

의료기관의 빅데이터 활용방안에 대한 연구

김성수*
청주대학교 의료경영학과*

Study on Big Data Utilization Plans of Medical Institutions

Sung-Soo Kim*
Dept. of Healthcare Management, Cheongju University*

요 약 의료정보의 급속한 발달로 인하여 막대한 양의 정보가 쌓이고 있다. 이러한 정보를 이용하여 임상연구를 하고자하는 욕구가 늘고 있으며, 고령화와 의료비의 가파른 상승을 해결해야하는 문제에 직면해 있다. 이에 대한 대안으로 빅데이터의 활용에 대한 목소리가 높다. 본 연구는 우리나라 의료기관들의 정보화 현황을 살피고 빅데이터 활용방안에 대한 정책적 시사점을 제공하고자 한다. 문헌조사와 의료기관의 의료정보전문가 면담을 통해 자료를 수집하였으며, 수집기간은 2013년 9월부터 2013년 11까지 4개월간 시행하였다. 연구결과 향후 병원정보시스템은 환자관리 및 행정에서 연구와 정보전략 측면으로 발전하고 있다. 따라서 빅데이터 활용을 위한 시스템 구축과 비정형 데이터의 효과적 활용을 고려하여 전문인력 양성과 더불어 의료비 절감을 위한 국가의 정책지원이 마련되어야 할 것이다.

주제어 : 빅데이터, 의무기록, 병원정보시스템, 의료정보, 정형화

Abstract Due to rapid development of medical information, a huge amount of information is being accumulated. Desires to conduct clinical researches by using this information are increasing, and medical institutions are encountering problems of aging society and drastic increase of medical expenses. Utilization of Big Data as an alternative is now being emphasized. The purpose of this study is to examine informatization of medical institutions and suggest political implications for Big Data utilization plans. Data was collected through literature searches and interviews with medical information professionals of medical institutions, from September to November, 2013, for four months. As a result of the study, it could be found that the hospital information system is improving from patient management and administration to researches and information strategies. Thus, national supports for medical expense reduction as well as fostering professional manpower should be provided, considering establishment of the system for utilization of Big Data and efficient application of unstructured data.

Key Words : Big data, Medical record, Hospital Information Management, Medical Information, Formalization

1. 서론

의료기관은 그 동안 의료정보시스템의 급속한 발전과 진료를 통해 얻은 데이터를 어떻게 관리하고 이를 다양

하게 활용할 것인가에 많은 고민을 하고 있다. 그 동안 의료기관은 신속하고 정확한 진료를 위하여 IT산업에서 활용되는 다양한 정보기술을 접목하기에 이르렀다. 지금까지는 이러한 정보시스템이 진료 및 행정업무의 편의를

*이 논문은 2012학년도에 청주대학교 보건의료과학연구소가 지원한 학술연구조성비(특별과제)에 의해 연구되었음

Received 2 December 2013, Revised 9 February 2014

Accepted 20 February 2014

Corresponding Author: Sung-Soo Kim(Dept. of Healthcare Management, Cheongju University)

Email: mra7033@naver.com

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. 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.

위해 개발되고 운영되어 왔다.

2008년 7월 스마트폰을 대표하는 아이폰 3G가 출시되고 기존의 SNS와 연동되면서 데이터가 급속도로 증가하였다. 이에 정보기술 시장조사기관인 가트너는 2012년 빅데이터가 미래전략기술 Top 10에 조사되었다고 보고하였다[1]. 세계 3대 저널로 대표되는 네이처는 2008년 빅데이터에 관한 저널로 관심을 모으기 시작하면서 이후 사이언스지에도 빅데이터에 관한 특별호가 게재되었다[2,3]. 또한 글로벌 경제전문지, 컨설팅 그룹에서 'Big Data' 관련 특집을 연이어 출간하면서 비중 있게 보도하기 시작하였다. 이중 2011년에 발표된 메켄지 보고서는 빅데이터 활용 시 미국 의료분야에서 매년 3,300억 달러 가치의 생산을 이룰 수 있을 것으로 예상하였다. 이는 미 정부 의료 예산의 약 8%에 해당하는 막대한 규모로 부수적으로 발생하는 가치도 약 1,000억 달러에 이른다고 한다. 의료기관 별 진료방법, 효능, 비용 데이터를 분석하여 보다 효과적인 진료방법을 파악하였을 경우 임상분야의 직접 효과는 약 1,650억 달러가 예상된다[4].

2013년 6월에 공개된 OECD의 국민의료비 통계자료에 따르면, 2011년 기준 우리나라의 GDP대비 국민의료비 지출 비율은 7.4%로서 OECD평균 9.3%보다는 낮은 실정이다. 하지만 2000년 4.3%와 비교하면 빠른 속도로 증가하는 것을 알 수 있다[5]. 이렇게 의료비가 증가하는 추세라면 국가재정에 막대한 영향을 미칠 수밖에 없다. 앞으로 의료계의 당면과제는 이러한 국가적 상황과 맞물려 살펴보아야 할 것이다. 의료계는 교통사고와 산업재해로 인한 사망을 증가하는 의료과오 및 의료과실로 인한 사망이 증가하고 있어 환자안전 문제가 시급하다. 또한 암 등 아직 완치에 가까운 의료기술을 이룰 수 있는 연구를 지속하는 것도 필요하다. IT산업의 발전이 의료계에도 적지 않은 영향을 끼쳐 의료정보시스템의 발전을 가져왔다. 앞서 밝힌 메킨지 보고서의 의료분야 가치는 국민의료비의 절감과 혁신적인 임상연구를 가능케 하는 것과 연관이 매우 깊을 것이다.

기존의 타 산업의 빅데이터 관련 연구는 주로 시스템과 분석프로그램 개발이 주 목적이다. 하지만 병원이나 보건 의료 서비스를 제공하는 기관들의 빅데이터 활용은 주로 임상 질환의 치료개선, 치료효과 입증, 재입원을 감소와 같은 3개 영역에서 중점적으로 다룰 수 있다. 따라서 이러한 차이점을 인지하고 의료기관의 빅데이터 활용

방안에 대한 접근이 필요할 것이다.

이에 본 연구에서는 우리나라 의료기관의 IT현황을 살피고 앞으로 보건의료비용을 줄이고 가치혁신을 이룰 빅데이터의 활용 현황과 연구방향을 탐색하고자 한다.

본 연구는 현재 우리나라의 의료기관의 의료정보시스템의 개발 수준을 살펴 이를 기반으로 빅데이터 활용실태를 점검하고 국내외 사례를 살펴봄으로써 앞으로 의료기관의 빅데이터 활성화 방향을 모색하고자 한다. 이에 본 연구에서는 다음과 같은 세부적인 목표를 달성하고자 한다.

첫째, 의료기관의 정보화 시스템의 근간이 되는 병원 정보시스템의 발전과정에 대하여 분석한다. 둘째, 각 병원정보시스템에 따른 정형화 실태를 파악한다. 셋째, 빅데이터의 분석과 활용에 대해 살펴본다. 넷째, 의료기관의 빅데이터 분석방법과 전망을 제시한다. 다섯째, 의료기관 빅데이터 이용사례를 살펴보고 마지막으로 우리나라의 의료기관 빅데이터 활용을 증진할 수 있는 정책제언을 하고자 한다.

2. 연구방법

본 연구에서는 의료기관 빅데이터 활용을 위한 의료정보시스템의 현황을 살피고 국내외 선행연구 및 사례연구를 고찰하여 효과적인 빅데이터 활용방안을 제시하고자 한다. 문헌고찰을 통하여 우리나라 의료기관들의 정보시스템현황을 살피고자 한다. 또한 해외 의료기관들의 활용사례를 분석한다. 빅데이터의 활용을 위해서는 의료기관에 존재하는 각각의 정보시스템에 따른 정형화에 대한 실태를 분석하는 것이 중요하다. 이러한 자료는 문헌조사를 통해 살펴보기 어려워 의료기관의 의료정보시스템 전문가 면담을 통하여 조사하였다.

위와 같은 방법을 통한 자료수집기간은 2013년 8월부터 2013년 11월까지 4개월간 시행하였다.

3. 연구결과

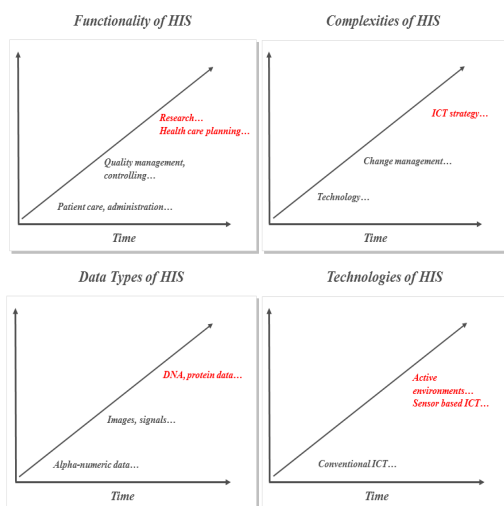
3.1 병원정보시스템의 발전과 빅데이터

의료기관의 빅데이터를 이용하기 위해서는 병원정보

시스템의 발전을 살펴볼 필요가 있다. 기반시설이나 시스템상황으로 많은 발전이 있음에도 불구하고 현재 우리나라는 매년 병원정보시스템의 현황을 조사하는 체계가 없는 실정이다. 하지만 앞서 몇몇 조사에서 발표된 현황을 살펴보고 이러한 시스템이 의료기관의 빅데이터를 이용할 수 있는 발판이 될 것이다.

Haux R(2006)의 연구에 따르면 앞으로 병원정보시스템은 과거 환자의 치료와 행정업무에서 연구중심과 국가보건정책으로 확대될 것이다. 또한 단순히 기술의 이용에 머물지 않고 정보기술을 전략적으로 활용할 수 있는 기반으로 발전하고 있다[6].

산업 각 분야의 빅데이터에서도 여러 종류의 데이터 형태를 이용하는 것과 마찬가지로 병원정보에서는 환자의 DNA와 같은 유전자 정보까지 효과적으로 활용할 수 있게 된다. 신체에 착용하는 컴퓨터의 등장이나 바이오센서 기반의 의료정보 기술도 발전되어 병원정보의 폭발적인 증가가 예상된다[Fig. 1].



[Fig. 1] Development of Hospital Information System

2005년 대한의료정보학회에서 발표된 국내 의료기관(요양기관)의 정보화 현황에 따르면 병원의 규모가 클수록 원무, 보험청구, 행정, 약제, 외래진료의 처방전달시스템(OCS) 등의 정보화가 상당한 수준에 이르고 있는 것으로 발표되었다[7]. 환자의 치료에 직접 관련 있는 처방전달시스템과 의료영상저장전송시스템(PACS)의 경우 중

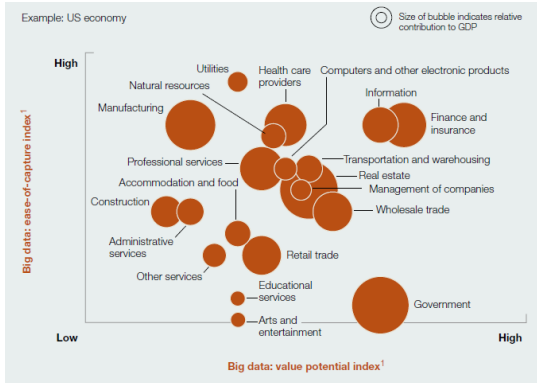
합전문병원은 90%이상 구축이 완료되었다. 그리고 종합병원도 80% 수준을 넘는 것으로 나타났다. 기존에 종이 의무기록이 전산화 되면서 전자의무기록 도입율은 낮은 상태이지만 그 동안 도입효과가 널리 알려지면서 현재는 많은 의료기관들이 상당히 도입한 것으로 알려져 있다. 원무업무나 행정업무의 경우 종합전문병원의 도입율이 100%에 이르러 빅데이터 활용의 기반이 될 것으로 보인다. 그리고 대형 종합병원(고려대학교 병원, 가천의과대학교 길병원, 건국대학교병원, 서울대학교병원, 서울아산병원, 삼성서울병원, 연세세브란스병원 등)을 중심으로 u-health 시범사업을 실시하고 있어 앞으로 빅데이터 활용의 폭이 더욱 커질 것이다.

최근의 국내 종합병원의 전자건강기록시스템(EHR) 연구에 따르면 2010년 상급종합병원과 종합병원의 전자건강기록시스템의 도입율이 37.2%, 처방전달시스템(OCS)은 87%로 세계최고 수준인 것으로 나타났다. 이것은 미국 11.9%(2009년), 오스트리아 11.9%(2007년), 독일 7.0%(2007년)에 비해 상당히 높은 수준인 것으로 볼 수 있다[8].

2012년 국가정보화백서에 따르면 공공의료의 핵심인 ‘보건기관 통합정보시스템 구축사업’은 전국 3,500여 보건기관(보건소, 보건지소, 보건진료소 등)에서 개별적으로 비표준 형태로 구축 운영 중이던 시스템을 전국 단위의 통합·표준화된 정보시스템으로 구축하여 보건행정 업무를 효율적으로 처리하고 양질의 지역보건의료 서비스를 제공하기 위한 목적으로 2006년부터 추진하였다. 2011년도에 전국 253개 보건소 등 모든 보건기관에 통합정보시스템을 보급 완료함으로써 지역보건의료서비스, 건강증진사업 실적 보고 및 정책적 통계 제공과 시스템 운영의 비용 효과성 제고 등이 가능한 체계를 갖추었다[9].

최근 의료기관의 의료정보화에 관심이 증가되면서 빅데이터활용에도 많은 관심을 보이고 있다. 이를 위해 건강보험심사평가원에서 실태조사를 2013년 8월에 실시할 계획을 가지고 있는 것으로 나타났다[10]. 최근 OECD 회원국을 대상으로 공통된 설문지를 가지고 의료기관의 보건의료정보기술 이용현황에 대한 실태조사가 이루어질 전망이다[11]. 현재까지 의료기관의 정보화에 대한 꾸준한 관심에도 불구하고 이에 대한 실태 파악 및 관련 자료의 구축이 아직도 미비한 것이 아쉬운 실정이지만 전자의무기록 등 정보화 수준이 상당히 발전한 것은 앞으로

빅데이터 활용을 통하여 국민의 건강수준 및 의학연구에 많은 발전을 기대할 수 있을 것이다. 2013년 발표된 메켄 지사의 조사결과는 의료기관의 빅데이터활용의 잠재가치가 매우 크다는 것을 알 수 있다[12][Fig. 2].



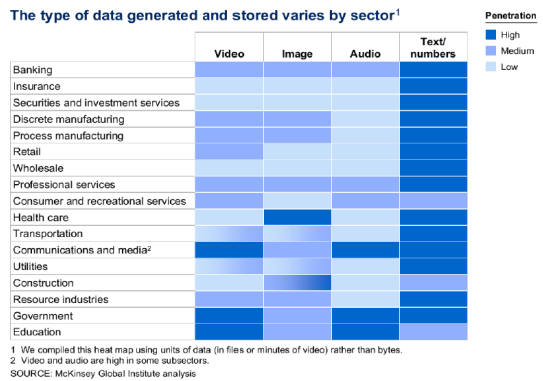
[Fig. 2] Big data : value potential index[12]

3.2 정형화

정형화된 데이터는 분석이 원활하여 신속한 결과를 보여주는데 매우 중요한 부분이다. 하지만 데이터의 양과 증가 속도가 빨라지면서 비정형화된 데이터를 정형화된 데이터로 변환하는 것은 그 만큼 결과의 지연을 초래하게 된다.

모바일 기술의 발달로 비정형화된 데이터가 급속도로 증가하면서 이를 위한 분석기법과 활용법이 다양한 현실

이다. 병원의 정보시스템도 각종 텍스트 데이터부터 숫자형 데이터, 이미지 데이터 등 다양한 데이터 형태를 보이고 있다. 특히 다른 산업분야에 비하면 보건의료분야는 이미지 데이터의 비중이 큰 것으로 나타났다[Fig. 3].



[Fig. 3] The type of data generated and stored varies by sector[12]

이를 각 정보시스템별로 살펴보면 다음과 같다<Table 2>. 종이기반의 환자진료기록을 전자화한 시스템이 전자 의무기록이다. 의무기록을 이루는 각 서식은 정형화를 위하여 템플릿으로 전자화한 경우 일정부분 정형화를 보이는 것은 하지만 대부분의 전자 의무기록은 진료결과를 서술하는 형식으로 기록하여 비정형화된 모습을 보이고 있다. 퇴원환자의 경우 의무기록을 전체적으로 검토한 후

<Table 1> Informatization of Healthcare Provider

| Work | Adoption and Impediment(%) | | | | | | | |
|---|----------------------------|-------|------------------|-------|----------|-------|-------|-------|
| | Special hospital | | General hospital | | Hospital | | Total | |
| | 1999년 | 2005년 | 1999년 | 2005년 | 1999년 | 2005년 | 1999년 | 2005년 |
| Outpatient OCS | 81.8 | 97.6 | 46.9 | 84.2 | 22.7 | 66.9 | 50.5 | 75.6 |
| Ward OCS | 81.8 | 97.6 | 43.6 | 84.2 | 18.2 | 58.4 | 47.9 | 70.6 |
| Medicine Work | 83.3 | 100 | 78.1 | 88.2 | 42.9 | 53.4 | 68.1 | 69.1 |
| Clinical Pathology Examination Work | 83.3 | 97.6 | 50 | 86.7 | 23.8 | 54.1 | 52.4 | 68.5 |
| Radioactive Examination Work | 83.3 | 97.6 | 51.6 | 88.2 | 23.8 | 53.4 | 52.9 | 68.7 |
| Special Treatment Management Work | 83.3 | 97.6 | 46.9 | 82.2 | 15 | 43.1 | 48.4 | 60.9 |
| Patient Administration Work | 100 | 100 | 96.9 | 98.7 | 96 | 95.4 | 97.6 | 96.8 |
| Administrative Work | 91.7 | 100 | 93.8 | 63.1 | 87.5 | 62.6 | 91 | 73 |
| Insurance Claim and Account Receivable | 100 | 100 | 96.9 | 100 | 87.5 | 89.4 | 94.8 | 93.5 |
| Medical Request Work | 75 | 97.6 | 41.4 | 69.1 | 9.5 | 36.4 | 42 | 53.8 |
| Picture Archiving Communication System (PACS) | 16.7 | 90.5 | 6.5 | 78.6 | 5 | 22.6 | 9.4 | 47.1 |
| Admission EMR | - | 21.4 | - | 14.5 | - | 21 | - | 19.6 |
| Outpatient EMR | 16.7 | 19.1 | 0 | 14.8 | 9.5 | 23.4 | 8.7 | 20.7 |
| Claim EDI | | 100 | | 90.5 | | 92.4 | | 94.8 |

〈Table 2〉 Actual condition of Data formalization of Hospital Information System

| Information System | Details Data | Actual condition of formalization |
|--|---|--|
| Electronic Medical Record | <ul style="list-style-type: none"> · Patient Records · Discharge Summary Records | unstructured structured |
| Management System | <ul style="list-style-type: none"> · Personnel Affairs/Wages · Budes/Accounting · Purchasing/Assets | structured |
| Picture Archiving & Communication System | <ul style="list-style-type: none"> · Radiography Images · Radiograph Read Results | unstructured unstructured |
| Order Communication System | <ul style="list-style-type: none"> · All Sorts of Prescriptions (operation·examination·injection, etc.) | structured |
| Laboratory Information System | <ul style="list-style-type: none"> · Clinical Pathology (blood, urine, sputum, etc.) · Tissue Pathology Slides · Tissue Pathology Read Results | structured unstructured unstructured |

진단, 수술 등 각종 정보를 국제 분류기준에 의하여 코드화하여 DB를 구축하고 있어 정형화된 형태를 띠고 있다.

경영관리시스템의 경우 병원의 경영정보와 관련한 데이터들이 축적된 시스템으로 볼 수 있다. 병원경영의 효율화를 위한 빅데이터기반 시스템이 될 수 있으며 이러한 시스템은 전자적 자원관리(ERP, Enterprise Resource Planning) 시스템의 도입으로 분석이 용이하고 자유로운 정형화된 형태를 가지고 있다고 볼 수 있다.

의료영상저장전송시스템(PACS, Picture Archiving & Communication System)은 주로 초음파, CT, MRI 등 방사선을 이용하여 진단을 위한 촬영장치에서 얻어지는 영상을 관리하는 시스템이다. 촬영결과로 발생하는 영상과 이를 판독한 진단 결과들이 생겨난다. 각종 영상은 비정형화된 데이터로 볼 수 있으며, 영상의 판독결과 또한 비정형화된 모습을 보이고 있다. 임상연구의 활성화와 방사선 판독결과의 활용을 위해 구조화된 양식으로 변환하는 프로젝트 추진이 가능하고, 서술된 판독결과를 자연어 처리를 통해 원하는 정보나 진단결과에 해당하는 환자를 검색하는 방안을 모색할 필요가 있다.

처방전달시스템(Picture Archiving & Communication System)은 환자를 치료하기 위하여 각종 처방을 입력하고 확인하는 시스템이다. 특히 수술, 검사, 투약 등으로 이루어져 있으며, 이들은 각 병원에서 정한 코드로 표준화되어 정형화된 모습을 보이고 있다. 하지만 병원마다 코드가 다르기 때문에 여러 병원의 연구나 조사를 위한 분석에는 용이하지 않다. 앞으로 전자건강기록시스템(EHR, Electronic Health Record)의 발전과 더불어 다기관 임상연구에 활용가치가 높을 것이다.

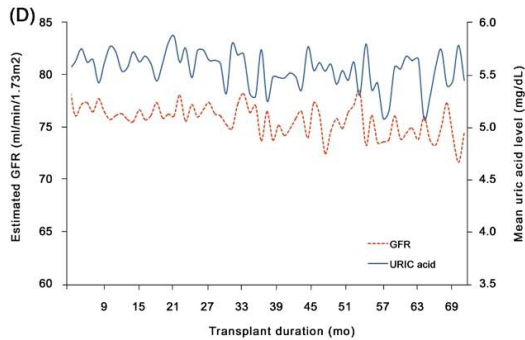
임상병리정보시스템은 혈액, 뇨, 객담 등을 통해서 병의 진단 및 분석을 위한 시스템이다. 각종 검체로부터 얻어진 결과가 숫자형인 경우는 정형화된 것이나, 조직의 병리적 소견을 기술한 결과는 비정형 데이터이다. 많은 환자의 경우에서 조직검사결과로 질병이 확진되기 때문에 임상연구에서 조직병리 판독결과는 매우 중요하다.

이 밖에도 각 의료기관에 따라서 병원정보시스템은 많은 차이를 보이고 있으며 정형화 정도 또한 다르다. 전자의무기록의 경우 많은 병원들이 각종 서식을 구분하고 서식의 특징을 잘 분석하여 템플릿을 통해 구조화 시킨 반면 그렇지 못한 병원들도 상당히 존재한다. 본 연구에서는 이러한 차이점을 인지하고 전문가의 자문을 통해 전체적인 병원정보시스템의 정형화실태를 분석하였다.

3.3 분석과 활용

의료기관의 정보화가 가속화되는 가운데 이의 활용이 향후 보건의료비용을 줄이고 국민의 건강을 향상 시킬 수 있을 것이다. 미래창조과학부는 공공 및 민간에서 빅데이터서비스 도입을 지원할 빅데이터 분석과 활용을 위한 센터를 구축하기 시작하였다[13]. 의료서비스도 공공 분야에 해당하며 어느 분야보다 빅데이터 활용에 대한 과급효과가 크기 때문에 국가차원의 지원책이 절실한 실정이다. 의료정보화가 정부의 지원에 비해 전 세계적으로 비교하여도 상당한 수준에 이른 것은 각 의료기관들의 꾸준한 노력의 일환이었다. 현재까지 의료기관들은 환자과 행정에 필요한 부분을 시작으로 각종 경영지표를 신속히 산출할 수 있는 시스템에 이르기까지 발전하여 왔다[14-16]. 일부 의료기관에서는 전자의무기록시스템

을 연구에도 활용이 가능한 차세대 전자의무기록시스템의 개발을 고심하고 있다. [Fig 4]는 임상병리 시스템에서 신장이식환자의 검사결과를 추출한 결과이다[17].



[Fig. 1] Change of Uric acid and GFR after Kidney Transplantation[17]

(Table 3) Utilization of Hospital in Big Data

| Infra | Operation | Utilization of Hospital |
|----------------------------------|-----------|---|
| Information Integration | 65% | <ul style="list-style-type: none"> ·Works to integrate information systems of each hospital ·Timely provide data for persons in need (doctor, nurse, administrator, etc.) ·Integrate independent data of each medical office |
| Storage Infrastructure to Expand | 64% | <ul style="list-style-type: none"> ·Presently, mostly management system-oriented storages ·In the future, establish infrastructures for utilization of Big Data such as research information |
| Mass Warehouse | 59% | <ul style="list-style-type: none"> ·Prepare for rapid increase of medical information ·Structuralize image and description text data ·Establish and integrate registries of each disorder |
| Safety Management | 58% | <ul style="list-style-type: none"> ·Systems fitting for international and domestic patient information protection |

2012년 IBM의 보고서에 의하면 빅데이터의 주요 인프라는 다음과 같다[18]<Table 3>. 빅데이터의 분석활용에 있어서 가장 시급한 것은 정보통합이다. 모든 분석 작업의 핵심 요소라고 할 수 있다. IBM Global Business

Service(2012)조사에 의하면 전체 응답자중 65%가 중요하다고 응답하였다. 의료기관에는 각각 분산된 많은 시스템이 존재한다. 이렇게 분산된 시스템은 데이터를 필요로 하는 사람에게 적시에 제공되어야 한다. 또한 임상각 진료과에 독립적으로 존재하는 데이터를 통합하여야 효과적인 임상지표 활용이 가능하다.

빅데이터활용에 있어서 두 번째로 중요한 것은 확장 가능한 스토리지 인프라이다. 현재 의료기관들은 운영체시스템의 활용이 주를 이루고 있으나 향후 연구정보 등 빅데이터 활용을 위한 인프라를 구축하기 위하여 확장성을 고려하여야 한다.

타 산업분야에 비하여 의료기관들의 데이터웨어하우스 구축은 저조한 실정이다. 의료정보의 급속한 증가에 대비하여 정보활용이나 주제별로 데이터웨어하우스나 마트를 구축하는 것이 필요하고 정형화 되지 않은 이미지 및 서술텍스트 데이터 활용을 위한 인프라 구축이 시급하다. 임상연구시 진단을 확인하는데 필요한 병리조직 검사 결과의 상세 정보를 조회할 수 있도록 구조화하여 대용량 데이터웨어하우스에 저장하는 것은 분석활용뿐 아니라 신속한 연구결과를 얻을 수 있는 중요한 시스템이 될 것이다. 또한 전자의무기록 시스템을 통해 각 질환별 레지스트리를 구축하여 임상병리, 진단 등 구조화된 데이터의 자동 추출도 다양한 임상연구에 활용될 수 있다. 타 산업의 어느 분야보다 정보보호가 중요시되는 만큼 국제 및 국내 환자정보보호에 적합한 시스템을 활용을 한층 강화할 수 있어야 한다[19].

빅데이터 활용을 위한 3대 요소가 자원, 기술, 인력을 꼽는다[20]. 앞서 <Table 3>에서 자원과 기술에 대한 부분을 살펴보았다. 어느 분야의 데이터 분석보다 빅데이터의 경우는 사람의 중요성이 크다고 할 수 있다. 수학, 공학, 경제학, 통계학 등 여러 분야의 통찰력과 비판적 시각이 기존의 정형화된 분석에서 찾아볼 수 없는 금맥을 찾을 수 있기 때문이다. 의료기관에서도 이들 인력이 매우 중요하게 될 것이다.

의료기관은 환자를 치료함으로써 수익이 발생되기 때문에 이들 치료와 관련된 인력이 대부분이다. 데이터를 분석하는 인력은 극히 드물고 각 부서의 행정 및 월간 통계 및 연보를 내는 수준에 불과하다. 앞으로 빅데이터의 효율적인 분석활용을 위하여 임상지식과 통계학, 데이터베이스 개념을 가지고 의료환경의 변화와 임상 트렌드를

읽어낼 데이터 사이언티스트 양성이 필요하다[21].

3.4 분석방법 및 전망

빅데이터 분석방향을 논하기 위하여 기존의 빅데이터 분석방법에 따른 의료기관의 분석방향을 살펴볼 필요가 있다<Table 4>.

<Table 4> Changes of analytical methods in Big Data

| Classification | | general industry (TO-BE) | hospital (TO-BE) |
|--------------------|------------------|--|---|
| Utilization Change | Data Open | Original/Analysis/Visualization: 3 stages | Management/Analysis/Clinical Departments: Local Data |
| | Issue Approach | Pre-responsive plans through real-time issue detection | Real-time searching of health and medical implications and research themes |
| | Utilization Type | Universal Departments/Civil Section: Horizontal Analysis | Analysis through horizontal collaboration among jobs |
| | Entity | Persons in charge of analyzing and planning | Collaboration among clinical professionals, analysis professionals, and medical information professionals |
| Analysis Change | Analysis Subject | Structured Data + Unstructured Data such as SNS, questions, etc. | Structuralize image and description text data Establish registries of each disorder |
| | Scale | Peta~Zettabyte | Terabyte |
| | Analysis Range | Multiple Storages | Within medical institutions, national data (HIRA) |
| | Application Time | In-Time Processing | In-Time Processing |
| | Data | Flow (on-the-fly) analysis | Utilize log files |

빅데이터 활용변화에 있어 주체는 임상전문가뿐 아니라 DB 분석전문가, 의료정보 전문가가 수평적 관계에서 협업이 무엇보다 중요하다. 규모는 테라바이트급 정도로 추산하고 있다. 분석범위는 의료기관내 데이터 분석뿐 아니라 전국민 건강보험자료의 활용이 가능하도록 다각화 하여야 할 것이다. 최근 보건산업진흥원의 연구에 의하면 환자/국민, 의료기관 그리고 기업 측면에서 빅데이터 보건산업 신산업 전망을 살펴볼 수 있다[22]<Table 5>.

그 동안 고 비용이었던 개인의 유전체 분석이 상당히 낮춰질 전망으로 이를 활용하여 개인별 맞춤 의학을 실현할 수 있을 것이다. 이는 병원, 의료진의 추천과도 연결될 수 있다. 실제로 KT에서도 이러한 유전체 분석의 기반 시스템을 활용할 수 있는 상용서비스가 이루어져 있다. 의료기관의 데이터는 정형화된 것보다 비정형화된

<Table 5> Prospect of New Healthcare Industry using the Big Data

| group | Utilization of Big data | Overseas Case |
|------------------|--|--|
| Patients/ People | Service to predict, through analyzing individual dielectric substances, disorders that are likely to develop and recommend appropriate hospitals/medical teams. | Genome Health Solutions Myriad Genetics |
| | Superhigh-speed dielectric substance analysis service & Service to provide new academic information of related fields and treatments for medical teams. | Genome Health Solutions |
| Hospitals | Superhigh-speed dielectric substance analysis service & Service to extract, visualize, and map data that medical teams are interested in, and provide it based on cloud. | DNAexus Appistry Inc |
| | Service to process a large amount of unstructured medical data (record, video, picture, etc.) in the structured shape that hospitals and medical teams request. | Health Fidelity |
| | Service to collect and process a large amount of data related to medical/hospital management and financial results, and provide it for hospitals and medical teams. | Explorys, Inc. |
| | Service to integrate clinical data owned by pharmaceutical/medical instrument companies and medical institutions, and provide data pool for researchers. | NextBio |
| | Information accumulated among patients having same diseases can be utilized for demand prediction pool of medical, pharmaceutical, and medical instrument development. | Patientslikeme |
| Companies | Service to collect data, which pharmaceutical/medical instrument companies request, from individuals, companies, and groups, and provide it through anonymity and standardization. | Qualcomm Life |

데이터들이 더 많이 존재하기 때문에 이들 비정형 데이터를 정형화하는 서비스가 각광 받을 것으로 보인다. 실제 미국의 한 기업에 의하면 병원에 저장된 각종 데이터를 결합하여 질환별 레지스트리를 구축하는 서비스를 하고 있다. 이를 통해서 임상연구를 시행하고자 하는 연구자는 분석을 위한 데이터 획득과정이 상당히 줄고 양질의 결과를 신속히 분석할 수 있게 된다. 결론적으로 양질의 데이터를 관리하는 것이 빅데이터 활용에도 도움이 될 것이다[23].

3.5 의료기관 빅데이터 이용사례

빅데이터를 활용하면 미국 의료부문은 연간 3,300억 달러(미 정부 의료 예산의 약 8%에 해당하는 규모)의 직간접적인 비용 절감 효과를 보일 것으로 전망된다[4]. 특히 임상분야에서는 의료기관 별 진료방법, 효능, 비용 데이터를 분석하여 보다 효과적인 진료방법을 파악하고 환자 데이터를 온라인 플랫폼화 하여 의료협회 간 데이터 공유로 치료 효과를 제고할 수 있고, 공중보건 영역에서는 전국의 의료 데이터를 연계하여 전염병 발생과 같은 긴박한 순간에 빠른 의사결정을 가능케 할 전망이다. 이들 사례를 국내와 국외로 구분하여 살펴보았다.

3.5.1 분당서울대병원

분당서울대병원은 의료계 첫 빅데이터 도입이라는 평가를 받고 있다. 2012년 4월 빅데이터 분석 기술을 도입한 차세대 임상데이터웨어하우스(CDW)구축 사업을 진행하였다[23]. 현행 데이터웨어하우스의 문제점을 개선하기 위해서 빅데이터 도입을 결정하였다. 현재 병원이 보유한 진료기록은 약 6TB 규모로, 현행 시스템으로 처리 속도가 더디다. 병원은 빅데이터 분석을 통해 시간 단축 및 자연어처리 기능을 강화 할 수 있을 것으로 보았다. 도입전 기술검증을 진행한 결과 기존 데이터웨어하우스 시스템에서 10분 이상 걸리던 작업이 10~20초 내로, 100배 이상 빨라졌다. 이러한 사업을 통해 분당서울대병원은 대용량 데이터 분석 및 자연어 검색 지원으로 의료진의 연구 검색 활성화, 자료제공 업무 요청 대기 시간 단축을 통한 효율성 증대, CDW 재설계를 통한 검색 정확성 개선, 연구지원 생산성 향상 등의 효과를 거둘 수 있을 것으로 보고 있다.

3.5.2 Blue Cross Blue Shield

SAS의 빅데이터 마이닝을 활용하여 병원 재입원률을 감소시킨 사례로 인정받고 있다[24]. 병원 재입원률이 그간 크게 올랐다. Blue Cross Blue Shield 병원은 노스캐롤라이나 주의 최대 보험사로, 환자의 삶의 질과 만족도 개선, 재입원률의 감소를 위해 노력해 왔다. 정부와 여러 기관들이 핵심적인 보건 이슈로 보고 있는 문제이다. 따라서 본질적으로 재입원은 발생하면 안 되는 상황이다. 재입원이라 하면 일반적으로 퇴원 후 30일 이내에 예정에 없이 병원에 다시 입원하는 것을 말한다. 데이터 모델 개발을 통해 재입원 발생 원인뿐만 아니라 발생 예상시점도 예측가능 하도록 하였다. 이 시스템을 활용해 보건 프로그램 자원을 활용해 재입원 발생을 예방할 수 있다. 자사 자료의 최초 검증 후, 이 보험사는 입원환자 재입원률 감소에 대한 엄청난 잠재력을 보였다. 문제는 방대한 가용 데이터에서 진짜 정보를 찾아내는 것이다. 재입원 모델을 통해 먼저 노스캐롤라이나의 당사 회원들을 대상으로 재입원 원인을 파악하였다. 어떤 원인들은 조치가 가능하기도 하다. 따라서 재입원 확률이 높은 증상을 가진 환자에 대해 정보를 가지고 있으면, 이들에게 맞는 프로그램을 제공할 수 있다. 이를 통해 재입원 방지에 기여할 수 있게 되는 것이다. 데이터의 면밀한 탐색이 아주 중요하다. 데이터 마이닝으로 찾기 훨씬 어려운 관계도 파악할 수 있다. 의사결정 트리나 신경망 모델을 이용해 표준 회귀모델 분석 시 의학적으로 명백하지 않은 관계도 찾아낼 수 있다. 이러한 관계와 데이터 분석능력은 업무프로세스에 중요한 정보를 제공하며 우리가 시도하고자 하는 프로세스의 속성을 이해할 수 있도록 지원한다. 노스캐롤라이나 Blue Cross Blue Shield는 재입원 모델을 활용해 환자와 의료서비스 제공자 간 이해의 격차를 줄이고 조정하였다. 환자와 의사, 간병인과 보험사 모두 좋은 의료서비스 보장과 현대 보건의료 시스템의 긍정적 변화에 한 축을 담당할 수 있게 되었다.

3.5.3 미국 국립보건원 '필박스(Pillbox) 프로젝트' 추진

필박스(Pillbox) 프로젝트는 약 검색을 지원하는 검색 통계프로그램으로, 주요 질병의 분포, 연도별 증가 등을 분석하고, 약의 효능 확인이 가능하며, 연간 100만 건의 알약 문의가 필박스로 대체되고, 연간 약 5,000만 달러의

비용 절감 효과 발생할 것으로 추산하고 있다[25].

4. 논의 및 결론

그 동안 의료분야는 고령화 사회와 질병발생의 다양화, 의료기관간의 경쟁심화로 변화가 필요한 시기에 접어들었다. 이러한 시점에 본 연구를 통해서 의료기관의 빅데이터활용의 잠재가치는 매우 큰 것으로 나타났다. 의료분야의 당면한 과제를 이를 통해 해결할 수 있을 것으로 기대한다[26-27].

병원정보시스템의 급속한 발전으로 이를 이용한 빅데이터 분석의 기반이 이루어졌다고 본다. 그럼에도 불구하고 이러한 시스템을 잘 활용하지 못하고 있다. 임상연구에 전자무기록 시스템이나 임상병리시스템을 활용하여 다차원적 분석을 통한 신속 정확한 데이터 분석이 가능한 반면, 대부분의 연구자들은 전자무기록 시스템에 있는 환자정보를 화면을 통하여 검토하고 각종 검사 결과를 연구에 필요한 별도의 DB에 입력하는 것이다 [27]. 이용할 자료는 존재하지만 효과적인 이용을 위한 분석시스템이나 체계가 미비한 것이 문제이다.

기존의 빅데이터 관련 연구 및 동향을 검토한 결과, 전반적으로 가치 있는 자료 및 사례들이 발견되었다. 그럼에도 불구하고 타 산업에 비해 여전히 비정형데이터가 차지하는 비중이 높고 전문적인 내용이라 분석 장벽이 높은 것으로 나타났다. 이제 빅데이터 활용을 위한 기반 시스템 활용을 높이고 여러 사례에서 나타난 분석을 각각의 의료기관에 맞게 리모델링하여 의료기관의 원활한 운영과 수준 높은 연구에 활용할 필요가 있다.

의료기관의 빅데이터 활용을 위하여 다음과 같은 정책제언을 하고자 한다.

첫째, 의료기관의 정보화 실태를 국가적 차원에서 주기적으로 점검할 필요가 있다. 현재 국가적 차원의 주기적 실태 파악이 되지 않아 정확하고 미래지향적인 빅데이터 실태를 확인하기 어렵다. 또한 이렇게 파악된 자료는 향후 통합작업을 통하여 의료기관내 이용에서 의료기관간 이용을 통해 활용 폭을 넓힐 수 있을 것이다.

둘째, 그 동안 의료기관은 낮은 수가에도 불구하고 의료정보화를 위해 많은 자원을 투자하였다. 빅데이터의 활용이 국가 의료비 절감에도 효과가 기대되는 바 이를

위한 의료기관에 대한 지원정책이 필요하다.

셋째, 의료기관은 빅데이터분석을 위해 적극 투자할 필요가 있다. 기반 시스템은 의료정보의 발전으로 많이 성숙해 있으나 고차원 적인 분석시스템은 아직 타 산업 분야에 비하여 부족한 실정이다. 빈도와 비율이 주를 이루는 보고서 형태의 분석에서 탈피하여 환자들이 소리 없이 외치는 요구를 데이터마이닝 등 빅데이터 분석기법을 활용하여야 한다.

넷째, 의료기관은 이들 분석과 활용에 있어 규모와 실정에 맞게 아웃소싱도 고려할 필요가 있다.

마지막으로 데이터 사이언티스트로 일컬어지는 인재 양성이 시급하다. 의료분야는 특성상 상당히 전문적인 데이터이다. 따라서 임상과 DB, 통계학 등 의료분야의 전반적 지식에 해박한 인재야 말로 숨겨진 데이터의 진실을 발견하고 원활한 활용을 이끌 견인차 역할을 할 수 있게 된다.

본 연구는 기존의 빅데이터 관련 연구 및 현황을 광범위하게 분석하여 활용방안 및 연구방향을 탐색하고자 하였다. 그러나 본 연구는 다음과 같은 몇 가지 제한점을 가지고 있어 향후 연구에는 이를 극복하는 추가 연구가 필요한 실정이다.

첫째, 관련 문헌고찰과 전문가의 자문을 통해 연구를 수행하였지만, 아직 의료기관의 빅데이터활용이 성숙하지 않은 시점에서 확대 해석된 문헌들로 인하여 장밋빛 전망을 내어 놓은 오류를 범할 수 있을 것이다.

둘째, 현재까지 우리나라 전체적으로 의료기관의 정보화 실태를 정확하게 조사한 자료의 부재로 향후 추가 연구에서 최근의 정보화 실태를 파악하고 빅데이터 활용방안에 대한 추가 탐색이 이루어져야 한다.

셋째, 의료기관에서 실제로 경영자료 분석이나 임상연구에 참여한 전문가를 광범위하게 조사하여 실증분석이 이루어져야 할 것이다.

ACKNOWLEDGMENTS

This work was supported by the research grant of Cheongju University in 2012.

REFERENCES

- [1] Gartner. CEO Advisory : Big Data Equals Big Opportunity, 2011.
- [2] <http://www.nature.com/nature/journal/v455/n7209>
- [3] <http://www.sciencemag.org/content/331/6018.toc>
- [4] McKinsey Global Institute. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, 2011.
- [5] Organization for Economic Cooperation and Development (OECD). OECD health data 2013 – frequently requested data. Downloadable Excel file. http://www.oecd.org/els/health-systems/oecdhealthdata_2013-frequentlyrequesteddata.htm. Accessed July 29, 2013.
- [6] Haux, R. Health information systems - past, present, future. *International Journal of Medical Informatics*, Vol. 75, No. 3, pp. 268-281, 2006.
- [7] Health Insurance Review & Assessment Service. Report on the informatization of healthcare provider, Korean Society of Medical Informatics. 2005.
- [8] D. Yoon, B. C. Chang, S. W. Kang, H. Bae, R. W. Park, Adoption of electronic health records in Korean tertiary teaching and general hospitals, *International Journal of Medical Informatics*, Vol. 81, No. 3, pp. 196-203, 2012.
- [9] National Information Society Agency. 2012 National Informatization White Paper. Ministry of Public Administration and Security, 2012.
- [10] http://www.rapportian.com/n_news/news/view.html?no=12844
- [11] <http://www.hkn24.com/news/articleView.html?idxno=126974>
- [12] McKinsey Global Institute, Game changers: Five opportunities for US growth and renewal, 2013.
- [13] <http://biz.newdaily.co.kr/news/article.html?no=20045>
- [14] J. G. Dewitt, P. M. Hampton, Development of a data warehouse at an academic health system: knowing a place for the first time. *Acad Med*, Vol. 80, No. 11, pp. 1019-1025, 2005.
- [15] J. R. Schubart, J. S. Einbinder, Evaluation of a data warehouse in an academic health sciences center. *International Journal of Medical Informatics*, Vol. 30, No. 3, pp. 319-333, 2000.
- [16] Silver M, Sakata T, Su, H. C, Herman C, Dolins S. B, O’Shea M. J. Case study: how to apply data mining techniques in a healthcare data warehouse. *Journal of Healthcare Information Management*, Vol. 15, No. 2, pp. 155-164, 2001.
- [17] K. M. Kim, S.-S. Kim, D. J. Han, W. S. Yang, J. S. Park, S. K. Park, Hyperuricemia in kidney transplant recipients with intact graft function. *Transplantation Proceedings*, Vol. 42, No. 9, pp. 3562-3567, 2010.
- [18] IBM press release. Analytics: The real world use of big data, 2012.
- [19] Hongro, L., Youngju, S., Jeongkyeom, K., Kiseok, C. & Jaesoo, K. A Study on the Analysis of Personal Information in the R&D Report. *International Conference on Convergence Technology*, Vol. 1, No. 1, pp. 165-167, 2012.
- [20] President’s Council on National ICT Strategies. Establishing a smart government by using big data. Seoul, Korea: President’s Council on National ICT Strategies; 2011.
- [21] S. M. Lee, R. W. Park, Basic Concepts and Principles of Data Mining in Clinical Practice. *J Kor Soc Med Informatics*, Vol. 15, No. 2, pp. 175-189, 2009.
- [22] S. M. Back, Prospects and policy direction of the New industries to take advantage of healthcare using Big Data. *KHIDI Brief*, Vol. 84, pp. 1-20, 2013.
- [23] http://www.ddaily.co.kr/news/news_view.php?uid=109365
- [24] <http://blogs.wsj.com/cio/2012/03/30/blue-cross-expects-cost-savings-from-big-data-dive/>
- [25] K. Jee, G. H. Kim, Potentiality of Big Data in the Medical Sector: Focus on How to Reshape the Healthcare System, *Healthcare Informatics Research*, Vol. 19, No. 2, pp. 79-85, 2013.

- [26] T. B. Murdoch, A. S. Detsky, The inevitable application of big data to health care. *Journal of American Medical Association*, Vol. 309, No. 13, pp. 1351-1352, 2013.
- [27] H. R. Keun, Heart Diseases Prediction based on Data Mining Techniques. *International Conference on Convergence Technology*, Vol. 1, No. 1, pp. 1-6, 2012.

김 성 수(Sung-Soo Kim)



- 2001년 8월 : 인제대학교 보건대학원 보건학과 (보건학석사)
- 2010년 8월 : 인제대학교 일반대학원 보건학과 (보건학박사)
- 2001년 7월~2012년 2월 : 서울아산병원 의료정보관리팀
- 2012년 3월~현재 : 청주대학교 의료경영학과 조교수

- 관심분야 : 보건정보, 보건통계, 의무기록, 임상연구
- E-Mail : mra7033@naver.com