CDASH and SDTM: Why do We Need Both Standards?

Shannon Labout, Interim Chief Standards Officer, CDISC
Kit Howard, Director Education, CDISC
Learning Outcomes

- Define SDTM and CDASH
- Identify how SDTM and CDASH are the same
- Discuss how and why they are different
- Explain why it is best to use both
Section 1: Define SDTM and CDASH
SDTM/IG

Defines predictable representation of collected data

- Logically groups data into topics
- Naming conventions, data types, controlled terminology
- All clean collected data plus some derived

Supports

- Data aggregation for submissions, data warehouses, registries
- Regulatory review, IRB reporting, safety surveillance
- Hypothesis generation
CDASH/IG

Data Collection & Management

- Consistent CRFs – questions, answers, assumptions
- Ease of use for investigative sites
- Data quality rules that enable SDTM’s quality assumptions

Supports

- Beginning with the end in mind
- Smooth flow of data from collection to submission
- Data integrity and traceability back to the source
Data Flow

CDASH

SDTM

Database Lock

Data Capture

Datasets

Analysis/reviewers’ Requirements

Warehouse Requirements

Investigators’ Requirements
Section 2: Similarities
Variable Overlap: Exact Match

- 34% Not identical
- 66% Identical
Variable Overlap: Exact Match + Standard Mappings

- 89% Identical + mappings
- 11% Not identical
**More Similarities**

<table>
<thead>
<tr>
<th>Same Controlled Terminology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same data assumptions</td>
</tr>
<tr>
<td>- Only include IE criteria not met</td>
</tr>
<tr>
<td>- DA is for tracking controlled materials, not for treatment compliance</td>
</tr>
<tr>
<td>Same organization of data into domains</td>
</tr>
</tbody>
</table>
Section 3: Differences

Philosophy drives design
SDTM assumes that if there is no record then nothing happened. This only works if it was checked in data capture, which requires a question and record (e.g., Were there any AEs?)
Human vs Machine Readable Data

**SDTM**
- USUBJID: ABC1201
- HOTERM: Hospitalization
- HOSTDTC: 2011-08-06
- HODUR: P6H

**CDASH**
- SUBJID: 1201
- HOTERM: Hospitalization
- HOSTDAT: 06-AUG-2011
- HOCDUR: 6
- HOCDURU: HOURS

**Rationale**
SDTM machine-readable formats for variables such as dates are good for data reusability but are not user-friendly for data capture. There is more chance for error when people record data in unfamiliar formats.
Each CRF should have the data that makes sense to collect together. Per CDASH, this can mix domains if standard variable names are used, and in the end the data appear in the right domains.

In SDTM, data MUST appear only in the correct domain.
**Horizontal vs Vertical Data**

**SDTM**

Findings data must be in a normalized or vertical structure; answers are already known.

**CDASH**

Findings data may have to be horizontal for each test to have its own code list; SDTM CT is used for variable names & CRF prompts.

---

**Rationale**

In normalized data, each test is on a different record and may need different controlled terms, called “value-level metadata,” which cannot be modeled in most data capture systems; some EDC systems can’t handle normalized data at all.

---

**EDC CRF**

<table>
<thead>
<tr>
<th>Hamilton Depression Scale 21 Items (HAM-D-21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. DEPRESSED MOOD (Sadness, hopelessness, help-</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>4. INSOMNIA EARLY</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>5. INSOMNIA MIDDLE</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

**CDASH**

<table>
<thead>
<tr>
<th>SUBJID</th>
<th>HAMD101</th>
<th>HAMD104</th>
<th>HAMD105</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Absent</td>
<td>No difficulty falling asleep</td>
<td>No difficulty</td>
</tr>
</tbody>
</table>

**SDTM**

<table>
<thead>
<tr>
<th>USUBJ ID</th>
<th>RTESTCD</th>
<th>RTEST</th>
<th>RSORRES</th>
<th>RSSTRE SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2324- P0001</td>
<td>HAMD101</td>
<td>HAMD1-Depressed Mood</td>
<td>Absent</td>
<td>0</td>
</tr>
<tr>
<td>2324- P0001</td>
<td>HAMD104</td>
<td>HAMD1-Insomnia Early - Early Night</td>
<td>No difficulty falling asleep</td>
<td>0</td>
</tr>
</tbody>
</table>
## Metadata Content

### CDASH Metadata

<table>
<thead>
<tr>
<th>Data Collection Scenario</th>
<th>CDASHIG Variable</th>
<th>Question Text</th>
<th>Prompt</th>
<th>CRF Completion Instructions</th>
</tr>
</thead>
</table>

### SDTM Metadata

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Label</th>
<th>Type</th>
<th>Controlled Terms Codelist or Format</th>
<th>Role</th>
<th>CDISC Notes</th>
</tr>
</thead>
</table>

### SDTM

Metadata includes variable labels and roles

### CDASH

Metadata includes question text/prompt, CRF completion instructions

### Rationale

SDTM focuses on tabulating data. CDASH addresses data capture needs, designed to ensure clear questions that produce consistent answers. Because they have different purposes, some of their metadata differs.
CDASH uses the same variable names and assumptions as SDTM except where data capture needs require a different approach. CDASH also provides mapping algorithms to SDTM.

SDTM
Provides traceability back to an SDTM-variable annotated CRF

CDASH
Structure can provide link to source data at the site
Some variables can only be derived after studies are complete, e.g., USUBJID, STRESC/STRESN, EXDOSE.

USUBJID is unique across a submission; STRESC & STRESN standardize findings results across the submission; In doubleblind studies, treatment dose is unknown until the blind is broken; SDTM primary focus is pooling data for submission; CDASH primary focus is individual studies.

Rationale:

<table>
<thead>
<tr>
<th>USUBJID</th>
<th>EXTRT</th>
<th>EXDOSE</th>
<th>EXDOSU</th>
<th>EXDOSFRQ</th>
<th>EXSTDTC</th>
<th>EXENDTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC1201</td>
<td>BestDrug</td>
<td>350 mg</td>
<td></td>
<td></td>
<td>2011-08-06</td>
<td>2011-09-01</td>
</tr>
</tbody>
</table>
Some Benefits of Using Both

- Ensure asking the same question using the same answer lists as analyzing
- Optimize site data requirements and structure for transmission and analysis
- Supports traceability back through data collection
- Minimizes programming and validation resources and increases quality when transferring data from capture to tabulation
- Can address the points in all the previous slides
- Helps to “future-proof” the data for warehousing
Unlocking Cures