Paper DH02

Data Standards for eCOA in Clinical Trials

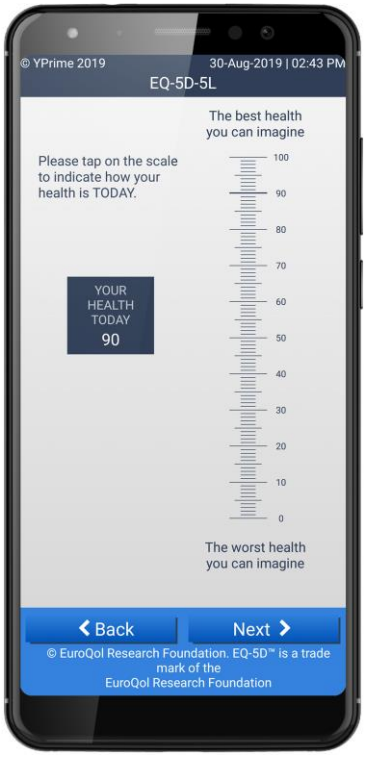
Terek Peterson, MBA and Frank Menius, YPrime LLC, Malvern, PA, USA

ABSTRACT

Adoption of data standards are necessary to underpin higher data quality, efficiencies and integrated applications across increasingly complex clinical research processes. The US Food and Drug Administration (FDA) and Pharmaceuticals and Medical Devices Agency (PMDA) have clearly embraced data standards for the submission of data, laying the groundwork for expanded industry adoption. Collecting data in a standardized way is the next step to reduce messy, time-wasting efforts that can negatively impact patient safety, development timelines and resource utilization. The adoption of data standards for electronic Clinical Outcomes Assessments (eCOA), the fastest growing technology in the eClinical space is critical to ensure quality and efficiency in this rapidly evolving discipline. The paper will discuss the benefits of implementing standards and which forms of data collection standards companies are requesting. In addition, it will explore work in development, technology selection and the uncommon roles needed for a standardized eCOA deployment.

INTRODUCTION

Over the last 20 years the work of the Clinical Data Interchange Standards Consortium (CDISC), a global standards development organization, has made notable progress in creating shareable, end-to-end data standards for clinical and nonclinical research. To date, these standards focus on core principles of data standardization including implementation models, domain definitions and template specifications for structured representation of results. While standards around data collection exist, adoption has been slow to follow. More work is needed, especially in the area of electronic clinical outcome assessments (eCOA). The implementation of eCOA is rapidly expanding, not only because it focuses on measures that are meaningful to patient populations but also because data collection relies on familiar consumer technologies like tablets and smartphones. Patient acceptance and familiarity also boost compliance and data integrity. The surge of eCOA adoption presents an ideal opportunity to apply data standards for even greater efficiencies, quality and interoperability across the clinical research environments. This paper provides observations based on data standards, their use with eCOA implementation, and future consideration. See below examples of devices used for capturing results for the EQ-5D-5L instrument.



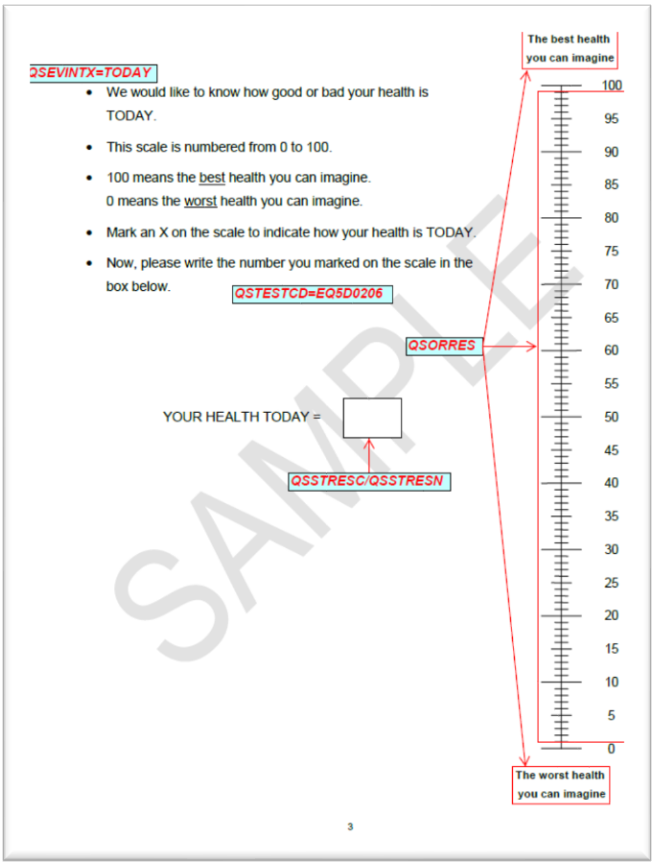
COMMON DATA STANDARDS IN USE

Mature standards created by CDISC already exist, with acceptance by regulatory bodies like FDA and PMDA. How do these standards influence eCOA data collection? Mature standards which are widely used include:

- The Study Data Tabulation Model (SDTM), the structured representation of the collected data and one of the first standards developed, has evolved to not only support clinical data but also medical devices and pharmaco-genomics data. SDTM Implementation Guide (SDTMIG) provides the organization, structure, and format of these standard tabulation datasets. “Flavors” of SDTM (e.g., SDTM+, SDTM-, SDTM-Inspired) are being implemented for eCOA in different ways across organizations that have well-structured standards governance and implementation strategies. Data transfers and metadata are requested in numerous ways, including delimited, .csv, .xpt, .sas7bdat, odm.xml, .json, and .xlsx, requiring further agility from vendors.

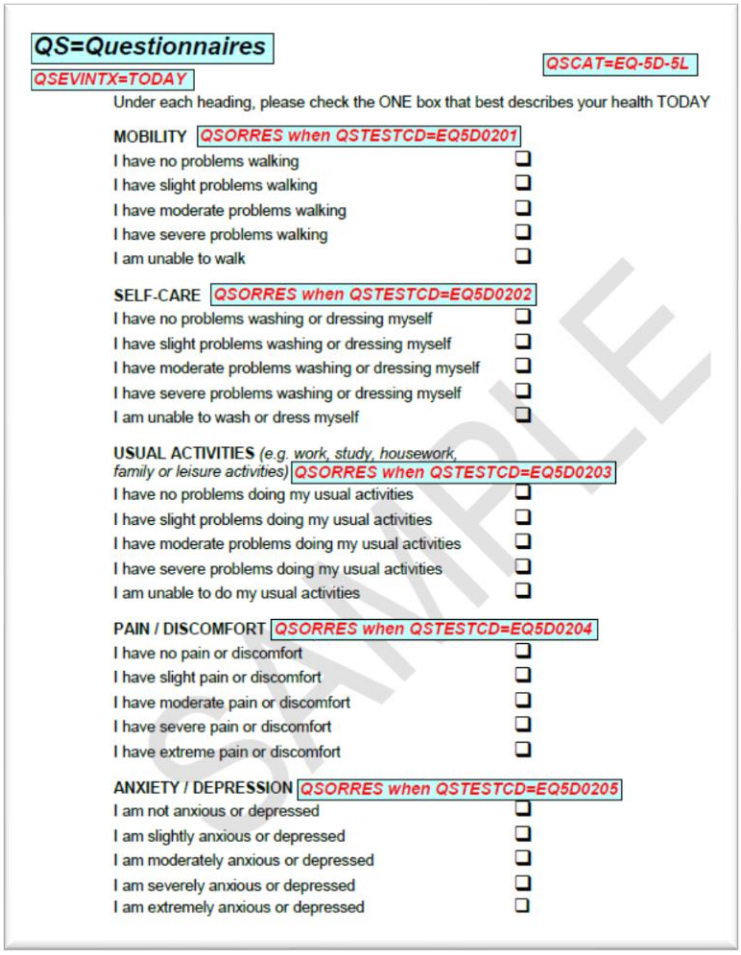
- The Analysis Data Model (ADaM) and the ADaM Implementation Guide (ADaMIG) provide a standard structure for analysis data derived from SDTM to support the creation of analyses, tables, listings, and figures as part of study reporting. As more critical data supporting efficacy endpoints and analysis populations are collected via eCOA devices, implementers cannot forget that devices move with the subject (e.g., work, travel, shopping, camping). Sponsors cannot assume that the subject is always using devices in the same setting, with consistent internet connection, and no distractions. This needs to be considered when developing compliance rules for inclusion/exclusion, analysis populations, censoring, or other analysis criteria.

- Questionnaires, Ratings and Scales (QRS) supplements are vital to clarify how implementation of an instrument should be performed. Each QRS instrument is a series of questions, tasks, or assessments used in clinical research to provide a qualitative or quantitative outcome of a clinical concept or task-based observation. The QRS team develops Controlled Terminology (CT), SDTM supplements, and implementation guidelines with detailed examples. Consistent use of these QRS supplements has not been adopted but is important for data collection groups to implement. The QRS team is working through a backlog of supplements to further this effort, which leads to the need for early and immediate implementation as the list of standards and knowledge needed to maintain these grows. eCOA vendors stand to gain efficiencies by adopting these already developed QRS supplements. See below the sixth question of the EQ-5D-5L from that QRS supplement.



CDISC – QRS SUPPLEMENTS

QRS supplements may be the industry's best-kept secret and is definitely one of CDISC's most under-used resource. Using these guidelines can significantly improve resource utilization and downstream data analysis. The QRS describes how to collect the data, includes variable naming and controlled terminology, helping to avoid re-inventing the wheel with every implementation. The example displayed (EQ5D) is also provided as part of that QRS supplement. When these suggestions are applied to a study's eCOA screen shots it can be provided as part of regulatory requirements to submit an annotated CRF (aCRF.pdf).



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 ❑

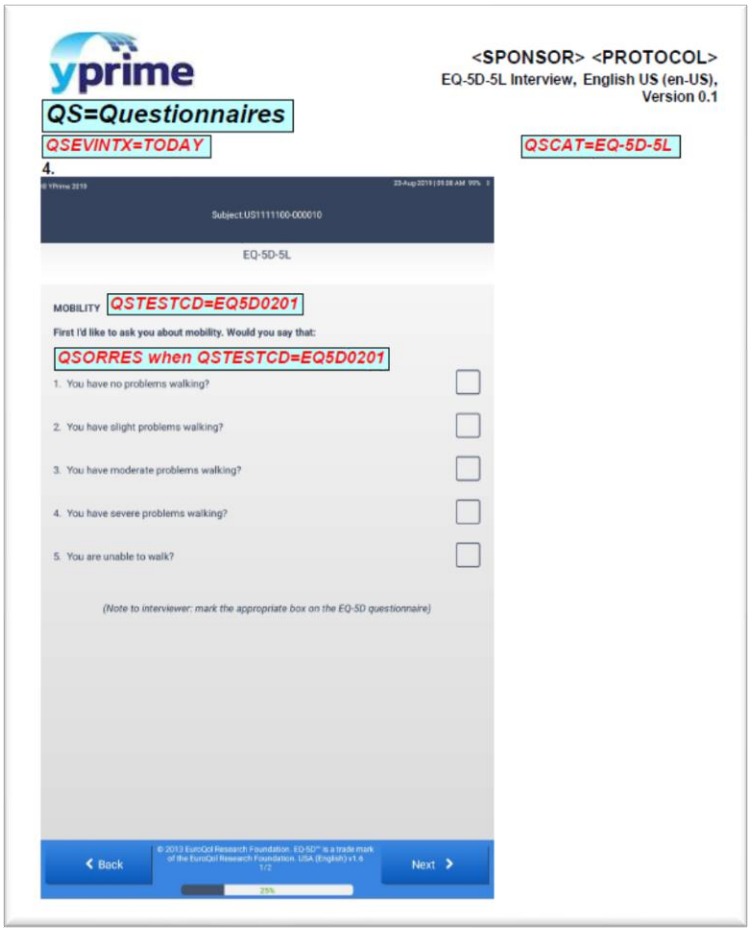
PAIN / DISCOMFORT QSORRES when QSTESTCD=EQ5D0204

- 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 ❑

In the next example, the application of the QRS supplement to a master screen report shows how the flow of the eCOA system will be seen by the user. Not only is this excellent documentation that can be used as part of a CRF to support regulatory submissions, it creates clarity between vendors and sponsor on expectation around data transfers. Since a flavor of SDTM is being used, the creation of fully compliant SDTM is more efficient.



VARIOUS DATA COLLECTION STANDARDS

While skills and roles are emerging to support the industry and regulatory calls for expanded adoption of data standards across the clinical data lifecycle, adoption of data collection standards have been slower than tabulation and analysis standards. Even implementation of these mature standards varies by company, and collection standards vary significantly. The most common observed forms include:

- Near-SDTM
  - SDTM – minus
    - Variable names and labels are as specified in the SDTMIG
    - Missing any kind of derived variables such as SEQ, EPOCH, Day Flags etc. These variables are put on later by the data management CRO or sponsor.
    - Benefit of replication efficiencies
    - Benefit use of industry standard conformance tool such as Pinnacle 21 can be used for validation and conformance checks
  - SDTM – plus
    - Variable names and labels are as specified in the SDTMIG
    - Derived variables are present but set to null
    - No SUPPQUAL variables but additional variables that are then transposed by the data management CRO or sponsor into SUPPQUAL
    - Benefit of replication efficiencies
    - Benefit use of industry standard conformance tool such as Pinnacle 21 can be used for validation and conformance checks

- SDTM inspired
    - Naming and labels are analogous to SDTM but do not match the SDTMIG
        - CAT instead of QSCAT or using the prefix QRS instead of QS
    - Controlled terminology applied
    - Benefit of replication efficiencies
    - Less benefit of standardization. Must use internally developed or custom tool for validation and conformance checks.
- Branching standards
    - Standards are copied from previous study
    - No SDTM conformance
    - Benefit of replication efficiencies
    - No benefit of standardization. Must use internally developed or custom tool for validation and conformance checks.
    - Manual processes
- No standard
    - Still common
    - No benefit of replication efficiencies
    - No benefit of standardization. Must use internally developed or custom tool for validation and conformance checks.
    - Highest risk of errors and lowest efficiency
    - Manual processes

DATA STANDARDS NOT YET COMMONLY ADOPTED

Starting with a solid foundation of standardization of the data, themselves, makes logical sense. However, other available standards which need additional work and are less commonly adopted in eCOA and data collection include the following:

- Protocol Representation Model (PRM) provides a standard for protocols with focus on study design, eligibility criteria, and other requirements from global health authorities. PRM assists in automating CRF creation and electronic health records (EHR) configuration to support clinical research and data sharing. Data collection clearly would benefit from a machine-readable protocol to take advantage of the standards that exist beyond the protocol and automate the creation of collection instruments. There continues to be collaboration from industry on a common protocol template which will become the foundation for the future where machine-readable protocols will lead to efficiencies in data collection and reporting.

- The Clinical Data Acquisition Standards Harmonization (CDASH) collection standard establishes standardized CRF questions, best practices, and controlled terminology that align with SDTM. Data collection formats and structures provide clear traceability of submission data and more transparency for regulators. In addition, the collection of data in CDASH format leads to great efficiencies in the programming resources needed to perform SDTM creation. Unfortunately, these standards do not seem to be utilized with many companies embracing near-SDTM and SDTM inspired company standards. Is the lack of regulatory mandate creating too much interpretation for data collection? Perhaps it is the lack of knowledge of CDISC standards in data management groups, as generally CDISC standards are thought to only apply to statistics and programming groups.

THE CASE FOR CDASH OR NEAR SDTM

Implementing CDASH or near SDTM in the eCOA environment would essentially mean SDTM is implemented as the operational standard across the full clinical trial process. This would bring increased interoperability and data lineage. The use of standards would increase efficiencies due to repetition and standardized specifications and forms. Since the SDTM model does have a series of derived variables, these would be created but not populated. Standardization means that commercially available industry tools can be used for data validation and conformance checks. Standardization also increases the value of the eCOA data transfer as it eliminates the need for the data to be manipulated by the data management CRO.

THE FUTURE

The future involves more data, automation, speed and the formation of a strong data science community to adopt and refine these new tools, collaboratively and transparently. The integration of data standardization is the next step in gaining efficiency, productivity, and cost savings in the eCOA environment.

NEW TOOLS AND APPROACHES

Data standards are the first step into anticipating the needs of the future of clinical research, and the oceans of data involved. As the volumes and sources of data connecting a typical clinical trial continue to expand, the skills needed

to analyze the data and tools used to collect and manage data will continue to rapidly evolve. The expanding role of data science and the use of open source technologies will help usher in these changes. Data science skills combine technology, statistics, business and in-depth clinical research knowledge. Increasingly, the industry is looking to alternative languages that employ the latest statistical techniques, such as R and Python.

METADATA

Management of metadata is an important consideration, especially as more automation is introduced. The spreadsheet approach will be replaced be electronic metadata repositories to which all systems will have access, ensuring accurate processing of data. The use of Application Programming Interfaces (APIs) will ensure that eCOA vendors can easily integrate with internal systems.

CONCLUSION

The adoption of data standards is necessary for eCOA to move to higher data quality, efficiency, and integrated applications. The data collected by eCOA is vital to the clinical trial and ultimately continues with the trial into submission to regulatory bodies. Regulatory bodies are increasingly implementing strict data standards for submissions. Work has been completed within those agencies as well as organizations like CDISC, and PHUSE to provide guidance on data standard implementation. By following that guidance and implementing CDASH and SDTM data standards at collection via the eCOA system, SDTM becomes the operational standard across the full clinical trial process. The eCOA vender is not only eliminating many people hours of rework in the life cycle of the clinical trial and possibly getting drugs to market faster, but they are also adding tangible value to their product deliveries.

REFERENCES AND SUGGESTED READING

QRS/CDISC, https://www.cdisc.org/foundational/qrs.

SDTM/CDISC, https://www.cdisc.org/standards/foundational/sdtm.

SDTMIG/CDISC, https://www.cdisc.org/standards/foundational/sdtmig.

CDASH/CDISC, https://www.cdisc.org/standards/foundational/cdash.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

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/