

임상시험정보의 효율적인 활용을 위한 통계모델링 구축

김동선*, 조성제*

*동방대학원대학교 문화정보학과

e-mail : dustin.kim.radar@gmail.com

The establishment of the statistics modeling for the effective utilization of the clinical trials information

Dong-seon Kim*, Sung-Je Cho*

*Dongbang University's Cultural information college

요 약

식품의약품안전청은 산발적으로 흩어져 있던 임상시험 관련 정보를 통합하여 서비스하기 위한 방안으로 임상시험정보방을 개설하게 되었다. 임상시험승인현황 등 원시자료는 종이형태로 생산, 관리된다. 이를 통계자료로 활용하기 위해서 식품의약품안전청의 관련 시스템인 KiFDA 시스템에 입력을 하게 된다. 이 시스템에 입력된 데이터는 실시간으로 임상시험관련 웹서비스 시스템에 구축된다. 이때에 실시간 통계정보의 정확성이 무엇보다 중요하다. KiFDA 시스템의 데이터베이스의 실시간관명과 임상시험계획승인제도(IND) 승인리스트 원본의 실시간관명이 틀린 것도 있었다. 이는 동일한 실시간관도 화면에서는 서로 다른 실시간관으로 보일 수 있다는 것을 의미한다.

결국 통계자료가 부정확하게 표출되는 데는 이런 원인들이 있었던 것이다. 본 논문에서는 기존의 문제점을 개선하기위해 임상시험정보방의 효율적인 통계모델링을 설계하여 물리적 데이터베이스를 구축하였다.

자 고려사항, 임상시험의 종류 및 요건과 관련용어도 정리해놓았다. 임상시험의뢰자에게 도움을 줄 수 있는 승인절차도 자세히 설명되어 있다.

특히 의약품 임상시험계획 승인현황을 [1]을 통해 매주 제공하고 있다. 그러나 연도별 의약품 임상시험 승인건수, 실시간관현황 등 당초 제공하려 했던 통계관련 정보는 확인하기 어려운 정보기술문제가 발생하여 정확한 정보제공에 어려움이 있었다.

식품의약품안전청의 '임상시험정보의 효율적 제공을 위한 임상자료 DB구축 및 활용방안' 연구용역에 참여하여 이와 같은 통계정보의 미흡한 부분을 분석하여 이를 개선할 수 있는 방안을 연구 하였다.

우리는 통계관련 정보의 수집과 처리과정이 일반 임상시험정보와는 다르다는 사실을 확인하였다. 또한 통계관련 정보는 데이터의 무결성이 상당히 요구되기 때문에 이를 충분히 감안하고자 노력하였다. 초기 [1]에서 제공하고자 시도했던 통계정보서비스의 문제점을 파악하여 이를 보완하는 작업을 병행하였다. 최종적으로 우리는 [1]에서 통계서비스에 대한 오류와 문제점을 찾아냈으며 이를 해결하게 되었다.

1. 서 론

임상시험계획승인제도(IND) 도입이후 다국가 임상시험을 포함한 국내 임상시험이 급증하면서, 식품의약품안전청 허가를 득하기 위한 임상시험 기준에 대한 관심이 고조되어 왔으며 공신력 있는 임상시험정보 제공이 절실히 요구되어 왔다. 또한 임상시험평가지침, 가교시험 질의응답 등 규정에 대한 설명이 필요해졌다.

이에 식품의약품안전청은 산발적으로 흩어져 있던 임상시험 관련정보를 통합하여 서비스하기 위한 방안으로 임상시험정보방[1]을 개설하게 되었다.

[1]에서는 임상시험 승인절차는 물론이고 국내임상시험 승인현황을 파악할 수 있으며 또한, 임상시험관련법규, 1990년 초반부터 식약청에서 제정된 임상시험가이드 라인을 포함하여 최근 제정된 효능군별 임상시험 평가지침 및 식약청에서 발표한 세미나자료 등을 검색하여 쉽게 확인할 수 있도록 하였다.

임상연구자 및 제약 산업체, 일반인, 환자들이 쉽게 이용할 수 있도록 임상시험에 대한 알기 쉬운 설명과 피험

*본 논문은 2009년도 식품의약품안전청 용역연구 과제 연구비에 의하여 연구됨

본 논문의 구성은 먼저 서론을 기술하고, 제2장에서는 국내외 임상시험 정보망의 관련현황, 제3장에서는 임상시험정보의 통계모델링 내용, 제4장은 물리적 데이터베이스 구축, 제5장은 결론에 대해 언급하고자 한다.

2. 관련 연구 현황

2.1 ClinicalTrials.gov[2]

[2]는 미국 NIH(National of Health)가 FDA에 보고하는 임상시험 정보를 바탕으로 그 산하기관인 NLM (National Library of Medicine)을 통하여 개발 운영 및 관리하는 사이트이다.

[2]는 NIH,연방정부, 기업체에 의한 후원으로 69,455 trials를 보유하고 있으며 연구된 데이터베이스는 50 개주와 162 나라에 제공되며 하루 약 5만 여명이 접속하여 자료를 활용 하고 있다.

[2]에서 asthma(천식)에 관한 임상시험 정보를 얻으려면 메인화면에서 임상시험 검색(search for clinical trial)을 클릭한 후 기본 검색화면에서 asthma를 치면 데이터베이스에 저장되어 있는 asthma와 관련된 임상시험 케이스들을 나열하여 보여 준다. 또한 모든 질병, 치료법 등을 입력하여 검색 할 수 있다. 기본검색에서 질병, 약물 등을 키워드로 검색 할 수 있다. 예를 들어 고혈압(hypertension), 아스피린(aspirin), 두통(headache)등을 입력 하면 이와 관련된 임상 시험케이스 등이 화면에 나열 된다.

상세검색, 일반 검색기능을 이용하여 임상시험 진행여부, 임상 시험 종류, 임상 시험을 진행한 나라, 치료법, 의뢰자, 연구 번호, 나이군, 후원자 등을 조건으로 더 쉽게 찾아 볼 수 있다. 기본 검색화면에서 asthma를 입력한 후 보여 지는 화면으로서 1,925개의 임상시험 케이스가 있다. 너무 많아 일일이 찾아 볼 수 없다면 상세검색 화면으로 임상시험 진행여부, 임상시험 종류, 임상 시험 진행 나라, 치료법, 의뢰자, 연구번호, 나이군, 후원자등을 입력하여 데이터를 검색할 수 있다. 항목을 입력할 때마다 케이스가 줄어들므로 많은 항목을 입력 할수록 원하는 임상 시험 케이스를 찾을 수 있다. 조건별 검색으로 질병상태, 치료법, 의뢰자, 시험 기관 위치, 등의 알파벳순과 카테고리 순으로 검색할 수 있다.

2.2 CenterWatch Clinical Trials[3]

[3]은 1994년에 설립된 Center Watch는 환자와 연구자들에게 임상시험에 관한 정보를 제공하고 있다. 메인 화면 상단의 임상시험(clinical trials)를 클릭하여 임상시험 결과 데이터베이스를 찾아 들어간다.

데이터베이스는 적응증, 약물명, 치료명 등으로 검색해 볼 수 있으며 정보는 일반적으로 임상시험에 관한 간단한 설명, 시험장소, 목적, 약물 사용여부 등이 제공된다. 이 사이트는 직접 입력하여 데이터를 찾지 않고 메디컬 컨디션(medical condition) 이나 drug, therapy name을 클릭하여 찾아들어 간다. 메디컬 컨디션에 따라 검색할 수 있고

록 분류하였다. 메디컬 조건을 선택 하여 보여 지는 화면으로서 알파벳 순서로 증상과 병명이 섞여서 보여 진다.

그 중에 하나를 선택하면 된다. 클릭하면 보여 지는 데이터베이스 결과화면으로서 임상시험 정보로는 임상시험 상태, 임상시험번호, 스폰서, 제목, brief summary, primary outcome, 등등 데이터를 보여 준다.

2.3 National Center Institute Clinical Trials[4]

[4]는 미국 국립 암협회에서는 가장 최근의 암 임상 시험결과를 제공 한다. cancer type과 cancer topic에 따라 찾을 수 있으며 왼쪽에 keyword를 입력하여 결과를 찾을 수 있다.

메인화면 상단의 임상시험을 클릭 한 후 clinical trails result를 선택하면 암 종류와 증상에 따라 결과물을 찾을 수 있다.또한 왼쪽의 search trial results칸에 암 종류를 입력하면 임상시험 케이스가 나열되어 보여 진다. 키워드와 암 유형 등으로 검색된 결과는 summary, source, background, study, result 등이 제공된다.

2.4 ClinicalStudyResult.org[5]

[5]는 미국 제약협회(PhRMMA)에서 제공하는 임상시험 결과에 대한 database로 online 상으로 쉽게 접근이 가능하다. 이 database에는 현재 미국시장에서 발매 중인 각종 의약품과 관련하여 미국 제약 협회의 회원사들이 진행한 임상시험을 자진해서 제공한 연구 결과가 구축되어 있다.

[5]에서는 회사명, 제품명, 적응증 등을 선택하거나 스테디 이름, 적응증 등을 키워드로 입력한 뒤 검색하면 등록되어 있는 임상시험 결과들을 확인 할 수 있으며 가능하다면 임상시험 원문에 대한 링크 또한 제공하고 있다. 이 사이트에서 일반인이 검색하기에는 다소 까다롭다. 메인화면에서 drug name, company name, generic name등을 모두 정확히 입력하여야 하므로 한 가지라도 잘 모르면 검색하여 찾아 들어 갈 수가 없다.

2.5 Current Controlled Trials[6]

[6]은 the Science Navigation Group의 일부로 전 세계에서 진행되고 있는 무작위 대조 시험 (randomized controlled trials)에 대한 정보를 제공하기 위하여 1998년에 시작한 데이터베이스이다.

검색할 수 있는 키워드로 검색하면 제목, 질병, 위치, 모집한 환자의 조건, 결과 등을 나열하여 보여 준다.고려되어야 할 사항을 연구하여, 임상시험계획 수립 단계에서 고려되어야 할 통계부분 계획에 대한 지침서를 작성하였으며 제2장에서는 국내·외 임상시험 가이드라인 연구개발의 현황, 제3장에서는 임상시험 현황과 임상시험계획서 및 결과보고서의 통계 검토의견 현황분석과 통계부분 작성 및 체크리스트에 대한 내용, 제4장은 연구결과의 고찰 및 결론, 제5장은 연구 목표달성도 및 관련분야에의 기여도를

설명하고 제6장에서는 연구개발 결과 활용계획에 대해 언급하고자 한다.

본 연구를 통해 임상시험 허가심사 자료의 심사 평가 도구 및 심사자료 관련 통계부분 계획 지침자료를 마련하는 성과를 기대할 수 있으며 임상시험 허가심사에 대한 통계부분 지침서를 통해 관련부서에서 임상시험계획서의 검토에 활용할 수 있다.

3. 임상시험정보의 통계모델링

임상시험정보 서비스시스템에서는 다음과 같은 문제점이 지적되었다.

- 가. 임상시험이 단일계획서에 따라 하나이상의 기관에서 시행될 경우, 시험기관별로 실시결과가 표출되게 되고, 중복된 자료를 계속 표현된다.
- 나. 임상시험관련 통계의 각종 현안자료들이 KiFDA[7]에 입력되기 전의 원시자료와 일치하지 않아 통계자료가 부정확하게 표출된다.
- 다. 2001년부터 2008년도까지 약1,400여건의 임상시험관련 원시자료 입력을, 현재자료의 부정확성으로 인해 작업 불가.
- 라. 임상시험 원문검색방법 중, 주제검색과 조건검색 및 지도검색방법이 없다.

3.1 원인분석

위의 4가지 문제들은 모두 데이터베이스와 관련이 있는 것들이다. 따라서 [7]의 데이터베이스를 기준으로 통계의 부정확성에 대한 원인을 살펴볼 필요가 있다. 먼저, 임상시험승인현황 등 원시자료는 종이형태로 생산, 관리된다. 이를, 통계자료로 활용하기 위해서 [7]의 시스템에 입력을 하게 된다. 이 시스템에 입력된 데이터는 실시간으로 임상시험관련 웹서비스시스템으로 보내지는 것으로 확인되었다. 또한 해당 쿼리 일부를 직접 발췌하여 확인한 결과, 문제가 없는 것으로 확인되었다. 즉, 부정확한 통계의 원인은 최초로 데이터가 입력되고 보관되는 과정에서 발생한 것으로 추정되었다. [7]의 시스템에서 불러오는 데이터는 [1]에서 가공하는 것이 아닌 것으로 확인되었다. 즉, KiFDA 시스템에서 불러온 데이터는 [1]에서 곧바로 출력하는 형태였다. 이는 원본을 갖고 있는 KiFDA 시스템에서 오류가 없다면 [1]의 통계화면에서도 오류가 없음을 의미하는 것이다.

보다 정확한 원인분석을 위해 [7]의 시스템에 검증을 실시하였다. 그 결과 몇 가지 문제점들이 새롭게 발견되었다. 화면에서 표출되는 건수와 자료의 건수가 서로 틀렸다. 어느 기관의 경우에는 한해 실시건수가 41건이 화면에서 누락되기도 하였다. 2007년 한해 [7]의 시스템에 있는 데이터베이스자료와 IND승인리스트 원본건수를 대조해본 결과 서로 맞지 않았다. 특히, 실시기관의 총 건수를 비교해본 결과, 2007년 한해에만 283건의 자료가 [7]의 데이

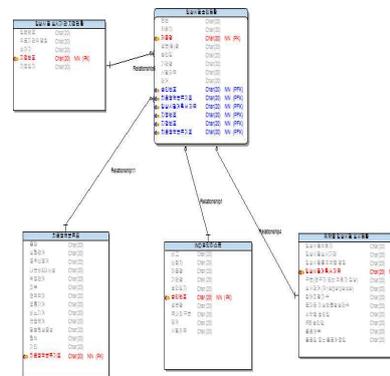
터베이스에 누락되어 있었다. 이는 원시자료의 입력이 덜 되었다는 것이고, 결국 실제 건수와 임상시험통계화면에서 보이는 실시건수가 부정확할 수밖에 없는 원인이 되었다.

이와 같이 실시건수라는 임상시험관련 데이터베이스 자료 중에서 기본적인 중요한 자료의 누락 이외에도, [7]의 데이터베이스의 실시기관명과 IND 승인리스트 원본의 실시기관명이 서로 달랐다. 이는 동일한 실시기관도 화면에서는 서로 다른 실시기관으로 보일 수 있다는 것을 의미한다. 결국 통계자료가 부정확하게 표출되는 데는 이런 원인들이 있었던 것이다.

3.2 통계 논리적 모델링

데이터베이스의 여러 자료들이 누락되거나, 또는 틀린 데이터가 수시로 입력되었던 이유를 살펴보았다. 이는 여러 요인이 복합되어 발생했을 것으로 추측해볼 수 있다. 체계적인 데이터베이스 설계 부족, 엔트리 간의 관계 설명 부족, 운영자의 이해를 도울 수 있는 설계 미흡 등이다.

이를 해결하기 위한 방안으로, 해당 엔트리들에 대한 논리모델링 도입을 고려하게 되었다. 만일 설계된 논리모델링이 시스템에 반영된다면 사용자가 엔트리 즉, 테이블간의 관계와 속성을 쉽게 확인할 수 있기 때문에 오류발생을 방지하는데 일차적인 기여를 할 수 있다고 판단하였다. 아래 (그림 1)은 논리모델링 진행과정 중 일부 화면의 모습이다.



(그림 1) 통계 논리적 모델링

임상시험관련 통계자료의 부정확성 및 오류 해결을 위해 시도된 논리적인 모델링 기법의 도입은 결과적으로 정확하고 오류 없는 자료 표출을 가능하도록 하였다. 확인 결과, 임상시험 승인현황 및 실시기관 현황은 곧바로 서비스 가능한 상태가 되었으며, 그 이외의 전문가용 자료 등도 해당 전문가들이 이용 가능할 수 있도록 부정확성과 오류가 해결되게 되었다.

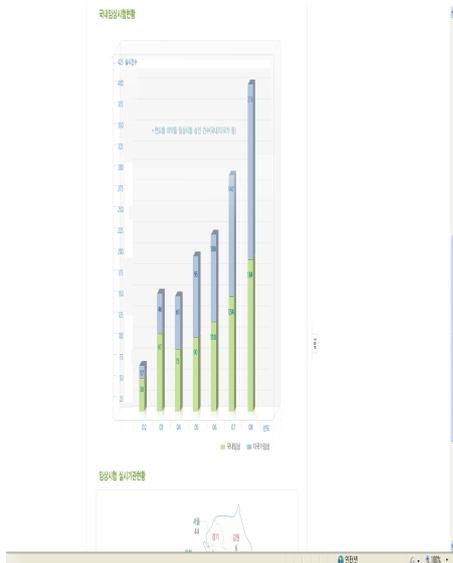
4. 물리적 데이터베이스 구축

임상시험관련 통계자료의 부정확성 및 오류 해결을 위해 시도된 논리적인 모델링 기법의 도입은 결과적으로 정확하고 오류 없는 자료 표출을 가능하도록 하였다.

[7]에 입력된 원시데이터를 확인해본 결과 임상시험현황

에 대한 데이터는 모두 정확하게 처리되는 것을 알 수 있었다.

(그림 2)는 실시기관 현황 데이터 중 실시기관의 불일치 및 중복등록문제를, 새로운 리스트를 시스템에



(그림 2) 오류수정 후 임상시험현황 화면

적용함으로써, 서비스되는 화면에서 정확한 데이터가 표출되고 있는 모습이다.



(그림 3) 오류수정 후 실시기관현황 화면

(그림 3) 화면은 실시기관 현황데이터로서 기존 실시기관 현황 데이터 중 불일치 및 중복문제를 논리모델링기법의 적용으로 극복한 후 표출한 화면이다. 보다 다양한 서비스 화면을 제공하기 위해서는 향후 보다 풍부하고 다양한 관련데이터 입력이 요망된다.

5. 결론

의약분야의 선진국은 신약의 개발에 천문학적인 투자를 하고 있다. 그 만큼 신약시장이 주요 시장중 하나이기 때문이다. 신약을 개발하는 과정에서 임상시험은 핵심과정 중 하나이다. 정부에서 심혈을 기울여 관련 제도와 시스템을 개선하고 정비하는 것은 그래서 당연하면서도 필수적인 과제라 할 것이다.

기존 임상시험정보를 제공하기 위해 구축된 시스템에 존재하고 있던 다양한 오류와 문제점들을 찾아내, 원인을 분석하는 데는 여러 제약조건들이 있었다. 그러나 마침내 원인을 찾아내 제거함으로써 극복해냈다. 이 과정에서 우리는 다시 한 번 데이터베이스 시스템을 설계하고 구현하는 과정에는 논리모델링 등 체계적이고 도식적인 기법과 도구의 활용이 절실히 요구된 다는 것을 알 수 있었다.

참 고 문 헌

- [1] <http://clinicaltrials.kfda.go.kr>
- [2] <http://clinicaltrials.gov>
- [3] <http://search.centerwatch.com>
- [4] <http://www.cancer.gov>
- [5] <http://www.phrma.org>
- [6] <http://www.controlled-trials.com>
- [7] <http://kifda.kfda.go.kr>