PharmaSUG 2022 - Paper SI-067

Finding Accord: Developing an eCOA Data Transfer Specification (DTS) ALL Can Agree On

Frank Menius, YPrime, USA
Monali Khanna, YPrime, USA
Vincent Allibone, YPrime, UK

ABSTRACT

When building a study where electronic Clinical Outcome Assessments (eCOA) are collected, the data transfer is often one of the last items considered and due to this can often be a source of headache for data providers and consumers alike. Assumptions are often made by one or more party about aspects of the collected data or their transfer that are easy to confuse or overlook. This paper will not only focus on bringing the topic of the data transfer closer to the beginning of the study design discussion but will also focus on the ins and outs of drafting a complete and concise transfer agreement that will work for all. We will discuss best practices for the industry, items to include or exclude, common sources of issues and errors, and real-world examples of the good and bad of a data transfer. Reusability, machine and human readability, and technical aspects will also be included. It is possible to make an eCOA data transfer clear, quick, easy, and painless. Let us show you how.

INTRODUCTION

Adoption of electronic Clinical Outcome Assessment (eCOA) for clinical trial data collection is happening at a faster rate than ever before with many healthcare organizations and pharmaceutical companies now electing to move away from paper data collection methods. eCOA utilizes technology including computers, tablets, and smartphones to allow patients, caregivers, and clinicians to directly report outcomes, providing high-quality data collection and near real-time insights. eCOA data collection improves patient engagement, compliance levels, and data quality but it does come with some challenges that need to be addressed to ensure that study teams receive data in the format and manor expected for analysis. To ensure that this is the case the eventual data transfer and the data transfer specification (DTS) must move from being a study after thought and be brought to the forefront of study design. From collection to analysis, the data transfer needs to be considered by all stakeholders including eCOA vendor design and configuration specialists, data managers from sponsor or clinical research organization (CRO) teams, standards experts, statistical programmers, and biostatisticians. It is imperative that these groups consider the data transfer early in the study lifecycle and align on expectations.

WHAT IS THE DTS? WHAT DOES IT INCLUDE?

The data transfer specification (DTS) is a document that outlines the agreed upon method that collected eCOA data will be transferred from the eCOA vendor to the sponsor or 3rd party. It details the mechanics of the transfer (the how) as well as detailed descriptions of all variables and values along with any formatting or mapping requirements needed (the what).

LET'S START WITH THE HOW

Any good data transfer specification should start by identifying the sponsor and protocol as well as contact information for signatories and any necessary stakeholders. This will prevent confusion an ensure that direct lines of communication are established for any questions or findings that might arise throughout the study. It is important that this contact information be updated as personnel transition on or off of the study team.

The DTS should identify the general type of transfer. Is this transfer to be an export from the vendor to the sponsor; is it an import of data from a site or 3rd party; or does this specification detail a two-way flow of data? It is also important to note if the transfer is expected to be a cumulative file encompassing all

study data to the time point of transmission or is it to provide incremental or transactional data in a more real time manor?

It is also very important to discuss the file type and mode of transfer early to ensure that capabilities and expectations are aligned and included in the DTS document. What file type is important and can vary from delimited files like .csv or .txt to more structured databases like .sas7bdat and .xpt and may even include other languages and datatypes like .JSON or .XML. It is important to discuss these items early to ensure that all parties have the technical capabilities to provide the files in the type and manor desired or that conversion of the files is possible. Not all vendors or sponsors will have the same technical or programming capabilities.

The DTS should identify how the file is to be transferred. Will this be placed on the vendor or sponsor's sFTP (secure File Transfer Protocol) site? Is it desired that data be emailed or placed directly into a database system like SAS Life Sciences Analytics Framework (LSAF)? What are the account and access credentials needed to place or retrieve the file? Should the file be encrypted? If direct communication between two systems or websites is desired, do both the vendor and the sponsor have the technical compatibility to perform that integration? All these points need to be discussed well in advance to make sure that expectations are met, or future development can take place. When possible, it is suggested that all data transfers be completed automatically eliminating the reliance on a single person or set of people who may take time off, get sick, or simply forget which causes a delay in the transfer of data.

The DTS should establish a when for the data transfer. This should be coordinated to be as specific as possible. It should not simply say "monthly" for example but specify the exact date or day the transfer is desired. It is much clearer to all parties if it specifies "the first Tuesday of each month" or "The 15th of each month". It may even be desired to nail down the transfer to a time, such as "6am the first Tuesday of each month" or "by EOD the 15th of the Month". If any multiple automated systems are being used to transfer files down stream it may be important to communicate any travel delays that may be present. For example, it may be necessary to state that the vendor posts the file on the 15th of the month but the file moves into the sponsor data warehouse on the 16th of the month. Items such as "Upon Request" should be avoided as they are not programmable. If an upon request transfer is needed it should be communicated that it would likely need a set number of days warning to the vendor so they can ensure the transfer is scheduled in time.

The DTS should also indicate if any messages or reports should be sent out to stake holders alerting them that a transfer was posted. If this transfer is taking place in a transactional manor between two systems, is there a provision for load or compliance errors and who should they be sent to? A detailed list of all possible alerts, warnings, or errors should be noted in the DTS with specific instructions on how to handle each.

Finally, the DTS should be sure to document any specific file naming conventions that need to be used and include any differences for test verses production data.

Figure 1 below shows an example of how to capture the mechanics of a data transfer in the data transfer agreement.

Figure 1

Item	Description	Type
1.	**Transfer Types**	☐ Export ☐ Import
2.	**File Will Contain**	☐ Cumulative Data ☐ Incremental Data

Item	Description	Type
3.	**Method of Transfer and File Format** • As a VENDOR standard, data exchanged with external customers must be transferred via a secure method or an encrypted format. • VENDOR does not consider password protected zipped files sent via e-mail over a public domain a secure method of transfer. • If client should choose, however, to use this method of securing data during transfer, VENDOR will accommodate the request.	☐ Email ☐ .CSV (Comma Delimited) ☐ .CSV (Pipe Delimited) ☐ .TXT (Text File) ☐ .XLS (Excel File) ☐ .XPT (SAS Transport) ☐ sFTP ☐ .CSV (Comma Delimited) ☐ .CSV (Pipe Delimited) ☐ .TXT (Text File) ☐ .XLS (Excel File) ☐ .XPT (SAS Transport) ☐ Web Services ☐ XML ☐ Other:
4.	**Frequency**	☐ 1xDaily at <> AM or <> PM ☐ Weekly on <> day ☐ Monthly on the <> of each month ☐ Upon Request ☐ Once Only ☐ Real Time ☐ Other:
5.	**Destination** ☐ URL ☐ Server ☐ Email Address	**<>** ☐ N/A
6.	**Access Credentials** (repeat for any integrated environments as needed)	User Name: Password:
7.	**FTPS Account Info Provided By**	☐ N/A ☐ VENDOR *(Provide access info to Client)* ☐ Client *(Provide access info to VENDOR)*
8.	**Number of Data Files**	<> ☐ N/A
9.	**File protection**	☐ Password protected archive, password: <> ☐ PGP Encryption: <PGP - work with sponsor/VENDOR IT to provide certificate> ☐ Other: ☐ N/A
10.	**Email Notification of File Transfer** **Only sent when files are generated*	☐ Required ☐ Not Required
11.	**Specific Email Subject Line Required** (for flat files sent via email only)	<> ☐ N/A
12.	**Specific Email Body Required** (for flat files sent via email only)	<> ☐ N/A

Item	Description	Type
13.	**Data File Name** (for flat files or if XML data sets via web services require a file name) (repeat for multiple data files as needed)	☐ N/A
14.	**Headers**	☐ Included ☐ Not Included

NOW LET'S CONSIDER THE WHAT

The Data Transfer Specification needs to include information on what is to be included in the data transfer and how it is to be formatted. Ideally this should include all collected data from the eCOA device. Each variable that should be delivered to the sponsor should be explicitly detailed including the Variable Name, Data Format, Length, Description, and Examples or Decodes for values. The DTS should also document where these data are stored in the vendor database for traceability and ease of programming. If possible, this portion of the DTS should be both human and machine readable. Any specific formatting such as date time formatting needs to be explicitly documented. If there are fields that are to be populated a null as a placeholder for programming that needs to be documented as well.

It is recommended that data be collected, stored, and transferred in near Standard Data Tabulation Model (SDTM) format for clear traceability and data linage throughout the study lifecycle. Figure 2 below shows some examples from a near SDTM data transfer. Note that ellipses were used to skip common SDTM variables in the example.

Figure 2

YPrime Data Point (Determined by YPrime database)	**Integration Field Name** (Provided by recipient)	**Data Format Type** (As needed for study)	**Field Length** (As needed for study)	**Field Description** (As needed for study) - 40 character limit	**Examples, Decodes or Comments** (As needed for study)
VENDOR.dbo.StudyCustom.Protocol	STUDYID	☒ Text Char ☐ Integer ☐ Variable Char ☐ Date *(specify format)* ☐ Other	40	Study Identifier	e.g. "ABC123456" ☐ N/A
Hard Code	DOMAIN	☒ Text Char ☐ Integer ☐ Variable Char ☐ Date *(specify format)* ☐ Other	2	Domain Abbreviation	"QS" ☐ N/A
VENDOR.dbo.Patient.PatientNumber	USUBJID	☒ Text Char ☐ Integer ☐ Variable Char ☐ Date *(specify format)* ☐ Other	40	Unique Subject Identifier	e.g. "123-4567" ☐ N/A

YPrime Data Point (Determined by YPrime database)	**Integration Field Name** (Provided by recipient)	**Data Format Type** (As needed for study)	**Field Length** (As needed for study)	**Field Description** (As needed for study) - 40 character limit	**Examples, Decodes or Comments** (As needed for study)
VENDOR.dbo.Site.Site Number	SITEID	☒ Text Char ☐ Integer ☐ Variable Char ☐ Date *(specify format)* ☐ Other	40	Study Site Identifier	e.g. “123” ☐ N/A
Hard Code	QSSEQ	☐ Text Char ☒ Integer ☐ Variable Char ☐ Date *(specify format)* ☐ Other	8	Sequence Number	*Leave Blank*
…	…	…	…	…	…
Vendor Custom Mapping	QSTESTCD	☒ Text Char ☐ Integer ☐ Variable Char ☐ Date *(specify format)* ☐ Other	8	Question Short Name	Refer to Value Level Metadata
Vendor Custom Mapping	QSTEST	☒ Text Char ☐ Integer ☐ Variable Char ☐ Date *(specify format)* ☐ Other	40	Question Name	Refer to Value Level Metadata
…	…	…	…	…	…
Vendor Custom Mapping	QSORRES	☒ Text Char ☐ Integer ☐ Variable Char ☐ Date *(specify format)* ☐ Other	200	Finding in Original Units	Refer to Value Level Metadata
...	…	…	…	…	…
YPrime.dbo.DiaryEntry. DiaryDate/StartDate	QSDTC	☐ Text Char ☐ Integer ☐ Variable Char ☒ Date *(specify format)* ☐ Other	ISO 8601, 19 characters	Date/Time of Finding	YYYY-MM-DDTHH:MM:SS
...	…	…	…	…	…

For variables where there are a large number of values to populate depending on the selections made in the eCOA instrument such as QSTESTCD, QSTEST and QSORRES above, it may be necessary to use

a Value Level Metadata table to map the collected data to the variables desired in the data transfer. Figure 3 below shows how that Value Level Metadata table may look and be used. The example shows the results from one question in the European Quality of Life Five Dimensions Five Level Scale (EQ-5D-5L), a very common instrument for use with eCOA. This example does not show all variables that may have value level mapping but illustrates how mapping changes based on the value selected by the user.

Figure 3

VENDOR	VENDOR	VENDOR	TRANSFER	TRANSFER	TRANSFER
QUESTIONNAIRE	QUESTION	VALUE	QSTESTCD	QSTEST	QSORRES
EQ-5D-5L	Please tap the ONE box that best describes your health TODAY MOBILITY	I have no problems walking	EQ5D0201	EQ5D02-Mobility	I HAVE NO PROBLEMS WALKING
EQ-5D-5L	Please tap the ONE box that best describes your health TODAY MOBILITY	I have slight problems walking	EQ5D0201	EQ5D02-Mobility	I HAVE SLIGHT PROBLEMS WALKING
EQ-5D-5L	Please tap the ONE box that best describes your health TODAY MOBILITY	I have moderate problems walking	EQ5D0201	EQ5D02-Mobility	I HAVE MODERATE PROBLEMS WALKING
EQ-5D-5L	Please tap the ONE box that best describes your health TODAY MOBILITY	I have severe problems walking	EQ5D0201	EQ5D02-Mobility	I HAVE SEVERE PROBLEMS WALKING
EQ-5D-5L	Please tap the ONE box that best describes your health TODAY MOBILITY	I am unable to walk	EQ5D0201	EQ5D02-Mobility	I AM UNABLE TO WALK

Finally, the DTS document should provide example data formatted to its specifications for developers to view. If the above figures 2 and 3 represented a DTS for a pipe delimited .txt file, the example provided in the DTS may look like figure 4 below.

Figure 4

STUDYID|DOMAIN|USUBJID|SITEID|QSSEQ|…|QSTESTCD|QSTEST|…|QSORRES|…|QSDTC|…|

ABC123456|QS|123-4567|123||…|EQ5D0201|EQ5D02-Mobility|…|I HAVE NO PROBLEMS WALKING|…|2021-04-01T14:23:45|…|

LOOKS SIMPLE, WHY DISCUSS AT THE BEGINNING AGAIN?

The technical aspects of the Data Transfer Specification may look simple and straight forward enough to put off until other pressing study priorities are met. It would be easy to see it as just a form that can be filled out at anytime and really shouldn't be an issue before the study has been built and patients are enrolled. The above mentality has been prevalent in eCOA studies to this point and is fraught with issues that can cause costly delays and loss of efficiency for the study. Below is a look at some of the common misconceptions and pit falls that plague data transfers when they are left as an afterthought.

IT TAKES TIME

eCOA data transfers take time to design, program, validate, and approve. The time period from the first draft of the DTS to the actual transfer of production data into a sponsor database can take weeks if not months. Data from vendors can not just be turned on and off like water from a spigot. The data transfer specification has to be designed, values mapped, and approved. Then programmers must develop programming to perform any formatting and mapping specified. Then testers must programmatically test and validate that all programming operates as expected. Then user acceptance testing (UAT) takes place with the sponsor, and finally the production transfer. Any findings or questions that come up anywhere during this process has the potential to delay additional steps as items are reworked. A great deal of time and energy is lost simply in waiting for questions and answers to get routed between different stakeholders in different knowledge silos. Getting these stakeholders together in a meeting early in study development is the key to getting the DTS designed and approved in a timely manor so that development can take place within desired timelines. When the DTS is put off there is a risk that data will not be available to the team for key analysis dates meaning costly delays.

VENDORS NEED TO SEND ALL DATA AS COLLECTED

Industry best practice for eCOA vendors is to send all data as collected. In the long run the eCOA vendor is the raw data collector and must provide raw data to the sponsors and sites. That does not mean that vendors can't provide formatting changes to meet casing or character needs, nor does it mean that vendors can not map multiple variables in the data transfer based off value choices. It also does not mean that vendors do not provide many other services such as on-screen calculations or reports to assist the clinical teams in decision making. What it does mean is when it comes to the data transfer eCOA vendors need to include all collected data in their original form in accordance with collection on the approved system screen reports. In addition, eCOA vendors should not calculate, derive, window, or impute any data in the data transfer. A discussion needs to be had during study design to make sure that there is a clear understanding of what will be collected and transferred in accordance with the approved instrument and what will not.

ECOA IS NOT THE SAME AS PAPER

While eCOA instruments often have their basis in paper questionnaires, that does not make them identical, and the data collected are often not exactly the same. This can also lead to assumptions by one party or another as to what data are collected and can be transferred. For example, one possible difference is the inclusion of sub-setting questions not necessarily captured in the original paper version of the questionnaire. A questionnaire may have had a section that said, 'If you had a headache today, please mark its severity' In the paper version a subject without a headache would just skip that section. In an eCOA version however this may be optimized to ease the burden on the user or to improve data quality by preventing erroneous entries. The eCOA version may have the sub-setting question 'Did you have a headache today?' 'Yes' or 'No'. If the subject chooses yes then they will be presented with the severity question as before, but if they select 'No' then the system will skip to the next suitable question section. This Sub-setting question 'Do you have a headache today?' and the values of 'Yes' or 'No' are collected subject data and should be included in the data transfer. If however on the sponsor side

standard metadata and programming was developed with only the paper entry in mind this sub-setting question would be left out and its inclusion in the data transfer could cause a load or compliance error. This is but one example of many nuanced scenarios where data collected by eCOA devices may differ from the original assumptions of the study team and is another reason to data transfer discussions early in the study life cycle.

WAIT, THESE AREN'T THE DATA I THOUGHT I WOULD ANALYZE

There is nothing worse than working on a data transfer and specification close to an analysis date and having the statistician brought into the conversation only to find out that the collected data are not what the analysis team thought they would be. Now data must be post processed to get them into a usable form in a short amount of time, causing delays and/or cost and scope increases. How can this happen? Imagine the case where the questionnaire asked a subject daily what time they went to bed and what time they got out of bed. These times were recorded in the eCOA device, mapped in the DTS, and transferred to the sponsor as Date/Time values, however the value needed for analysis was sleep duration. It may sound like a simple calculation for a statistical programming team to make but if they already have the SDTM database built, and quality controlled then this is a significant amount of rework. The question can also be asked if the difference between the time a patient went to bed and the time a patient got out of bed can clearly be extrapolated to sleep duration. Is this making too broad of an assumption? Perhaps this situation could have been avoided if the statistician was involved early on in discussions about how data were being collected and would be transferred. Could the questionnaire have been updated to have the patient enter the number of hours that they slept?

OK, WE'LL START THE DTS EARLY, WHAT ELSE CAN WE DO?

In addition to bringing all data transfer stakeholders together, in a meeting and not through email, early in study design, there are a couple other steps that can be utilized to make sure all parties understand the data transfer and that it occurs as smoothly as possible.

PROVIDE DATA STANDARDS EARLY

Many sponsors have internal data standards for regularly used questionnaires, values, or codelists. Other sponsors stick to industry standards like SDTM. Regardless of the type of standards used, if they exist, they should be provided to the eCOA vendor and other stakeholders as soon as possible. If it is known that a certain questionnaire has a standard set of internal metadata or compliance values that will need to be transferred in the data transfer, then the eCOA vendor can design the eCOA instrument to collect and store values in accordance with those standards. This eliminates future mapping as well as any question about including raw data in the data transfer as the raw data already meets the standard.

ANNOTATE THE SCREEN REPORTS

Using the same standard approach and resources as the annotated Case Report Form (aCRF), approved Screen Reports can also be annotated to provide traceability and understanding to data collected by eCOA devices and how those data relate to future datasets and the data transfer. Annotated Screen Reports (aSR) can also be an excellent tool to help build the data transfer specification as they graphically illustrate the relationship between collected subject entered data and transferred values. The annotations can also be used to illustrate how values are stored in vendor databases or intermediate datasets and data warehouses. Figure 5 below shows the EQ-5D-5L mobility question used in Figure 2 above as a screen report, while Figure 6 below shows that same screen report with annotations indicating SDTM values (Blue), Vendor Data Base Storage (Gold), and location in intermediate sponsor data warehouse (Gray). These annotations not only help stakeholders understand the data but also contribute to a clear data lineage.

Figure 5

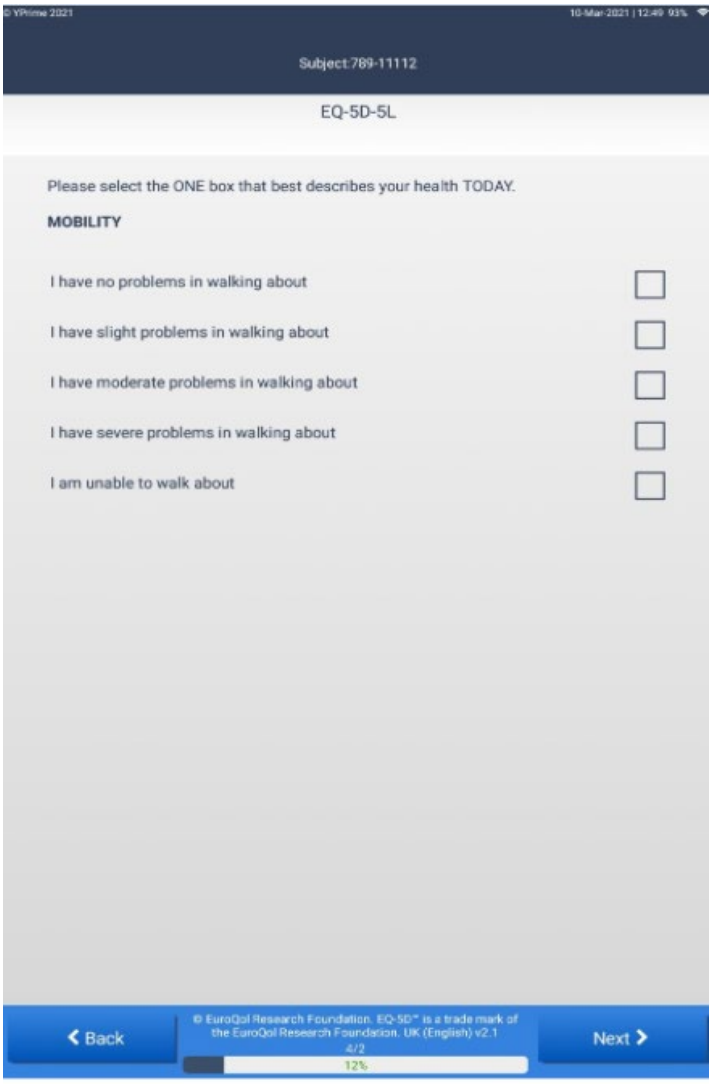
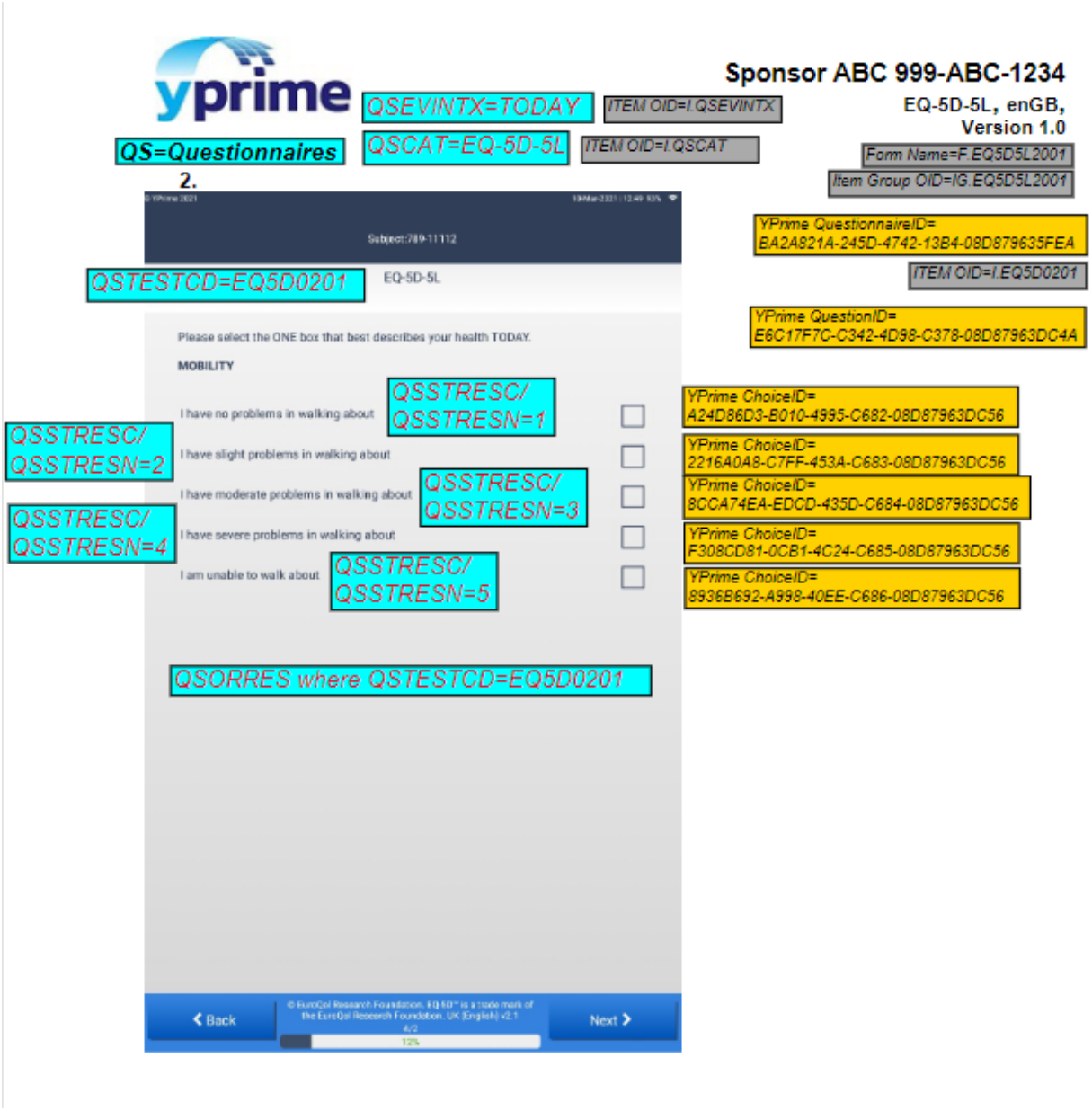

© YPrime 2021
10-Mar-2021 | 12:49 93%
Subject:789-11112
EQ-5D-5L
Please select the ONE box that best describes your health TODAY.
MOBILITY
I have no problems in walking about
I have slight problems in walking about
I have moderate problems in walking about
I have severe problems in walking about
I am unable to walk about
Back
© EuroQol Research Foundation. EQ-5D™ is a trade mark of the EuroQol Research Foundation. UK (English) v2.1
4/2
12%
Next

Figure 6



CONCLUSION

The Data Transfer Specification is a vital piece of study documentation ensuring that data collected by eCOA vendors is transferred in an expected manor to study sponsors and 3rd party partners. It is vital that the data transfer does not remain a study afterthought but becomes an item for discussion and planning at the start of study design. It is important for all stakeholders to come together early to understand the collected data and the data transfer needs as well as insure that the DTS is thorough and complete. By having clear lines of communication early, providing standards at onset of design, and

annotating approved screen reports all parties can have a clear understanding of transferred data without the fallacy of assumptions.

REFERENCES

The Clinical Data Interchange Standards Consortium (CDISC), 2022, *QRS*, Accessed 07MAR2022, < https://www.cdisc.org/standards/foundational/qrs>

The Clinical Data Interchange Standards Consortium (CDISC), 2022, *EQ-5D-5L v1.2 Approved.pdf*, Accessed 07MAR2022, <https://www.cdisc.org/system/files/members/standard/foundational/qrs/EQ-5D-5L%20v1.2%20Annotated%20CRF.pdf>

RECOMMENDED READING

The Federal Drug Administration (FDA), 2022, *Electronic Source Data in Clinical Investigations,* Accessed MAR2022, <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/electronic-source-data-clinical-investigations>

The Federal Drug Administration (FDA), 2022, *Clinical Outcome Assessment (COA): Frequently Asked Questions,* Accessed MAR2022, <https://www.fda.gov/about-fda/clinical-outcome-assessment-coa-frequently-asked-questions >

The Federal Drug Administration (FDA), 2022, *Clinical Outcome Assessment: A Regulatory Perspective,* Accessed MAR2022, < https://heal.nih.gov/files/2020-11/Papadopoulos.%20NINDS%20presentation_9.18.2020_Final_508C.pdf >

The National Health Council, 2022, *FDA's Roadmap to Patient-Focused Outcome Measurement in Clinical Trials*, Accessed MAR2022, < https://nationalhealthcouncil.org/webinars/coa-series-fdas-roadmap-to-patient-focused-outcome-measurement-in-clinical-trials/>

YPrime, 2022, *White Papers*, Accessed MAR2022, < https://www.yprime.com/white-papers/ >

CONTACT INFORMATION

Frank Menius

YPrime

3301 Benson Dr.

Raleigh, NC 27609

919-412-6537

fmenius@yprime.com

YPrime: https://www.yprime.com/

LinkedIn: https://www.linkedin.com/in/frank-menius/

Monali Khanna

YPrime

9 Great Valley Parkway, Malvern, PA 19355

Malvern, PA 19355
mkhanna@yprime.com
YPrime: https://www.yprime.com/
LinkedIn: https://www.linkedin.com/in/monali-k-a259a5178/

Vincent Allibone
YPrime
Sandwich Discovery Park
Sandwich, Kent
vallibone@yprime.com
YPrime: https://www.yprime.com