



BACPAC Modified SDTM Guide

Version 1.1, July 2021

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

Table of Contents

1. Identifier Variables	4
1.1 STUDYID: Study ID	4
1.2 USUBJID: Unique Subject ID	4
1.3 --SEQ: Variable to Identify Unique Observations in a Dataset	5
2. Timing Variables	7
2.1 Formatting of Date Variables and Use of Partial Dates	7
2.2 VISIT/VISITNUM: Variables for Visit	9
2.3 RFSTDTC: Reference Start Date	12
2.4 RFENDTC: Reference End Date	12
2.5 --DY: Study Day of Finding	12
2.6 --DTC: Date/Time of Finding	13
2.7 --EVLINT: Evaluation Interval	13
3. Grouping and Synonym Qualifiers	14
3.1 --CAT: Category (Grouping Qualifier)	14
3.2 --SCAT: Subcategory (Grouping Qualifier)	14
3.3 --TEST: Name of Measurement, Test or Examination (Synonym Qualifier)	14
3.4 --TESTCD: Abbreviated Test Code (Grouping Qualifier)	14
4. Result Qualifiers	15
4.1 --STRESC: Character Result or Finding in Standard Format	15
4.2 --STRESN: Numeric Result or Finding in Standard Format	16
4.3 --ORRES: Result or Finding in Original Units	16
5. Record Qualifiers	18
5.1 --DRVFL: Derived Flag	18
5.2 --METHOD: Method of Test or Examination	18
5.3 AGE	19
5.4 SEX	19
5.5 RACE	19
5.6 RACEMULT	20
5.7 ETHNIC	20
5.8 BRTHDTC	20

6.	Variable Qualifiers	21
6.1	--STRESU: Standard Units of Result or Finding	21
6.2	--ORRESU: Units of Original Result or Finding.....	21
7.	Variables Specific to Treatment Categories (EX Domain)	22
7.1	EXTRT: Name of Treatment.....	22
7.2	EXTRTOTH: Other Treatment.....	22
7.3	EXACN: Action Taken with Treatment.....	23
7.4	EXADJ: Reason for Treatment Adjustment.....	23
7.5	EXDOSE: Dose of Treatment.....	23
7.6	EXDOSEU: Units of Treatment Dose	23
7.7	EXROUTE: Route of Administration	23
8.	Variables Specific to Functional Tests (FT Domain)	24
8.1	FTPOS: Position of Participant During Observation.....	24
8.2	FTLOC: Location Used for the Measurement	24
8.3	FTLAT: Laterality.....	24
8.4	FTREPNUM: Repetition Number	24
8.5	FTSTAT: Completion Status.....	25
8.6	FTREASND: Reason Not Done.....	25
8.7	FTAID: Aid Used During Test	25
8.8	FTAIDOTH: Other Aid Used During Test.....	25

1. Identifier Variables

1.1 STUDYID: Study ID

Study ID is the text descriptor used to identify each study.

- Can be no longer than 8 characters.
- Cannot contain spaces or hyphens but can contain numbers (e.g., REACH01).

Research units that conduct multiple studies will have a Study ID for each study.

1.2 USUBJID: Unique Subject ID

The format of BACPAC unique subject IDs will be as follows:

[Study ID]-[Site ID]-[Participant ID]

Site ID

The Site ID for BACPAC studies has two components that combine to create a 4-digit numeric representation.

The zero-padded, 4-digit number has the form XXYY where

- XX is a zero-padded two-digit ID specific to each BACPAC research unit (e.g., 01 for the UCSF MRC as in Table 1.2.1).
- YY is a zero-padded two-digit ID that is used to enumerate the site(s) that enroll participants in a given research study.

Table 1.2.1. BACPAC Research Unit Site ID Assigned Numbering

Research Unit	Type	PI(s)	Site ID
University of California San Francisco	MRC	Lotz	01
University of Michigan	MRC	Clauw/Hassett	02
University of Pittsburgh	MRC	Sowa/Vo	03
Cedars-Sinai	P2	Spiegel	04
University of Pittsburgh	P2	Wasan	05
Brigham Young University	Tech	Bowden	06
Harvard University	Tech	Walsh	07
Massachusetts General Hospital	Tech	Wey	08
Ohio State University	Tech	Marras	09
University of California San Francisco	Tech	Fields/Krug	11
University of California San Francisco	Tech	Majumdar	10
University of Utah	Tech	Rieke/Shah	12
University of Utah	CT	Fritz	13
University of North Carolina Chapel Hill	DAC	LaVange/Ivanova	14

Each BACPAC research unit should enumerate Site IDs starting with XX01.

- Single site studies will use XX01 for the Site ID.
- Example: a two-site study would use Site IDs XX01 and XX02.

Enrollment Site ID numbers should be kept consistent across multiple studies under a given BACPAC research unit. Therefore, two *different* enrollment sites contributing data to a single BACPAC research unit conducting multiple studies should not have the same two-digit ID (YY).

Example:

The University of Michigan research unit conducts 2 studies – one is a randomized trial and the other is a cohort study. They use 5 enrollment sites across the 2 studies and 2 of these sites recruit participants for both studies.

The Site IDs for these enrollment centers would take the values 0201 – 0205 and two of the sites would have the same Site ID for each study. Here, “02” represents the University of Michigan research unit and 01 – 05 represent the individual enrollment sites. For example, if the two overlapping sites are labeled as 01 and 04, we may see StudyID-SiteID combinations as shown in Table 1.2.2.

Table 1.2.2. Examples of Study ID, Site ID and USUBJID

STUDYID	SITEID	USUBJID for first participant enrolled
COHORT	0201	COHORT-0201-00001
COHORT	0202	COHORT-0202-00001
COHORT	0203	COHORT-0203-00001
COHORT	0204	COHORT-0204-00001
SMART	0201	SMART-0201-00001
SMART	0204	SMART-0204-00001
SMART	0205	SMART-0205-00001

Additional enrollment centers can be added, and issued Site IDs, as needed.

Participant ID

- Participant IDs should be nested within Site IDs and increment from 00001 in consecutive order.
- Participant IDs should have a 5-digit length and be padded with zeros as shown above.

1.3 --SEQ: Variable to Identify Unique Observations in a Dataset

Several of the dataset domains have multiple rows (observations) for each participant (USUBJID). In these domains, the --SEQ variable is used to index each unique row for a USUBJID, where “--” is replaced by the domain abbreviation (e.g., QSSEQ for the Questionnaires findings domain). This variable can take any integer value and does not have to start with 1. The ordering of the --SEQ variable represents how the data should be sorted before submission. The sort order is defined in the specifications under Key Variables (Figure 1.3.1).

Dataset	Description	Class	Structure	Purpose	Key Variables
DM	Demographics	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID
QSMD	Questionnaires for the Minim	FINDINGS	One record per questionnaire per question per time point per vi	Tabulation	STUDYID, USUBJID, QSCAT, QSSCAT, VISITNUM, QSTESTCD
SC	Subject Characteristics	FINDINGS	One record per characteristic per subject	Tabulation	STUDYID, USUBJID, SCTESTCD

Figure 1.3.1. The Datasets tab from the BACPAC CDISC modified-SDTM specifications file provides the sort order of each dataset (i.e., Key Variables). The --SEQ variable value is assigned after sorting is complete.

For example, the QSMD dataset should be sorted in the order displayed in the Key Variables column: STUDYID, USUBJID, QSCAT, QSSCAT, VISITNUM, QSTESTCD. Subsequently, values for QSSEQ are assigned for each row for each USUBJID in the order they appear in the sorted dataset (Figure 1.3.2).

	Study Identifier	Domain Abbreviation	Unique Subject Identifier	Sequence Number	Category of Question	Subcategory for Question	Question Short Name	Question Name
1	SMART	QS	SMART-0201-00001	1	Anxiety	GAD	GAD01	How often bothered by feeling nervous, anxious or on edge
2	SMART	QS	SMART-0201-00001	2	Anxiety	GAD	GAD02	How often bothered by not being able to stop or control worrying
3	SMART	QS	SMART-0201-00001	3	Anxiety	GAD	GAD2RAW	GAD-2 Raw Score
4	SMART	QS	SMART-0201-00001	4	Anxiety	GAD	GAD01	How often bothered by feeling nervous, anxious or on edge
5	SMART	QS	SMART-0201-00001	5	Anxiety	GAD	GAD02	How often bothered by not being able to stop or control worrying
6	SMART	QS	SMART-0201-00001	6	Anxiety	GAD	GAD2RAW	GAD-2 Raw Score
7	SMART	QS	SMART-0201-00001	7	Anxiety	PROMIS Emotional Distress - Anxiety	EDANX01	I felt fearful
8	SMART	QS	SMART-0201-00001	8	Anxiety	PROMIS Emotional Distress - Anxiety	EDANX40	Hard to focus on anything other than anxiety
9	SMART	QS	SMART-0201-00001	9	Anxiety	PROMIS Emotional Distress - Anxiety	EDANX41	My worries overwhelm me
10	SMART	QS	SMART-0201-00001	10	Anxiety	PROMIS Emotional Distress - Anxiety	EDANX53	I felt uneasy
11	SMART	QS	SMART-0201-00001	11	Anxiety	PROMIS Emotional Distress - Anxiety	PRANX4AR	PROMIS-Anxiety 4a Raw Score
12	SMART	QS	SMART-0201-00001	12	Anxiety	PROMIS Emotional Distress - Anxiety	PRANX4AT	PROMIS-Anxiety 4a T-Score

Figure 1.3.2. An example QS dataset where the Sequence Number (QSSEQ) variable is indexed for the first 8 QSTEST/QSTESTCD values in order of QSCAT, QSSCAT, VISITNUM and QSTESTCD.

2. Timing Variables

2.1 Formatting of Date Variables and Use of Partial Dates

The BACPAC modified SDTM data standard requires dates and times in ISO 8601 format. Implementation of the ISO 8601 standard means that date/time variables are character/text data types. The SDTM fragment employed for date/time character variables is DTC. These variables record dates, times, or timing information. This guide also provides information on how partial dates (e.g., only month and year or only year known) should be recorded, and how to report duration of time in hours, minutes and seconds.

The SDTM implementation guide template uses ISO 8601 for calendar dates and times of day, which are expressed as follows:

- YYYY-MM-DDThh:mm:ss

where:

- [YYYY] = four-digit year
- [MM] = two-digit representation of the month (01-12, 01=January, etc.)
- [DD] = two-digit day of the month (01 through 31)
- [T] = (time designator) indicates time information follows
- [hh] = two digits of hour (00 through 23) (am/pm is NOT allowed)
- [mm] = two digits of minute (00 through 59)
- [ss] = two digits of second (00 through 59)

Other characters defined for use within the ISO 8601 standard are:

- [-] (hyphen): to separate the time Elements "year" from "month" and "month" from "day" and to represent missing date components.
- [:] (colon): to separate the time Elements "hour" from "minute" and "minute" from "second"
- [/] (solidus): to separate components in the representation of date/time intervals
- [P] (duration designator): precedes the components that represent the duration

Spaces are not allowed in any ISO 8601 representations.

Key aspects of the ISO 8601 standard are as follows:

- ISO 8601 represents dates as a text string using the notation YYYY-MM-DD.
- ISO 8601 represents times as a text string using the notation hh:mm:ss.
- The SDTM and SDTMIG require use of the ISO 8601 Extended format, which requires hyphen delimiters for date components and colon delimiters for time components. The ISO 8601 basic format, which does not require delimiters, should not be used in SDTM datasets.
- When a date is stored with a time in the same variable (as a date/time), the date is written in front of the time and the time is preceded with "T" using the notation YYYY-MM-DDThh:mm:ss (e.g., 2001-12-26T00:00:01).

For variables ending in --DTC, the YYYY-MM-DD format is used. The following examples illustrate how to report partial dates.

Example 1: How to report the date if only year and month are known.

Example 2: How to report the year if month and date are unknown.

	STUDYID	DOMAIN	USUBJID	RFSTDTC	RFPENDTC	BRTHDTC	AGE
1	SMART	DM	SMART-0201-00001	2019-12-22	2020-01-01	1962-06-17	58
2	SMART	DM	SMART-0201-00002	2018-03-23	2018-08	1989-01-11	31
3	SMART	DM	SMART-0201-00003	2019-01-23	2019-07-08	1977-11-05	42
4	SMART	DM	SMART-0201-00004	2019-03-07	2020-01-25	1971-07-08	49
5	SMART	DM	SMART-0201-00005	2018-05-11	2019-06-03	1960	60

The BAPAC Minimum Dataset contains a question asking participants to report the average time slept at night. This variable represents a duration of time, not a specific point in time, indicated by the letter P.

Example 3: Format for how QSSTRESC should be reported when QSTESTCD equals PSQI4. The “PT” designates a duration of time. The format is the number of hours, then the letter H, then the number of minutes, then the letter M.

QSCAT	QSSCAT	QSTESTCD	QSTEST	QSSTRESC	QSSTRESN
Sleep	Sleep Duration	PSQI4	How many hours and minutes of actual sleep at night	PT3H11M	3.1833333333
Sleep	Sleep Duration	PSQI4	How many hours and minutes of actual sleep at night	PT4H57M	4.95

2.2 VISIT/VISITNUM: Variables for Visit

CDISC SDTM standards prescribe how nominal study visit data should be stored. The following two variables are used to represent each visit.

- VISIT: Visit Description (text)
- VISITNUM: Numeric (for ordering data by nominal visit)

For BACPAC, these variables must apply uniformly to all studies to facilitate integration of data (e.g., VISIT has to be consistent across projects). Therefore, VISIT[NUM] will be represented based on a number of weeks. One month is assumed to be 4 weeks.

Table 2.2.1. BACPAC Nominal Study Visit Variable Structure

Nominal Timing of Collection	Visit Description (VISIT)	Visit Number (VISITNUM)
X weeks prior to protocol-specific reference visit	Week -X	-X
Protocol-specific reference visit (e.g., randomization visit)	Week 0	0
7 weeks post baseline	Week 7	7
3 months post baseline	Week 12	12
6 months post baseline	Week 26	26
30 weeks post baseline	Week 30	30
1 year post baseline	Week 52	52
2 years post baseline	Week 104	104

If multiple data collection days are planned for a single study visit, the following strategy should be employed:

- VISIT = "Week X - Visit 1", VISITNUM = X.01 (for first planned data collection visit associated with anchor Week X)
- VISIT = "Week X - Visit 2", VISITNUM = X.02 (for second planned data collection visit associated with anchor Week X)
- In summary, the "- Visit Y" prefix should be added to the VISIT label to indicate the anchor week (i.e., target for first component of data collection) and VISITNUM should use the hundredths place to indicate sequence of the separate visits (e.g., 0.01, 0.02, 0.03 and so on).

If more frequent visits are planned for a study, the following strategy should be employed:

- VISIT = "Week X - Day Z"
- VISITNUM = X.Z

If unscheduled visits occur after the Week X visit but before the Week Y visit, the following strategy should be employed:

VISIT = "Week X - Unscheduled" for all unscheduled visits

VISITNUM = X.01 (for first unscheduled visit in Week X)

VISITNUM = X.02 (for second unscheduled visit in Week X, and so on)

Example: A participant of study ABCDE who completed the study per protocol has planned visit schedule listed in Table 2.2.2.

Table 2.2.2. Visit Schedule for Participants in study ABCDE

Event	Planned Study Visit	VISIT	VISITNUM	Description
Screening week	Screening week	Week -1	-1	7-day period prior to date of randomization
Randomization, PRO and demographic data intake	Baseline visit 1	Week 0 - Visit 1	0.01	Date of randomization (Day 0); baseline visit including PRO data intake
Second collection of PRO data	Baseline visit 2	Week 0 - Visit 2	0.02	Second baseline visit where same measures are taken; should be at least 1 day but no more than 2 weeks after date of randomization
3-Month follow-up, PRO measures	Month 3 visit	Week 12	12	3-Month follow-up visit should be at least 10 days/weeks but no more than 14 weeks after date of randomization
4-Month follow-up, physical assessment	Month 4 visit	Week 16	16	4-Month follow-up visit should be at least 14 weeks but no more than 18 weeks after date of randomization
6-Month follow-up	Month 6 visit	Week 26	26	6-Month follow-up visit should be at least 24 weeks but no more than 26 weeks after date of randomization
Long-term follow-up	Follow-up period	Week 104	104	Longer-term follow-up period lasting up to 2 years after reference start date

PRO=participant-reported outcome

One participant of study ABCDE has the actual visit schedule listed in Table 2.2.3 because they were unable to complete all evaluations at the Week 16 visit and needed an unscheduled visit to complete them. Additionally, the participant knew they would be moving out of the area before the end of the 2-year long-term follow-up period and elected to have data collected after only 70 weeks.

Table 2.2.3. Actual Study Visits for Participant ABCDE-0101-00001

Event	Planned Study Visit	VISIT	VISITNUM	Description
Screening week	Screening week	Week -1	-1	7-day period prior to date of randomization
Randomization, PRO and demographic data intake	Baseline visit 1	Week 0 - Visit 1	0.01	Date of randomization (Day 0); baseline visit including PRO data intake
Second collection of PRO data	Baseline visit 2	Week 0 - Visit 2	0.02	Second baseline visit where same measures are taken; should be at least 1 day but no more than 2 weeks after date of randomization
3-Month follow-up, PRO measures	Month 3 visit	Week 12	12	3-Month follow-up visit should be at least 10 days/weeks but no more than 14 weeks after date of randomization
4-Month follow-up, physical assessment	Month 4 visit	Week 16	16	4-Month follow-up visit should be at least 14 weeks but no more than 18 weeks after date of randomization
4-Month follow-up, physical assessment	Month 4 visit	Week 16 – Unscheduled	16.01	Only half of the tests were performed at the first visit, so the participant had to return for an unscheduled visit
6-Month follow-up	Month 6 visit	Week 26	26	6-Month follow-up visit should be at least 24 weeks but no more than 26 weeks after date of randomization
Long-term follow-up	Follow-up period	Week 70 - Unscheduled	70.01	Longer-term follow-up period lasting up to 2 years after reference start date

2.3 RFSTDTC: Reference Start Date

The participant Reference Start Date (RFSTDTC) is generally defined as the “Week 0” visit. It is the date from which the timing of other study visits and study day will be derived.

Example: BACPAC Minimum Dataset to be collected 3 months (+/- 2 weeks) after the reference start date.

Reference Start Date Specific to RCTs

- Phase 2 and Clinical Trials: date of randomization visit
- Michigan SMART Trial: date of first randomization visit
- Collaborative Trial: TBD

Example Definitions of Reference Start Date Specific to Observational Studies

- Date of study visit corresponding to baseline data collection for items not extracted from EHR (e.g., baseline outcome measures).
- Date of consent when baseline data are collected across multiple days.
- Date of the first treatment if that date will be used as reference for the 3-month follow-up visit.

2.4 RFENDTC: Reference End Date

The participant Reference End Date (RFENDTC) is the date when the participant was determined to have ended the study. In the case of controlled randomized trials, it may correspond to the last date that a participant received treatment.

2.5 --DY: Study Day of Finding

The --DY variable is used in Findings Domains to report the day of the study that the data were collected, relative to the Reference Start Date as Day 1.

Study day of finding = [Current Date – Reference Start Date + 1]

The Subject Reference Start Date (RFSTDTC) is designated as Study Day 1. The Study Day value is incremented by 1 for each date following RFSTDTC.

Dates prior to RFSTDTC are decremented by 1, with the date preceding RFSTDTC designated as Study Day -1 (there is no Study Day 0).

This algorithm for determining Study Day is consistent with how people typically describe sequential days relative to a fixed reference point but creates problems if used for mathematical calculations because it does not allow for a Day 0.

Study Day is not suited for numerical computations (e.g., durations). The raw date values should be used rather than Study Day in those calculations.

2.6 --DTC: Date/Time of Finding

Within a given findings domain, the --DTC variable represents the exact date of data collection.

2.7 --EVLINT: Evaluation Interval

This is the duration of the interval associated with an observation, sometimes otherwise referred to as a recall period for participant reported outcome data. Variable values should be reported in ISO 8601 format. See section 2.1 for details on ISO 8601 formats for duration.

In the case of --EVLINT variables, the “-“ indicates a recall period prior to the time of data collection or measurement (e.g., "past 2 years" would translate to -P2Y).

Low-back Pain Specific Pain Intensity											
1. In the past 7 days... QSEVLINT = -P7D											
How would you rate your <i>low-back pain</i> on average?											
0	1	2	3	4	5	6	7	8	9	10	
No Pain						Worst imaginable pain					

Figure 2.7.1. Example from the BACPAC Minimum Dataset for QSTESTCD = LBPPINT1 where the evaluation interval is the past 7 days, represented as -P7D

3. Grouping and Synonym Qualifiers

3.1 --CAT: Category (Grouping Qualifier)

Units are required to use categories defined by the DSWG for the BACPAC Minimum Dataset and all broadly collected measures, including participant-reported outcome (PRO) data (QS domain), Treatment Categories Questionnaire data (EX domain), physical function and QST data (FT domain) and any harmonized biomechanics feature data.

For other data that are mapped to SDTM, use DSWG-recommended categories, or create categories according to working group recommendations.

3.2 --SCAT: Subcategory (Grouping Qualifier)

Units are required to use subcategories defined by the DSWG for the BACPAC Minimum Dataset and broadly collected measures.

For PRO data, the subcategory is the name of the CRF.

For other data that are mapped to SDTM, use DSWG-recommended subcategories, or create subcategories according to working group recommendations.

3.3 --TEST: Name of Measurement, Test or Examination (Synonym Qualifier)

For BACPAC datasets submitted to the BACPAC Data portal in BACPAC modified SDTM format, --TEST is the exact or truncated wording of the questions or prompts for measures. This variable is subject to controlled terminology on the Codelists sheet of the specifications file. Units are required to use the values for --TEST developed by the DSWG for BACPAC Minimum Dataset and other broadly collection data.

For tabular data outside of DSWG-developed standards, the maximum length for any value of --TEST may not exceed 100 characters.

3.4 --TESTCD: Abbreviated Test Code (Grouping Qualifier)

For BACPAC datasets, this is an up to 8-character code with a one-to-one mapping to a --TEST variable in the same domain. This variable is subject to controlled terminology. Units are required to use the values for --TESTCD developed by the DSWG for BACPAC Minimum Dataset and other broadly collected data.

For tabular data outside of the DSWG-developed standards, the maximum length for any value of --TESTCD may not exceed 8 characters. The first character cannot be a number, and the values of --TESTCD cannot contain characters other than letters, numbers and underscores.

4. Result Qualifiers

Result qualifiers describe the specific results associated with the topic variable in a Findings dataset.

4.1 --STRESC: Character Result or Finding in Standard Format

This is the standardized character value of the result or finding.

If --ORRES (see Section 4.3) is populated, --STRESC is the value of --ORRES converted to standard format. For example, in the BACPAC Minimum Dataset, if the value of SCORRES is 'Transgender Female' where SCTESTCD = GENIDENT, the value of SCSTRESC for that observation will be 'Female.' If the value of SCORRES where SCTESTCD = WEIGHT is recorded in pounds, the value of SCSTRESC will be the value of SCORRES converted to kilograms.

	SCTESTCD	SCTEST	SCORRES	SCORRESU	SCSTRESC
1	GENIDENT	Gender Identity	Transgender Female		Female

Figure 4.1.1. Data from a simulated BACPAC QS dataset where the original value “Transgender Female” maps to the standardized value “Female”

	SCTESTCD	SCTEST	SCORRES	SCORRESU	SCSTRESC	SCSTRESU
	WEIGHT	Weight at Baseline	194.6 LB	428.1	KG	

Figure 4.1.2. Data from a simulated BACPAC QS dataset where the original value, 194.6 pounds, is converted to standard units, 428.1 kilograms

Otherwise, values of --STRESC reflect reported findings and will match one of the code list values assigned to --STRESC at that observation's value of --TESTCD. For example, the following code list describes possible values of QSSTRESC where QSTESTCD = GAD01:

ID	Name	NCI Codelist Code	Data Type	Order	Term
LIKERT4A	Likert Scale 4-Point (A)		text		1 Not at all
LIKERT4A	Likert Scale 4-Point (A)		text		2 Several days
LIKERT4A	Likert Scale 4-Point (A)		text		3 More than half the days
LIKERT4A	Likert Scale 4-Point (A)		text		4 Nearly every day

Figure 4.1.3. LIKERT4A code list in the Codelists tab from the BACPAC CDISC modified-SDTM specifications file

When QSTESTCD = GAD01, QSSTRESC takes on values from the code list's "Term" column:

	QSTESTCD	QSTEST	QSSTRESC
1	GAD01	How often bothered by feeling nervous, anxious or on edge	Nearly every day
2	GAD01	How often bothered by feeling nervous, anxious or on edge	Several days
3	GAD01	How often bothered by feeling nervous, anxious or on edge	Nearly every day
4	GAD01	How often bothered by feeling nervous, anxious or on edge	Nearly every day
5	GAD01	How often bothered by feeling nervous, anxious or on edge	Several days
6	GAD01	How often bothered by feeling nervous, anxious or on edge	Nearly every day
7	GAD01	How often bothered by feeling nervous, anxious or on edge	Several days
8	GAD01	How often bothered by feeling nervous, anxious or on edge	Nearly every day
9	GAD01	How often bothered by feeling nervous, anxious or on edge	More than half the days

Figure 4.1.4. Data from a simulated BACPAC QS dataset

4.2 --STRESN: Numeric Result or Finding in Standard Format

This is the standardized numeric result or finding.

If --ORRES (see Section 4.3) is populated and contains numeric values, --STRESN is the value of --ORRES converted to standard units. In the BACPAC Minimum Dataset, for example, if the value of SCORRES where SCTESTCD = WEIGHT is recorded in pounds, the value of SCSTRESN will be the value of SCORRES converted to kilograms.

Otherwise, values of --STRESN will match one of the code list values assigned to --STRESN at that observation's value of --TESTCD.

4.3 --ORRES: Result or Finding in Original Units

This is the value of a result or finding as it was originally recorded. The value of --ORRES is used to assign a standardized value to --STRESC, which is then used to assign a value to --STRESN when appropriate.

For the BACPAC SC dataset, SCORRES is populated only for records with data that needs to be standardized, i.e., when SCTESTCD = GENIDENT, HEIGHT, or WEIGHT.

When SCTESTCD = HEIGHT or WEIGHT, the value of SCORRES is the value from the completed case report form:

20. *Height 234.8 [inches] cm [measured] [self-reported]

21. *Weight 220.2 pounds [kilograms] [measured] [self-reported]

	SCTESTCD	SCTEST	SCORRES	SCSTRESC	SCSTRESN
1	HEIGHT	Height at Baseline	234.8	234.8	234.8
2	WEIGHT	Weight at Baseline	220.2	484.4	484.4

Figure 4.3.1. How data from a CRF translates to the simulated data from the BACPAC Minimum Dataset when units are standardized

When SCTESTCD = GENIDENT, the value of SCORRES is the value selected on the case report form, or the value entered in the “Other” blank on the case report form:

4. *Gender identity ☒ Male ☐ Female ☐ Unknown Other, Specify _____

	SCTESTCD	SCTEST	SCORRES	SCSTRESC	SCSTRESN
1	GENIDENT	Gender Identity	Male	Male	.

Figure 4.3.2. How data from a CRF translates to the simulated data from the BACPAC Minimum Dataset when standardized values are selected

4. *Gender identity ☐ Male ☐ Female ☐ Unknown Other, Specify Transgender Female

	SCTESTCD	SCTEST	SCORRES	SCSTRESC	SCSTRESN
3	GENIDENT	Gender Identity	Transgender Female	Female	.

Figure 4.3.3. How data from a CRF translates to the simulated data from the BACPAC Minimum Dataset when values are entered manually and standardized

5. Record Qualifiers

Record qualifiers define additional attributes of the observation record as a whole, rather than describing a particular variable within a record.

5.1 --DRVFL: Derived Flag

This variable is used to indicate a derived record, such as an average or score. The value of --DRVFL is either Y or null.

The derived flag variable is not intended to indicate processed data, but instead indicates a record is derived from other submitted data, based on an algorithm or calculation in the case report form, SOP, or additional documentation available to analysts.

	QSTESTCD	QSTEST	QSSTRESC	QSSTRESN	QSDRVFL
5	PC27R	I have been able to remember to do things like take medicine or buy something I needed	Somewhat	3	
6	PC6R	I have been able to concentrate	A little bit	2	
7	PRCF2AR	PROMIS-Cognitive Function 2a Raw Score	5.0	5	Y
8	PRCF2AT	PROMIS-Cognitive Function 2a T-Score	41.2	41.2	Y

Figure 5.1.1. Values of QSDRVFL in a simulated BACPAC Minimum Dataset

The Codelists spreadsheet for the BACPAC Minimum Dataset specifications file contains additional columns which are not necessary for submission to the BACPAC Data Portal or for creation of a Define XML file. One of these optional columns, column K, is an indicator for whether a --TESTCD measure is derived from other data in the dataset.

A	E	F	G	K
ID	Order	Term	NCIT	Deriv
QSTESTCD	133	PC6R		
QSTESTCD	134	PC27R		
QSTESTCD	135	PRCF2AR		Y
QSTESTCD	136	PRCF2AT		Y

Figure 5.1.2. An optional column in the example BACPAC Minimum Dataset specifications file is used to indicate derived records

If there are additional derived records in the dataset being submitted to the BACPAC Data Portal, the --DRVFL variable should be used to indicate these records and the calculations for these records should be reported in accompanying documentation.

5.2 --METHOD: Method of Test or Examination

The --METHOD variable describes the method of a test or examination. In the BACPAC Minimum Dataset, --METHOD is only used in the SC domain, and is only populated where SCTESTCD = HEIGHT or WEIGHT. Height and weight are both either MEASURED or SELF-REPORTED.

	SCTESTCD	SCTEST	SCMETHOD	SCORRES	SCORRESU
1	HEIGHT	Height at Baseline	SELF-REPORTED	234.8	
2	WEIGHT	Weight at Baseline	SELF-REPORTED	220.2	LB
3	HEIGHT	Height at Baseline	SELF-REPORTED	179.3	CM
4	WEIGHT	Weight at Baseline	SELF-REPORTED	73.2	KG
5	HEIGHT	Height at Baseline	MEASURED	60	IN
6	WEIGHT	Weight at Baseline	MEASURED	69.6	KG
7	HEIGHT	Height at Baseline	MEASURED	58.1	IN
8	WEIGHT	Weight at Baseline	MEASURED	62	KG

Figure 5.2.1. Example values of SCMETHOD in the simulated BACPAC Minimum Dataset

For other tabular data submitted to the BACPAC Data Portal, the variable --METHOD can be used to distinguish between similar measures with differing SOPs or methods of measurement. These differences should be noted in accompanying documentation.

5.3 AGE

The AGE variable is specific to the DM domain, and contains the subject's age. In the BACPAC Minimum Dataset, age is reported in years and stored as an integer.

5.4 SEX

The SEX variable is specific to the DM domain, and contains each subject's sex assigned at birth.

Table 5.4.1. Possible values of SEX

SEX
Female
Male
Intersex
Unknown

5.5 RACE

The RACE variable is specific to the DM domain and contains each subject's race. If multiple values are selected, the value of RACE is Multiple.

Table 5.5.1. Possible values of RACE

RACE
American Indian or Alaska Native
Asian
Black or African American
Native Hawaiian or Pacific Islander
White
Unknown
Not reported
Multiple

5.6 RACEMULT

RACEMULT is specific to the DM domain. If RACE=Multiple, RACEMULT contains all selected values of race separated by semicolons. Otherwise, RACEMULT is left blank.

	USUBJID	RACE	RACEMULT
38	SMART-0201-00038	Black or African American	
39	SMART-0201-00039	Multiple	Asian;White

Figure 5.6.1. Example values of RACE and RACEMULT in the simulated BACPAC Minimum Dataset

5.7 ETHNIC

ETHNIC is specific to the DM domain and contains each subject's ethnicity.

Table 5.7.1. Possible values of ETHNIC

ETHNIC
Hispanic or Latino
Not Hispanic or Latino
Unknown
Not reported

5.8 BRTHDTC

BRTHDTC is specific to the DM domain and contains each subject's date of birth. BRTHDTC is stored in the ISO 8601 format described in Section 2.1.

	USUBJID	BRTHDTC	AGE
1	SMART-0201-00001	1962-06-17	58
2	SMART-0201-00002	1989-01-11	31
3	SMART-0201-00003	1977-11-05	42
4	SMART-0201-00004	1971-07-08	49
5	SMART-0201-00005	1960-05-19	60
6	SMART-0201-00006	1980-01-03	40
7	SMART-0201-00007	1998-03-19	22
8	SMART-0201-00008	2000-01-28	20

Figure 5.8.1. Values of BRTHDTC in the simulated BACPAC Minimum Dataset

6. Variable Qualifiers

Variable qualifiers are used to further modify or describe a specific variable within an observation and are only meaningful in the context of the variable they qualify.

6.1 --STRESU: Standard Units of Result or Finding

The units of --STRESC/--STRESN are stored in --STRESU. In the BACPAC Minimum Dataset, SCSTRESU will be populated only when SCTESTCD = HEIGHT or WEIGHT. The standardized units for height and weight are centimeters (CM) and kilograms (KG), respectively.

	SCTESTCD	SCSTRESC	SCSTRESN	SCSTRESU
1	HEIGHT	234.8	234.8	CM
2	WEIGHT	484.4	484.4	KG
3	HEIGHT	179.3	179.3	CM
4	WEIGHT	73.2	73.2	KG

Figure 6.1.1. Example values of SCSTRESU when SCTESTCD = HEIGHT or WEIGHT in a simulated BACPAC Minimum Dataset

6.2 --ORRESU: Units of Original Result or Finding

The units of --ORRES are stored in --ORRESU. In the BACPAC Minimum Dataset, SCORRES will be populated when SCTESTCD = HEIGHT or WEIGHT. Possible values for SCORRESU when SCTESTCD = WEIGHT are LB and KG, and possible values when SCTESTCD = HEIGHT are IN and CM.

	SCTESTCD	SCORRES	SCORRESU
1	HEIGHT	234.8	CM
2	WEIGHT	220.2	LB
3	HEIGHT	179.3	CM
4	WEIGHT	73.2	KG
5	HEIGHT	60	IN

Figure 6.2.1. Example values of SCORRESU when SCTESTCD = HEIGHT or WEIGHT in a simulated BACPAC Minimum Dataset

7. Variables Specific to Treatment Categories (EX Domain)

7.1 EXTRT: Name of Treatment

The treatment names are specific to the EX domain and are specific to the Treatment Categories Questionnaire CRF.

Table 7.1.1. Possible values of EXTRT

Treatment (EXTRT)	Category (EXCAT)
Spinal fusion	Surgery
Non-spinal fusion	Surgery
Low back pain injection	Injection
Opioids	Medication
SSRI_SNRI	Medication
Gabapentin or pregabalin	Medication
Tricyclic antidepressants	Medication
NSAIDs	Medication
Adjustment or manipulation	PT, OT, or chiropractic
Other	Medication; PT, OT, or chiropractic
Exercise	Diet and exercise
Acupuncture	Alternative medicine
Therapy or counseling	Mental health
Mindfulness or meditation or relaxation	Mental health
Diet or weight loss program	Diet and exercise
Active PT or OT	PT, OT, or chiropractic
Other passive PT	PT, OT, or chiropractic

7.2 EXTRTOTH: Other Treatment

Questions 3 and 4 of the Treatment Categories Questionnaire have an option for “other” treatments within the treatment categories of Medication and PT, OT or chiropractic. EXTRTOTH is the free-text response for other treatment up to 100 characters.

DOMAIN	USUBJID	EXSEQ	EXCAT	EXTRT	EXTRTOTH	EXACN
EX	COHORT01-0101-00002	1	Diet and exercise	Exercise		
EX	COHORT01-0101-00002	2	PT, OT, or chiropractic	Other	Warm baths/ heating pad	
EX	COHORT01-0101-00002	3	Diet and exercise	Exercise		
EX	COHORT01-0101-00002	4	Diet and exercise	Exercise		
EX	COHORT01-0101-00002	5	Diet and exercise	Exercise		

Figure 7.2.1. Example of EXTRTOTH where EXCAT equals “PT, OT, or chiropractic”

7.3 EXACN: Action Taken with Treatment

EXACN indicates whether a treatment was started, or the dose was increased, decreased, or unchanged since the last time the Treatment Categories Questionnaire was completed.

Table 7.3.1. Possible values of EXACN

EXACN
New prescription/medication
Dose increased
Dose decreased
Dose unchanged

7.4 EXADJ: Reason for Treatment Adjustment

EXADJ indicates that a treatment was recommended or prescribed. If the treatment was not recommended or prescribed, this variable is left blank.

Table 7.4.1. Possible values of EXADJ

EXADJ
Recommended or prescribed

7.5 EXDOSE: Dose of Treatment

When the value of EXTRT is "Exercise," EXDOSE is an integer ranging from 0-7, indicating how many days in the past week the participant exercised. This variable is left blank if the participant answered "No" to question 5 in the Treatment Categories Questionnaire.

If additional EXTRT values are added to the specifications, EXDOSE can take a floating-point value for other values of EXTRT.

7.6 EXDOSEU: Units of Treatment Dose

When the value of EXTRT is "Exercise," the value of EXDOSEU is "Days per week." Additional values of EXDOSE and EXDOSEU may be added to the specifications in the future. If EXDOSE is missing, EXDOSEU is left blank.

7.7 EXROUTE: Route of Administration

EXROUTE indicates that a treatment was recommended, prescribed, or delivered via telehealth. If the treatment was not recommended, prescribed, or delivered via telehealth, this variable is left blank.

Table 7.7.1. Possible values of EXROUTE

EXROUTE
Telehealth

8. Variables Specific to Functional Tests (FT Domain)

8.1 FTPOS: Position of Participant During Observation

FTPOS is the position of the participant during the test.

Table 8.1.1. Possible values of FTPOS

FTPOS
Sitting
Standing
Prone
Supine
Side-lying

8.2 FTLOC: Location Used for the Measurement

FTLOC is the anatomical location of the participant relevant to the collection of the measurement.

Table 8.2.1. Selected example values of FTLOC

FTLOC
L4: Anteromedial Lower Leg
L5: Distal Anterior Tibia and Dorsum of Foot
S1: Posterolateral Calf and Foot
Lumbar L5
Trapezius
Lower Back

8.3 FTLAT: Laterality

FTLAT is a qualifier variable for the anatomical location of the participant further detailing laterality.

Table 8.3.1. Possible values of FTLAT

FTLAT
Right
Left

8.4 FTREPNUM: Repetition Number

FTREPNUM is the incidence number of a test that is repeated within a given timeframe for the same test. For BACPAC Physical Function and QST measures, the repetition number is recorded for trials of the measure within the same VISIT/VISITNUM. FTREPNUM takes an integer value.

8.5 FTSTAT: Completion Status

Completion status is used to indicate that a test was not done, or a test was attempted but did not generate a result. If the test was not attempted or a result was not generated, the value of FTSTAT = Not done, otherwise the value is left blank.

Table 8.5.1. Possible values of FTSTAT

FTSTAT
Not done

8.6 FTREASND: Reason Not Done

FTREASND is the reason the test was not done or did not generate a result. This variable should only be used if the value of FTSTAT = Not done. FTREASND is a free text character variable. Some values of FTREASND are indicated on annotated case report forms and other values may be taken from comment fields from case report forms.

Scoring (CRF):	FTSTAT = Not Done, if Q1=No, where FTTESTCD=SIT2ST5X
1. Was test attempted?	<input type="checkbox"/> Yes <input type="checkbox"/> No
2. Safe to stand without help (<i>practice trial</i>)	<input type="checkbox"/> Yes <input type="checkbox"/> No
3. Time to complete five stands: (Only enter if participant completes 5 stands)	<input type="text"/> seconds.

If Q2=No, FTREASND=Not safe to stand without help

Figure 8.6.1 Example of a text value of FTREASND indicated on the annotated case report form for the Five Times Sit to Stand Test.

8.7 FTAID: Aid Used During Test

FTAID is a variable unique to the Functional Tests domain and is not a CDISC SDTM standard variable. FTAID indicates an aid was used during a test. Not all case report forms indicate whether an aid was used, so this variable can remain blank.

Table 8.7.1. Possible values of FTAID

FTAID
None
Cane
Other

8.8 FTAIDOTH: Other Aid Used During Test

FTAIDOTH is a free text character variable that indicates which aid was used during a test when FTAID = Other. If FTAID takes a value other than "Other" then FTAIDOTH can be blank. If an aid was used that is listed in the comment section of the case report form, that aid can be listed as the value of FTAIDOTH.



NIH HEAL Initiative and Helping to End Addiction Long-term are service marks of the U.S. Department of Health and Human Services.