

Data Standards for eCOA in Clinical Trials

Terek Peterson, Vice President Customer Experience and Data Science, YPrime

Frank Menius, Data Scientist, YPrime



Agenda

- 1 Intersection of eCOA and CDISC standards
- 2 Benefits of implementing collection standards
- 3 Observed forms of data collection standards
- 4 Technology selection
- 5 Uncommon roles for standardized eCOA deployment

Today's Discussion: Concepts

- eCOA benefits to patients
 - Simplification
 - Improved compliance
- Data standards boost
 - Interoperability
 - Efficiencies
 - Automation
 - Quality
- New technology and roles
 - Open source choices
 - Skilled work force changing

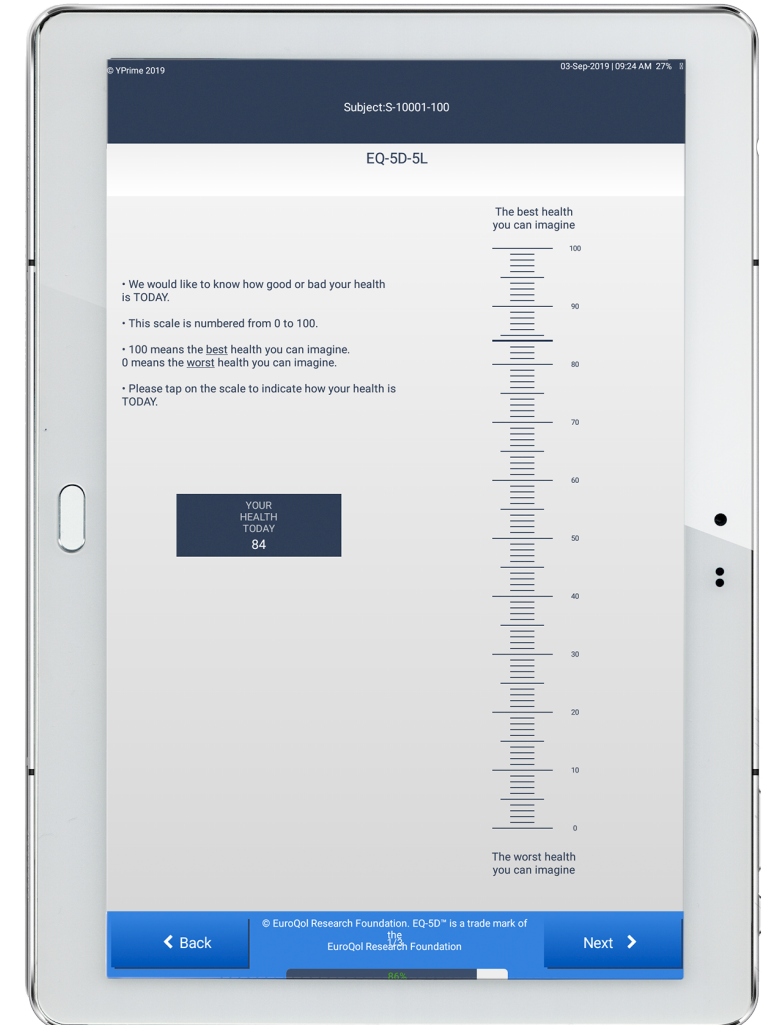
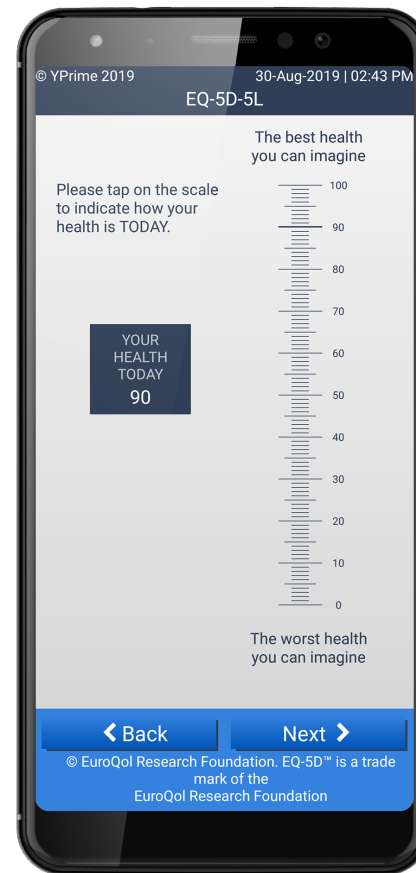


We All Know...

eCOA Motivates Patients and Simplifies Data Collection

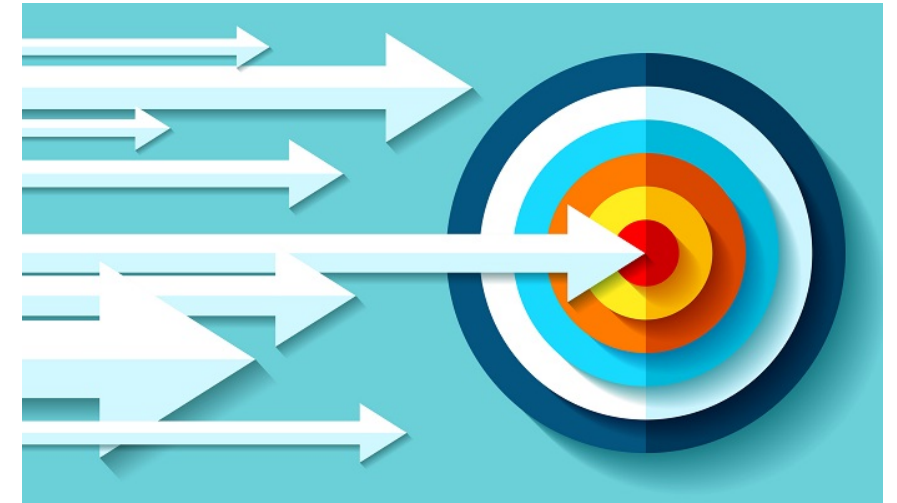
The use of electronic clinical outcomes assessments (eCOAs) ... solutions to capture patient data has been found to improve compliance levels.

This is because electronic devices can be used to motivate and engage patients and because ePROs simplifies the questionnaire process for patients.



Why are Data Standards Important?

- **US and Japan requiring data standards**
 - China and Canada have requested for some submissions
- Research professionals wasting time working with inconsistent data which causes delays and potentially affects patient safety
- Data collected in differing formats requires manual transformations which adds to costs, erodes timelines, and increases likelihood of errors
- Because converting and transforming is problematic, data should be collected in a standardized way too



Why Data Standards?

More than Consistent Formats

- Adoption on the rise, but many knowledge gaps remain about key benefits
 - Higher data quality
 - Efficiency gains
 - Interoperability
 - Regulatory directives

Providing Regulatory Submissions In Electronic Format — Standardized Study Data

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

December 2014
Electronic Submissions

Why Data Standards?

Implementation Challenges

- Rising clinical trial complexity and data volumes present unique challenges
 - Data transformation and management requires experts
 - Companies embracing collection standards differently
 - Additional coordination and training required
 - New skills and roles emerging to support end to end implementation for regulatory submissions



Observed Forms of Data Collection Standards

Like early days of SDTM before requirements

- Near-SDTM
 - SDTM – minus
 - Less than exact SDTM
 - SDTM – plus
 - Additional variables
- SDTM inspired
 - Naming analogous to SDTM
 - Controlled Terminology applied
- Branching standards
 - Copy from previous study
- No Standard
 - Still common



**Study Data Tabulation Model
Implementation Guide:
Human Clinical Trials**
Version 3.2

Prepared by the
CDISC Submission Data Standards Team

Case for CDASH or Near-SDTM

- Essentially implementation of SDTM as the operational standard across the full clinical trial process
- Increases interoperability and data lineage
- Since SDTM model does have a series of derived variables, these are created but not populated
- Use of conformance tools possible



CDISC – QRS Supplements

Clinical Data Interchange Standards Consortium

- Standardized questionnaire supplement
- Terminology that aligns with SDTM
- Clear instructions on how to implement certain questionnaires, ratings, and scales

QS=Questionnaires

QSCAT=EQ-5D-5L

QSEVINTX=TODAY

Under each heading, please check the ONE box that best describes your health TODAY

MOBILITY **QSORRES when QSTESTCD=EQ5D0201**

- I have no problems walking ☐
- I have slight problems walking ☐
- I have moderate problems walking ☐
- I have severe problems walking ☐
- I am unable to walk ☐

SELF-CARE **QSORRES when QSTESTCD=EQ5D0202**

- I have no problems washing or dressing myself ☐
- I have slight problems washing or dressing myself ☐
- I have moderate problems washing or dressing myself ☐
- I have severe problems washing or dressing myself ☐
- I am unable to wash or dress myself ☐

USUAL ACTIVITIES (e.g. work, study, housework, family or leisure activities) **QSORRES when QSTESTCD=EQ5D0203**

- I have no problems doing my usual activities ☐
- I have slight problems doing my usual activities ☐
- I have moderate problems doing my usual activities ☐
- I have severe problems doing my usual activities ☐
- I am unable to do my usual activities ☐

PAIN / DISCOMFORT **QSORRES when QSTESTCD=EQ5D0204**

- I have no pain or discomfort ☐
- I have slight pain or discomfort ☐
- I have moderate pain or discomfort ☐
- I have severe pain or discomfort ☐
- I have extreme pain or discomfort ☐

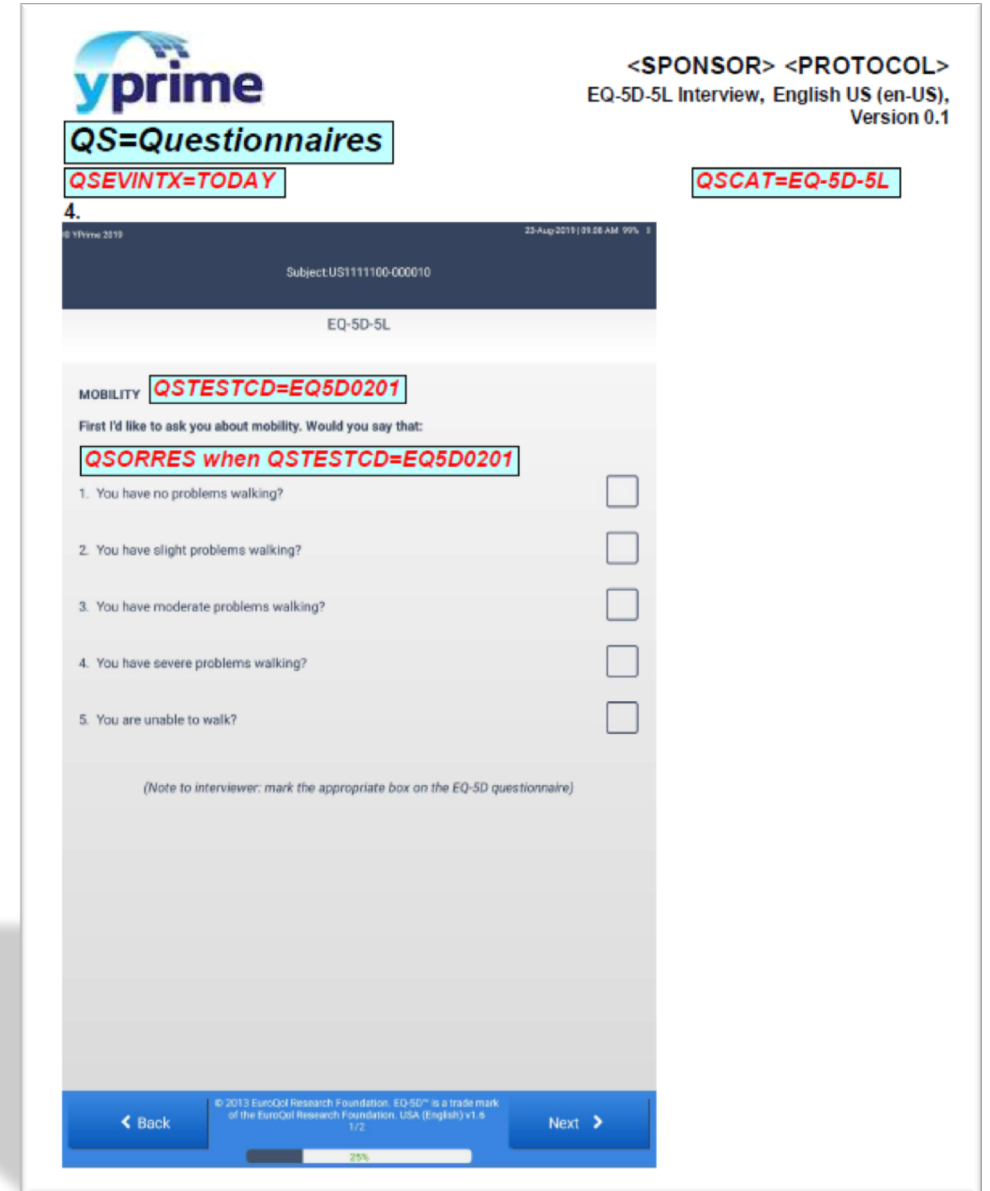
ANXIETY / DEPRESSION **QSORRES when QSTESTCD=EQ5D0205**

- I am not anxious or depressed ☐
- I am slightly anxious or depressed ☐
- I am moderately anxious or depressed ☐
- I am severely anxious or depressed ☐
- I am extremely anxious or depressed ☐

eCOA Implementation Example

Clinical Data Interchange Standards Consortium

- Overlaying CDISC SDTM QRS supplements to YPrime eCOA master screen report
 - Increases communication of metadata across functions
 - Could be used as part of a CRF needed at time of submission



The screenshot displays the YPrime eCOA interface. At the top left is the YPrime logo. At the top right, it shows "<SPONSOR> <PROTOCOL> EQ-5D-5L Interview, English US (en-US), Version 0.1". Below the logo, there are three highlighted boxes: "QS=Questionnaires" (blue), "QSEVINTX=TODAY" (red), and "QSCAT=EQ-5D-5L" (red). The main content area is titled "4." and shows a subject ID "Subject US111100-000010" and "EQ-5D-5L". Below this, the "MOBILITY" section is highlighted with a red box "QSTESTCD=EQ5D0201". The text "First I'd like to ask you about mobility. Would you say that:" is followed by a red box "QSORRES when QSTESTCD=EQ5D0201". There are five questions with corresponding checkboxes: "1. You have no problems walking?", "2. You have slight problems walking?", "3. You have moderate problems walking?", "4. You have severe problems walking?", and "5. You are unable to walk?". A note at the bottom states: "(Note to interviewer: mark the appropriate box on the EQ-5D questionnaire)". The bottom navigation bar includes "Back", "Next", and a progress indicator showing "25%".

Technology Selection

Pinnacle 21 Community (aka P21C)

- By not following SDTM or CDASH, use of industry compliance tools not possible or meaningful

Domain	Variables	Values	Pinnacle 21 ID	Publisher ID	Message	Category	Severity
QS	QSCAT	CDAI	CT2002	CG0021	QSCAT value not found in extensible codelist	Terminology	Warning
QS	QSEVAL	Investigator	CT2002	CG0021	QSEVAL value not found in 'Evaluator' extensible codelist	Terminology	Warning
QS	VARIABLE	QSDTC	SD0057	CG0016	SDTM Expected variable QSDTC not found	Metadata	Warning
QS	VARIABLE	QSDT	SD0058	CG0013	Variable appears in dataset, but is not in SDTM model	Metadata	Error
QS	VARIABLE, LABEL	QSORRES, Result or Finding in Original Units	SD0063	CG0303	SDTM/dataset variable label mismatch	Metadata	Warning
QS	VARIABLE, LABEL	QSSTRESC, Character Result/Finding in Std Format	SD0063	CG0303	SDTM/dataset variable label mismatch	Metadata	Warning
QS	VARIABLE, LABEL	QSSTRESN, Numeric Result/Finding in Standard Units	SD0063	CG0303	SDTM/dataset variable label mismatch	Metadata	Warning
QS	VARIABLE, DATASET	EPOCH, QS	SD1077	FDAB022	Regulatory Expected variable EPOCH not found	Metadata	Error
QS	Variable, Length	QSDRVFL, 0	SD1078	CG0015	Permissible variable with missing value for all records	Presence	Notice
QS	Variable, Length	QSELTN, 0	SD1078	CG0015	Permissible variable with missing value for all records	Presence	Notice
QS	Variable, Length	QSEVLINT, 0	SD1078	CG0015	Permissible variable with missing value for all records	Presence	Notice
QS	Variable, Length	QSGRPID, 0	SD1078	CG0015	Permissible variable with missing value for all records	Presence	Notice
QS	Variable, Length	QSORRESU, 0	SD1078	CG0015	Permissible variable with missing value for all records	Presence	Notice
QS	Variable, Length	QSREASND, 0	SD1078	CG0015	Permissible variable with missing value for all records	Presence	Notice
QS	Variable, Length	QSRFTDTC, 0	SD1078	CG0015	Permissible variable with missing value for all records	P	
QS	Variable, Length	QSSTAT, 0	SD1078	CG0015	Permissible variable with missing value for all records	P	
QS	Variable, Length	QSSTRESU, 0	SD1078	CG0015	Permissible variable with missing value for all records	P	
QS	Variable, Length	QSTPT, 0	SD1078	CG0015	Permissible variable with missing value for all records	P	
QS	Variable, Length	QSTPTNUM, 0	SD1078	CG0015	Permissible variable with missing value for all records	P	
QS	Variable, Length	QSTPTREF, 0	SD1078	CG0015	Permissible variable with missing value for all records	P	
QS	Excess, Variable	1, QSBLFL	SD1082	FDAB018	Variable length is too long for actual data	M	

P21

Technology Selection

Open Source Languages

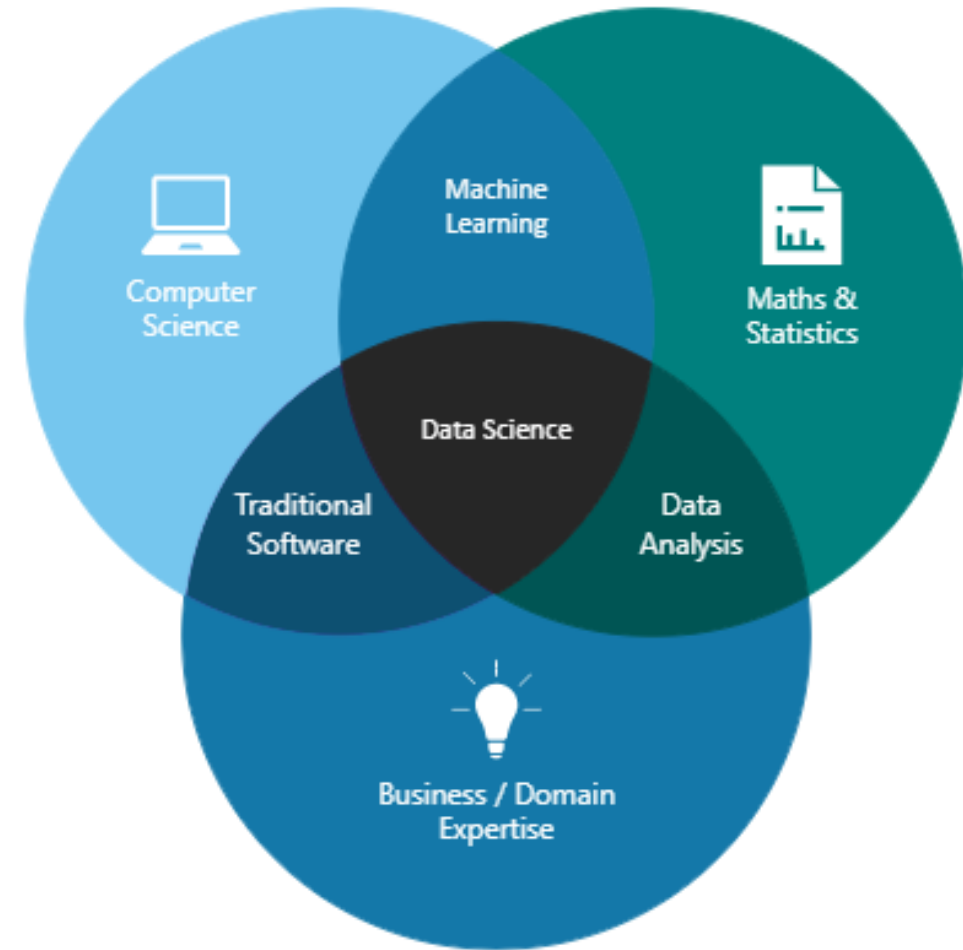
- Open source languages and tools are becoming more common across the data science field
 - R and Python
 - Cutting edge, employing the latest statistical techniques
 - Used to handle Big Data
 - Realtime dashboard and analytics
 - Machine learning possible



Emerging Role: Data Science

The intersection of technology, statistics and business

- Clinical Trial Expertise
- Data Preparation/Integration
- Scientific Rigor
- Collaborative
- Statistics
- Visualization
- Hacker Mindset



The Future

- Even more data
- More integrations
- Automated data monitoring



Closing and Questions



Contact:

Terek Peterson, MBA

YPrime

9 Great Valley Parkway

Malvern, PA 19355

484-560-8960 tpeterson@yprime.com

<https://www.yprime.com>

<https://www.linkedin.com/in/terekpeterson/>

Frank Menius

YPrime 3301 Benson Dr #300

Raleigh, NC 27609

267-536-3268 fmenius@yprime.com

<https://www.yprime.com>

<https://www.linkedin.com/in/frank-menius/>