CDISC/SDTM 표준으로 변환된 eCRF 생성

홍영표*, 유용빈**, 김석일***
*한국국제대학교 건강보건대학 병원관리학과,
**가톨릭대학교 보건대학원 보건정보학과,
***가톨릭대학교 의과대학 예방의학교실
e-mail:yphong@hanmail.net

Generation of eCRF converted to CDISC/SDTM standard

Yeong-Pyo Hong*, Yong-Bin Yu**, Sukil Kim***
*Dept of Hospital Management, College of Health, The HanKook International University of Korea.
**Department of Health Information, Graduate School of Public Health, The Catholic University of Korea
*** College of Medical, The Catholic University of Korea,

요 약

본 연구에서는 임상시험 데이터를 수집하는 CRF를 웹의 표준인 XML 기반으로 설계하고, 임상시험 데이터 표준인 SDTM으로 변환된 eCRF를 생성하는 XML 기반 프로세스을 개발했다, 데이터구조와 폼이 분리되는 XML 기반 개발 프로세스는 XML Schema로 eCRF를 개발하고, SDTM을 적용한 XML 문서를 생성한다, 생성된 XML 문서에서 SDTM 규약에 맞는 도메인별로 변환방법을 제시한다. 이에 eCRF 분석과 데이터베이스 설계 및 구현이 동시에 이루어진다.

Key Words : 임상시험 데이터표현 표준, CDISC, SDTM, eCRF, Clinical Data Management,

1. 서론

임상시험용의약품의 안정성과 유효성은 임상시험(Clinical Trial/Study)에 의해서만 확립될 수 있어서, 임상시험은 신약개발에 있어서 가장 중요한 과정이다[1-2]. 최근의 임상시험은 다국적・다기관간의 협력에 의해 운영되는 것이 보편화되고 있으며, 임상시험 기준강화로 임상시험 기간이 길어지고 있어서, 데이터 입력뿐 아니라. 데이터 쿼리 및 정리에 많은 시간이 소요되므로, 데이터 표현과 데이터 처리 등 데이터 관리를 위한 국제적 표준화가 요구된다[3].

CDISC(Clinical Data Interchange Consortium)는 1997년 이래 전자적인 임상시험 정보수집과 교환과정을 향상시키고자, 표준데이터모델의 필요성에 의해 조직되었다. CDISC Technical Roadmap에서 SDTM(Study Data Tabulation Model)은 신약승인 신청시 관련된 임상자료를 제출할 때 표현하는 공통형식을 정의하며, 20004년 FDA(Food and Drug Administration)는 SDTM 채택함을 발표했다[4-5].

임상시험 데이터를 수집하는 전자적임상증례기록지(Electronic Case Report Form, eCRF)는 기존에 전통적인 프로세스로 개발할 경우, 데이터베이스 설계가 잘못되었으면 CRF 분석설계 단계부터 다시 해야만 하고, SDTM을 적용할 경우 역시 새로운 시스템으로 개발해야만 한다.

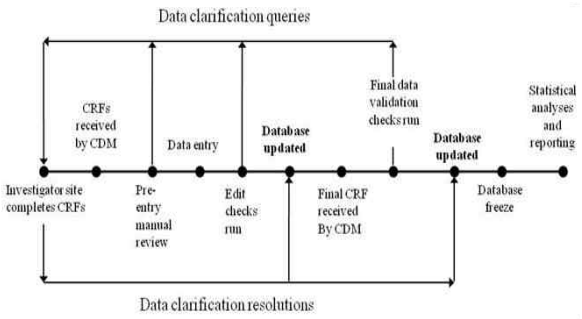
본 연구에서는 eCRF를 데이터와 Form을 별도로 관리할 수 있는 웹의 표준인 XML 기반으로 개발하고, 임상데이터 표준인 SDTM으로의 변환 방법을 제시한다.

1.1. 연구배경

1.1.1 XML 기반 eCRF 개발 프로세스

임상데이터관리(Clinicla Data Management, CDM)는 임상시험 데이터수집에 있어서 근본적인 목적인 통계적 분석에서 요구되는 데이터와 보다 더 좋은 임상 실습을 만족시킬 수 있는 질 높은 데이터를 적시에 전달될 수 있도록 하는 과정이다(Fig. 1)[6].

이 과정을 전자적인 데이터수집(Electronic Data



[Figure 1] CDM data validation activities

capture, eEDC)을 위해 eCRF를 웹기반에서 전통적인 소프트웨어 개발방법으로 개발할 경우, 객체와 상호작용 모델링을 포함하는 수준높은 설계과정에서부터 시작되어, 분석단계에서부터 데이터베이스 구조를 고려하여 설계해야 하므로 많은 시간과 비용이 소요된다[7].

W3C XML(eXtensible Markup Language)은 확장 가능한 마크업 언어인 웹 표준으로. XML 기반은 웹상에서 단순히 데이터를 eForm에 입력하는 형태가 아니라, 데이터와 eForm을 분리하여 관리가 가능하다. XML 기반 eForm은 전통적인 개발 프로세스와는 다르게, 문서 구조를 작성하면서 데이터베이스 설계가 가능하며, 문서가 유효(validation)한지 여부를 점검하면서 구현되므로, XML 기반의 eCRF 개발 프로세스 과정이 요구된다[8-9].

1.1.2. 임상시험 데이터 표준화

정보기술 발달로 EDC 사용이 증가하며, 전자적임상시험 프로세스(electronic clinical trial process)를 향상시키고 있다. 그러나, 임상시험을 주관하는 수탁기관이나 제약회사, 심사기관 등 관련 기관들마다 서로 다른 데이터 구조와 표현을 사용하고 있어, 관련기관들 사이에 정보를 주고 받을 경우, 서로 다른 상대방의 데이터 구조와 표현에 맞추어 다시 작성 후 전달해야 하는 번거로움이 있다. 또한 임상시험마다 각각 다른 프로세스와 요구사항들로 각각 다른 시스템을 개발해야하며, FDA에서 요구하는 국제 표준으로의 변환 개발이 어렵다[10].

SDTM은 임상시험 기간동안 수집되는 데이터를 도메인 또는 주요 주제에 대한 조직화로, 임상시험 연구 데이터베이스 표준을 만들어 임상응용 표준개발을 지원한다. 2004년 New Drug Applications (NDAs)에 SDTM standard format을 채택함을 발표했으며, 2009년 5월 현재, 2005년 8월 발표한 SDTM Version 3.1.1 Final을 채택하고 있다[4-5]. 이에 eCRF를 SDTM으로 개발하거나, 기존 시스템을 SDTM으로 변환되어야 함이 요구된다.

1.2. 연구목적

본 연구에서는 데이터 구조와 Form을 별도 관리 가능한 W3C XML기반으로 임상시험에서 전자데이터를 수집하는 eCRF 시스템을 개발하고, 개발된 eCRF에서 임상데이터 표준인 SDTM으로의 변환방법을 제시하며, 이를 실제 구현한 예를 제시한다.

2. 연구범위 및 방법

2.1 연구범위

기존 CRF를 분석해서 Schema를 설계하고, XML 문서에서 SDTM에 명시되어 있는 도메인별 데이터 표현을 부분적으로 적용하여, SDTM이 별도의 XML 문서로 생성되는 과정으로 연구 범위를 제한다. 개발 환경은 XML 개발 통합 도구인 XMLSpy와 StyleVision을 사용하였으며[11, XML parse는 Libxml 20630과 Libxslt 10122 버전을 사용한 xsltproc tool을 사용하여 검증하였다.

2.2 연구방법

CDM에서 SDTM을 구현하기 위해 XML 데이터를 처리하기 위한 XSL 스타일시트와 SDTM 도메인을 분석하여. eCRF를 구현한다.

2.2.1 W3C XSL

XSL은 XML 데이터 처리용 스타일시트를 웹에 보여주기 위한 XML 기반 언어이며, XLS은 XSLT(XSL Transformation)와 XPathfh 구성된다. 기술적으로 메모리 상에 있는 원본 트리를 또 다른 결과 트리로 어떻게 변환시킬 것인가를 지정한다. 즉, XSLT 자체로는 데이터 변환 과정의 로직을 기술하고, XML 문서에서 필요한 데이터를 얻기 위해 사용한 XPath 표현식는 XML 데이터에 접근할 때 사용하는 쿼리(query) 언어이다.

XPath 문법은 SQL과 전혀 다르게 계층적인 데이터 트리를 처리하기 위해 설계되어, XML 문서에 내재되어 있는 계층적인 구조의 위치로 데이터를 쉽게 참조할 수 있다[12]. XPath 형식은 노드축을 중심으로 상대/절대 위치경로를 따라 표현되며, 본 연구에서 사용된 기본적인 구문(syntax)은 Table 1과 같다.

[Table 1] XPath syntax examples

XPath Syntax	Comment
/	모든 절대 경로는 사선(/)으로 시작
.	문맥노드 자신
..	문맥 노드의 부모노드선택
@domain	문맥 노드의 Domain 노드 선택
/child::*	노드의 모든 자식들인 요소노드를 선택
child::*paragraph	문맥의 모든 자식노드들을 선택
ancestor-or-self::*	문맥 노드의 자신과 조상 노드 선택

2.2.2 Analysis of SDTM Domains

SDTM은 임상시험표 데이터셋(clinical trial tabulation data sets)의 내용 (content)과 설계(design)에 관한 임상시험 자료처리과정에서의 표준이다. 연구 데이터셋과 연구설계 정보는 임상연구의 계획・설계인 프로토콜 정보를 포함하고 있으며, 구조화된 프로토콜은 검토자에게 실제적인 데이터를 효과적으로 비교할 수 있는 기준을 제공한다.

SDTM 발표되기 이전의 임상시험 데이터들은 도메인만 있어서, 수집된 데이터를 관리하는데 많은 어려움이 있었으나, SDTM은 도메인이름, 변수, 변수명과 구조를 잘 정의하고 있다[5].

SDTM은 General Observation Classes와 Special-Purpose Datasets으로 분류된다. Special-Purpose Datasets은 도메인 데이터셋(Domain datasets)(Table 2)과 시험설계 데이터셋(Trial Design Datasets)(Table 3), Representing Relationship and Data 로 구성된다.

General Observation Classes는 임상연구에서 단계(subject-level)적으로 수집되는 데이터이며, 중재(Interventions), 이벤트(Events), 조사(Findings)로 구성된다. 중재데이터는 연구 주제에 관해 계획적으로 관리된 데이터이며, 복용 약물(CM), 약물 Dose (EX), 약물 물질(SU)등의 데이터이다. 이벤트 데이터는 계획된 프로토콜에서의 이상반응(AE), 프로토콜 범주(DV), 병력(MH) 등이 있다. 조사 데이터는 계획된 평가에서부터 명시된 질문까지 결과를 수집하며, ECG 테스트 결과(EG), 실험실 테스트 결과(LB), 대상 특성(SC), 신체 계측(VS) 등이 있다. 데이터 소스는 메타데이타가 근원이 되며 도메인 수를 결정한다[12].

Special-Purpose Datasets 중 Damain Datasets 인 Table 3의 DM, CO와 같은 두 글자는 도메인 그 자체를 구분하여 부여된 식별자이며, 인구학적 자료(DM), 설명문(CO), 대상 요소(SE), 대상 방문(SV) 등에 관한 데이터가 기록된다. Table 3의 Structure 항목에 기술된 'One record per subject '은 대상마다 한 개의 레코드로 구성되는 레코드 구조를 기술하고 있다.

[Table 2] Domain Datasets and their Data Structure

dataset	Description	Data Structure
DM	Demographics	One record per subject
CO	Comments	One record per comment per subject
SE	Subject Elements	One record per actual Element per subject.
SV	Subject Visits	One record per subject per actual visit

[Table 3] Trial Design datasets and their Propose data Structure

dataset	Description	Data Structure
TA	Trial Arms	One record per planned Element per Arm
TE	Trial Elements	One record per planned Element
TV	Trial Visits	One record per planned Visit Arm
TI	Trial Inclusion /Exclusion Criteria	One record per I.E criterion

시험설계모델(Table 3)은 시험설계에 관한 정보로 시험시점(TA), 시험요소(TE), 시험 포함/배제 기준(TI) 등의 데이터가 있다[14]

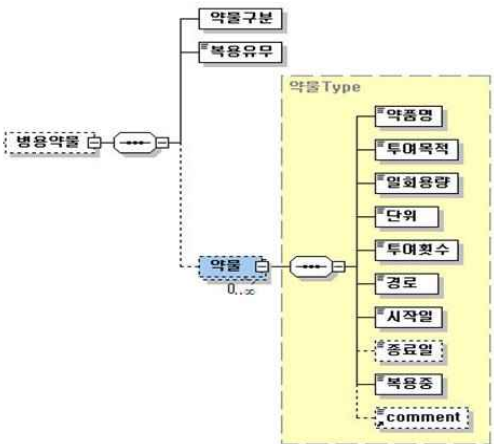
3. 연구결과

3.1 XML Schema Design

Fig. 3은 XMLSpy를 사용하여 작성한 병용약물 CM 도메인에 대해 도식화된 Schema 이며, Prog.1은 Fig.3 Schema의 글로벌 (global) "약물 Type" 이외의 병용약물에 대한 Text 프로그램 소스이다.

3.2. SDTM XML

Prog. 2는 Fig. 3.의 Global CM Type인 Schema를 적용한 XML 문서이며, 속성값으로 SDTM을 적용하였다.



[Figure 2] Design of CD Domain in XML Schema

```
<xs:element name="병용약물" minOccurs="0">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="약물구분">
        <xs:complexType>
          <xs:attribute name="varName" use = "required" fixed="CMCAT"/>
          <xs:attribute name="code" fixed="병용약물"/>
        </xs:complexType>
      </xs:element>
      <xs:element name="복용유무">
        <xs:complexType>
          <xs:simpleContent>
            <xs:extension base="boolean">
            <xs:attribute name="varName" use="required" fixed="CMOCCUR"/>
          </xs:extension>
          </xs:simpleContent>
        </xs:complexType>
      </xs:element>
    <xs:element name="약물" type="약물Type" minOccurs="0" maxOccurs="
unbounded"/>
```

[Prog. 1] Text in XML Schema of the CM domain

```
<약물>
  <약품명 varName="CMTRT">ASPIRIN</약품명>
  <투여목적 varName="CMINDC">headache</투여목적>
  <일회용량 varName="CMDOSE">5</일회용량>
  <단위 varName="CMDOSU">mg</단위>
  <투여횟수 varName="CMDOSFRQ">once</투여횟수>
  <경로 varName="CMROUTE">OR</경로>
  <시작일 varName="CMSTDTC">20070501</시작일>
  <종료일 varName="CMENDTC">20070501</종료일>
  <복용중 varName="CMENRF">NO</복용중>
</약물>
```

[Prog. 2] Text SDTM SML of the Global CM Type

Table 4는 Prog. 2에서 글로벌 "약물 Type"에 사용된 SDTM 변수명이다. 그 이외 항목에서 반드시 필요한 항목과, 데이터 타입 등을 규정하고 있다.

3.3. Extract SDTM from the XML Schema by XSL

Prog. 3은 Prg 2.에서 SDTM을 추출하기 위한 XSL 프로그램이다. Prog. 4는 Prog.3의 XSL에 의해 SDTM이 추출된 XML 문서로, SDTM의 도메인명이 태그명으로 부여되어 있다,

4. 고찰

본 연구는 정보기술 표준과 임상시험 표현 표준으로 개발되어, 데이터 교환에 많은 기여를 할 것으로 기대된다, 그러나, CDMS 상의 데이터베이스와

[Table 4] Varible Name of CM domain in Figure 3

Variable Name	Variable label
CMTRT	Reported Name of Drug, Med, Therapy
CMINDC	Indication
CMDOSE	Dose per Administration
CMDOSU	Dose Units
CMDOSFRQ	Dosing Frequency per Interval
CMROUTE	Route of Administration
CMSTDTC	Start Date/Time of Medication
CMENDTC	End Date/Time of Medication
CMENRF	End Relative to Reference Period

```
<?xml version="1.0"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
  <xsl:template match="/">
    <xsl:element name='result'>
    <xsl:for-each select='//*[@varName]/..'>
      <xsl:variable name='domain'>
        <xsl:value-of select='ancestor-or-self::*[@domain]/@domain' />
      </xsl:variable>
      <xsl:element name="{$domain}">
        <xsl:for-each select='*'>
          <xsl:variable name='sdtm'>
            <xsl:value-of select='./@varName' />
          </xsl:variable>
          <xsl:element name="{$sdtm}">
            <xsl:value-of select='.' />
```

[Prog. 3] XSL source for the SDTM Extract

```
- <result>
  - <CM>
      <CMTRT>ASPIRIN</CMTRT>
      <CMINDC>headache</CMINDC>
      <CMINDC>5</CMINDC>
      <CMDOSU>mg</CMDOSU>
      <CMDOSFRQ>once</CMDOSFRQ>
      <CMROUTE>OR</CMROUTE>
      <CMSTDTC>20070501</CMSTDTC>
      <CMENDTC>20070501</CMENDTC>
      <CMENRF>NO</CMENRF>
    </CM>
```

[Pro. 4] Extract SDTM from the Prog 3

SDTM 도메인과 적절히 연결시키는 매핑규칙에 대한 연구와 CDMS 상에서의 SDTM을 적용한 데이터베이스를 SDTM 도메인으로 자동 변환하는 연구가 필요하다.

참고문헌

[1] Shin SG. Clinical Trials of Drug. J Korean Med Assoc 1998; 41(8):806-9(Korean)

[2] FDA, FDA Standaeds initiatives: An Overview. 2003.

[3] CDISC. Available from: http://www.cdisc.org.

[4] Armando Oliva, M., The Revie of SDTM Datasets at CDER : A Clinical Reviewer's Perspective .FDA CDER, 2005.

[5] Dimena S, D.J., Mikael P, Michael W, Clinical Operation Online (COOL) - A World Wide Web - based Approach to Running Clinical Trials : Results from two International multicenter gastrointestinal trials..Drug Inf J, 2001.35:p.745-753.

[6] 윤청, 소프트웨어 개발 방법론. 2000.

[7] IBM(2005) IBM Workplace Forms.

[8] R., K., DataStandsrds for Clinical Development -Progress Update APPLIED CLINICALTR IALS ,2002: p.35-44.

[10] W3C, .http://www.w3.org/XML/Schema

[11] CDISC, Study Data Tabulation Model Implementation Guidem, 3.1.2 Final. 2008