Data Standards: What Are They and Who Needs Them?

NIH Data Standards Forum
October 19, 2012

Dr. Carolyn Compton, M.D., Ph.D.
Chief Executive Officer
Critical Path Institute
Topics for the Data Standards Forum

- What are data standards for clinical studies?

- What kinds of things can be done with standardized data sets vs. unstandardized data sets?

- What kinds of bias or limitations are introduced when using unstandardized data sets?

- What kinds of unique data sets could be constructed if built on standardized data (opportunities)?

- What is the value proposition for data standards in clinical studies?
C-Path: What We Do

DEVELOP “STANDARDS”

- **Measurement standards**
  - Molecular biomarkers for toxicity, efficacy and patient stratification
  - Imaging biomarkers for efficacy and stratification
  - Patient-, observer-, clinician- reported outcomes

- **Methods standards**
  - Disease models and clinical trial simulation tools
  - *In vitro* models

- **Data standards**
  - With CDISC, clinical data standards for therapeutic areas

ACQUIRE REGULATORY QUALIFICATION

- Recognition, endorsement for a given context of use
Act as trusted neutral third party

Convene consortia of industry, academia, and government for pre-competitive collaboration
- The best science
- Shared risk and costs

Iteratively involve FDA in the development process
- Regulatory participation, guidance
- Official recognition through “qualification” of Drug Development Tools
  - DDTs = biomarkers, clinical outcome assessments, *(animal models)*
What Are Data Standards?

- Data standards provide the rules for structuring information, so that the data entered into a system can be:
  - Reliably read
  - Sorted
  - Indexed
  - Stored and retrieved
  - Communicated between systems
  - Pooled

- They promote the consistent recording of information and are fundamental to the efficient exchange of information.

- They help protect the long-term value of data.
What Are Data Standards?

- **Data standards address structure:**
  - Defines how data cataloged in a consistent, logical manner

- **Data standards address content:**
  - Defines the allowable values for each piece of data
  - Specifies what terminology / dictionaries are used

- **Data standards are *not* clinical protocol standards...**
  - They do not tell you how to conduct an assay nor administer a clinical test or trial; they define how to categorize the results and relevant parameters
Who needs to care about standards?

- Anyone who designs data collection instruments
- Anyone who analyzes data
- Anyone dealing with or pooling data from multiple sources
- Anyone developing an application or information system to review/manage/analyze/compile data
- Anyone who is responsible for the operations, interoperability and security of information systems
What are the desired characteristics of a clinical data standard for research?

- Freely available, not proprietary
- Developed and maintained through expert consensus
- Published by a standards development organization (example: ISO quality standards)
- Compatible with any information technology system
- Implemented widely with an established track record
An available solution: CDISC

One example:
Clinical Data Interchange Standards Consortium Study Data Tabulation Model (CDISC SDTM)
What are the components of the CDISC SDTM standard?

- **Study Data Tabulation Model (SDTM)**
  - SDTM foundational standard: Organizational framework for clinical data
  - Therapeutic Area Supplements expand SDTM into specific disease areas

- **Controlled Terminology**
  - Defines allowable values for each SDTM
Study Data Tabulation Model (SDTM)

Provides a general framework for describing the organization of the information collected during a study

- **Domains**
  - Collection of observations on a particular topic
  - Examples include: Demography, Laboratory Test Results, Clinical Events, Medical History, etc.
  - Groups of logically related domains are grouped into “classes”

- **Standard variables within each domain**
  - Demography: **BRTHDTC**=Birth date
  - Laboratory Test Results: **LBSPEC**=Specimen Type
  - ....
Why is controlled terminology important?

When many different terms are used to describe identical or similar concepts, it becomes very difficult to combine data.

- Dates: MM/DD/YYYY vs. YYYY/MM/DD
- Frequency: QD vs. Daily
- Gender: M/F vs. male/female vs. 0/1 vs. 1/2 etc.

CDISC collaborates with the NCI Enterprise Vocabulary Services team to develop and maintain terminology used in CDISC data models.
Example of CDISC data standard development for Tuberculosis

- Based on work done by a National Institute of Health Roadmap project team*
  - Tuberculosis Data Standards Project
  - Defined clinically vetted data elements for TB

- A working group in the Critical Path to TB Drug Regimens (CPTR) consortium at C-Path developed the CDISC SDTM Therapeutic Area supplement for TB
  - C-Path / CDISC / Duke / FDA / NCI-EVS / CDC collaboration
  - Result was a formally published CDISC therapeutic area standard for TB

*NIH Roadmap Project for Tuberculosis Principal Investigator: Carol Dukes-Hamilton, Duke University. Stakeholders, Companies and Institutions: Aeras Global TB Vaccine Foundation, Center for Disease Control and Prevention, CDISC, Division of TB Control, Refugee and Migrant Health, Duke University Medical Center, Food and Drug Administration, Foundation for Innovative New Diagnostics (FIND), Global Alliance for TB Drug Development, Health Level 7 (HL7), KNCV, National Cancer Institute, National Heart, Lung & Blood Institute (NHLBI), National Institutes of Health (NIH), National TB Controllers Association, Quintiles, WHO’s Stop TB Partnership
How data elements become SDTM standards: TB example

Data Element: Phase of TB treatment

Data Element: TB Symptoms

Data Element: Tuberculin Skin Test Result
Definition: The number of millimeters in diameter of the induration, or raised hardening, at the tuberculin skin test site.
Permissible value set: mm of induration.

~139 Data Elements

Model to CDISC domains

CDISC Domains

Skin Response (SR)
Clinical Events (CE)
Exposure (EX)

CDISC Variables

USUBJID | EXTRT | EXDOS | EXDOSU
---------|-------|-------|-------
USUBJID  | CETERM | CEPRESP | CEOCCUR

Examples of Controlled Terminology

USUBJID | SRTESTCD | SRTEST | SORORRES | SORORRESU
---------|----------|--------|----------|----------
12345    | INDURDIA | Induration Diameter | 16 | mm

Tuberculosis SDTM User Guide

Examples compiled into user guide
Published CDISC Standard

Tuberculosis Therapeutic Area Supplement to the Study Data Tabulation Model User Guide

Prepared by the Critical Path to TB Drug Regimens (CPTTR)

v1.0 published June 29 2012
All documents are publically available

**SDTM**
http://www.cdisc.org/sdtm

**TB Terminology**
http://www.cdisc.org/terminology

**TB User Guide**
http://www.cdisc.org/therapeutic

**Published Controlled Terminology**
http://www.cancer.gov/cancertopics/cancerlibrary/terminologyresources/cdisc
How has C-Path applied CDISC data standards?
C-Path CAMD: Developing tools to Advance Effective Treatments for Alzheimer’s and Parkinson’s Disease

- Qualify biomarkers (FDA Draft Guidance 2010)
- Develop common data standards
- Create integrated databases of clinical trial data
- Develop “accepted for use” quantitative disease models

The first CDISC therapeutic area data standards were developed for Alzheimer’s disease, published September 2011

Nonmember participants: Academic key opinion leaders, CROs
Nine member companies agreed to share data from 22 trials.

The data were not in a common format.

The data needed to be combined in a consistent manner.

All data were remapped to the CDISC AD standard and pooled.

A new in silico modeling tool was created through the application of data standards and is under review by the FDA.
What Was Learned?
ADAS-Cog Variability

- Cognition tests are used to assess Alzheimer’s patients
- Patients are asked to perform a set of tasks
  - Word recall
  - Follow a series of commands
  - Naming of objects
  - ....
- Different implementations of the test were found
  - Different number of questions
  - Different order of questions and tasks
  - Different scoring of same item
- These differences were identified and reconciled as a result of the Alzheimer’s data standards and mapping project
Sources of Data for Building AD Model: Integration from Diverse Sources

- **Natural History**
  - Inter-patient variability
  - Patient specific factors
  - Imaging and CSF biomarkers

- **Placebo Effect**
  - 9 trials, 3223 patients
  - Inter-patient variability
  - Patient Specific Factors

- **Treatment Effect**
  - Estimate data on drug treatment effects (magnitude, onset, offset)
  - 73 Trials (1990 to Present)
  - Inter-study variability
Model Allows for Accurate Quantitative Predictions of Defined Patient Populations

10 year prediction of disease progression as a function of baseline MMSE scores

Mean (line) and 90% Credible Intervals (gray shaded area)

65 year old males
non ApoE4 carriers

Mild
Moderate
Severe
The CAMD Data Challenge

Key Insights Gained

- Legacy data conversion is resource intensive but worthwhile for specific projects

- Assurance is needed that a specific dataset will be useful in achieving research/regulatory qualification objectives

- Selectivity is beneficial: convert only the needed data

- New insights can be obtained from data converted to a common standard and aggregated to enable queries and analysis

- Addition of standardized data from other sources (prospective, retrospective) becomes simplified and expands the power and utility of a standardized data resource
Preferred Approach

1. Define research goal
2. Identify data needed
3. Apply data standards
4. Pool data
5. Develop new research tools
Drug Development Pipeline: Applicability of Data Standards

Primary application of CDISC clinical data standards

http://www.nature.com/horizon/chemicalspace/background/odyssey.html
Clinical Terminology Standards (Section XII E pg 28):

Using a public process that allows for stakeholder input, FDA shall develop standardized clinical data terminology through open standards development organizations (i.e., the Clinical Data Interchange Standards Consortium (CDISC)) with the goal of completing clinical data terminology and detailed implementation guides by FY 2017.

FDA Priorities for Therapeutic Area Data Standards

Priority Therapeutic Areas for Development

An initial inventory of data standards needs, resulted in the identification of 57 therapeutic areas prioritized into three tiers[1]. Further standardization of clinical study data specific to these and other therapeutic areas will facilitate the evaluation of medical products. To identify the preliminary priority areas several factors were considered: (1) areas of particular need, (2) areas with existing data standardization projects underway, and (3) areas with greater drug development pipeline activity. We encourage interested stakeholders to engage in and, whenever possible, sponsor these data standardization efforts.

Priority Disease/Domain Areas for Data Standardization

<table>
<thead>
<tr>
<th>Tier 1</th>
<th>Tier 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acne</td>
<td>Pain*</td>
</tr>
<tr>
<td>Alzheimer’s Disease*</td>
<td>Parkinson’s Disease*</td>
</tr>
<tr>
<td>Anti-diabetic agents*</td>
<td>Prevention of pregnancy</td>
</tr>
<tr>
<td>Crohn’s Disease</td>
<td>Psoriasis</td>
</tr>
<tr>
<td>Infections of skin and/or subcutaneous tissue</td>
<td>QT Studies</td>
</tr>
<tr>
<td>Oncology: time to efficacy event other than overall survival*</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td></td>
<td>Urinary tract infections</td>
</tr>
<tr>
<td>Tuberculosis*</td>
<td></td>
</tr>
</tbody>
</table>

TransCelerate has defined five specific initiatives, one is focused on data standards, working with CFAST
Enable pooling of data that can generate new insights and surface subtle signals

When used from the start, reduce the time and thus cost of conducting clinical studies

Improve the efficiency and effectiveness of regulatory reviews
C-Path thanks the Food and Drug Administration and Science Foundation Arizona for their significant funding of our work.