BACPAC
Back Pain Consortium

Required Documentation Guide for BACPAC Modified SDTM Standard

Version 1.0, July 2021

Prepared by the BACPAC Data Integration, Algorithm Development, and Operations Management Center (DAC)

Table of Contents

1 Metadata Documentation

1.1 Specifications

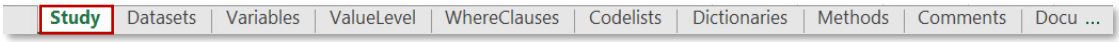
1.1.1 Study Information

Study information is found in the “Study” tab in the specifications file. The values of study name, study description, and protocol name should all be filled in by each research unit to describe the study.

StudyName is the full text name of the study (e.g., Low Back Pain: Biological, Biomechanical, Behavioral Phenotypes) and *ProtocolName* is the acronym of the study (e.g., LB3P).

If a study does not have a formal acronym then (1) *ProtocolName* can be the same as the *STUDYID*, or (2) StudyName and ProtocolName can take the same value (e.g., *StudyName* and *ProtocolName* could be the same for UCSF’s ComeBack study).

StudyDescription is a description of the purpose of the study, as shown in the example below.



Attribute	Value
StudyName	Virtual Reality Therapy for Chronic Low Back Pain
StudyDescription	This study will test the effectiveness of an evidence-based virtual reality therapy program as a non-pharmacological supplement to managing patients with pain due to chronic lower back pain. Outpatients will be randomized to receive one of three virtual reality (VR) programs: skills-based VR, distraction VR, or sham VR.
ProtocolName	VR123
StandardName	BACPAC Modified SDTM Standard
StandardVersion	3.1.2 (Modified)
Language	en

1.1.2 Code Lists

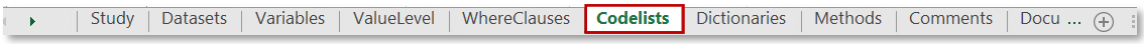
Code lists describe possible variable values. They can be used to assign a value or set of values directly to a variable. For example, the following code list describes possible values of QSCAT from the Minimum Dataset.

ID	Name	NCI Codelist Code	Data Type	Order	Term
QSCAT	Question Category		text	1	Pain Duration and Frequency
QSCAT	Question Category		text	2	Pain Location
QSCAT	Question Category		text	3	Widespread Pain
QSCAT	Question Category		text	4	Pain Somatization
QSCAT	Question Category		text	5	Pain Catastrophizing
QSCAT	Question Category		text	6	Pain Intensity
QSCAT	Question Category		text	7	Pain Interference
QSCAT	Question Category		text	8	Pain Intensity and Interference
QSCAT	Question Category		text	9	Physical Function
QSCAT	Question Category		text	10	Sleep
QSCAT	Question Category		text	11	Depression
QSCAT	Question Category		text	12	Anxiety
QSCAT	Question Category		text	13	Patient Satisfaction
QSCAT	Question Category		text	14	Substance Use
QSCAT	Question Category		text	15	Opioid Use

Code lists can also be used in conjunction with where clauses (see example at end of section) to assign values to variables under certain conditions. For example, the following code list, LIKRTN4A, describes possible values of QSSTRESN when QSTESTCD = GAD01 or GAD02. The Decoded Value represents the corresponding values of QSSTRESC, and has its own code list in the specifications file, LIKRT4A.

ID	Name	NCI Codelist Code	Data Type	Order	Term	NCI T	Decoded Value
LIKRTN4A	Likert Scale 4-Point (A)		integer	1	0		Not at all
LIKRTN4A	Likert Scale 4-Point (A)		integer	2	1		Several days
LIKRTN4A	Likert Scale 4-Point (A)		integer	3	2		More than half the days
LIKRTN4A	Likert Scale 4-Point (A)		integer	4	3		Nearly every day

To add a code list, first navigate to the “Codelists” tab in the specifications file.



To create a code list,

- Assign an appropriate ID and name to the code list.

ID	Name
QSTESTCD	Question Short Name

- Fill in the “Data Type” column with either “text” or “integer.”

Data Type
text

- Fill in the “Order” column. For the first row, order=”1.” Order should increase by 1 with each additional row.

Order
1
2
3

- Enter the first possible value of the variable in the “Term” column.

Term
GAD01

- If needed, add to the “Decoded Value” column. Decoded values are needed for integer code lists if they have corresponding character values, and for QSTESTCD code lists. The decoded values of the QSTESTCD code list should be the matching values of QSTEST.

Term	NCI T	Decoded Value
GAD01		How often bothered by feeling nervous, anxious or on edge

- In the example Minimum Dataset specifications file, the QSTESTCD code list includes values of Category (QSCAT), Subcategory (QSSCAT), Derived (QSDRVFL), and Evaluation Interval (QSEVLNT). These columns are not required for submission, but may be used as a guide and do not interfere with Define XML creation.

Category	Subcategory	Derived	Evaluation Interval
Anxiety	GAD		-P2W
Anxiety	GAD		-P2W
Anxiety	GAD	Y	-P2W

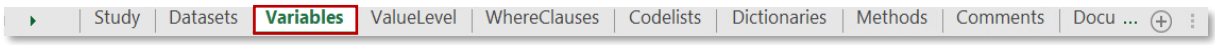
- Repeat the above steps for each additional value of “Term.”

Example code list:

ID	Name	NCI Codelist Code	Data Type	Order	Term	NCI T
QSTESTCD	Question Short Name		text	1	GAD01	
QSTESTCD	Question Short Name		text	2	GAD02	
QSTESTCD	Question Short Name		text	3	GAD2RAW	

Decoded Value	Category	Subcategory	Derived	Evaluation Interval
How often bothered by feeling nervous, anxious or on edge	Anxiety	GAD		-P2W
How often bothered by not being able to stop or control worrying	Anxiety	GAD		-P2W
GAD-2 Raw Score	Anxiety	GAD	Y	-P2W

To add a code list directly to a variable, navigate to the “Variables” tab in the specifications file.



Find the desired dataset and variable, and add the code list ID to the “Codelist” column.

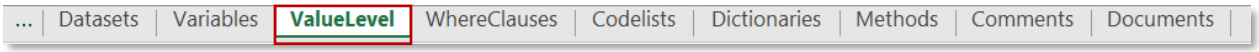
Order	Dataset	Variable	Mandatory	Codelist
5	QSMD	QSCAT	Yes	QSCAT

To add a code list to a variable only under certain conditions, navigate to the “WhereClauses” tab in the specifications file.

Add a where clause describing the conditions.

ID	Dataset	Variable	Comparator	Value
QSGAD_WC1	QSMD	QSTESTCD	IN	GAD01, GAD02

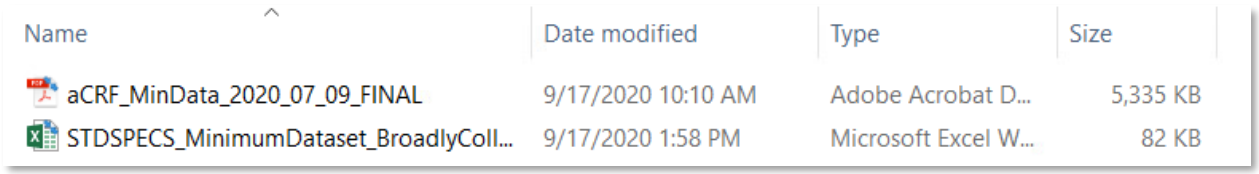
Navigate to the “ValueLevel” tab. Use the where clause created previously to assign the code list to the variable by adding the code list ID to the “Codelist” column.



Where Clause	Description	Data Type	Length	Significant Dig	Format	Mandator	Codelist
QSGAD_WC1	GAD	text	30			Yes	LIKERT4A

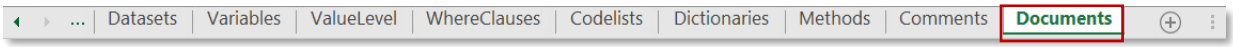
1.1.3 Documentation

The documents described in Section 2 must be provided as PDF files and linked in the Define-XML. To add a PDF, first save the PDF to the same folder as the specifications file.



Name	Date modified	Type	Size
aCRF_MinData_2020_07_09_FINAL	9/17/2020 10:10 AM	Adobe Acrobat D...	5,335 KB
STDSPECS_MinimumDataset_BroadlyColl...	9/17/2020 1:58 PM	Microsoft Excel W...	82 KB

Navigate to the “Documents” tab within the specifications file.

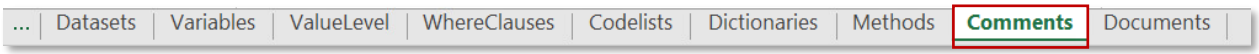


Add a document ID in the ID column. This can be a sequential ID number or a longer ID consisting of letters and/or numbers with a length of 10 or less. Add the title of the PDF to the “Title” column. Add the name of the file, including the “.pdf” file extension, to the “Href” column.

ID	Title	Href
1	Minimum Dataset Annotated Case Report Form	aCRF_MinData_2020_07_09_FINAL.pdf
2	CPAQ Annotated Case Report Form	CPAQ-8_aCRF_V0.3.pdf
3	FABQ Annotated Case Report Form	FABQ_aCRF_V0.3.pdf
4	FABQ-PA Annotated Case Report Form	FABQ_PA_aCRF_v1.0.pdf
5	PSS Annotated Case Report Form	PSS-4_aCRF_v1.0.pdf
6	Description of Complex Algorithms	complex_algorithms.pdf
7	Reasons for Unexpected Data	unexpected_data.pdf
GENDOC	Gender Identity Proposal	Gender_Identity_Proposal_200709.pdf

1.1.4 Comments

Comments may be added to rows in the “Datasets,” “Variables,” and “ValueLevel” tabs. To create a comment, navigate to the “Comments” tab in the specifications file.



Assign the comment an ID in the “ID” column.

ID	Description	Document	Pages
GENIDENT	This is a comment about gender identity mapping in the SC dataset.	GENDOC	

Add the desired text of the comment to the “Description” column.

ID	Description	Document	Pages
GENIDENT	This is a comment about gender identity mapping in the SC dataset.	GENDOC	

The “Document” and “Pages” columns may be left blank. However, if the comment you are adding is long and you would like to attach the comment as a separate document, or you would like to link a document providing further explanation, this can be done using the “Document” and “Pages” columns. Text is still required in the “Description” column in order for the comment to display correctly in the Define-XML. To add a document:

- Follow the steps in Section 1.2.6 to add the document to the specifications.
- Return to the “Comments” tab and fill in the “Document” column with the ID from the “Documents” tab.

ID	Title	Href
GENDOC	Gender Identity Proposal	Gender_Identity_Proposal_200709.pdf

ID	Description	Document	Pages
GENIDENT	This is a comment about gender identity mapping in the SC dataset.	GENDOC	

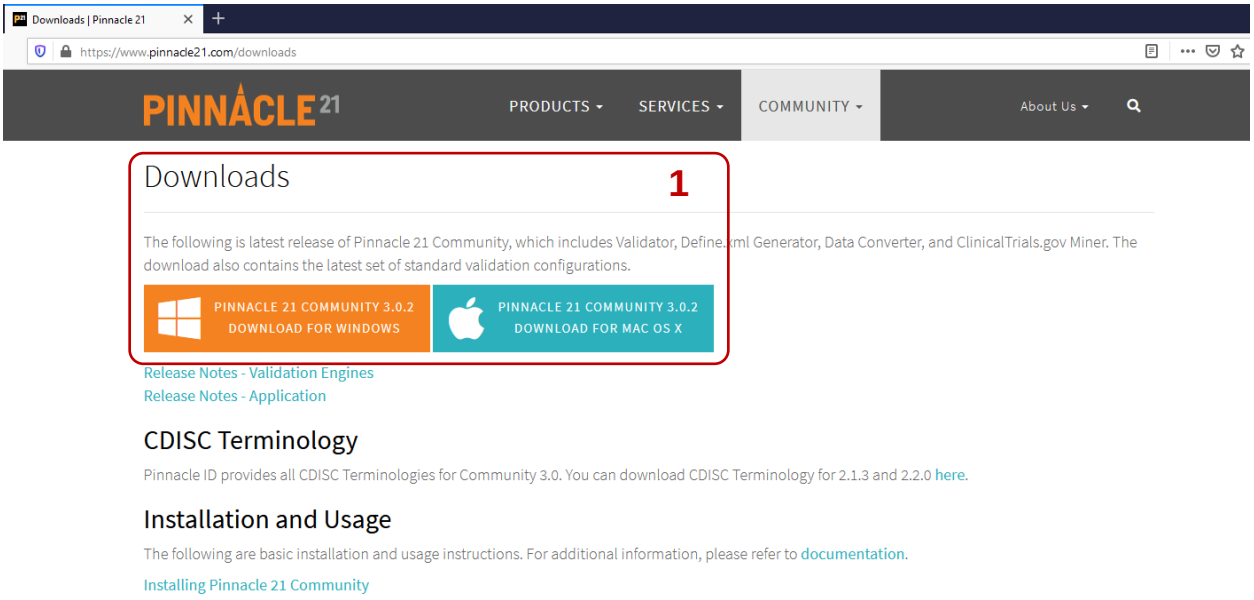
- If there is a specific page in the document you would like to reference, add it in the "Pages" column. To reference more than one page, list the page numbers delimited by spaces. A hyphen may be used to indicate a page range, and the Define-XML will link to both the first and last page.

Document	Pages
GENDOC	
1	5-24
1	1
1	1 2 3 4

1.2 Define-XML

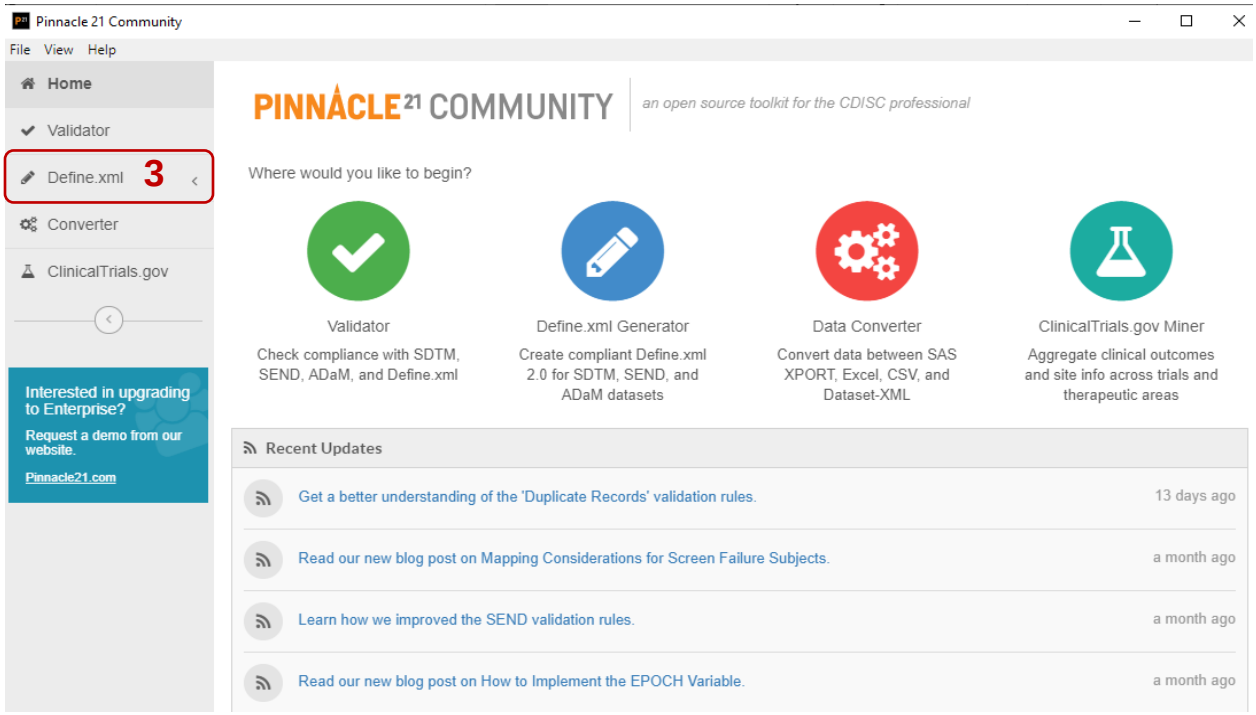
1.2.1 How to Create a Define-XML File

1. Download the version of the Pinnacle 21 Community software for your operating system (Windows or Mac OS X) from the following URL: https://www.pinnacle21.com/downloads

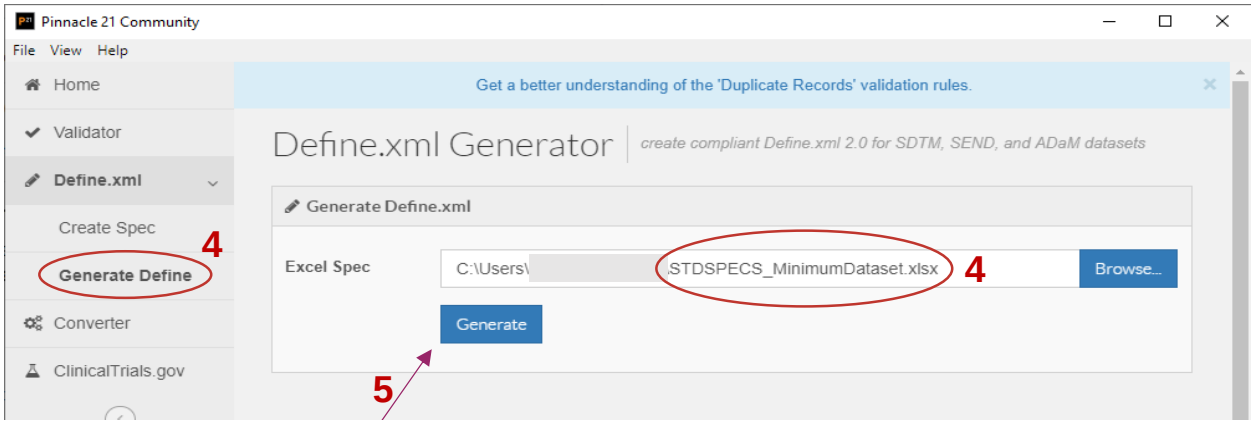


2. Install the software on your computer.

3. Once installed, open Pinnacle 21 Community and click on the "Define.xml" tab.



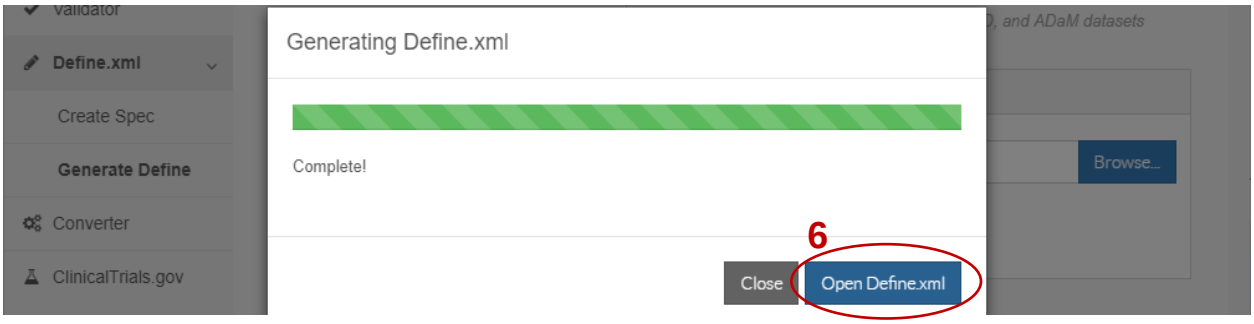
4. Click on the "Generate Define" tab, then click "Browse" to select the specification file.



5. Click on the blue "Generate" button to create the Define-XML file.

   By default, the XML file and associated stylesheet are saved in the folder [*…/Documents/Pinnacle 21 Community/defines*]. You can move or copy the output files to a preferred folder location.

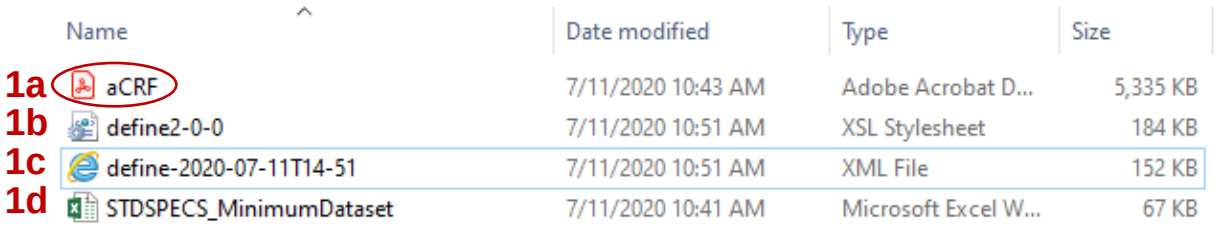
6. To view the Define-XML file, you can click the "Open Define.xml" button. Note: This will only work if the XML file default is to open in your web browser. Otherwise, it will open as a text file.



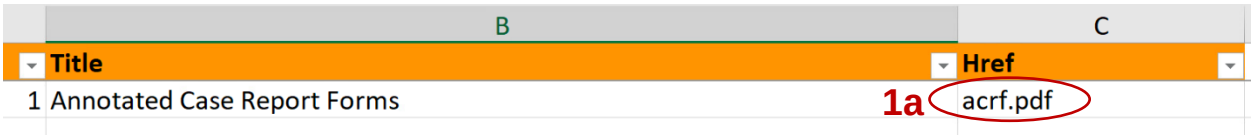
For additional information on viewing a Define-XML, go to Section 1.2.2.

1.2.2 How to View a Define-XML File

1. Confirm all relevant files are saved in the same folder. At minimum, this includes the:
    a. Annotated CRF (PDF),
    b. XSL Stylesheet,
    c. Define-XML file, and
    d. Specifications file (XLSX)

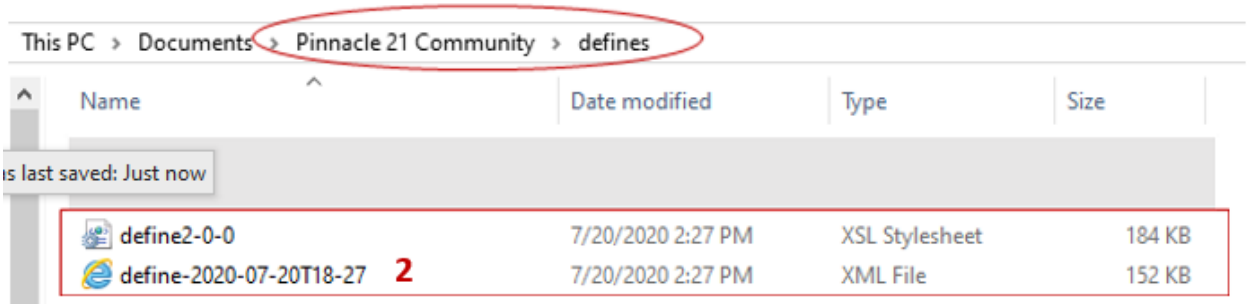


Reminder: The annotated CRF must have the same file name as listed in the specifications!



2. To open and view the Define-XML file, go to the folder where the Define-XML is saved, right click on the XML file, then click "Open with" and select a compatible browser.

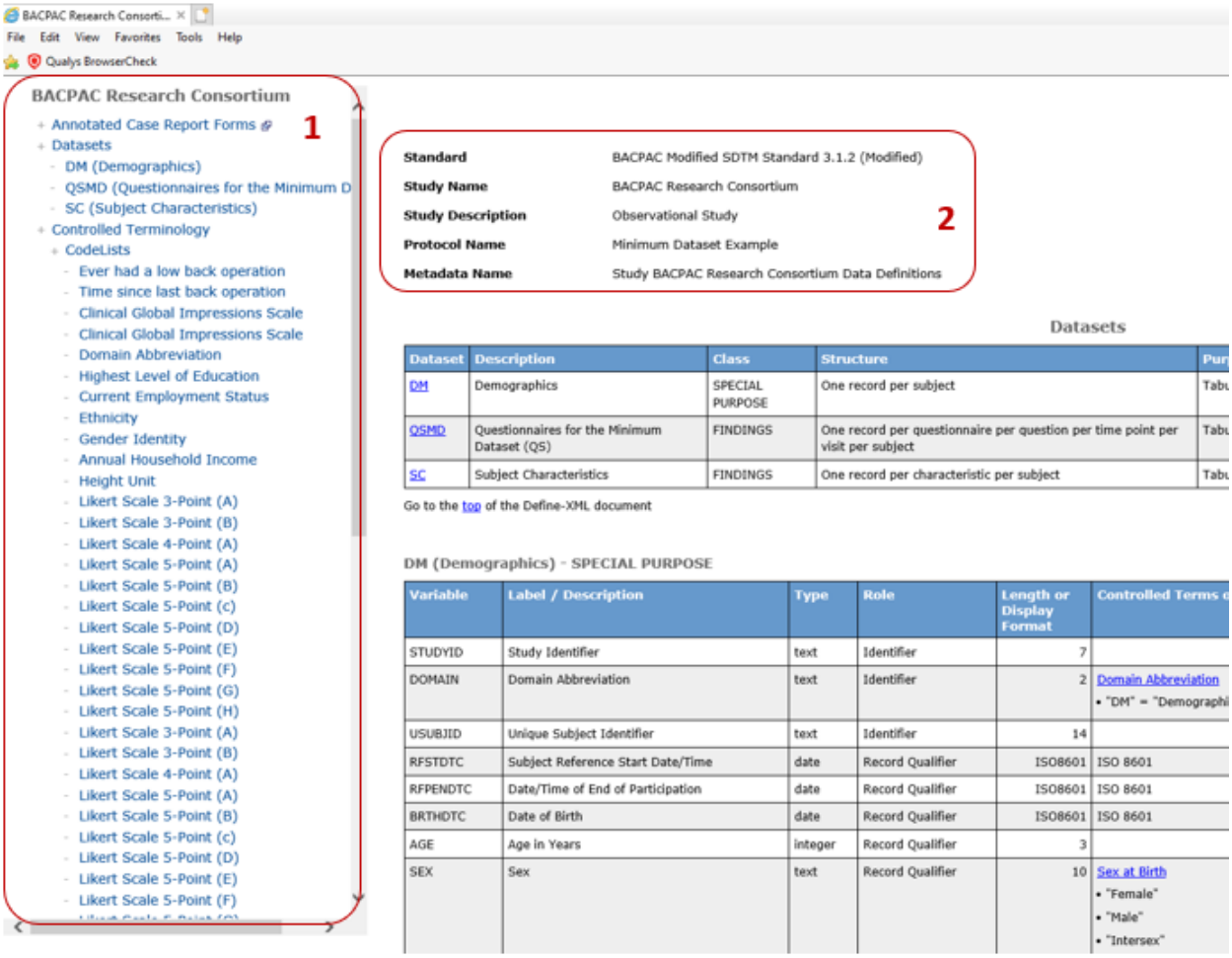
    Note: XML files can be read using Internet Explorer, but other browsers may require an extension.



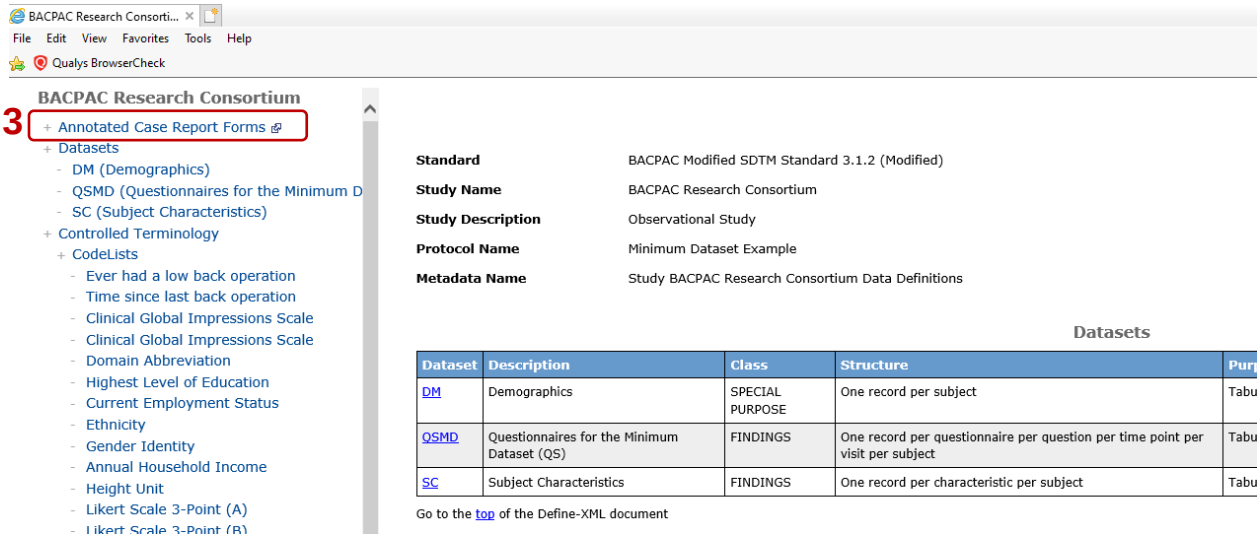
1.2.3 Navigating a Define-XML File

1. A navigation pane is available on the left-hand side of the page. You can scroll through the sections and click on the hyperlink to the different sections.

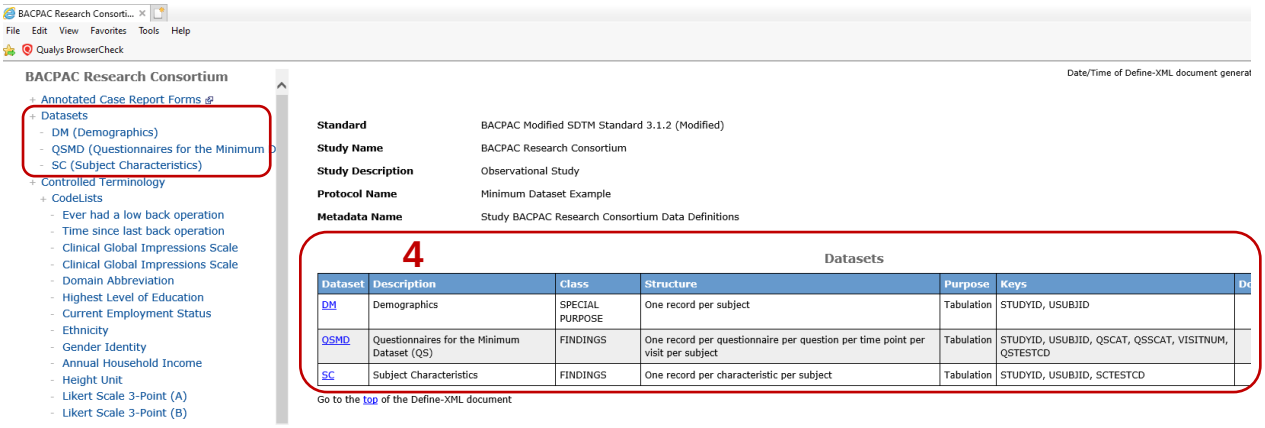
2. Find information describing the study information contained within this Define-XML document at the top.



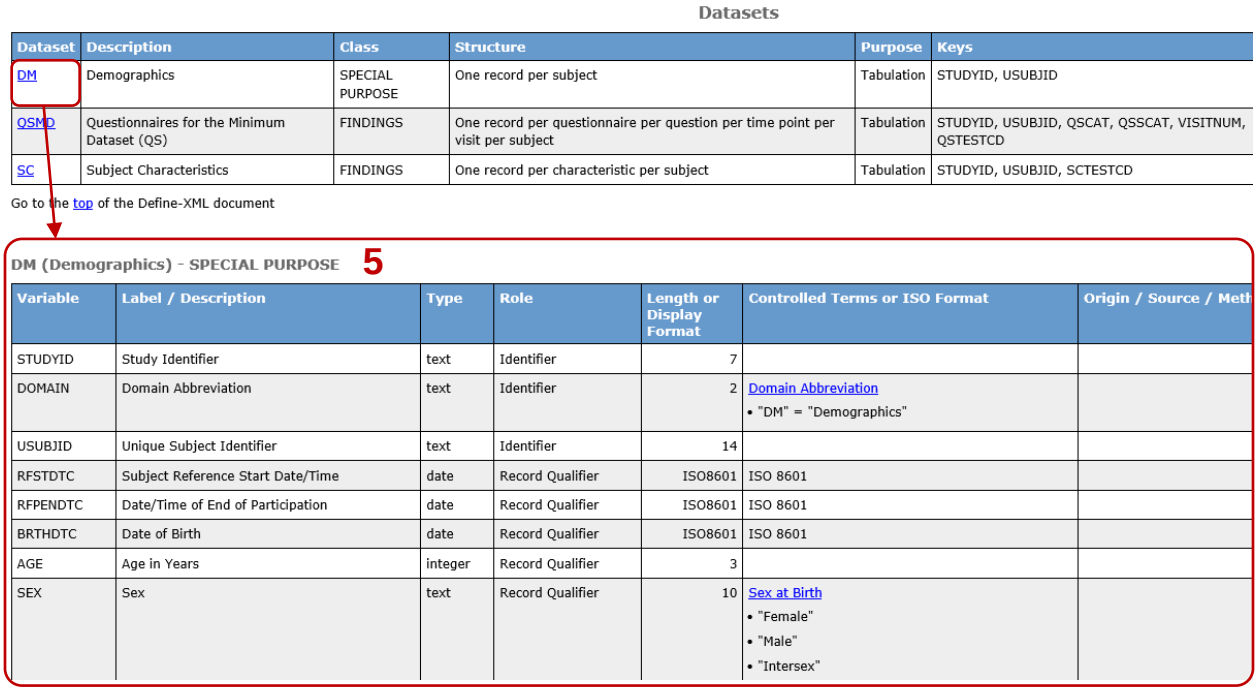
3. Click on “Annotated Case Report Forms” and a new browser tab will open with the annotated CRF file referenced in the specifications.



4. The "Datasets" section shows the list of datasets described within the Define-XML file.



5. Each dataset is hyperlinked to its own section of the Define-XML document. Each of these sections include the name, label, type, role, and format of the variables contained within the dataset. Within the formats, there are links to code lists descriptions when relevant.



Datasets

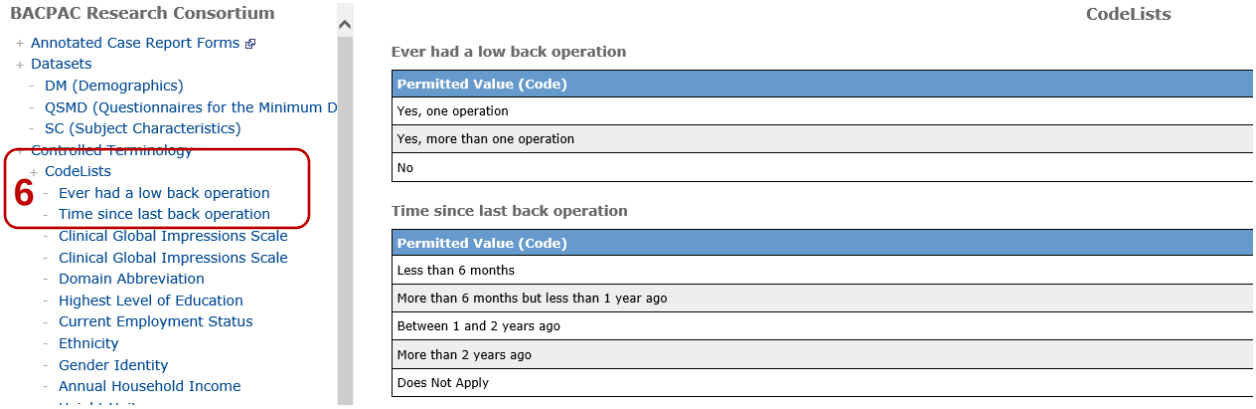
Dataset	Description	Class	Structure	Purpose	Keys
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID
QSMD	Questionnaires for the Minimum Dataset (QS)	FINDINGS	One record per questionnaire per question per time point per visit per subject	Tabulation	STUDYID, USUBJID, QSCAT, QSSCAT, VISITNUM, QSTESTCD
SC	Subject Characteristics	FINDINGS	One record per characteristic per subject	Tabulation	STUDYID, USUBJID, SCTESTCD

Go to the top of the Define-XML document

DM (Demographics) - SPECIAL PURPOSE **5**

Variable	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Meth
STUDYID	Study Identifier	text	Identifier	7		
DOMAIN	Domain Abbreviation	text	Identifier	2	Domain Abbreviation • "DM" = "Demographics"	
USUBJID	Unique Subject Identifier	text	Identifier	14		
RFSTDTC	Subject Reference Start Date/Time	date	Record Qualifier	ISO8601	ISO 8601	
RFPENDTC	Date/Time of End of Participation	date	Record Qualifier	ISO8601	ISO 8601	
BRTHDTC	Date of Birth	date	Record Qualifier	ISO8601	ISO 8601	
AGE	Age in Years	integer	Record Qualifier	3		
SEX	Sex	text	Record Qualifier	10	Sex at Birth • "Female" • "Male" • "Intersex"	

6. The "CodeLists" section shows controlled terminology (permitted values) for responses to questions in the datasets.



BACPAC Research Consortium

+ Annotated Case Report Forms
+ Datasets
- DM (Demographics)
- QSMD (Questionnaires for the Minimum D
- SC (Subject Characteristics)
- Controlled Terminology
+ CodeLists
- Ever had a low back operation
- Time since last back operation
- Clinical Global Impressions Scale
- Clinical Global Impressions Scale
- Domain Abbreviation
- Highest Level of Education
- Current Employment Status
- Ethnicity
- Gender Identity
- Annual Household Income

6

CodeLists

Ever had a low back operation

Permitted Value (Code)
Yes, one operation
Yes, more than one operation
No

Time since last back operation

Permitted Value (Code)
Less than 6 months
More than 6 months but less than 1 year ago
Between 1 and 2 years ago
More than 2 years ago
Does Not Apply

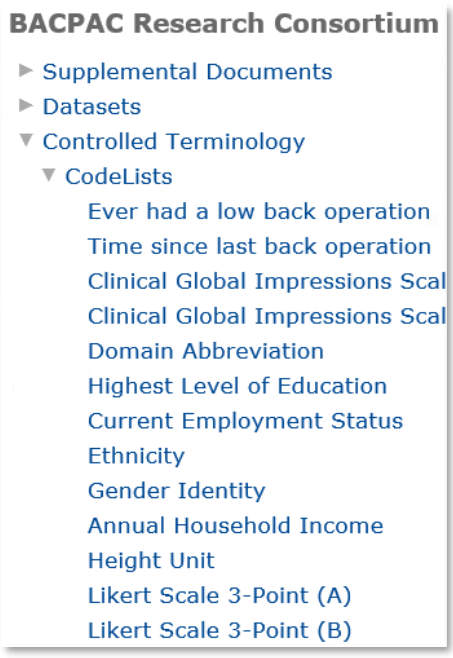
1.2.4 Study Information

Study information can be found at the top of the Define-XML, above the “Datasets” table. Study Name, Study Description, and Protocol Name all reflect the values of the corresponding attributes in the specifications file. Metadata Name is automatically generated from “StudyName.”

Standard	BACPAC Modified SDTM Standard 3.1.2 (Modified)
Study Name	BACPAC Research Consortium
Study Description	Observational Study
Protocol Name	Minimum Dataset and Broadly Collected PROs Example
Metadata Name	Study BACPAC Research Consortium Data Definitions

1.2.5 Code Lists

A list of all code lists can be found at the end of the Define-XML. In the navigation pane on the left, select “Controlled Terminology,” revealing the “CodeLists” drop-down list. Click “CodeLists” and a list of all code lists from the specifications file will appear. Code lists are listed in alphabetic order by the ID assigned in the specifications, but the name of the code list is the value that will display in the navigation pane. Clicking on a code list name hyperlink will take you to the corresponding code list in the Define-XML.



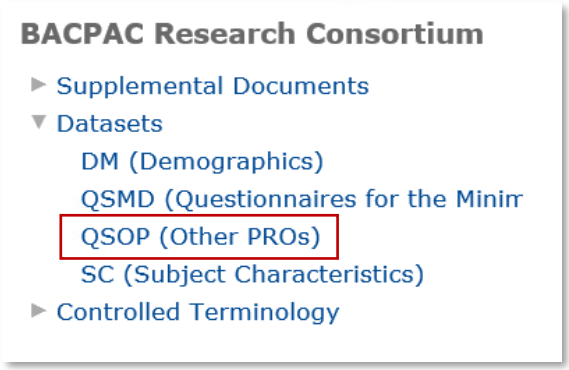
For example, the “Height Unit” hyperlink will take you to the following table:

Height Unit

Permitted Value (Code)	Display Value (Decode)
IN	Inches
CM	Centimeters

Values from the “Term” column in the specifications will appear in the “Permitted Value (Code)” column. If decoded values were assigned in the specifications, they will appear in the “Display Value (Decode)” column. If not, the code list table in the Define-XML will only include the “Permitted Value (Code)” column.

Code lists will also appear in the dataset-specific tables. In Section 1.1.2, the QSCAT code list was assigned to the QSCAT variable in the QSOP dataset. To view this code list, select “Datasets” in the navigation pane of the Define-XML, then click on the “QSOP” hyperlink.



The dataset hyperlink will take you to a dataset-specific table. The assigned code list will display in the “Controlled Terms or ISO Format” column. The value in the column will begin with a hyperlink to the code list table, followed by the values in the code list (see the code list for DOMAIN in the table below). If the code list is too long to fit in the table, the number of terms in the code list will display instead of the values of the code list.

QSOP (Other PROs, QS) - FINDINGS Location: qsop.xpt

Variable	Where Condition	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
STUDYID		Study Identifier	text	Identifier	8		
DOMAIN		Domain Abbreviation	text	Identifier	2	Domain Abbreviation • "QS" = "Questionnaires"	
USUBJID		Unique Subject Identifier	text	Identifier	19		
QSSEQ		Sequence Number	integer	Identifier	3		
QSCAT		Category of Question	text	Grouping Qualifier	50	Question Category [15 Terms]	

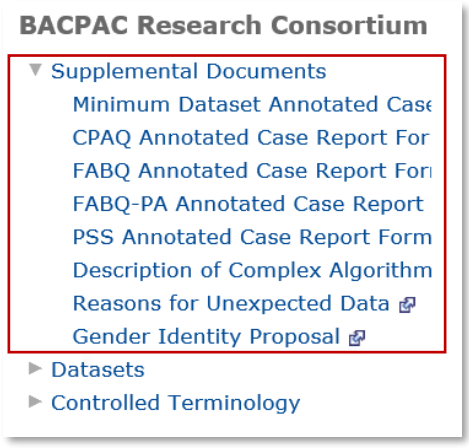
If the code list was assigned using a where clause, it will appear in the row(s) below the variable to which it was assigned. If more than one code list is assigned to a variable, a separate will appear for each individual where condition.

For example, the LIKRTN4A code list depicted in Section 1.1.2 appears in the row below the QSSTRESN variable in the QSMD table. The “Variable” column is blank, and the “Where Condition” describes the where clause assigned in the specifications. The “Label / Description” contains the description of the where clause, from the “ValueLevel” tab of the specifications. “Type” and “Length or Display Format” are the values of “Data Type” and “Length” from the “ValueLevel” tab, respectively. The code list will still appear in the “Controlled Terms or ISO Format” column. The other code lists assigned to QSSTRESN under different conditions follow the row for the LIKRTN4A code list.

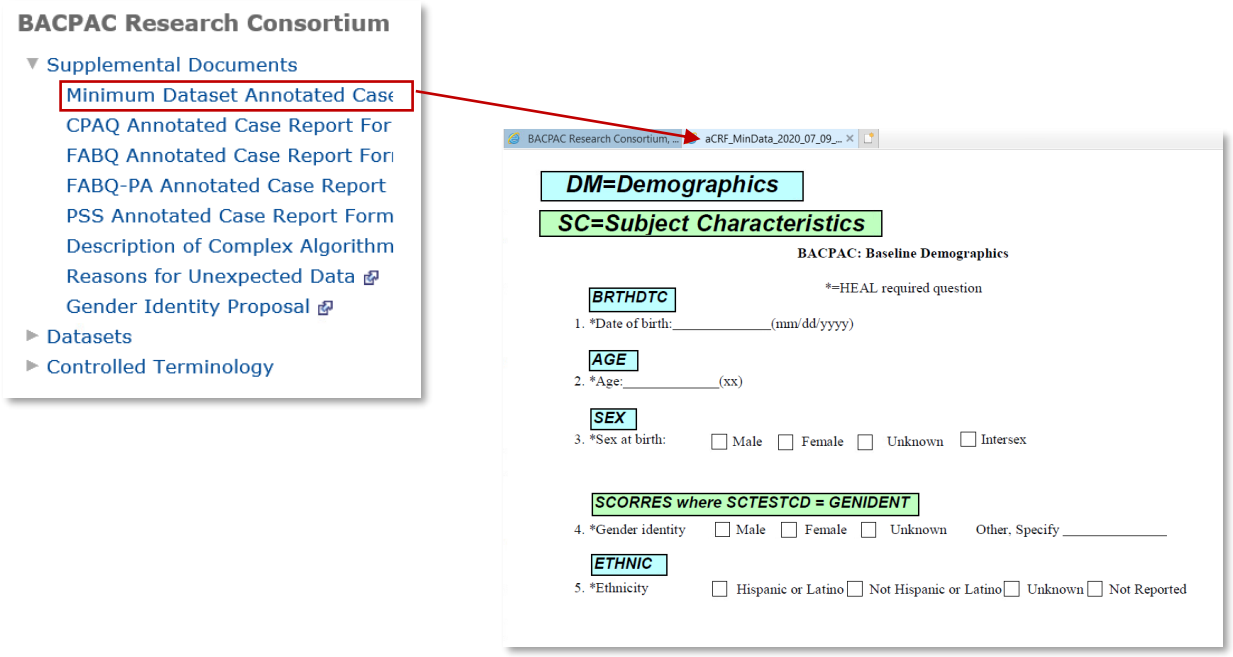
Variable	Where Condition	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format
QSSTRESN VLM		Numeric Finding in Std Units	float	Result Qualifier	8	
	QSTESTCD IN ("GAD01" (How often bothered by feeling nervous, anxious or on edge), "GAD02" (How often bothered by not being able to stop or control worrying))	GAD	integer		8	Likert Scale 4-Point (A) • 0 = "Not at all" • 1 = "Several days" • 2 = "More than half the days" • 3 = "Nearly every day"
	QSTESTCD IN ("EDANX01" (I felt fearful), "EDANX40" (Hard to focus on anything other than anxiety), "EDANX41" (My worries overwhelm me), "EDANX53" (I felt uneasy))	PROMIS Emotional Distress - Anxiety	integer		8	Likert Scale 5-Point (A) • 1 = "Never" • 2 = "Rarely" • 3 = "Sometimes" • 4 = "Often" • 5 = "Always"

1.2.6 Documentation

All documents in the specifications are listed in the left navigation pane under “Supplemental Documents.”



Clicking on the document hyperlink will open the document. PDFs will open as a new tab in the browser.



Reminder: As described in the Define-XML Guide, the XSL stylesheet, XML file, specifications file, and documents referenced in the specifications file should all be saved in the same folder. If not, the document hyperlink in the Define-XML will not open the document.

1.2.7 Comments

Dataset Comments

Dataset comments will be visible in the “Datasets” table located at the beginning of the Define-XML. Comments will be located in the “Documentation” column.

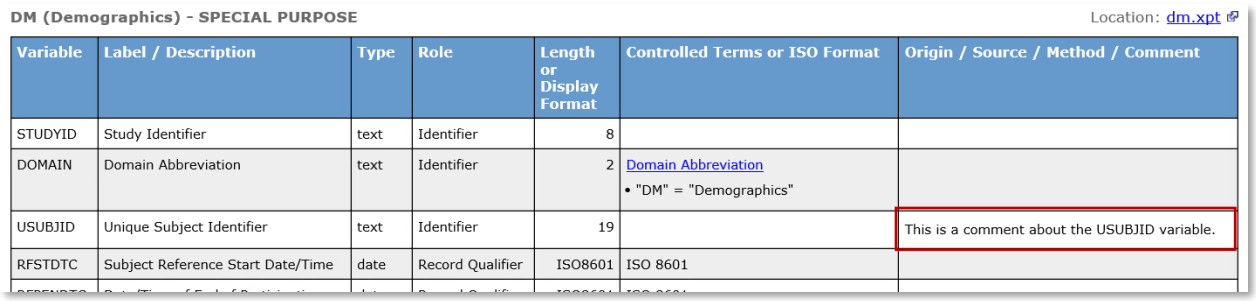
Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	This is a comment about the DM dataset. Minimum Dataset Annotated Case Report Form [1]	dm.xpt
QSMD	Questionnaires for the Minimum Dataset (QS)	FINDINGS	One record per questionnaire per question per time point per visit per subject	Tabulation	STUDYID, USUBJID, QSCAT, QSSCAT, VISITNUM, QSTESTCD	This is a comment about the QSMD dataset. Minimum Dataset Annotated Case Report Form [5 - 24]	qsmd.xpt
QSOP	Other PROs (QS)	FINDINGS	One record per questionnaire per question per time point per visit per subject	Tabulation	STUDYID, USUBJID, QSCAT, QSSCAT, VISITNUM, QSTESTCD		qsop.xpt
SC	Subject Characteristics	FINDINGS	One record per characteristic per subject	Tabulation	STUDYID, USUBJID, SCTESTCD	This is a comment about the SC dataset. Minimum Dataset Annotated Case Report Form [1 2 3 4]	sc.xpt

Variable Comments

In the Define-XML's navigation pane, select the dataset containing the variable. The hyperlink will take you to a dataset-specific table.

Comments will appear in the "Origin / Source / Method / Comment" column, in the row of the variable with which they are associated.

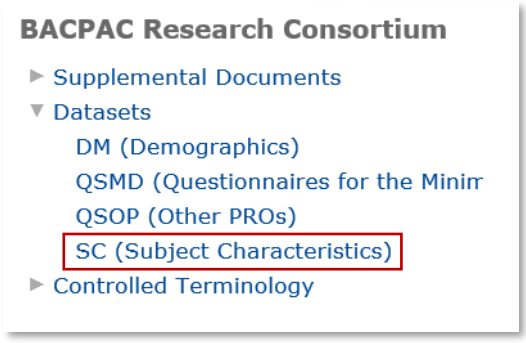


DM (Demographics) - SPECIAL PURPOSE — Location: dm.xpt

Variable	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
STUDYID	Study Identifier	text	Identifier	8		
DOMAIN	Domain Abbreviation	text	Identifier	2	Domain Abbreviation • "DM" = "Demographics"	
USUBJID	Unique Subject Identifier	text	Identifier	19		This is a comment about the USUBJID variable.
RFSTDTC	Subject Reference Start Date/Time	date	Record Qualifier	ISO8601	ISO 8601	

Value Level Comments

In the navigation pane, select the relevant dataset. The hyperlink will take you to a dataset-specific table.



In the table, scroll to the variable associated with the comment. Find the “Where condition” associated with the comment.

SCSTRESC VLM		Character Result/Finding in Std Format	text	Result Qualifier	100		
	SCTESTCD = "GENIDENT" (Gender Identity)		text		10	Gender Identity • "Female" • "Male" • "Unknown" • "Other"	This is a comment about gender identity mapping in the SC dataset. Gender Identity Proposal

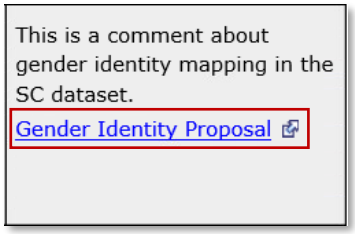
The comment will display in the “Origin / Source / Method / Comment” column.

SCSTRESC VLM		Character Result/Finding in Std Format	text	Result Qualifier	100		
	SCTESTCD = "GENIDENT" (Gender Identity)		text		10	Gender Identity • "Female" • "Male" • "Unknown" • "Other"	This is a comment about gender identity mapping in the SC dataset. Gender Identity Proposal

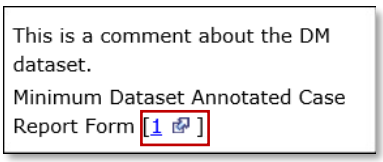
Document Links in Comments

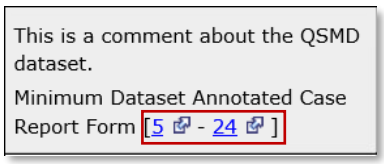
If a document was added to a comment, the hyperlink will follow the text of the comment.

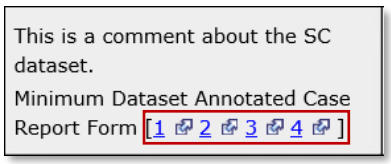
If there are no page numbers referenced, the document name will link to the document.



If page numbers are referenced, they will be listed in square brackets after the document name. Each page number will link to the corresponding page in the document.







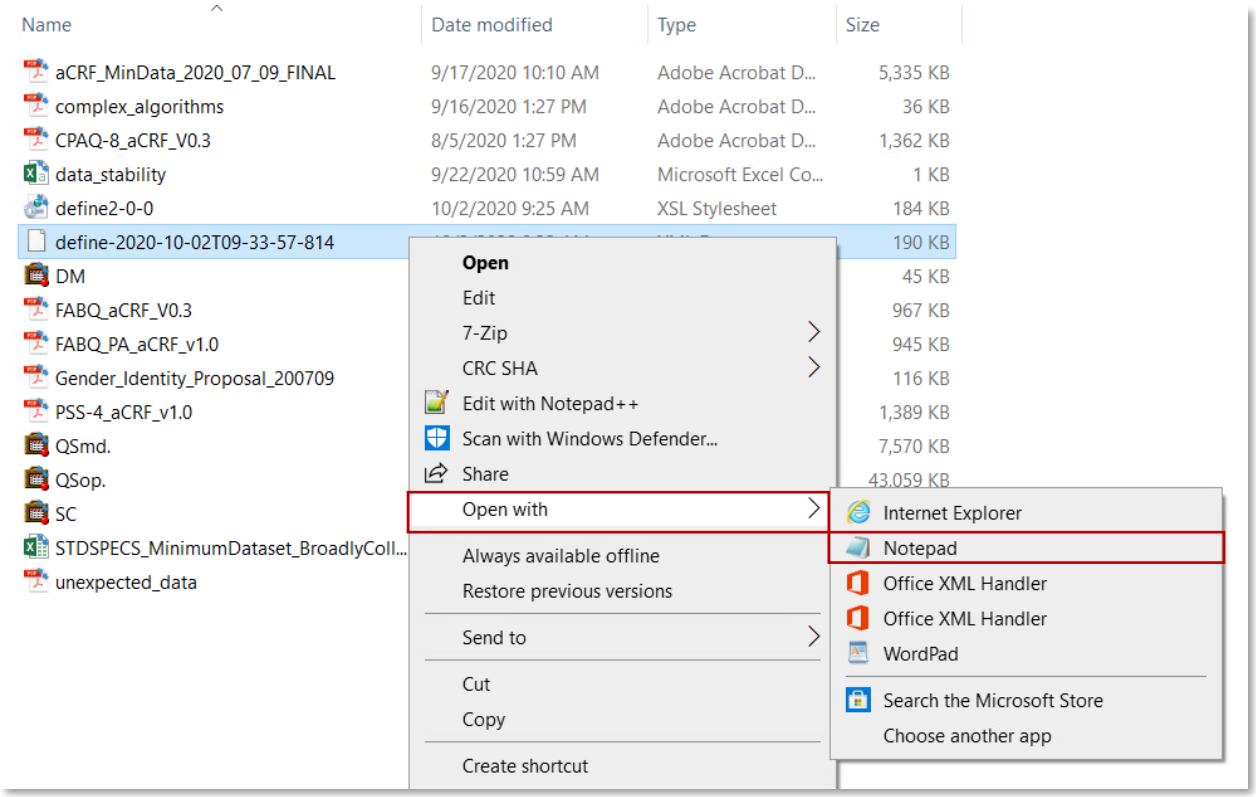
1.2.8 Dataset Hyperlinks

The dataset hyperlinks in the Define-XML are automatically generated with the ".xpt" file extension, and are found in the "Location" column of the "Datasets" table.

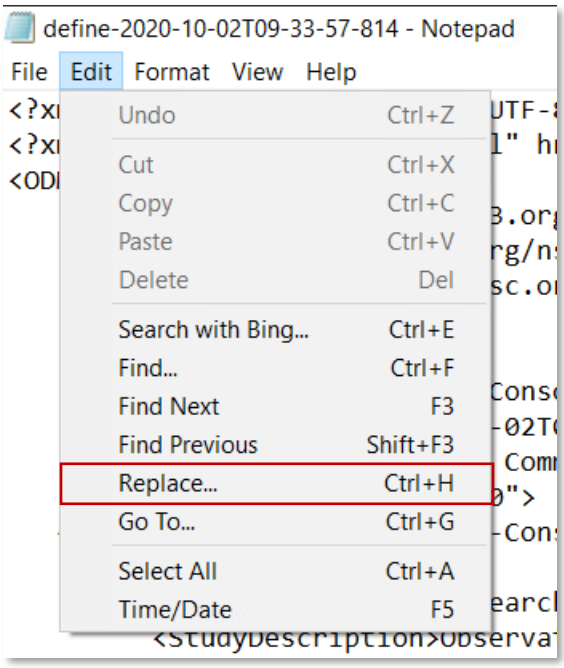
Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	This is a comment about the DM dataset. Minimum Dataset Annotated Case Report Form [1]	dm.xpt
QSMD	Questionnaires for the Minimum Dataset (QS)	FINDINGS	One record per questionnaire per question per time point per visit per subject	Tabulation	STUDYID, USUBJID, QSCAT, QSSCAT, VISITNUM, QSTESTCD	This is a comment about the QSMD dataset. Minimum Dataset Annotated Case Report Form [5 - 24]	qsmd.xpt
QSOP	Other PROs (QS)	FINDINGS	One record per questionnaire per question per time point per visit per subject	Tabulation	STUDYID, USUBJID, QSCAT, QSSCAT, VISITNUM, QSTESTCD		qsop.xpt
SC	Subject Characteristics	FINDINGS	One record per characteristic per subject	Tabulation	STUDYID, USUBJID, SCTESTCD	This is a comment about the SC dataset. Minimum Dataset Annotated Case Report Form [1 2 3 4]	sc.xpt

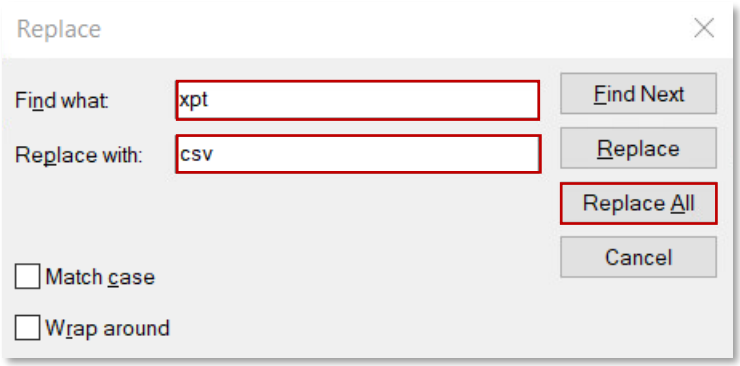
If you are submitting data in CSV files instead, you will need to edit the Define-XML with the correct file extension. To do so, open the Define-XML in a text editor, such as Notepad, by right clicking on the file and selecting "Open with…"



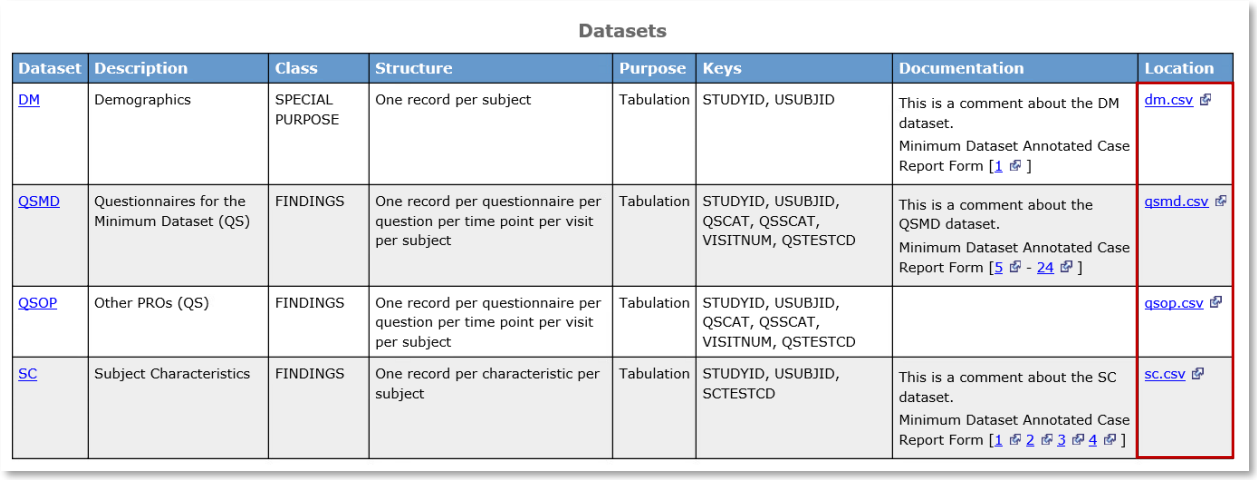
Replace all occurrences of ".xpt" with ".csv." Press Ctrl + H or go to the Edit menu and select "Replace…" to open the Replace box.



Search for "xpt" and replace with "csv." Click "Replace All" and save the file.



Open the Define-XML again to confirm the file extensions have been replaced and that the hyperlinks work correctly.



Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	This is a comment about the DM dataset. Minimum Dataset Annotated Case Report Form [1]	dm.csv
QSMD	Questionnaires for the Minimum Dataset (QS)	FINDINGS	One record per questionnaire per question per time point per visit per subject	Tabulation	STUDYID, USUBJID, QSCAT, QSSCAT, VISITNUM, QSTESTCD	This is a comment about the QSMD dataset. Minimum Dataset Annotated Case Report Form [5 - 24]	qsmd.csv
QSOP	Other PROs (QS)	FINDINGS	One record per questionnaire per question per time point per visit per subject	Tabulation	STUDYID, USUBJID, QSCAT, QSSCAT, VISITNUM, QSTESTCD		qsop.csv
SC	Subject Characteristics	FINDINGS	One record per characteristic per subject	Tabulation	STUDYID, USUBJID, SCTESTCD	This is a comment about the SC dataset. Minimum Dataset Annotated Case Report Form [1 2 3 4]	sc.csv

2 Supporting Documentation

As described in the Data Transfer SOP, the documents in this section are required to be provided as PDF files and listed in the specifications file.

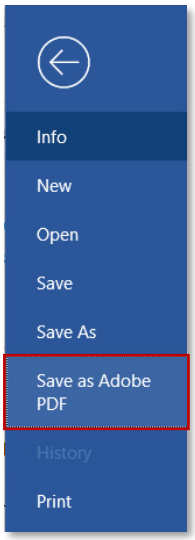
Example document list in the specifications file:

ID	Title	Href
1	Minimum Dataset Annotated Case Report Form	aCRF_MinData_2020_07_09_FINAL.pdf
2	CPAQ Annotated Case Report Form	CPAQ-8_aCRF_V0.3.pdf
3	FABQ Annotated Case Report Form	FABQ_aCRF_V0.3.pdf
4	FABQ-PA Annotated Case Report Form	FABQ_PA_aCRF_v1.0.pdf
5	PSS Annotated Case Report Form	PSS-4_aCRF_v1.0.pdf
6	Description of Complex Algorithms	complex_algorithms.pdf
7	Reasons for Unexpected Data	unexpected_data.pdf
GENDOC	Gender Identity Proposal	Gender_Identity_Proposal_200709.pdf

2.1 Annotated Case Report Forms

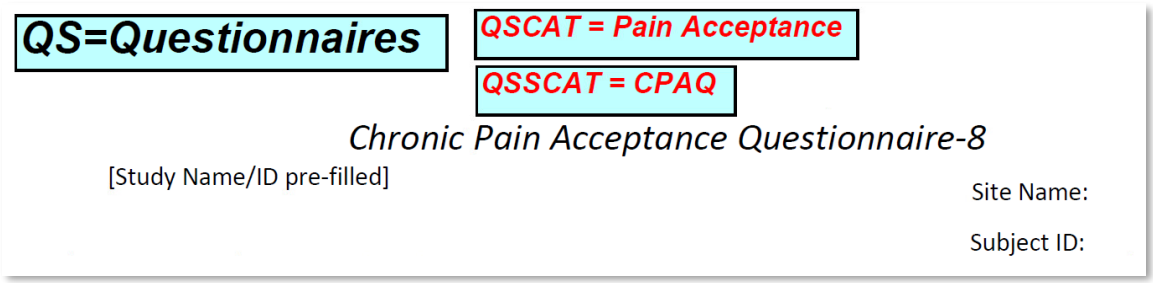
The following example walks through how to create an annotated CRF for a dataset in the QS domain.

If the CRF is saved as a Word document, save a new copy as a PDF. Go to “File,” then “Save as Adobe PDF.”

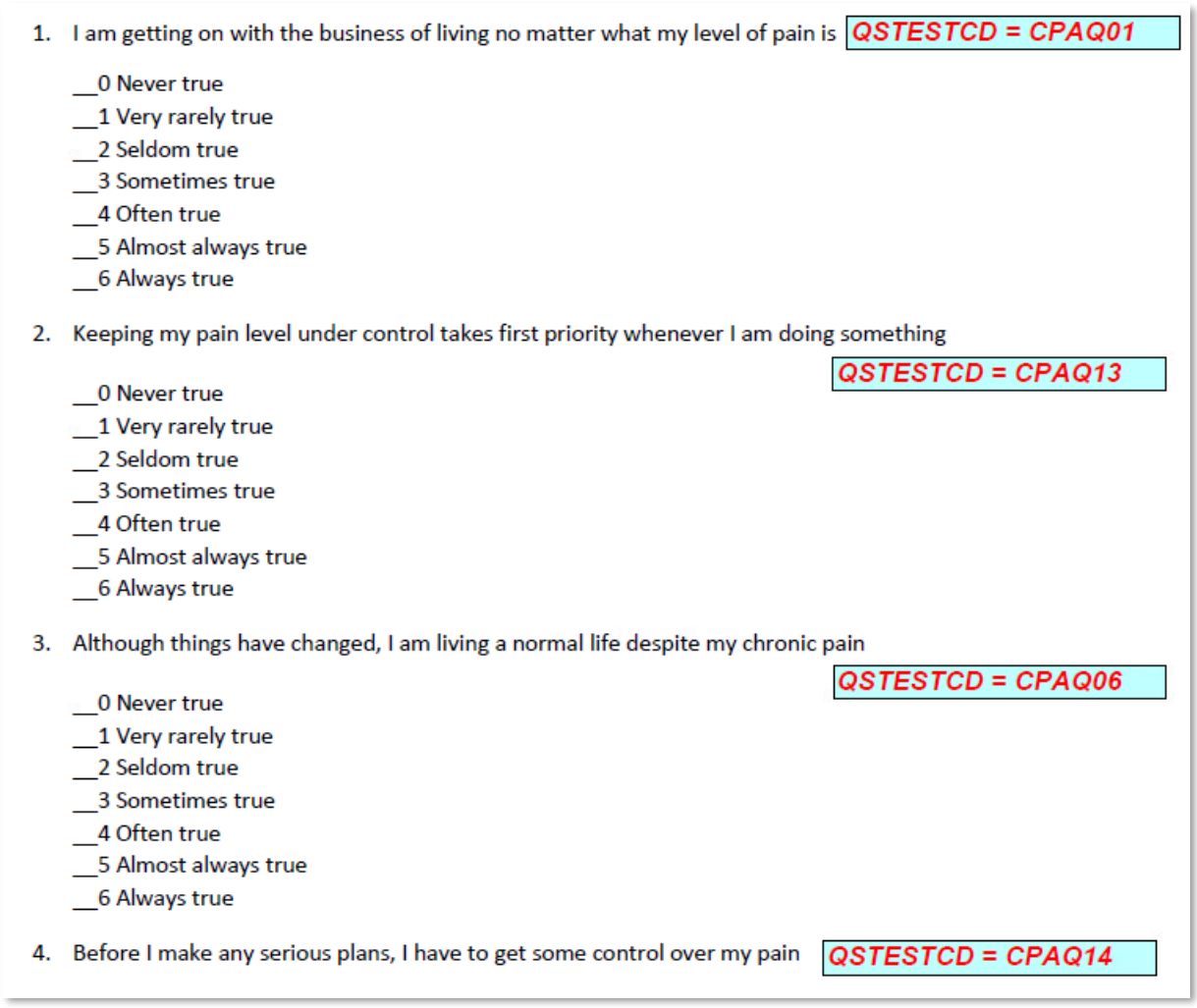


Open the PDF. At the top of the file, add text boxes including the values of domain, QSCAT, and QSSCAT. Text boxes can be copied and pasted from aCRFs provided by the DAC. aCRFs for the Minimum Dataset, CPAQ-8, FABQ, FABQ-PA, and PSS-4 are currently available in Microsoft Teams on the Data Sharing, Management, and Standards Wiki. After copying and pasting, edit the text boxes as necessary.

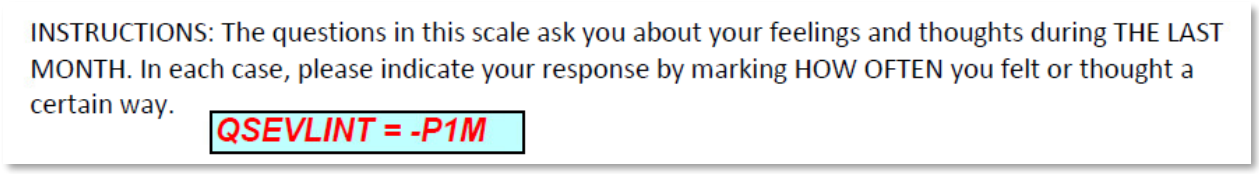
The domain box will include the domain abbreviation, followed by the domain name. The QSCAT and QSSCAT boxes include the variable followed by the value of the variable.

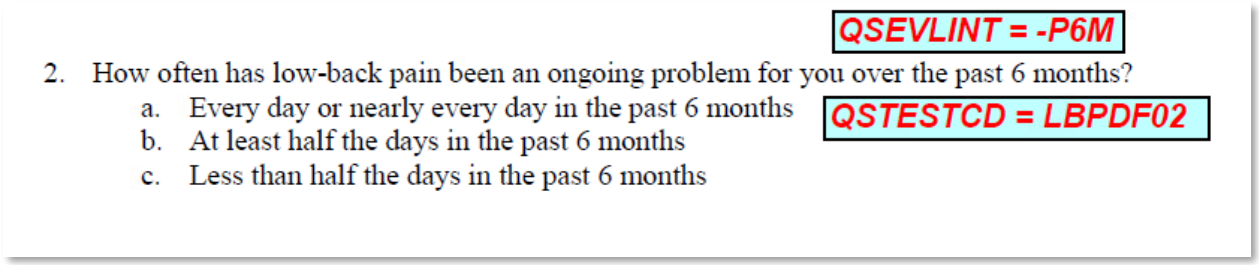


In the same manner as above, create QSTESTCD annotations for each question.

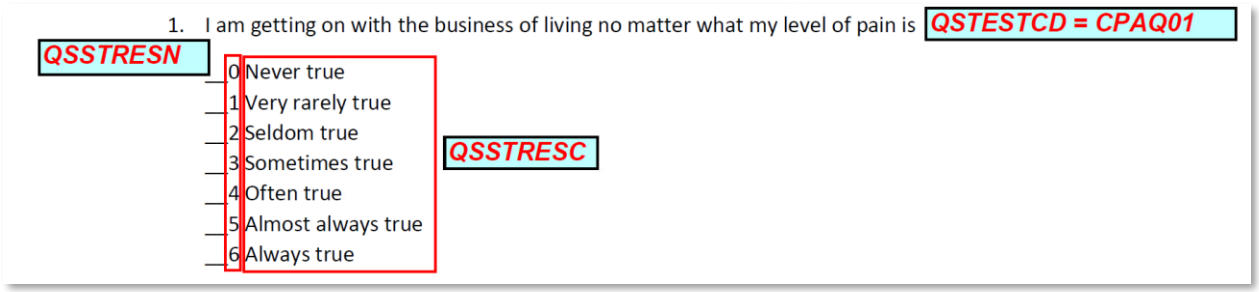


If any questions have values of EVLINT, add an annotation for each unique value.

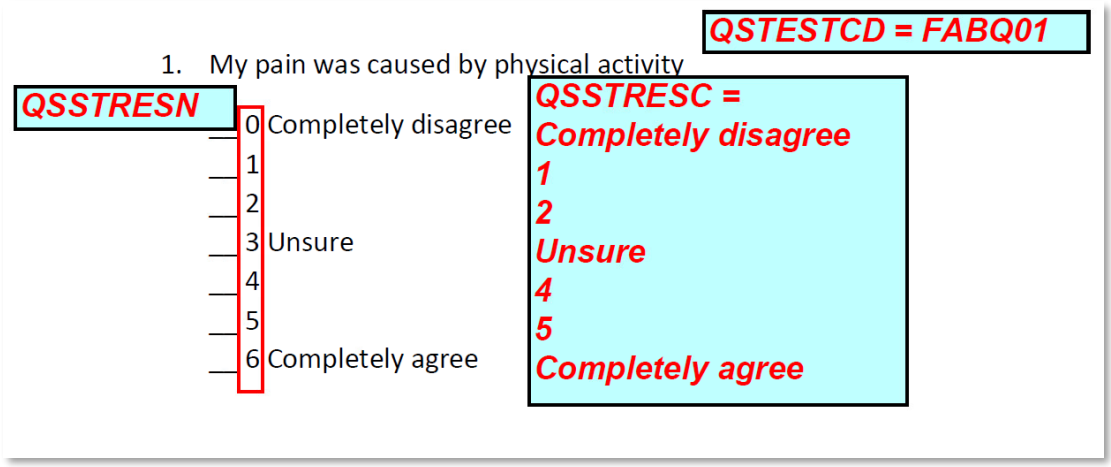


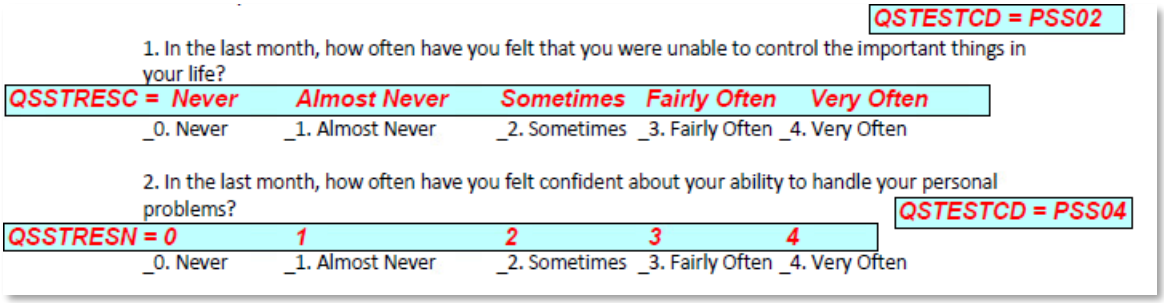


Create annotations for each distinct set of QSSTRESC and QSSTRESN values. Depending on the format of the CRF, you may be able to use the values as listed on the form. Create red boxes around the value. This can be done by copying and pasting from one of the DAC's CRFs and resizing the boxes as necessary. Add QSSTRESC and QSSTRESN annotations to label the values in the boxes.



Alternatively, you may need to type out the QSSTRESC and/or QSSTRESN, as illustrated below.





2.2 Description of Complex Algorithms

Complex algorithms should be described in a PDF and included in the specifications file. Follow the instructions in Section 1.2.6 to add the PDF to the specifications.

The complex algorithms document should describe calculations more complex than simple summary scores. Examples of algorithms that should be included:

- Z-score and t-score calculations, particularly if the calculations are age- or gender-specific
- A calculation that requires a macro to execute
- A formula or algorithm that requires more explanation than is appropriate in the comments column in the specifications file

2.3 Reasons for Unexpected Data

Reasons for unexpected data should be explained in a PDF and included in the specifications file. Follow the instructions in Section 1.2.6 to add the PDF to the specifications.

Examples of unexpected data:

- A participant in the Demographics (DM) domain was randomized, but did not receive the study treatment. The participant was assigned a treatment arm, but does not have any records related to that treatment.
- A participant in the Demographics (DM) domain does not appear in any of the other domains.

3 Identification of Stable Data

3.1 Data Stability Identifier

The *Data Stability Identifier* is required to have one of three values for each participant:

- "CC" → Indicates that data collection is Complete for the participant and data are fully Cleaned in accordance with the data management plan (DMP).
  For example:
  - The subject has completed the study, all data are complete and clean.
- "CI" → Indicates that data collection is Complete for the participant but that data issues are outstanding or that formal review of the data in accordance with the DMP is Incomplete.
  For example:
  - The subject has completed the study and all data have been collected. The research unit has completed QC and doesn't have changes to make, but a formal review of the data that has been entered has not been conducted.
  - The subject has completed the study and all data have been collected. A formal review of the data has been conducted, but the issues identified have not yet been corrected.
- "II" → Indicates that data collection is Incomplete for the participant and as a result data quality control is also Incomplete.
  For example:
  - The subject has not completed the study. A formal review of the data has not been conducted.

The *Data Stability Identifier* relates to participant-reported outcome data, demographics data, and other baseline characteristic data but does not pertain to data derived from biospecimens (e.g., laboratory parameters, omics data, imaging data). Methods for documenting data cleanliness/stability for data derived from biospecimens will be described separately.

3.2 Supplementary CSV File

A supplementary CSV File containing the following two fields is required for each data transfer:

- Unique Participant ID
- Data Stability Identifier

See below for an example of the data included in the supplementary CSV file.

Unique Subject ID	Data Stability Identifier
ABCD-0101-0001	CC
ABCD-0101-0002	CI
ABCD-0101-0003	CC
ABCD-0101-0004	CC
ABCD-0101-0005	II