

건강보험공단 자료를 이용한 빅데이터 자료분석 방법: 부비동염수술과 천식사이의 관련성에 대한 고찰과 분석을 위한 방법론

서울대학교 공과대학 전기정보공학부,¹ 서울대학교 의과대학 분당서울대학교병원 이비인후과²

유승학¹ · 위재운² · 김정훈² · 윤성로¹

Methodology for Big Data Analysis Using Data from National Health Insurance Service: Preliminary Methodologic Study and Review about the Relationship between Sinus Surgery and Asthma

Seunghak Yu, MS¹, Jaewoon Wee, MD², Jeong-Whun Kim, MD² and Sungroh Yoon, PhD¹

¹Department of Electrical and Computer Engineering, Seoul National University, Seoul; and

²Department of Otorhinolaryngology, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, Korea

ABSTRACT

Background and Objectives : Sinus surgery has been reported to improve pulmonary function and decrease the use of asthma medications in patients with chronic rhinosinusitis and asthma. Most studies, however, have used a small number of patients and were conducted over a short period. To demonstrate a causal relationship between sinus surgery and asthma, a sufficient amount of patient data observed over a long period is required. To address the limitations of the existing approaches, we conducted a preliminary methodological study for large-scale data analysis using data from the National Health Insurance Service (NHIS) to suggest a basis for the effect of sinus surgery on asthma. **Materials and Methods** : The data from NHIS consisted of unidentified medical histories of a sample cohort representing the whole nation over a period of nine years. We selected the following types of study samples: 1) patients with surgical codes for sinus surgery; 2) patients with disease codes for sinusitis; 3) patients with disease codes for asthma; and 4) patients with medication codes for asthma treatment. **Results** : In this study, we applied a methodology for selection of subjects from big data to investigate the effect of sinus surgery on improving asthma in the future. We could include 152 subjects after the four-stage selection method from 1,025,340 patients. **Conclusion** : We could establish a method to select patients with history of sinus surgery and asthma treatment from a big data. This methodology using big data may contribute to identify relationship between sinus surgery and asthma treatment in the future.

KEY WORDS : Big data · Health insurance service · Relation study · Sinus surgery · asthma.

서 론

헬스케어 산업은 환자 관리, 자료 보관, 규제 요구사항 등

으로 인해 전통적으로 대용량의 데이터를 생산하는 분야 중 하나이다.¹⁾ 대부분의 헬스케어 데이터는 하드카피 형태로 저장되어 왔지만 최근 디지털화가 가속화됨으로써 헬스케

논문접수일: 2014년 11월 18일 / 수정완료일: 2015년 1월 25일 / 심사완료일: 2015년 2월 5일

교신저자: 윤성로, 151-744 서울 관악구 관악로 1 서울대학교 공과대학 전기정보공학부

Tel: +82-2-880-1401, E-mail: sryoon@snu.ac.kr

교신저자: 김정훈, 463-707 경기도 성남시 분당구 구미로173번길 82 서울대학교 의과대학 서울분당병원 이비인후-두경부외과학교실

Tel: +82-31-787-7405, Fax: +82-31-787-4057, E-mail: kimemall@naver.com

본 연구는 SNUBH grant # 12-2013-004에 의해서 지원되었음.

어 빅데이터 시대가 도래하였다. Institute for Health Technology Transformation(iHT²)의 보고에 따르면 2011년 기준으로 미국에서만 150 엑사바이트(10^{18} 기가바이트)의 헬스케어 데이터가 생성되었으며 이러한 증가속도를 유지한다면 곧 제타바이트(10^{21} 기가바이트)에 도달할 것으로 예상된다.²⁾

헬스케어 빅데이터 시대 이전의 기존 연구들은 소규모의 집단을 이용한 단기간의 연구가 대부분이며, 큰 집단을 대상으로 수술 전후 장기간의 자료를 이용하여 시행한 대규모의 연구는 없는 실정이다. 연구자가 계획하고 있는 대규모 자료를 이용 시 질환 간의 명확한 인과관계를 제시할 수 있을 정도의 충분한 수의 데이터를 통해 확고한 근거를 제시할 수 있고 또한 데이터의 특성 상 장기간 연속적인 질환의 경과 관찰이 가능한 유용한 연구가 될 것으로 사료된다.

2012년 Global Initiative for Asthma(GINA)³⁾ 가이드라인에 따르면 천식의 유병률은 1~18%로 추산되고 있으며, 많은 나라에서 유병률의 증가가 관찰되고 있다. 우리나라 성인의 경우 1998년 8,823명을 대상으로 천식에 대한 설문조사에서 천명 등 천식증상의 빈도는 12.8%였고, 2001년 2,467명의 성인 대상으로 설문지 및 메타콜린 기관지유발시험을 이용한 조사에서 천명과 함께 기도과민증이 있는 현증 천식이 40세 이하에서는 2%, 40~54세에서는 3.8%, 55~64세에서는 7.7%, 65세이상의 노인에서는 12.7%로 보고되었다.

다양한 연구들이 이와 같이 높은 유병율을 보이며 많은 사회적 비용을 초래하는 천식이 만성부비동염과 상관관계가 있을 것으로 가설을 설정하고 진행되었다. 실제로 천식을 동반한 만성 비부비동염 환자에서 부비동 내시경 수술은 폐기능을 의미있게 향상시키고 천식약물의 사용을 감소시키는 것으로 보고되어왔다.⁴⁻⁸⁾ 하지만 기존의 연구는 대부분 서양에서 진행된 소규모 집단을 이용한 단기간 연구이기 때문에 한국인의 유전적 특성에 의거한 질환의 발생 및 연관관계를 파악할 수 있는 한국적인 부비동염과 천식치료의 가이드라인을 제시할 수 있는 새로운 연구가 필요하다.

따라서 본 논문에서는 만성부비동염의 적극적인 치료 시 천식의 호전 정도를 정량적으로 평가하여 임상적으로 연관관계가 높은 만성부비동염과 천식의 상관관계를 규명하고 치료자의 입장에서 적극적인 질환의 치료를 가능케 하기 위해 빅데이터 분석의 방법론적인 준비과정을 수행하고자 한다. 또한 향후 국민건강보험공단의 자료의 전 국민 대표성에 대한 통계학적 근거가 최종적으로 공식화된 이후 본 연구의 방법론이 사용될 수 있도록 사전 예비연구를 진행함을 목표로 한다.

대상 및 방법

실험 데이터

본 연구에서는 국민건강보험공단에서 제공한 1,025,340명의 9년간의 진료내역 자료와 검진결과 자료를 이용하였다. 이 데이터는 국민건강보험공단 측에서 2002년말 기준의 전 국민 대상 데이터에서 상실자를 제외한 뒤 1,025,340명의 데이터를 표본 추출한 것이다. 표본 추출 방식은 연령대, 성별, 자격구분, 소득분위계 층화(1,476개)된 비례배분 층화 계통 무작위추출 방식이다. 연령은 18개층(0, 10~80+, 5세 연령군), 자격구분은 지역가입자, 직장가입자, 의료급여자로 나뉘며 소득분위는 지역 및 직장가입자 각 20분위, 의료급여자(0분위)로 나뉜다.

대상자 선정

대상자는 20세 이상 성인이자 만성부비동염 및 비용으로 부비동 내시경 수술을 받았으며 천식으로 인해 약물을 처방 받은 환자로 정의한다. 대상자 선별을 위해 다음과 같은 4단계의 선별과정을 거쳤다. 1단계로는 20세 이상 성인 중 부비동내시경 수술을 받은 환자를 선별하고자 하였다. 2단계에서는 부비동내시경수술을 받은 환자 중 부비동염 상병코드를 가진 환자를 선별하고자 하였다. 3단계에서는 부비동염으로 부비동내시경수술을 받은 환자 중 천식환자를 선별하고자 하였다. 마지막 4단계에서는 부비동내시경수술을 받은 천식환자 중 천식으로 약물 처방을 받은 병력이 있는 환자를 선별하고자 하였다.

분석도구

헬스케어 빅데이터 분석은 전통적인 분석방식과 비교하여 사용되는 모델과 알고리즘은 동일하지만 분석환경에서 차이를 가진다.¹⁾ 일반적으로 전통적인 방식은 데스크탑 등 단일 시스템 상에서 사용자 친화적 인터페이스를 가진 상용 분석 프로그램으로 이루어진 반면에 빅데이터 헬스케어 분석은 복수개의 노드 혹은 클라우드 환경에서 오픈소스 플랫폼을 이용하여 분산되어 처리되며 프로그램이 매우 복잡하고 관련 전문지식이 필수적이다.

High-Availability Distributed Object-Oriented Platform(Hadoop)이 대표적인 오픈소스 플랫폼으로 크게 Hadoop Distributed File System(HDFS)과 MapReduce로 이루어져 있다. HDFS는 Hadoop이 설치된 환경에서 데이터를 잘게 나누어 노드들에 분산하여 저장, 관리 하는 기능을 하고 MapReduce는 실제 분석을 분산처리 할 수 있는 인터페이스를 제공한다. 이 밖에도Hadoop과 연계하여 사용할 수 있

는 많은 오픈소스 도구들이 개발되고 있어 다채로운 분석을 수행할 수 있지만 관련 전문 프로그래밍 지식이 반드시 필요하며 상대적으로 많은 시간이 소요된다.

본 연구에서는 건강보험공단 데이터의 보안유지 관련 조약에 의해 제한된 실험환경에서 실험을 진행하였으며 서울대학교 빅데이터 센터(서울특별시 관악구 관악로 1 서울대학교)에 위치한 서버 상에서 Statistical Analysis Software(SAS) 프로그램을 활용하였다. 향후 보안관련 이슈가 해결된다면 샘플링 데이터가 아닌 전수 데이터를 이용하여 대량의 자료 처리를 위한 인프라 구축 및 실험을 다음과 같이 계획한다.

먼저 SAP 社의 인메모리 데이터베이스인 HANA를 이용하여 분석방법 및 알고리즘을 설정한다. 인메모리 데이터베이스는 디스크에 설치되는 기존 방식과 다르게 메인 메모리에 설치되어 운영되는 데이터베이스 관리 시스템으로 처리 속도가 빠르다. 또한 다양한 기본 기계학습 알고리즘들이 구현되어 있어 쉽게 적용가능 하지만 사용자가 알고리즘의 최적화를 진행하기는 어려우며 상용 도구로서 높은 가격대를 형성하고 있어 일반 사용자들의 접근성이 떨어진다. 따라서 연구실에서 확보하고 있는 HANA를 이용하여 기본 기계학습 알고리즘들을 적용하는 초기 분석을 통해 분석방법 및 알고리즘 적용 방향성을 확보하고 오픈소스 플랫폼인 Hadoop을 통해 분석의 최적화를 진행하는 것을 목표로 한다.

Apache Mahout은 Hadoop 플랫폼 상에서 사용할 수 있는 기계학습 라이브러리로 일반적인 기계학습 알고리즘 들을 Hadoop의 MapReduce를 적용하여 분산 처리 가능하며 확장성 있는 알고리즘들로 수정하여 제공한다. 따라서 HANA를

통해 선별된 기계학습 알고리즘들을 Apache Mahout의 라이브러리를 건강보험공단 데이터에 맞도록 수정하여 최적화하는 방식을 통해 Hadoop프로그래밍에 소요되는 시간을 효과적으로 절약할 수 있을 것으로 기대한다.

결 과

대상자의 1단계 선별: 부비동내시경 수술

2005~2007년(3개년) 진료내역 자료에서 분류코드(DIV_CD)가 부비동내시경수술 관련 코드인 20세 이상 성인 환자를 추출하였다. 사용한 부비동내시경수술 코드는 Table 1을 기본으로 하여 내시경하에서 실시한 경우, 야간, 공휴일, 제2의 수술, 재수술(주된수술), 재수술(제2의수술)의 경우를 모두 포함한 348가지이다. 1단계에서 선별된 데이터는 1,025,340의 환자 데이터 중 1,990명의 환자데이터였다.

대상자의 2단계 선별: 부비동염

부비동내시경수술을 받은 환자 중 부비동염 상병코드를 가진 환자를 선별하여 부비동염으로 부비동내시경수술을 받은 환자를 선별하였다. 1단계에서 선별된 환자를 대상으로 명세서 일반자료에서 주상병(MAIN_SICK)과 부상병(SUB_SICK)이 Table 2의 급성관련 코드 8개(병명 33개), 만성관련 코드 8개(병명 20개), 비용종 관련 코드 5개(병명 12개)인 부비동염 관련 상병코드 환자데이터를 선별하여 부비동내시경수술 코드가 있는 1,990명의 환자데이터중 1,810명의 환자데이터가 선별되었다.

Table 1. Data extraction based upon the presence of cost code of sinus surgery

Code	Serial number	Operation (Korean)	Operation (English)
O0951	자95가	비용적출술(단발성)	Single nasal polypectomy
O0952	자95나	비용적출술(범발성)	Diffuse nasal polypectomy
O1050	자105	상악동비내수술	Intranasal antrostomy
O1070	자107	상악동근치수술	Caldwell-Luc operation
O1091	자109가	전두동비외수술(단순 천공술)	Operation of frontal sinus (Trephine operation)
O1092	자109나	전두동비외수술(이마절개하 골성형)	Operation of frontal sinus (Osteoplasty following brow incision)
O1093	자109다	전두동비외수술(관상봉합절개하 골성형)	Operation of frontal sinus (Osteoplasty following coronal incision)
O1100	자110	사골동비내수술	Intranasal ethmoidectomy
O1110	자111	사골동비외수술	Extranasal ethmoidectomy
O1120	자112	접형골동비내수술	Intranasal sphenoidectomy
O1130	자113	전부비강근본수술	Radical operation of pansinusitis
O1140	자114	상악동사골동근본수술	Radical operation of maxillary and ethmoid sinus
O1150	자115	상악동사골동접형골동근본수술	Radical operation of maxillary, ethmoid and sphenoid sinus
O1160	자116	전두동사골동근본수술	Radical operation of frontal and ethmoid sinus
O1170	자117	전두동사골동접형골동근본수술	Radical operation of frontal, ethmoid and sphenoid Sinus
O1175	자117-1	전두동, 사골동, 상악동근본수술	Radical operation of frontal, ethmoid and maxillary Sinus
O1180	자118	사골동접형골동수술	Ethmoidectomy and sphenoidectomy

대상자의 3단계 선별: 천식

부비동염으로 부비동내시경수술을 받은 1,810명의 환자 데이터 중에서 수술일 기준으로 전후 3년간의 명세서 일반

Table 2. Data extraction based upon the presence of disease code of sinusitis

Code	Diagnosis (Korean)	Diagnosis (English)
Acute		
J01	급성 부비동염	Acute sinusitis
J010	급성 상악동염	Acute maxillary sinusitis
J011	급성 전두골동염	Acute frontal sinusitis
J012	급성 사골동염	Acute ethmoidal sinusitis
J013	급성 접형골동염	Acute sphenoidal sinusitis
J014	급성 범부비동염	Acute pansinusitis
J018	기타 급성 부비동염	Other acute sinusitis
J019	상세불명의 급성 부비동염	Acute sinusitis unspecified
Chronic		
J32	만성 부비동염	Chronic sinusitis
J320	만성 상악동염	Chronic maxillary sinusitis
J321	만성 전두골동염	Chronic frontal sinusitis
J322	만성 사골동염	Chronic ethmoidal sinusitis
J323	만성 접형골동염	Chronic sphenoidal sinusitis
J324	만성 범부비동염	Chronic pansinusitis
J328	기타 만성 부비동염	Other chronic sinusitis
J329	상세불명의 만성 부비동염	Chronic sinusitis, unspecified
Nasal polyps		
J33	코폴립	Nasal polyp
J330	비강의 폴립	Polyp of nasal cavity
J331	폴립모양부비동변성	Polypoid sinus degeneration
J338	부비동의 기타 폴립	Other polyp of sinus
J339	상세불명의 코폴립	Nasal polyp, unspecified

자료에서 주상병(MAIN_SICK)과 부상병(SUB_SICK)이 Table 3의 17개의 천식관련 상병 코드(병명 29개)가 있는 환자데이터를 선별하여 부비동내시경수술코드와 부비동염 질병코드가 있는 1,810명의 데이터 중 368명의 데이터가 선별되었다.

대상자의 4단계 선별: 천식 약물 처방

부비동염으로 부비동내시경수술을 받은 환자 중 수술일 기준 전후 3년간 천식을 앓고 천식 약물을 처방 받은 환자데이터를 최종 대상자 데이터로 선정하였다. 천식 약물은 흡입 스테로이드제와 흡입 베타2항진제로 정의하고 이를 처방받은 환자를 선별하여 152명의 최종 대상자데이터를 선정할 수 있었다. 이때 수술로 인한 일시적인 약물 사용량의 변화를 배제하기 위해 수술 전후 3년간의 관찰기간 중 수술 직전3개월과 직후 3개월은 대상에서 제외하였다.

고 찰

기도는 비전정에서부터 폐포까지 연속된 구조물이다. 이들 점막은 지속적으로 외부에 노출되어 있으며, 자극물질, 알레르기 항원, 미생물 등과 같은 다양한 외부위험에 대해 첫번째 방어를 담당한다. 이러한 방어작용은 선천면역과 적응면역을 통하여 효과적으로 일어나며, 이들의 이상은 알레르기성 비염, 만성 비부비동염, 천식 등의 질환을 일으킬 수 있다. 최근에는 기도를 상기도와 하기도를 하나의 구조물로 생각하고 상하기도 질환과의 연관성에 대한 연구가 활발히

Table 3. Data extraction based upon the presence of disease code of asthma

Code	Diagnosis (Korean)	Diagnosis (English)
F54	천식에 연관된 심리적 요소 및 행동적 요소	Examples of the use of this category are asthma (F54 and J45)
I501	심장성 천식	Cardiac asthma
J44	기타 만성 폐색성 폐질환	Other chronic obstructive pulmonary diseases
J45	천식	Asthma
J450	주로 알레르기성 천식	Predominantly allergic asthma
J451	비알레르기천식	Nonallergic asthma
J458	혼합형 천식	Mixed asthma
J459	상세불명의 천식	Asthma, unspecified
J46	천식지속 상태	Status asthmaticus
J460	주로 알레르기성 천식을 동반한	With predominantly allergic asthma
J461	비알레르기천식을 동반한	With nonallergic asthma
J468	혼합형 천식을 동반한	With mixed asthma
J469	상세불명의 천식을 동반한	With unspecified asthma
J82	달리 분류되지 않은 폐호산구증가증	Eosinophilic asthma
T486	달리 분류되지 않은 항천식제에 의한 중독	Poisoning by antiasthmatics, NEC
Y556	치료시 유해작용을 나타내는 달리 분류되지 않은 항천식제	Antiaesthmatics, NEC causing adverse effects in therapeutic use
Z825	천식 및 기타 만성적 하부호흡기질환의 가족력	Family history of asthma and other chronic lower respiratory diseases

진행되고 있으며, 특히 상기도의 구조물인 코와 하기도의 구조물인 기관지와와의 연관성에 대한 연구가 활발하게 진행되고 있다.

천식으로 인한 사회적 비용은 상당한 것으로 알려져 있다. 따라서, 국민경제적, 보건학적으로도 이에 대한 정확한 통계학적 자료 및 학술적 근거를 마련하는 것은 중요하다. 건강보험공단 공식 통계를 기초로 하여 2004년부터 2005년까지 조사한 바에 따르면, 천식으로 인한 직접비용, 간접비용, 무형비용을 산출하였을 때 천식으로 인한 총 사회적 비용이 4조 1,148억원에 이른다. GINA에서는 천식을 제대로 치료하지 않았을 경우 발생하는 비용이 크고, 천식과 관련된 비의료적인 경제적 비용 역시 상당함을 강조하였다.

만성 비부비동염과 천식 사이의 밀접한 연관성은 지난 수십년 간 보고되어 왔다. 천식 환자에서 7~15%에서 비용종이 동반되고 있는 것으로 알려져 있다.⁹⁾ 최근 Global Allergy and Asthma European Network(GA2LEN)는 유럽 12개국, 50,000명 이상을 대상으로 시행한 대규모 설문조사에 따르면 천식의 유병률은 8.6%로 관찰되었고, 천식과 만성 비부비동염 간의 위험비(odds ratio)가 3.48로 강한 연관성이 있는 것으로 보고되었으며, 이는 만성 비부비동염과 알레르기 비염 사이의 연관성보다도 더 높은 것이었다. 최근 18,000명을 대상으로 시행된 또 다른 설문조사는 천식 환자에서 만성 비부비동염의 유병률이 8.4%로 조사되었고, 만성 비부비동염 환자의 24.4%에서 천식이 동반되어 있음을 보고하였다.¹⁰⁾ 천식의 중증도가 심해질수록 비부비동의 방사선학적 이상이 증가하며, 비용종의 유병률도 증가하는 것으로 알려져 있다.¹¹⁾¹²⁾ 비용종을 동반한 만성비부비동염 환자의 31%에서 천명음이, 42%에서 호흡곤란이 동반된다는 연구도 있다.¹³⁾ 여러 연구를 종합해보면 대략적으로 만성 비부비동염 환자의 26~60%에서 천식이 동반되는 것으로 알려져 있다.¹⁴⁻¹⁶⁾

만성 비부비동염의 수술적 치료와 천식의 인과관계에 있어서, 만성 비부비동염의 수술적 치료가 천식의 경과에 영향을 미친다는 여러 연구들이 발표되어왔다. 최근 발표된 European Position Paper on Rhinosinusitis and nasal Polyps (2012)에서 부비동염의 수술적 치료가 천식 증상완화에 긍정적인 영향을 줄 것이라고 보고되었다.¹⁷⁾ Ehnage 등은 51명의 천식을 동반한 만성 비부비동염 환자를 대상으로 시행한 double-blind randomized placebo-controlled study를 통하여 만성 비부비동염의 수술적 치료가 천식의 증상을 완화시키고, 천식관련 지표들이 개선됨을 보고하였다.¹⁸⁾ Proimos 등도 부비동 내시경 수술을 받은 86명의 천식환자에서 폐기능 검사상의 의미있는 개선과 천식 약물의 사용감소를 보고하였다.¹⁹⁾ Ragab 등은 90명의 천식을 동반한 만성 비

부비동염 환자를 대상으로 시행한 무작위대조군연구에서 내과적 그리고 외과적인 치료가 천식 약물의 사용을 감소시키고, 천식으로 인한 입원을 감소시키는 것으로 보고하였다.²⁰⁾ Senior 등은 천식을 동반한 부비동염 환자에서 부비동 내시경 수술 후 90%의 환자에서 천식 증상이 개선되는 것을 관찰하였고, 천식 발작이나 천식 약물의 사용이 감소됨을 보고하였다.²¹⁾ 소아에서도 천식이 동반된 만성 비부비동염 환자에서 부비동 내시경 수술은 천식의 임상양상을 완화시키고 결석을 줄이는 것으로 보고되었다.²²⁾

본 연구에서는 만성비부비동염과 천식의 실제 상관관계를 규명하고 만성비부비동염의 적극적인 치료 시 천식의 호전 정도를 대규모 데이터 분석을 통하여 정략적으로 평가하기 위한 사전작업으로 대상자 선정 방안을 제시하였다. 이를 위해 건강보험공단에서 제공한 1,025,340명의 9년간의 진료내역 자료와 검진결과 자료를 이용하여 4단계의 걸친 선별과정을 통해 최종 대상자 152명을 선정하였다.

그 동안 진단명의 정확성과 타당도를 조사한 연구에서 청구자료의 주상병과 의무기록의 일치도가 약 70%로 그 동안 보험공단의 자료는 진단명의 정확성 문제가 큰 단점 중의 하나로 지적되어왔다. 하지만 이번 연구에서는 진단명뿐만 아니라 약물의 처방과 수술내역을 진단명과 결합해 사용하여 진단명의 정확성 문제를 보완하고자 하였다. 또한 단기간의 약물사용이나 입원내역이 아닌 6년이라는 장기간의 데이터를 사용하여 장기간의 결과를 일반화하여 분석을 진행할 수 있었다. 따라서 본 연구에서는 진단명뿐만 아니라 4 단계에 걸쳐 부비동염 수술내역, 부비동염 관련 상병 진단여부, 천식관련 상병 진단여부, 천식관련 약물 처방여부를 순차적으로 살펴봄으로써 대상자를 선별하여 진단명의 정확성 문제를 보완하고자 하였다. 청구자료의 주상병만을 통해 대상자를 선정한다면 청구자료의 주상병과 의무기록이 일치하지 않는 경우가 많기 때문에 대상자가 누락되거나 잘못된 대상자가 선별될 가능성이 높다. 따라서 본 연구에서 제안하는 방식과 같이 진단명에만 의존하지 않고 실제 수술내역, 처방내역 등을 활용하여 대상자를 선별한다면 대상자 선별 정확도를 높일 수 있을 것으로 기대된다. 전 국민을 대표하는 건강보험공단 자료의 특성과 상대적으로 드물게 발생하기에 실제 임상자료로는 시험하기 힘든 주제에 대하여 대규모 자료를 이용하여 연구를 진행할 수 있었다. 또한 제한적인 실험실 환경이 아닌 실제 보건의료를 반영해 현황과 추세를 관찰할 수 있을 것으로 기대되며, 이미 자료가 구축되어 있기에 별도의 비용과 시간 소요를 줄일 수 있는 장점이 있을 것으로 예상된다.

본 연구에서는 서울대학교 빅데이터 센터 내의 서버 상에

서 SAS를 이용한 국가 단위의 건강보험공단 빅데이터를 초기 분석 함으로써 일반적인 임상 환경에서는 연구하기 힘들었던 조건에 관하여 체계적으로 연구할 수 있는 방법론적인 초석을 다질 수 있었다. 비록 아직까지 국민건강보험공단의 빅데이터가 국민전체를 대표하는 통계적으로 안정화된 자료가 아니며 보안 관련 이슈가 해결되지 않아 방법론적인 연구에 그친 한계가 있지만, 향후 이번 연구에서 구축된 대상자 선별 방법 및 알고리즘 등을 확장하여 후속연구를 진행할 예정이다. 상용 인메모리 데이터베이스 HANA에서 전수 데이터에 대한 초기분석을 진행하여 적용 가능 방법 및 알고리즘의 방향성을 잡고 오픈소스 플랫폼인 Hadoop과 오픈소스 기계학습 라이브러리 Apache Mahout을 연계하여 선별된 기계학습 알고리즘의 최적화를 진행함으로써 연구기간을 효과적으로 단축하여 부비동염수술과 천식간의 관계에 대한 통계적으로 의미 있는 임상 자료분석이 가능할 것으로 기대된다.

중심 단어 : 부비동염 · 부비동수술 · 천식 · 빅데이터.

REFERENCES

- 1) Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;2:3.
- 2) iHTT Report, Transforming Health Care Through Big Data Strategies for leveraging big data in the health care industry <http://ihealthtran.com/wordpress/2013/03/ihtr%C2%B2-releases-big-data-research-report-download-today/>
- 3) GINA Report, Global Strategy for Asthma Management and Prevention <http://www.ginasthma.org/documents/4>
- 4) Batra PS, Kern RC, Tripathi A, Conley DB, Ditto AM, Haines GK 3rd, et al. Outcome analysis of endoscopic sinus surgery in patients with nasal polyps and asthma. *The Laryngoscope* 2003;113:1703-6.
- 5) Dunlop G, Scadding GK, Lund VJ. The effect of endoscopic sinus surgery on asthma: management of patients with chronic rhinosinusitis, nasal polyposis, and asthma. *American Journal of Rhinology* 1999;13:261-5.
- 6) Garrel R, Gardiner Q, Khudjadze M, Demoly P, Vergnes C, Makeieff M, et al. Endoscopic surgical treatment of sinonasal polyposis-medium term outcomes (mean follow-up of 5 years). *Rhinology* 2003;41:91-6.
- 7) Park AH, Lau J, Stankiewicz J, Chow J. The role of functional endoscopic sinus surgery in asthmatic patients. *J Otolaryngol* 1998;27:275-80.
- 8) Uri N, Cohen-Kerem R, Barzilai G, Greenberg E, Doweck I, Weiler-Ravell D. Functional endoscopic sinus surgery in the treatment of massive polyposis in asthmatic patients. *J Laryngol Otol* 2002;116:185-9.
- 9) Larsen K. The clinical relationship of nasal polyps to asthma. *Allergy and asthma proceedings: the Official Journal of Regional and State Allergy Societies* 1996;17:243-9.
- 10) Eriksson J, Bjerg A, Lötval J, Wennergren G, Rönmark E, Torén K, et al. Rhinitis phenotypes correlate with different symptom presentation and risk factor patterns of asthma. *Respir Med* 2011;105:1611-21.
- 11) Lin DC, Chandra RK, Tan BK, Zirkle W, Conley D, Grammer LC, et al. Association between severity of asthma and degree of chronic rhinosinusitis. *Am J Rhinol Allergy* 2011;25:205-8.
- 12) ten Brinke A, Grootendorst DC, Schmidt JT, De Bruïne FT, van Buchem MA, Sterk PJ, et al. Chronic sinusitis in severe asthma is related to sputum eosinophilia. *J Allergy Clin Immunol* 2002;109:621-6.
- 13) Klossek JM, Neukirch F, Pribil C, Jankowski R, Serrano E, Chantal I, et al. Prevalence of nasal polyposis in France: a cross-sectional, case-control study. *Allergy* 2005;60:233-7.
- 14) Downing E, Braman S, Settipane G. Bronchial reactivity in patients with nasal polyps before and after polypectomy. *Journal of Allergy and Clinical Immunology* 1962;69:102.
- 15) Matsuno O, Ono E, Takenaka R, Okubo T, Takatani K, Ueno T, et al. Asthma and sinusitis: association and implication. *Int Arch Allergy Immunol* 2008;147:52-8.
- 16) Ragab A, Clement P, Vincken W. Objective assessment of lower airway involvement in chronic rhinosinusitis. *American Journal of Rhinology* 2004;18:15-21.
- 17) Fokkens W, Lund V, Mullol J, European Position Paper on Rhinosinusitis and Nasal Polyps group. European Position Paper on Rhinosinusitis and Nasal Polyps 2012. *Rhinology. Supplement*
- 18) Ehnhage A, Olsson P, Kolbeck KG, Skedinger M, Stjärne P. One year after endoscopic sinus surgery in polyposis: asthma, olfaction, and quality-of-life outcomes. *Otolaryngology--head and neck surgery: Otolaryngol Head Neck Surg* 2012;146:834-41.
- 19) Proimos E, Papadakis CE, Chimona TS, Kiagiadaki D, Ferekidis E, Yiotakis J. The effect of functional endoscopic sinus surgery on patients with asthma and CRS with nasal polyps. *Rhinology* 2010;48:331-8.
- 20) Ragab S, Scadding GK, Lund VJ, Saleh H. Treatment of chronic rhinosinusitis and its effects on asthma. *The European Respiratory Journal* 2006;28:68-74.
- 21) Senior BA, Kennedy DW, Tanabodee J, Kroger H, Hassab M, Lanza DC. Long-term impact of functional endoscopic sinus surgery on asthma. *Otolaryngol Head Neck Surg* 1999;121:66-8.
- 22) Manning SC, Wasserman RL, Silver R, Phillips DL. Results of endoscopic sinus surgery in pediatric patients with chronic sinusitis and asthma. *Arch Otolaryngol Head Neck Surg* 1994;120:1142-5.