문제를 발견해낸 결과를 보여주면서, 자발적 보고자료를 이용한 실마리정보 검색에 대규모 자료원을 이용한 분석이 보완적으로 활용될 수 있음을 시사하였다[7].

의약품안전성 관련 정책결정은 약물유해반응으로 인한 환자들의 피해와 직결되는 만큼 공중보건에 상당한 영향을 미치게 되므로 시의적절한 근거중심의 의사결정(evidence-based decision making)이 대단히 중요하다. 그런데 의약품안전성 문제는 초기에는 정규 약물역학적 연구결과로 얻은 과학적 증거를 기반으로 제기되기 보다는 자발적 부작용보고를 통하여 다수의 유해사례 발생이 파악되거나, 환자사례보고 등의 경로로 제기되기 때문에, 일반적으로 관련 전문위원회에서 인과성을 평가하고자 할 때는 의심되는 약물과 제기된 유해사례 사이의 인과성을 평가할 수 있는 근거가 되는 충분한 데이터가 축적되지 않은 상태에서 결정을 내릴 수 밖에 없는 상황이 빈번히 발생한다. 즉, 해당 약물의 사용으로 인한 부작용의 발생률이 일반인구에 비해서 얼마나 높은지, 부작용 발생에 영향을 미칠 수 있는 다른 요인을 모두 배제하였을 때에도 부작용 발생위험이 여전히 의미 있게 존재하는지, 부작용 발생으로 인한 피해가 약물사용으로 인한 유익성의 정도와 비교할 때 얼마나 중대한지 등을 결정하기에 충분한 수준을 갖춘 근거가 마련되어 있지 않은 상황에서 공중보건에 막대한 영향을 미치는 결정을 내려야 하는 경우가 많기 때문에, 의약품안전성에 대한 의사결정의 과학화가 쉽지만은 않다.

실제로, 부작용 발생위험 때문에 시판이 철회된 약물이 사후에 위험성이 없었던 것으로 밝혀지는 사례도 있다. 1956년 허가 이후 미국 임신부의 3분의 1이 사용한다고 알려질 만큼

입덧치료제로 널리 쓰였던 벤벡틴의 경우 1980년대 기형유발 가능성에 대한 여론 때문에 자발적으로 시판중지를 하였으나 이후 다수의 약물역학연구를 통하여 기형유발 위험성이 없음이 밝혀졌다[8]. 프로폭시펜은 1957년에 허가된 이후 마약성 진통제로 널리 사용되었으나, 과량사용으로 인한 사망 위험과 심장부작용 가능성에 대한 증례보고와 QT간격 연장이라는 대리결과변수(surrogate endpoint)를 사용한 연구를 근거로 2010년 시판중지 결정이 내려졌다. 그러나 추후에 미국 테네시주 메디케이드 데이터베이스를 분석한 결과, 프로폭시펜을 사용한 환자들이 비사용자 또는 하이드로코돈을 사용한 환자에 비하여 급성 심장사와 총 사망의 위험이 높게 나타나지 않았다는 연구결과가 발표되었다[9].

의약품안전성 관련 정책적 의사결정의 과학화를 위하여 빅데이터를 활용하여 조기에 약물부작용 발생에 관한 정보를 파악하여 시기적으로 빠른 결정을 내릴 수 있도록 하여, 부작용으로 인한 피해의 확산을 방지할 수 있다. 또한, 빅데이터 상에서 나타나는 다양한 요인들과의 상관성을 이용하여 부작용 발생가능성을 예측하여 약물로 인한 위험으로부터 환자들을 보호하는 중요한 역할을 담당할 수 있다. 특정 약물로 인한 중대한 부작용 발생을 효과적으로 예측할 수 있는 예측모형을 개발함으로써 향후 임상현장에서 약물유해반응에 대한 임상적 의사결정지원시스템(clinical decision support system)에 탑재하여 환자진료에 직접적인 도움을 줄 수도 있다.

빅데이터는 임상시험에 비하여 실제 진료환경을 반영하고, 모든 연령층 및 질환군을 포함하며, 의약품의 장기적인 사용에 의한 영향을 저비용-고효율적으로 분석할 수 있다. 또한 의료보험청구자료 등의 대규모 자료는 자발적 부작용신고자료에 비해 의약품을 사용하는 분모집단에 대한 정보를 알 수 있고, 의약품처방정보, 진단명, 처치내용 등에 관한 객관적인 정보가 존재하며, 인과성 분석에 필요한 다양한 공변량을 확보할 수 있다는 강점이 있다[10]. 최근에는 과거에 알려지지 않았던 약물유해반응을 발견하는 과정에서 의료진의 보고뿐 아니라 환자가 직접 보고한 안전성문제(patient-reported safety outcomes) 역시 중요하게 다루어져야 한다는 의견이 있는데, 웹 상의 게시물이나 소셜네트워크서비스 상의 언어 등 비정형 빅데이터를 활용하게 되면 실제 환자들의 목소리

를 약물 안전성 모니터링에 이용할 수 있다[11]. 소셜미디어 게시물 등을 분석함으로써 기존의 자료로는 파악할 수 없었던 의약품의 사용패턴을 파악함으로써 의약품 오남용 관리 등을 위한 전략을 수립하는데 큰 도움을 받을 수 있다.

### 의약품안전관리를 위한 빅데이터 활용 동향

최근 수 년간 선진국에서는 각기 흩어져 있는 대규모 데이터베이스를 통합하여 약물과 부작용 발생간의 관련성을 분석할 수 있는 능동적 약물감시체계를 구축하는 추세로 빠르게 변화하고 있다. 미국에서는 시판 후 약물감시체계를 강화하는 내용을 포함하는 2007년 Food and Drug Administration (FDA) 강화법(FDA Amendments Act) 통과 이후, 약물안전성 모니터링을 위한 국가전산시스템 구축 프로젝트인 센티넬 이니셔티브(Sentinel Initiative)의 운영을 의무화하여, 국내에 흩어져 있는 건강관련 자료원을 통합하여 2012년 7월까지 1억 명 환자에 대한 자료를 구축하고, 약물안전관리시스템을 구축하기 위한 인력 및 재원을 확충하도록 하였다. 이에 따라 2009년부터 2013년까지 시범사업으로서 미니센티넬(Mini-Sentinel)을 운영하여, 특정 약물에 대한 안전성 이슈가 발생하였을 때 FDA가 협연센터인 하버드필그림건강연구소에 이 문제를 의뢰하면, 협연센터에서 각 데이터협력기관에서 분석한 결과를 통합하는 분산망 형태의 분석체계를 구축하였다[12]. 또한, 2009년부터 2013년까지 제약회사, 학계, 비영리단체, FDA 등의 네트워크로 구성된 민관협력 프로젝트인 OMOP (Observational Outcomes Medical Partnership)을 운영하여 데이터베이스를 이용한 분석방법론을 정립하도록 하고, 현재는 Observational Healthcare Data Sciences and Informatics 프로젝트로 발전하여 운영하고 있다. 유럽의약품청(European Medicines Agency, EMA)에서는 2006년부터 유럽 내 약물역학전문가들과의 협력을 통하여, European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP)를 운영하여, 유럽 전역의 약물역학연구 자료원 간 연계, 학계와 규제기관 간 협업을 통한 시판 후 의약품모니터링을 강

화하고 있다. ENCePP에서는 질적으로 표준화된 연구가 수행될 수 있도록 연구계획 수립과 수행 등에 대한 가이드라인을 개발하고 교육프로그램을 제공하고 있다. 미국 FDA에서 주제를 지정하는 중앙집중형식의 센티넬시스템과는 달리, 이 네트워크에서는 각 연구자들이 주제를 제안하여 ENCePP의 연구표준에 맞추어 등록하고, ENCePP에서는 이를 검토하여 ENCePP연구로 인증하고 수행하도록 하고 있다[12]. 또한 EMA와 제약업계 간 협력체계인 IMI (Innovative Medicines Initiative)에서 2009년부터 운영하고 있는 PROTECT (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium) 프로젝트 내에서도 전자의무기록을 통한 실마리정보의 신속한 감지 방법론에 대한 논의가 이루어지고 있다. 이와 같이 미국과 유럽에서는 전자데이터기반 약물감시에 대한 국가의 전폭적인 재정적 지원과 정부, 학계, 산업계의 협력을 기반으로 한 국가프로젝트를 통하여 미국 내, 또는 유럽 각국에서 다양한 형태로 구축된 데이터베이스에서 공통수집변수체계가 마련되었고, 각 자료원에서 약물노출, 진단명 및 처치명, 동반질병 등 약물-부작용 간 인과관계평가에 영향을 미치는 교란변수들을 보정한 심층분석이 빠른 시간 내에 수행될 수 있는 기반이 구축되었다. 이는 기존에 안전성 이슈가 나타났을 때 새로운 연구를 수행하여 결과를 얻는 데까지 오랜 기간이 소요되어 온 것에 비하여 획기적인 변화이며, 앞으로는 약물안전성 이슈를 다루는 인프라와 기술력이 본격적으로 발휘될 전망이다.

대규모 전자의무기록데이터베이스뿐만 아니라, 웹 상의 정보와 같은 비정형자료를 활용하는 방법론의 개발도 가속화되고 있다. 환자들이 건강정보를 인터넷으로 검색하는 빈도가 높다는 점을 감안하여, 2013년에는 구글 검색로그에서 약물 또는 부작용에 대한 검색을 추출하여 웹을 통하여 실마리정보를 찾아내는 연구결과가 보고되었다. 전자의무기록을 활용한 연구 및 동물실험 등으로 확인된 약물유해반응으로 ‘항우울제 파록세틴과 지질저하제 프라바스타틴 간 약물상호작용으로 인한 고혈당증’을 검증모델로 하여, 고혈당증과 관련된 증상이나 용어를 검색한 사용자를 정의하고 이것이 파록세틴과 프라바스타틴을 검색한 사람과 일치하는 지를 데이터마이닝지표로 산출하여 평가한 것

이었다[13]. 약물부작용의 경우 웹 상의 정보뿐 아니라 환자사례보고가 이루어지는 경우가 많아 전문분야정보가 중요하다라는 점을 감안하여, 최근에는 전자의학학술정보원인 메드라인(Medline) 내 2천만 여건의 문헌정보로부터 약물-유해사례 정보를 추출한 뒤 이를 미국 자발적 부작용보고자료와 함께 분석함으로써 실마리정보 도출의 효율성을 증대시킬 수 있다는 연구결과가 보고되었다[14].

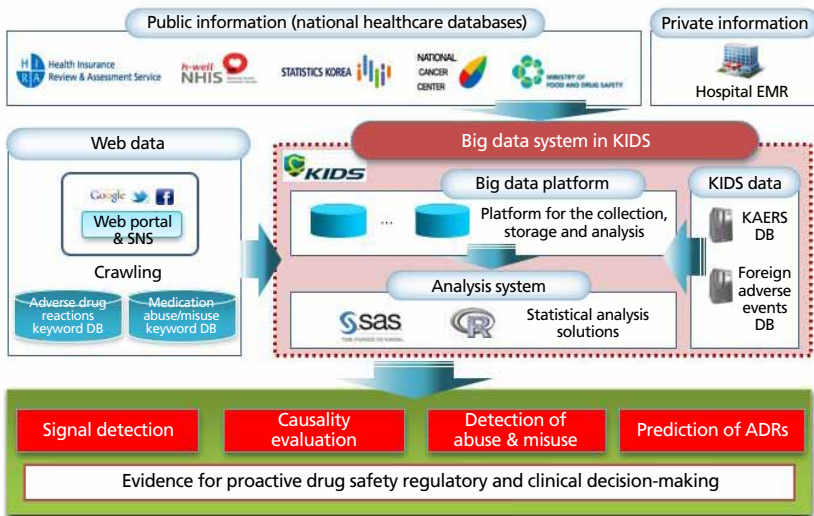
국내 역시 빅데이터를 활용한 안전성 감시방안이 개별 연구차원으로는 다양한 시도가 이루어지고 있다. 의약품부작용의 실마리정보검색을 위한 데이터마이닝 분야로는 건강보험심사평가원 보험청구자료를 활용하여 의약품을 처방받은 이후에 입원하는 건을 대상으로 한 실마리정보 검색의 유용성을 검증한 바 있고[15], 국내 병원전자의무기록 자료를 이용한 실마리정보 검색알고리즘을 개발하기도 하였다[16]. 건강보험심사평가원 요양급여청구자료를 이용하여 항정신병약[17], 대장내시경 전 장세척제[18], 혈압강하제[19,20] 등 의약품과 부작용 간 인과성 평가를 실시하기도 하였다. 시사프라이드가 심장부작용 위험성으로 시판철회된 이후, 우리나라 보험청구자료와 통계청 사망자료를 연계하여 후향적으로 분석한 결과 시사프라이드와 대사효소를 같이 함으로써 상호작용을 유발하는 약물을 병용한 경우 사망률이 14배 증가함을 확인하기도 하였다[21]. 시사프라이드와 병용금기약물의 사용에 따른 사망위험 분석은 건강보험심사평가원 자료와 통계청 사망자료를 연계하여 분석한 사례이나, 최근에는 개인정보보호법의 강화로 자료원 간 연계분석이 가능하지 않게 된 것은 공익적 목적의 연구를 수행하지 못하게 만든 문제가 있다.

비정형데이터 분석과 관련하여 한국의약품안전관리원에서는 2013년 미래창조과학부와 한국정보화진흥원 시범사업의 하나로 '의약품 안전성 조기경보서비스' 과제를 수행한 바 있다. 웹을 통하여 파악되는 의약품 부작용의 실마리정보를 병원전자의무기록자료를 통하여 검증하고 인터넷에서 언급되는 의약품 오남용의 실태를 실시간으로 파악할 수 있는 가능성을 확인하였으나, 실제 적용을 위해서는 다양한 사례에 확대 적용하는 방법론의 개발과 심층적인 타당도 검증이 필요 한 것으로 파악되었다. 식품의약품안전처에서는 2014년 주

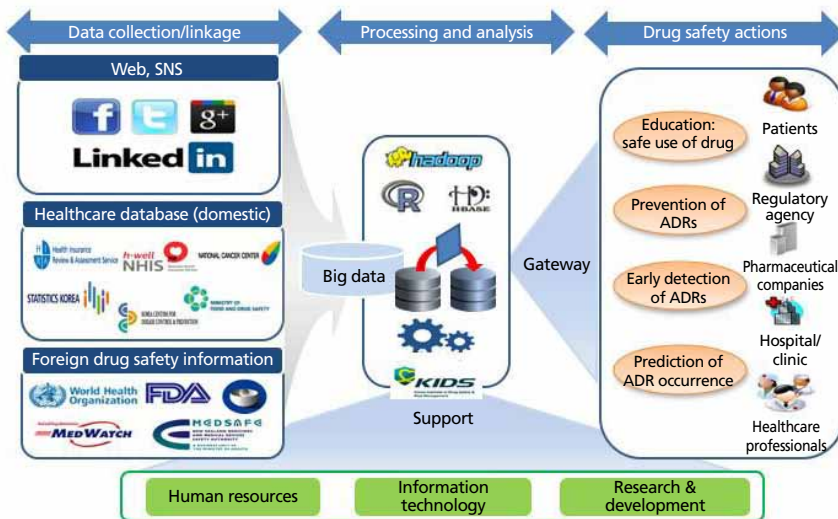
요 사업으로 '의약품의료정보 연계시스템 구축'을 포함하여 의약품 허가정보, 부작용 및 진료기록 등 의료정보를 통합 연계하여 부작용 인과성 평가에 이용할 수 있는 빅데이터 활용 의약품안전관리플랫폼을 구축하고자 시범사업을 추진하고 있다. 이 사업에서는 건강보험심사평가원, 건강보험공단, 통계청, 국립암센터, 식품의약품안전처 등 국가단위의 공공정보와 병원전자의무기록, 웹으로부터 수집된 정보 등을 국내 자발적 보고자료 및 국외 자발적 보고자료와 함께 포괄적으로 분석하여 의약품부작용의 실마리정보를 감지하고 의약품-부작용 간 인과관계를 평가하며, 의약품오남용 양상을 파악하고 부작용 발생을 예측함으로써 의약품 안전성 관련 정책적 및 임상적 의사결정에 필요한 근거자료를 생성할 수 있도록 플랫폼을 구축하고 있다(Figure 1). 우리나라에는 전국 민간건강보험데이터베이스와 각 병원의 전자의무기록자료, 각종 질병등록자료 등이 이미 잘 구축되어 있고 우수한 IT인프라를 갖추고 있다. 이를 연계분석함으로써 미국이나 유럽과 같은 빅데이터 활용체계를 갖출 수 있도록 정책적, 제도적 뒷받침이 필요하다. 미국의 센티넬이니셔티브와 같이 인프라를 충분히 구축하고 본격적인 운영의 기반을 닦기 위한 장기 시범사업프로젝트, 정부, 학계, 산업계의 상호협력체계를 도입하는 방안도 적극 추진할 필요가 있다. 의약품 안전성에 대한 빅데이터 분석에는 의약 전문지식과 정보기술 전문성이 결합된 전문인력이 필요하나 아직까지는 턱없이 부족한 실정이므로, 전문인력 양성을 위한 교육프로그램의 개발 역시 함께 추진되어야 할 것이다. 궁극적으로는 국내 각종 비정형 및 정형 자료원을 활용하여, 시의적절한 근거중심의 의약품안전성 의사결정이 가능하도록 빅데이터 기반의 의약품안전성 관리체계가 마련되어야 할 것이다(Figure 2).

## 의약품안전관리를 위한 빅데이터 활용 시 고려사항

빅데이터를 의약품안전관리를 위하여 사용하고자 할 때, 자료생성의 일차목적이 의약품 안전성을 평가하기 위한 것이 아닌 데이터일 경우, 각 자료의 특성을 충분히 고려하여야 한다.



**Figure 1.** Big data platform model by Korea Institute of Drug Safety and Risk Management (KIDS). EMR, electronic medical record; SNS, social network service; DB, database; KAERS, Korea Adverse Event Reporting System; ADR, adverse drug reaction.



**Figure 2.** A framework of future drug safety monitoring system based on the big data. SNS, social network service; ADR, adverse drug reaction.

의료보험청구자료를 이용하는 경우 진단명의 타당도, 보험급여체계, 청구관행 등이 연구결과에 영향을 미칠 수 있다. 비정형자료를 이용하여 실마리정보를 감지하는 단순한 상관성 분석결과로는 약물에 의한 부작용인지, 부작용에 따른 의약품 사용인지 선후관계를 따지기 어려울 수 있으므로, 이를 정의하기 위한 다양한 쿼리를 개발할 필요가 있을 것이다. 또한 인터넷 기반의 비정형 빅데이터를 활용하는 경우, 분석에 활용되는 검색쿼리의 정의에 대한 신중한 검토가 필요하다. 대중들이 다빈도로 사용하는 검색쿼리는 미디어의 동향, 인터넷

검색의 행태, 지역적 문화특성 등에 강하게 영향을 받는다. 예를 들어 Google flu trend가 여러 국가에서 인플루엔자 모니터링에 활용될 수는 있지만, 국내 인플루엔자환자 감시체계에 활용되기 위해서는 국내 환자들의 문화적 특성과 국내 보건 의료체계 특성을 반영한 한국어 검색어를 검색쿼리에 포함하여야 한다[22]. 마지막으로 의약품 안전성 분야에는 데이터화 되지 않은 문제 역시 존재하므로, 이를 파악하기 위한 추가적인 노력은 지속되어야 할 것이다.

## 결론

빅데이터는 약물안전성 문제를 조기에 발견하고 예측하며, 새로운 약물사용양상이나 부작용 발생양상을 발견해 내는 등 자발적 부작용보고의 한계점을 보완하면서 의미있는 안전성 문제를 파악하는 저비용, 고효율적인 능동적 약물감시방법의 재료로 사용될 수 있다. 국외에서는 대규모 자료원을 활용한 의약품안전관리체계가 활발히 구축되어 이미 운영단계로 이행되고 있으며, 국내에도 개별 연구의 차원을 넘어서 국가 의약품안전관리에 빅데이터를 활용하려는 변화의 바람이 불기 시작하였다. 그러나 아직까지는 국내 대규모 자료원들을 의약품안전관리를 위한 공식적 목적으로 사용하기에는 개인정보보호 등의 이유로 제한되는 등 한계가 있으며, 인적, 물적 인프라가 부족한 실정이다. 빅데이터 기반의 효율적인 한국형 약물안전관리체계 구축을 위하여 국회에서는 개인정보보호법의 개정 등 관련 법제도를 정비하고, 미래창조과학부, 기획재정부 및 보건복지부와 식품의약품안전처는 전문인력의 양성과 연구개발을 위한 예산을 지원

하여야 하고, 이러한 기반이 구축되면서 학계, 산업계 및 정부 간 효과적인 의사소통과 협업체계를 구축하여 운영함으로써 소기의 목적을 달성할 수 있도록 함께 노력하는 것이 필요하다.

**찾아보기말:** 빅데이터; 약물감사; 약물역학

**ORCID**

Sun-Young Jung, <http://orcid.org/0000-0003-2032-112X>  
 Nam-Kyong Choi, <http://orcid.org/0000-0003-1153-9928>  
 Joongyub Lee, <http://orcid.org/0000-0003-2784-3772>  
 Byung-Joo Park, <http://orcid.org/0000-0003-4630-4942>

**REFERENCES**

1. Beyer MA, Douglas L. The importance of 'big data': a definition. Stamford: Gartner; 2012.
2. Issa NT, Byers SW, Dakshanamurthy S. Big data: the next frontier for innovation in therapeutics and healthcare. *Expert Rev Clin Pharmacol* 2014;7:293-298.
3. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009;457:1012-1014.
4. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, Rothman RE. Influenza forecasting with Google flu trends. *PLoS One* 2013;8:e56176.
5. Chung SY, Jung SY, Shin JY, Park BJ. The role of the KIDS for enhancing drug safety and risk management in Korea. *J Korean Med Assoc* 2012;55:861-868.
6. Wysowski DK, Corken A, Gallo-Torres H, Talarico L, Rodriguez EM. Postmarketing reports of QT prolongation and ventricular arrhythmia in association with cisapride and Food and Drug Administration regulatory actions. *Am J Gastroenterol* 2001;96:1698-1703.
7. Wahab IA, Pratt NL, Kalisch LM, Roughead EE. Comparing time to adverse drug reaction signals in a spontaneous reporting database and a claims database: a case study of rofecoxib-induced myocardial infarction and rosiglitazone-induced heart failure signals in Australia. *Drug Saf* 2014;37:53-64.
8. McKeigue PM, Lamm SH, Linn S, Kutcher JS. Bendectin and birth defects: I. A meta-analysis of the epidemiologic studies. *Teratology* 1994;50:27-37.
9. Ray WA, Murray KT, Kawai V, Graham DJ, Cooper WO, Hall K, Stein CM. Propoxyphene and the risk of out-of-hospital death. *Pharmacoepidemiol Drug Saf* 2013;22:403-412.
10. Hall GC, Sauer B, Bourke A, Brown JS, Reynolds MW, LoCasale R. Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2012;21:1-10.

11. Basch E. The missing voice of patients in drug-safety reporting. *N Engl J Med* 2010;362:865-869.
12. Choi NK, Lee J, Park BJ. Recent international initiatives of drug safety management. *J Korean Med Assoc* 2012;55:819-826.
13. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. *J Am Med Inform Assoc* 2013;20:404-408.
14. Xu R, Wang Q. Large-scale combining signals from both biomedical literature and the FDA Adverse Event Reporting System (FAERS) to improve post-marketing drug safety signal detection. *BMC Bioinformatics* 2014;15:17.
15. Choi NK, Chang Y, Kim JY, Choi YK, Park BJ. Comparison and validation of data-mining indices for signal detection: using the Korean national health insurance claims database. *Pharmacoepidemiol Drug Saf* 2011;20:1278-1286.
16. Park MY, Yoon D, Lee K, Kang SY, Park I, Lee SH, Kim W, Kam HJ, Lee YH, Kim JH, Park RW. A novel algorithm for detection of adverse drug reaction signals using a hospital electronic medical record database. *Pharmacoepidemiol Drug Saf* 2011;20:598-607.
17. Shin JY, Choi NK, Jung SY, Lee J, Kwon JS, Park BJ. Risk of ischemic stroke with the use of risperidone, quetiapine and olanzapine in elderly patients: a population-based, case-crossover study. *J Psychopharmacol* 2013;27:638-644.
18. Choi NK, Lee J, Chang Y, Jung SY, Kim YJ, Lee SM, Lee JH, Kim JY, Song HJ, Park BJ. Polyethylene glycol bowel preparation does not eliminate the risk of acute renal failure: a population-based case-crossover study. *Endoscopy* 2013;45:208-213.
19. Song HJ, Lee J, Kim YJ, Jung SY, Kim HJ, Choi NK, Park BJ.  $\beta$ 1 selectivity of  $\beta$ -blockers and reduced risk of fractures in elderly hypertension patients. *Bone* 2012;51:1008-1015.
20. Jung SY, Choi NK, Kim JY, Chang Y, Song HJ, Lee J, Park BJ. Short-acting nifedipine and risk of stroke in elderly hypertensive patients. *Neurology* 2011;77:1229-1234.
21. Choi NK, Hahn S, Park BJ. Increase in mortality rate following coprescription of cisapride and contraindicated drugs. *Ann Pharmacother* 2007;41:667-673.
22. Cho S, Sohn CH, Jo MW, Shin SY, Lee JH, Ryoo SM, Kim WY, Seo DW. Correlation between national influenza surveillance data and google trends in South Korea. *PLoS One* 2013; 8:e81422.

**Peer Reviewers' Commentary**

본 논문은 보건의로 분야 측면, 특히 약물의 안전성 분야에서 빅데이터 활용의 유용성 및 활용 동향과 관련된 내용을 기술하고 있다. 보건의로에서 빅데이터의 사용의 중요성이 강조되고 있는 시점에 외국의 사례와 국내에서의 활용방안을 기존에 보고된 연구와 자료를 근거로 체계적으로 기술하였다는데 의의가 있다고 생각된다. 본 연구는 향후 약물의 안전성 분야에서 빅데이터 활용의 필요성과 방향을 제시했다는 점에서 의의가 있는 논문이라 판단된다.

[정리: 편집위원회]