빅데이터 시대의 임상데이터 활용을 위한 데이터 표준화 및 상세화

윤현미* · 손호선** · 이영성*** · 김영규****

I. 서론

대한민국에는 지금 빅데이터 열풍이 불고 있다. 일상생활에서 일어나는 모든 일들이 데이터화 된다고 해도 과언이 아닐 정도로 우리가 살아가는 이 시간에도 어마어마한 양의 데이터가 축적되고 있다. 하지만 데이터 그 자체로는 어떠한 의미가 없으므로 데이터의 가치는 데이터 마이닝 및 통계적 방법 등 여러 기법을 활용한 분석 결과물로 나타난다. 현대의학의 발전으로 인간의 평균수명이 늘어남에 따라 사람들은 오래 삶과 동시에 더 나은 질의 삶을 원하게 되었다. 따라서 사람들의 관심은 자연스레 의료 관련 분야에 집중되었고, 의학 관련 데이터에 또한 집중 하게 되었다.

의학 관련 데이터에 일반 사람들의 집중도가 높아진 만큼 의료 데이터는 비전문가가 보았을 때에도 충분히 이해가 가능한 수준이어야 한다. 그러나 현재까지의 의학데이터는 전문가들만이 알아볼 수 있는 수준의 어떠한 설명도 없는 의학용어들로 가득 하며 비전문가들의 수준에서는 이해하기 힘든 수준의 데이터이다.

본 논문에서는 양질의 데이터 수집을 위한 데이터의 표준화와 각 임상 항목의 설명, 기록 단위, 정상치 범위 등의 내용이 포함된 데이터의 상세화의 필요성을 설명하였다.

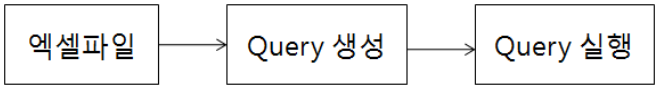
II. 본문

1. 임상데이터의 수집

1) 기존의 임상데이터 수집 방식

국립중앙인체자원은행에 근무하며 17개의 단위 인체자원은행(이하 병원)으로부터 임상데이터를 수집하는 업무를 하다 보니 수집 방식의 개선과 데이터의 표준화는 필수 요소라는 것을 알게 되었다.

기존 임상데이터 수집 방식의 절차는 그림 1과 같다.



(그림 1) 엑셀파일을 이용한 임상데이터 삽입 절차

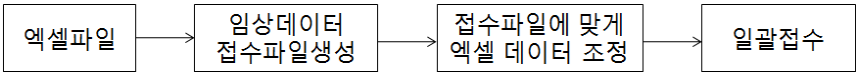
* 윤현미, 충북대학교 의생명과학경영융합대학원 대학원생, 010-8410-7422, yhm7422@gmail.com
** 손호선, 충북대학교 의생명과학경영융합대학원 교수, 043)249-1770, shon0621@gmail.com
*** 이영성, 충북대학교 의생명과학경영융합대학원 교수, 043-261-2869, lee.medric@gmail.com
**** 김영규, 충북대학교 의생명과학경영융합대학원 교수, 043-261-2830, ygk@chungbuk.ac.kr

표준화가 이루어지지 않은 상태의 임상 데이터는 같은 항목에 대한 의미가 각 병원마다 달라서 의미가 모호해지는 상황이 발생한다. 그만큼 데이터 삽입 시에 오류율이 높아지고 자료의 크기가 정해지지 않아 메모리가 낭비됨에 따라 업무 처리 속도가 현저하게 낮아진다. 각 병원의 임상데이터를 한데 수집하여 데이터베이스에 저장 및 관리해야 하는 연구자의 업무 효율이 낮아질 수밖에 없는 구조이다.

이러한 문제를 해결하기 위해 인체자원정보관리시스템(BIMS)가 개발되었다.

2) BIMS를 이용한 임상데이터 수집

BIMS(Biospecimen Information Management System)는 인체자원 정보를 한 프로그램에서 통틀어 관리하고자 개발된 시스템이다. BIMS의 일괄접수 기능을 이용한 임상데이터 수집 절차는 그림 2와 같다.

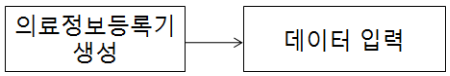


(그림 2) BIMS를 이용한 임상데이터 삽일절차)

각 병원으로부터 수집한 엑셀파일을 BIMS에서 생성한 임상데이터 접수파일 형식에 맞게 조정하고 BIMS의 일괄접수 기능을 이용하여 임상데이터를 데이터베이스에 삽입한다. 엑셀의 파일 그대로를 쿼리로 만들어서 데이터베이스에 삽입하는 방식 보다는 오류율이 적다는 장점이 있으나 일괄접수 방식도 데이터베이스 삽입에 있어 오류를 완전히 제거하지 못했다는 단점이 있다.

3) 의료정보등록기를 이용한 임상데이터 수집

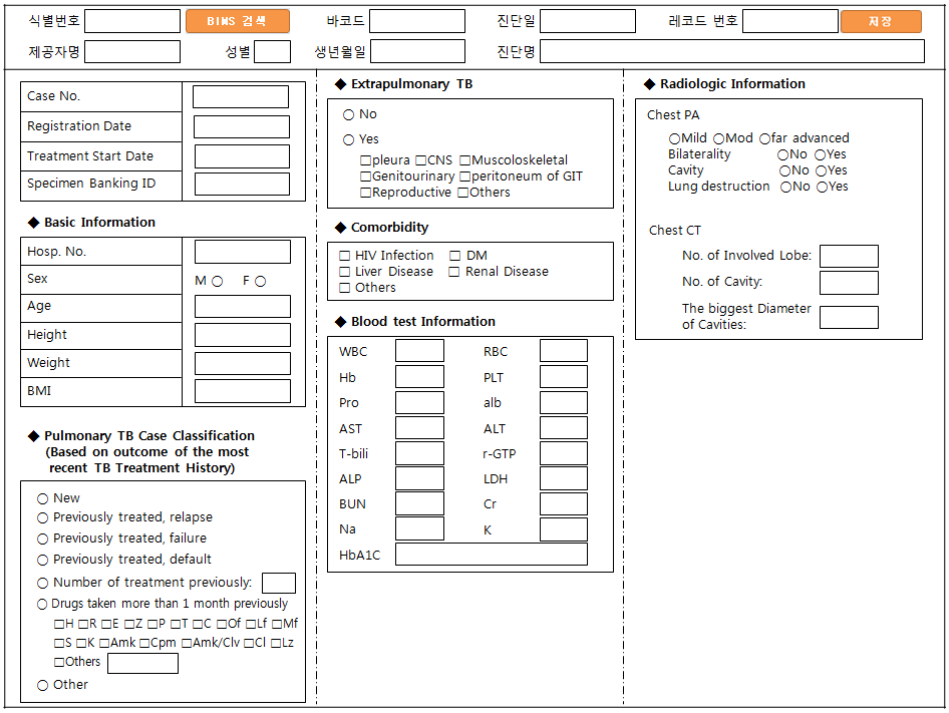
의료정보등록기는 Microsoft Access를 이용한 프로그램이다. 의료정보등록기를 이용한 임상데이터 수집절차는 그림 3과 같다.



(그림 3) 의료정보등록기를 이용한 임상데이터 수집 삽입절차

의료정보등록기의 장점은 입력란 하나하나가 데이터베이스와 직접 연동되어 있어 입력하는 대로 데이터베이스에 삽입된다는 장점이 있다. 엑셀로 쿼리를 생성하는 방식이나 BIMS 일괄접수 기능처럼 임상데이터를 엑셀 프로그램을 한 번 거쳐서 데이터베이스에 삽입하는 방식보다 데이터베이스로의 접근성이 우수하다는 것 또한 의료정보등록기의 장점이다. 그림 4는 의료정보등록기의 구성도이다

구성도에서 알 수 있듯이 의료정보등록기는 각 병원의 임상 항목을 입력란으로 구성하여 항목을 하나하나 입력하고, 저장 버튼을 눌러 데이터베이스에 삽입하는 방식이다. 의료정보 등록기에서 데이터 수정 또한 가능하다. 그러나 의료정보등록기의 단점은 환자 한명 한명의 데이터를 한 건씩 입력하는 구조라서 연구자의 실질적인 업무량이 앞선 두 방식보다 더 많다고 볼 수 있는 것이다. 따라서 각 병원마다 임상데이터량이 방대한 경우 BIMS 일괄접수 기능을 더 많이 사용하는 추세이다..



(그림 4) 의료정보등록기 구성도

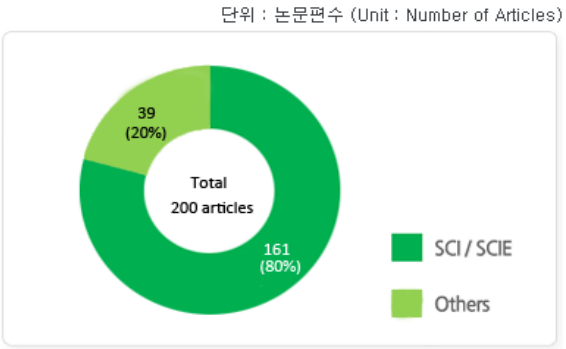
2. 임상데이터의 활용

1) 활용연구과제

(표 1) 활용연구과제, 출처 : 한국인체자원은행네트워크

병원명	인체자원 분양 vial	논문 건수
강원대병원	1044	4
경북대병원	2256	12
경상대병원	2624	11
계명대동산병원	2449	13
고려대구로병원	1886	17
부산대병원	3253	19
서울대병원	6177	12
서울아산병원	612	11
순천향대부천병원	4674	14
아주대병원	2436	10
원광대병원	3263	10
인제대부산백병원	1216	12
화순전남대병원	2246	16
전북대병원	1957	13
제주대병원	1100	2
충남대병원	2567	14
충북대병원	8161	19

2) 활용성과



2008년부터 2013년까지 총 200편 논문 발표
- SCI/SCIE 논문 161편
- 기타 논문 39편
2013년 총 66편 논문 발표
- SCI/SCIE 논문 55편
- 기타 논문 11편

(그림 5) 활용성과, 출처 : 한국인체자원은행네트워크 홈페이지

Ⅲ. 결론

임상데이터의 공공데이터로써의 가치와 활용도는 매우 높다. 하지만 데이터의 정확성이 낮은 연구 과제라면 타 연구자로부터의 신뢰 또한 낮을 것이다. 데이터가 정확하게 수집되었는지에 대한 사후확인도 중요하지만 사전적으로 좀 더 정확한 데이터를 수집하는 것이 더 중요하다고 생각한다.

임상데이터의 표준화는 연구 과제의 신뢰도를 중점으로 보았을 때 중요한 요소가 아닐 수가 없다. 표준화되지 않은 채 수집된 데이터는 낮은 질의 데이터로 수집되었을 가능성이 높다. 임상데이터를 활용한 연구 결과는 우리 삶의 질을 더 높이는 일과 직접적으로 연관되어 있기 때문에 좀 더 정확한 데이터를 활용해야 공공데이터로써의 신뢰할 수 있는 연구 결과가 도출되기 때문이다.

국민으로부터의 의료 데이터에 대한 관심이 높아진 만큼 앞으로의 수집 방식에 있어 정보의 표준화와 상세화는 꼭 필요한 요소이다. 앞으로 연구 결과에 대한 신뢰도를 높일 수 있는 임상데이터 수집 방식을 좀 더 강구해야 한다.

사사표기

본 논문은 2013년도 미래창조과학부의 재원으로 과학벨트기능지구지원사업의 지원을 받아 수행된 연구임 (2013K001552).

참고문헌

남세종 (2008), “임상 데이터에 대한 정보처리와 시스템에 대한 연구”, 세종대학교 대학원 석사학위 논문.

염지현 (2012), “전자건강기록에서의 임상시험 데이터 수집 (Acquiring Clinical Trial Data From Electronic Healthcare Records)”, 「기술혁신학회지」, 정보과학회논문지 : 컴퓨팅의 실제 및 레터 제 18

권 제 7 호(2012.7).
이춘열 (2014), “데이터 표준화의 절차”, 이춘열 외 1인 공저,「데이터 관리 및 활용 전략」.
조순로 (2011), “생물자원의 관리와 생명윤리정책연구”, 한남대학교 대학원 박사학위 논문.
예정훈 (2009), “한국인체자원은행에서의 임상문서표준구조 등록 시스템 구현 = Implementation of Clinical Document Architecture Registry for the Biobank of Korea”, 경북대학교 대학원 석사학위 논문.