

이비인후과 질환에서 빅데이터의 활용

가톨릭대학교 의과대학 이비인후과학교실

주 영 훈

Application of Big Data for Otolaryngologic Diseases

Young-Hoon Joo, MD, PhD

Department of Otolaryngology-Head and Neck Surgery, College of Medicine, The Catholic University of Korea, Seoul, Korea

서 론

빅데이터는 여러 가지 형태로 정의되고 있으나 대표적으로 미국 IT리서치기업인 가트너에 따르면, "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⁾ 빅데이터는 단순히 방대한 자료가 아니라 큰 크기로 인하여 처리와 분석이 용이하지 않으나, 이를 이용하여 의사결정에 도움이 되는 유용한 정보를 생산해낼 수 있는 자료, 그리고 그 자료를 활용할 수 있는 일련의 시스템을 뜻한다.¹⁾ 최근 빅데이터를 분석하여 유용한 정보를 생성할 수 있는 기술이 발전하면서 마케팅, 금융, 교육, 치안, 환경, 교통, 농업, 보건 의료 등 다양한 분야에 활용되고 있으며, 가트너는 빅데이터를 '21세기의 새로운 원유'라고 지칭하기도 하였다.

최근 전 세계적으로 정부차원에서 빅데이터에 대한 투자가 활발하며, 우리 정부도 2013년부터 창조경제, 정부 3.0을 국정과제로 내세우고 공공분야 빅데이터 활용을 위하여 노력하고 있다. 미국은 2012년 'Big Data R&D Initiative'에서 적극적인 투자지지를 밝혔고, 일본은

2012년 '빅데이터 활용 기본전략'을 발표하고 정부주도의 발전전략을 실행하고 있다. 우리나라는 미래창조과학부 중심의 '빅데이터 산업 발전 전략'을 발표하고 주요 공공분야의 자료를 공개하고 인력과 인프라 양성에 박차를 가하여, 현재 다수의 공공분야 빅데이터 프로젝트가 진행 중에 있다.

보건의료 분야에서 빅데이터에 대한 공공분야 투자는 비감염성질환이 증가하면서, 이로 인한 의료비지출 증가를 억제하기 위한 최선의 과학적 근거를 가진 의료를 찾는 노력에서 촉발되었다.²⁾ 미국의 경우 오바마 행정부가 빅데이터 활용의 하드웨어를, 비영리학술단체가 소프트웨어를, 그리고 비영리·비 정부단체가 이들의 활동을 조율하고 있다. 특히, 기존 구축 자료 간의 네트워크를 개발하고 행정기관과 의료기관의 개별환자를 연결시킨 데이터베이스를 구축하는 것이 보건의료 분야 문제해결을 위한 주된 인프라로서 강조되고 있다. 우리나라에서도 정보화진흥원 산하에 '빅데이터 분석 활용 센터'를 설립, 운영하는 등 미래창조과학부와 정보화진흥원을 주축으로 빅데이터를 효과적으로 활용하기 위한 방안을 마련하기 위하여 국가차원에서 적극적으로 노력하기 시작하였다. 대용량 자료로부터 일정 패턴을 파악하여 새로운 통찰을 제공하고 예측에 필요한 근거를 제공함으로써 의사결정에 도움을 줄 수 있다는 유용성이 밝혀지면서, 보건의료 분야에서도 빅데이터 기술을 적용하여 질병발생 양상을 파악하여 확산을 방지하고, 개인 특성에 맞는 맞춤형 치료, 예측, 위험요인 규명 등에 필요한 근거기반을

교신저자 : 주영훈, 14647 경기도 부천시 소사로 327
가톨릭대학교 의과대학 이비인후과학교실
전화 : (032) 340-7090 · 전송 : (032) 340-2674
E-mail : joodoct@catholic.ac.kr

제공할 수 있으리라는 기대로 이 분야에 대한 관심이 크게 증가하고 있다.³⁾

빅데이터의 활용이 보건의료 분야에 가질 수 있는 가치는 매우 커서, 본 종설을 통하여 빅데이터의 국내외 현황을 살펴보고, 이비인후과 질환에서의 응용사례를 살펴보고, 보건의료 분야에서 적극적 활용 방안을 살펴보기로 한다.

본 론

의학 연구 분야에 불어온 빅데이터의 영향

현대의학에서 빅데이터는 정밀의료, 개인 맞춤형 의료로 가기 위한 전제조건이다. 전국적인 EMR의 보급, 기술 발달(의료영상의 해상도 확대 등) 및 개인 건강관리 기기의 보급 확대는 의료 데이터양의 폭발적 증가를 낳고 있다. 데이터의 종류는 매우 다양하여 로봇 수술기, 내시경 등에서 촬영되는 영상 데이터, 음성 데이터와 웨어러블 디바이스로 인한 다양한 센서의 측정데이터, 마지막으로 의사의 진료기록인 문자 데이터 등이 있다. 의료 빅데이터의 활용은 향후 의료기술 발전에 큰 영향을 끼칠 것으로 전망되는 상황으로 한국이 직면한 노령 인구 증가에 따른 의료비 지출 증가 압박에 대응할 수 있는 방법으로서 가치가 있다고 할 수 있을 것이다.

미래형 첨단 의료기술에 빅데이터를 이용한 대표적 사례가 바로 IBM이 개발한 슈퍼컴퓨터 Watson이다. Watson은 의학을 비롯한 다양한 분야에 진출하여, 빅데이터를 활용한 최적의 암 치료법 도출에 활용되고 있다. 이를 위해 지난 수년 간 Memorial Sloan-Kettering Cancer Center와 협력하여 폐암, 유방암 등의 분야에, MD Anderson Cancer Center와의 협력을 통해 백혈병 분야에 집중하고 있다. 최근 개최된 미국임상암학회(ASCO)에서 MD Anderson이 발표한 종양학 전문가 어드바이저(Oncology Expert Adviser)로서의 Watson의 치료 정확도에 대한 연구결과에서 overall accuracy가 82.6%에 달하여 Watson이 상당한 수준의 정확도를 확보했음을 짐작할 수 있다.

빅데이터는 개인 맞춤형 의료에도 주요한 영향을 미친다. 최근까지 질병치료는 하나의 방식이 두루 적용되도록 하는 접근 방식이었다. 모든 사람들이 질병이 서로 상이하기 때문에 분자 차원에서 어떤 현상이 발생하는

지를 이해하는 것은 환자에게 가장 적합한 치료 방식을 결정할 수 있는 요인으로 작용한다. 인간의 게놈 안에 대략 60억개의 염기쌍이 존재하고, 20,300개의 단백질 코딩 유전자, 수천개의 RNA 분자, 최소 2,900개의 대사 물질이 있다. 이전의 방식과 같이 이 플랫폼 중 하나의 데이터를 분석하는 것에서 벗어나, 현재 연구자들은 슈퍼컴퓨터에서나 다룰 수 있었던 엄청난 데이터량을 생산하는 다양한 플랫폼에 주목하고 있다. 이 데이터는 질병의 진단에 있어 유용하게 사용될 수 있고, 암 발생을 예측하고 예방할 수 있으며, 환자의 프로테오미나 대사체를 프로파일링 함으로써 임상 의사들이 환자의 증세 변화를 쉽게 예측할 수 있다. 이처럼 빅데이터는 생명 과학 분야를 완전히 변화시키고 있으며 이는 인간 생물학에 있어 깊고 넓은 통찰을 제공할 뿐 아니라 대부분 연구의 응용에 있어 과학자들이 질병 과정을 이해하고 인류 질병의 새로운 치료법을 개발하는데 큰 도움을 주고 있다.

이비인후과 질환에서 빅데이터의 활용

이비인후과 질환에서의 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 조기에 예측에도 활용될 것으로 기대된다.^{3,4)}

최근 들어 국내 이비인후과 의사들이 관심을 갖고 연구를 진행하는 빅데이터는 국민건강영양조사 자료 및 국민건강보험공단 자료이다. 국민건강영양조사는 1995년 제정된 국민건강증진법에 근거하여 시행하는 전국 규모의 건강 및 영양조사이다. 1998년부터 2005년까지 3년 주기로 시행하였으며, 국가통계의 시의성 향상을 위해 2007년부터 매년 시행하고 있다. 국민건강영양조사의 목적은 국민의 건강수준, 건강행태, 식품 및 영양섭취 실태에 대한 국가 및 시도 단위의 대표성과 신뢰성을 갖춘 통계를 산출하는 것이며, 이를 통해 국민건강증진종합계획의 목표 설정 및 평가, 건강증진 프로그램 개발 등 보건정책 기초자료로 활용되고 있다. 국민건강영양조사 원시자료는 전수조사가 아닌 표본조사 자료이므로, 목표모

집단인 대한민국에 거주하는 국민에 대한 결과를 추정하기 위해서 복합표본설계를 반영하여 분석해야 한다. 대한이비인후과 학회에서는 2008년 하반기부터 국민건강영양조사사업에 참여하여 5년간 검진 사업을 지속하였다. 전국에 걸쳐 매년 수십명의 전공의와 교수들이 참여하여 검진이 실행되었고, 이를 바탕으로 “Prevalence of otolaryngologic diseases in South Korea : data from the Korea national health and nutrition examination survey 2008” 등 이비인후과 질환에 대해 50편 이상의 SCI(E) 논문들이 발표되었으며 이는 대국민 홍보 및 의료 정책 결정에 이용되고 있다.⁶⁾

최근들어 국민건강보험공단 데이터를 이용한 연구가 다양한 분야 및 전공에서 새로이 진행되고 있으며, 새로운 개념의 빅데이터가 아닐지라도 그 자체로도 많은 유용한 정보를 포함하고 그 활용가치가 다양하다. 건강보험 빅데이터인 ‘건강보험 표본 코호트 DB’와 ‘환자데이터셋’은 전 국민의 진료내역을 담고 있어 세계에서도 찾아보기 힘든 상당한 가치를 지닌 데이터들이다. 공단은 4대 사회보험의 통합징수, 의료급여 관리, 건강검진, 요양급여비용 지급 및 사후관리, 장기요양관리 등의 다양한 업무수행과정에서 생성된 1조 4,779억 건의 데이터를 보유하고 있다. 이 데이터들을 요양기관 정보, 병원 평가 정보, 의약품 생산·공급 데이터, 의약품 처방 및 조제 데이터, 의료기기정보 데이터, 질병, 인구, 소득 등의 통계 데이터, SNS 질병정보 등이 있다. 전 국민의 건강보험 취득, 상실, 성별, 연령, 사업장, 의료급여종별, 장애유형, 보험료 분위 등의 자료로 구성되어 있다. 진료내역으로는 진료명세서 일반내역, 진료명세서 상세내역, 진료상병내역, 진료처방내역이 있고 진료처방상세내역을 의료·보건기관, 치과·한방, 약국의 세 가지 유형으로 분리하여 환자기준으로 재구성하여 구축되어 있다. 요양기관내역 자료는 진료를 실시한 의료, 보건기관, 치과, 한방기관, 약국 등의 의사, 인원, 진료과목, 장비, 시설, 산재지정 여부 등의 정보를 담고 있다. 건강검진 자료에는 일반건강검진, 생애전환기 건강진단, 암검진(위, 대장, 간, 유방, 자궁경부), 구강검진, 영유아검진, 영유아 구강검진의 자료가 있고 검진일자, 검사결과 및 문진 내역을 연도별로 구축하고 있다. 중증질환자인 암, 희귀난치성 질환자, 화상환자의 산정특례환진일자, 산정특례상병코

드, 확진의사 등의 정보도 2005년 9월 제도 시행 후 중증 질환자(암)에 대해 등록되어 있고, 2009년 7월부터는 희귀난치성질환자가 등록되어 있다. 2016년 4월 대한이비인후과학회와 국민건강보험공단은 건강보험 빅데이터를 활용한 이비인후과 질환 현황 파악 및 관리방안 마련을 위하여 업무협력 협약(MOU)을 체결하였으며, 이후 7개 이비인후과 분과학회 및 유관학회의 연구 및 정책 활동을 지원하는 목적으로 공동 연구를 진행하고 있다.

결론

최근 의료 환경은 정밀의료, 개인 맞춤형 의료라는 커다란 명제와 더불어 국민의료비의 증가, 병원 수익구조 악화, 신약개발의 어려움 등을 극복하기 위한 산업계의 전반적인 효율성 제고 노력이 이루어져야 하는 상황이며, 이와 같은 문제점을 극복하고 경쟁력을 확보하기 위해 정부 및 개별 연구자들은 빅데이터의 활용에 많은 관심을 보이고 있다. 우리나라도 연구자들이 원하는 다양한 데이터를 적시에 제공받을 수 있는 플랫폼과 관리체계가 마련되어 있어 누구든 아이디어가 있으면 의료 빅데이터를 활용해 연구를 할 수 있는 상황이다. 이비인후과 내에서도 적극적으로 이 데이터들을 분석하고 활용하여 학문적 발전과 의료 상황 개선에 광범위하게 활용할 필요가 있다.

중심 단어 : 빅데이터 · 이비인후과 · 보건산업.

REFERENCES

- 1) Beyer MA, Douglas L. *The importance of 'big data': a definition.* Stamford: Gartner;2012.
- 2) Bae JM, Park BJ, Ahn YO. *Perspectives of clinical epidemiology in Korea.* J Korean Med Assoc 2013;56(8):718-23.
- 3) Issa NT, Byers SW, Dakshanamurthy S. *Big data: the next frontier for innovation in therapeutics and healthcare.* Expert Rev Clin Pharmacol 2014;7(3):293-8.
- 4) Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. *Detecting influenza epidemics using search engine query data.* Nature 2009;457(7232):1012-4.
- 5) 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.
- 6) Cho YS, Choi SH, Park KH, Park HJ, Kim JW, Moon JJ, et al. *Prevalence of otolaryngologic diseases in South Korea: data from the Korea national health and nutrition examination survey 2008.* Clin Exp Otorhinolaryngol 2010; 3(4):183-93.