

Subject-Level Variables for Oncology in the Subject-Level Analysis Dataset (ADSL)

Madhusudhan Reddy Papasani, Ginger Redner
Department of Biostatistics and Research Decision Sciences (BARDS), Merck & Co., Inc., Kenilworth, NJ, USA

INTRODUCTION AND PURPOSE

Individual subject characteristics are important for the analysis of oncology data, and there is great interest in understanding the initial diagnosis, cancer staging, prior disease progression, pathology, prior treatments/therapies, and performance status at baseline. These subject characteristics play a vital role in determining treatment options and evaluating patient response to treatments. ADSL dataset includes subject-level information. Currently, the ADaM model does not define standard variables for oncology therapeutic areas but recommends creating sponsor-specific subject-level variables following the ADaM guidelines. Sponsor-defined variables may work well within a given submission; however, there is a need for industry-wide standardization. The purpose of this presentation is to explain the concepts behind these subject-level characteristics in the context of analysis, as well as to propose standard variables for common cancer characteristics and associated metadata in alignment with the ADaM guidelines and oncology Therapeutic Area User Guides (TAUG).

Oncologic-Specific Subject-Level Variables

Important subject-level characteristic variables are listed in this section. It is to be noted that the subject-level data is usually presented in ADSL, though the presented standard variables can be used irrespective of the analysis dataset in which they are derived. The code lists presented in this section may not contain all possible values.

Initial Diagnosis

For clinical trials, data collected on initial diagnosis of cancer may include the date of diagnosis.

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
IDIAGDT	Date of initial diagnosis	Integer		Diagnosis integer; set to date part of MH.MHSTDTC, where MH.MHEVDTYPE = "INITIAL DIAGNOSIS" captured as a numeric variable.

Cancer Staging

The cancer stage is determined based on the primary tumor and the extent of cancer spread within the body. It is common to use Tumor-Node-Metastasis (TNM) methodology from the American Joint Committee on Cancer (AJCC) to assign stages to a cancer that indicate severity/extent of cancer at screening. Cancer staging is an important factor for cancer prognosis, population-based screening, and early detection methods. In TNM, T stands for extent of primary tumor, N stands for regional lymph node involvement, and M stands for distant metastasis. The following table presents the variable-level metadata for these variables.

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
PSTAGE	Stage of primary tumor	Char	TX, T0, ST1, T1A, T1B, T2, T2A, T2B, T3, T3A, T3B, T4, T4A, T4B	This is as per TNM classification pertaining to primary tumor stage. Please note that primary tumor stage is collected in SDTM and can have a code list that reflects the values of SDTM. It is anticipated that TNM staging will be represented in the RS domain. For example: TX: Can now be assessed. T0: No evidence of primary tumor. The following indicates increasing size or extent of primary tumor involvement in the order it is given: TX, T0, ST1, T1A, T1B, etc.
NSTAGE	Stage of nodal involvement	Char	NX, N0, N1, N2, N3	Stage of nodal involvement is collected in SDTM and can have a code list that reflects the values of SDTM. It is anticipated that TNM staging will be represented in the RS domain. For example: NX: Cannot be assessed. N0: No evidence for lymph node metastasis. N1, N2, and N3 indicate increasing extent of nodal involvement.
MSTAGE	Stage of distant metastasis	Char	MX, M0, M1, M1B	Stage of distant metastasis involvement is collected in SDTM and can have a code list that reflects the values of SDTM. It is anticipated that TNM staging will be represented in the RS domain. For example: MX: Cannot be assessed. M0: No evidence for distant metastasis. M1: Distant metastasis and increase as below. M1B

Cancer Staging (continued)

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
STAGE	Stage of overall cancer	Char	Stage 1, Stage 2, Stage 3, Stage 4, IA, IB, IIA, IIB, IIIA, IIIB, IV	This is the stage of cancer that the patient was assessed with at the point of inclusion in the study. Stage is normally captured in an SDTM domain. It is anticipated that staging will be represented in the RS domain. The code list can be a custom code list that reflects what is in the SDTM domain. It can have as its controlled terminology (Stage 1, Stage 2, Stage 3, Stage 4), or it can use the TNM classification (T, N, M) as below: IA, IB, IIA, IIB, IIIA, IIIB, IV

Pathology: This data includes carcinoma detection, grading, and assessment of biomarkers using molecular biology methodologies.

Molecular Markers by Microscopy

Assessment of biomarkers is usually done by microscopic observations and usually presented with grading or positive/negative. Biomarker data variable names should follow the convention of having the biomarker indicator within the first 5 characters, followed by the suffix STT. **NOTE: Breast cancer TAUG uses STAT.**

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
HER2STT	HER2 status	Char	1+; 2+; 3+; Positive; Negative	This is the HER2 assessment of the subject at randomization. HER2STAT is equal to MI.MISTRESC when MI.MITEST represents HER2 status. MI.MITEST is subject to sponsor-controlled terminology.
ERSTT	Estrogen receptor status	Char	1+; 2+; 3+; Positive; Negative	Receptor status Char 1+; 2+; 3+; Positive; Negative. This is the ER assessment of the subject at randomization. ERSTAT is equal to MI.MISTRESC when MI.MITEST represents estrogen receptor status. MI.MITEST is subject to sponsor-controlled terminology.
PRSTT	Progesterone receptor status	Char	1+; 2+; 3+; Positive; Negative	Receptor status Char 1+; 2+; 3+; Positive; Negative. This is the PR assessment of the subject at randomization. PRSTAT is equal to MI.MISTRESC when MI.MITEST represents estrogen receptor status. MI.MITEST is subject to sponsor-controlled terminology.
TNBCSTT	TNBC status		Positive, Negative, Missing	If HER2STAT, PRSTAT, and ERSTAT are all 'NEGATIVE', then TNBCSTAT = 'Positive'; else if all three are missing or 'UNKNOWN', then Missing; else 'Negative'.
CDX2STT	CDX2 status	Char		Lack of CDX2 expression has been shown to have poor prognostic significance for 5-year survival. The test is conducted on tumor sample using IHC method. CDX2STAT is equal to MI.MISTRESC when MI.MITEST represents CDX2 status. MI.MITEST is subject to sponsor-controlled terminology.
PDL1STT	PDL1 status			If MOREES = 0, then PDL1STAT = 'Negative'. Else if MOREES >0 then PDL1STAT = 'Positive'.
*STT	{Status}			Biomarker levels that are either presented as with grading or positive or negative.

Molecular Markers by Laboratory

Laboratory assessments that have continuous data such as PSA, CTC, or testosterone are conducted at baseline for cancer subjects.

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
PSABL	PSA (unit) at baseline	Integer		LB.LBSTRESN where LB.LBTESTCD = 'PSA' and LB.LBBLFL = 'Y'.
CTCBL	Circulating tumor cells (unit) at baseline	Integer		LB.LBSTRESN where LB.LBTESTCD = 'CTC' and LB.LBBLFL = 'Y'.
TESTOSBL	Testosterone (unit) at baseline	Integer		LB.LBSTRESN where LB.LBTESTCD = 'TESTOS' and LB.LBBLFL = 'Y'.
*BL	{Baseline}			Numeric data at baseline.

Overall Histological Classification

The following variables define subject data indicating histological classification at diagnosis.

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
HISTOLGY	Histopathology	Char		Histology of the tumor is normally captured in the MI domain. Examples may include the following: adenocarcinoma, squamous, adenosquamous, etc.

Prior Disease Progression

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
PROGTYBL	Progression type at baseline	Text	PSA; radiographic	This is the progression type in which the patient was assessed at the point of inclusion in the study. The code list can be a custom code list that reflects the types of progression that are of interest in the study (for example, at the baseline visit, progression type = PSA). If RS.RSTESTCD = 'IMRESP' and RSSTRESC = 'PD' and RS.RSTESTCD = 'RDIORESP' and RSTRESC = 'NON-PD' and VISIT = 'SCREENING' Radiographic: If RS.RSTESTCD = 'RDIORESP' and RSSTREC = 'PD' and VISIT = 'SCREENING'.

Prior Treatments/Therapies

If the subjects are not newly diagnosed with cancer when they join the study, any prior antineoplastic therapies (ie, anticancer treatments) may be of special interest to the study. Even if subjects are newly diagnosed, it may still be relevant to collect prior treatments for any other cancers they may have had.

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
PRADIAFL	Prior radiation flag	Text	Y; N	Y if there is at least one record with PR.PRSTAT = 'RADIATION' and PR.PRSTDTCTC is prior to DM.RFXSTDTCTC, otherwise N.
PSURGFL	Prior surgery flag	Text	Y; N	Y if there is at least one record with PR.PRSTAT = 'SURGERY' and PR.PRSTDTCTC is prior to DM.RFXSTDTCTC, otherwise N.
PCHEMOFL	Prior chemotherapy flag	Text	Y; N	Y if there is at least one record with PR.PRSTAT = 'CHEMOTHERAPY' and PR.PRSTDTCTC is prior to DM.RFXSTDTCTC, otherwise N.
PRADIOFL	Prior radiopharmaceutical flag	Text	Y; N	Y if there is at least one record with PR.PRSTAT = 'RADIOPHARMACEUTICAL' and PR.PRSTDTCTC is prior to DM.RFXSTDTCTC, otherwise N.
PTRTEDT	Prior treatment end date	Integer		If a patient has had prior cancer treatment, capture the date that the last treatment ended. For example: latest CM.CMENDTCTC where CM.CMSTAT = 'ANTINEOPLASTIC THERAPY' or latest PR.PRENDTCTC where PR.PRSTAT = 'RADIATION THERAPY' where --DTC variables are converted to numeric date format. The date part of CM.CMENDTCTC captured as a numeric variable.

Performance Status at Baseline – Eastern Cooperative Oncology Group Performance Status (ECOG)

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
ECOGBL	ECOG at baseline	Integer	0; 1; 2; 3; 4	QS.QSSTRESN where QS.QSTESTCD = 'ECOG101' and QSBLFL = 'Y'.

Substance status

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
SMOKSTT	Smoking Status	Char		Note that one subject may have multiple records in SU. If there is only one record with SUTRT = 'TOBACCO' and SUOCCUR = 'N', then 'Never Smoker'; else if there are multiple records, if there is at least one record with SUTRT = 'CIGARETTE', 'OTHER' and SUENRF = 'DURING/AFTER' then 'Current Smoker'; else if all records are with SUTRT in 'CIGARETTE', 'OTHER' and SUENRF = 'BEFORE' then 'Former Smoker'.
*STT	{Status}	Char		Substance abuse status examples include alcohol status, drug use status, and ending with *STT. This information usually derived from SDTM.SU domain.

Tumor Load and Metastasis by Location

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
TLSDBL	Target lesions sum of diameters at baseline	Integer		TR.TRSTRESN where TR.TRGRPID = 'GEFCF' and TR.TRTESTCD = 'SUMDIAM' and TRCAT = 'TARGET' and TRSTAT ne 'NOT DONE' and choose the latest observation where TRDY ≤1 and VISIT in ('Screening', 'Cycle 1 Visit 1', 'OTHER') and SUENRF = 'DURING/AFTER' then 'Current Smoker'; else if all records are with SUTRT in ('CIGARETTE', 'OTHER') and SUENRF = 'BEFORE' then 'Former Smoker'.
MSBONE	Metastasis at bone	Char		MH.external data where MH.MHGRPID = 'CDDB', 'Yes' if 'BONE' in MHL.MET1-MHL.MET10 or lowcase (MHL.METOT) has record with BONE = 'Y' in METLOC_CDDB.xlsx file. 'Yes' if 'BONE' in MH.LOCMET1-MH.LOCMET10 or lowcase (MH.LOCMETOT) has record with BONE = 'Y' in METLOC_CDDB.xlsx file. 'No', otherwise. Set to NULL if no CDDB record.
MSBRAIN	Metastasis at bone	Char		Where MH.MHGRPID = 'CDDB', 'Yes' if 'BRAIN' in MHL.MET1-MHL.MET10 or lowcase (MHL.METOT) has record with BRAIN = 'Y' in METLOC_CDDB.xlsx file. 'Yes' if 'BRAIN' in MH.LOCMET1-MH.LOCMET10 or lowcase (MH.LOCMETOT) has record with BRAIN = 'Y' in METLOC_CDDB.xlsx file. 'No', otherwise. Set to NULL if no CDDB record.
MSCHEST	Metastasis at bone	Char		Where MH.MHGRPID = 'CDDB', 'Yes' if 'CHEST WALL' in MHL.MET1-MHL.MET10 or lowcase (MHL.METOT) has record with CHEST_WALL = 'Y' in METLOC_CDDB.xlsx file. 'Yes' if 'CHEST WALL' in MH.LOCMET1-MH.LOCMET10 or lowcase (MH.LOCMETOT) has record with CHEST_WALL = 'Y' in METLOC_CDDB.xlsx file. 'No', otherwise. Set to NULL if no CDDB record.
MSLUNG	Metastasis at lung	Char		Where MH.MHGRPID = 'CDDB', 'Yes' if 'LUNG' in MHL.MET1-MHL.MET10 or lowcase (MHL.METOT) has record with LUNG = 'Y' in METLOC_CDDB.xlsx file. 'Yes' if 'LUNG' in MH.LOCMET1-MH.LOCMET10 or lowcase (MH.LOCMETOT) has record with LUNG = 'Y' in METLOC_CDDB.xlsx file. 'No', otherwise. Set to NULL if no CDDB record.
MSLNL	Metastasis at lymph nodes	Char		Where MH.MHGRPID = 'CDDB', 'Yes' if 'LYMPH NODE' in MHL.MET1-MHL.MET10 or MHL.METOT has record with LYMPH_NODE = 'Y' in METLOC_CDDB.xlsx file. 'Yes' if 'LYMPH NODE' in MH.LOCMET1-MH.LOCMET10 or lowcase (MH.LOCMETOT) has record with LYMPH_NODE = 'Y' in METLOC_CDDB.xlsx file. 'No', otherwise. Set to NULL if no CDDB record.
MSOTHR	Metastasis at other sites	Char		Where MH.MHGRPID = 'CDDB', 'Yes' if lowcase (MHL.METOT) has record that has Number_of_Locations_Other_Metastasis >0 in METLOC_CDDB.xlsx file. 'No', otherwise. Set to NULL if no CDDB record.
MS*	{Metastasis at}	Char		Prefixing MS followed by abbreviated location or site.
NMSBL	Number of metastatic organs at baseline	Integer		Derived by summing all the sites that have metastasis.
NMSBLGRY	Number of metastatic organs at baseline Group 1	Char		Grouped as per study specifications.

Other Dates of Interest

Variable Name	Variable Label	Type	Controlled Terms or Format	Notes
NCTXSDT	Start date of new anticancer therapy	Integer	Date format	The earliest date converted to numeric from the following records (after day imputation if necessary): CM.CMSTDTCTC where CM.CMGRPID = 'ODBF'. If day is missing, impute with day = 01 (ie, use CM.CMSTDTI instead of CMSTDTCTC) and flag in NCTXSDTIF = 'D'. Please generate a list with partial dates and send to DM for query.

CONCLUSION

The variables and associated metadata described here in the context of oncology therapeutic area would allow for industry-wide standardization. These oncology subject-level variables are currently being discussed within the ADaM oncology team to be incorporated in Analysis Data Model (ADaM) Common Methods Used in Oncology Analytics.