Paper AS02

Unravelling the complexity: Efficacy Datasets in Oncology Basket Trials Across Multiple Disease Tumors

Venkatesh Burla, Merck & Co., Inc., Rahway, NJ, USA
Manasa Ravi, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

Oncology basket trails are designed to assess the efficacy of treatments across various cohorts. The ADaM standards established by CDISC provide a framework to standardize the structure and format of analysis datasets in clinical trials, ensuring consistency and facilitating data exchange among researchers and regulatory authorities. This paper aims to outline the development of ADaM datasets within the context of a basket trial, focusing on a cohort with Von Hippel-Lindau (VHL) disease, where subjects exhibit multiple diseases, including Renal Cell Carcinoma (RCC), Pheochromocytoma and Paraganglioma (PPGL), pancreatic Neuroendocrine Tumor (pNET), and Central Nervous System (CNS) Hemangioblastoma. Key endpoints include Progression-Free Survival (PFS), Overall Survival (OS), Best Overall Response, and Tumor Percentage Change from baseline. Importantly, these endpoints are to be derived for each disease indication within a subject, reflecting the varied response dynamics observed across different tumor types within the same individual.

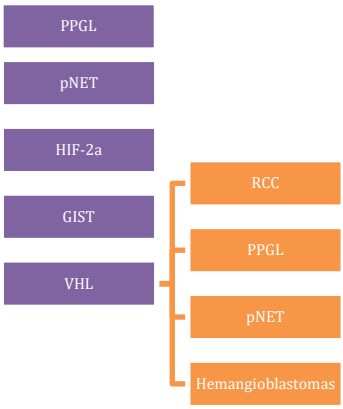
INTRODUCTION

In clinical trials, the ADaM (Analysis Data Model) standards play a crucial role in structuring datasets for analysis. These standards ensure that data is organized consistently and is analysis ready, which is essential for regulatory submissions and the evaluation of trial results. In the context of oncology basket trials, where multiple tumor types are studied simultaneously under a common treatment, ADaM datasets help capture and manage the complexity of such trials, particularly when it comes to efficacy endpoints.

In oncology basket trials, subjects may have multiple tumors with each tumor indicating different disease. This introduces challenges in defining and capturing data related to time-to-event outcomes, response rates, and progression across these tumor types. ADaM datasets like ADTL (Tumor Lesions), ADRS (Response Assessments), ADINTDT (Intermediate Dataset for Dates), and ADTTE (Time-to-Event) are pivotal in organizing and analyzing these endpoints. In this paper, we will discuss different approaches taken when developing efficacy ADaM datasets for subjects with multiple tumors, focusing on the derivation of endpoints for each subject and tumor.

STUDY OVERVIEW

Study to Evaluate the Efficacy and Safety of a study treatment in Participants with Advanced Pheochromocytoma/Paraganglioma (PPGL), Pancreatic Neuroendocrine Tumor (pNET), Von Hippel-Lindau (VHL) Disease-Associated Tumors, Advanced Gastrointestinal Stromal Tumor (wt GIST), or Advanced Solid Tumors With HIF-2a Related Genetic Alterations.



DATA COLLECTION (TR AND RS)

When multiple tumor types are treated under a treatment arm, it is essential to derive efficacy endpoints separately for each tumor type as discussed. This requires the collection of data in key datasets such as the Tumor Results (TR) and Response Assessments (RS) domains. For accurate analysis, data on tumor measurements and response evaluations must be captured distinctly for each tumor type. Therefore, it is crucial that the TUMORTYP variable is

included in these datasets to differentiate between the various tumor types being studied. This approach ensures that the efficacy assessments are specific to each tumor, allowing for a more precise evaluation of the treatment's impact on different tumor types within the trial.

Example for TR:

USUBJID	TUMORTYP	TRTESTCD	TRTEST	TRTSTDTL	TRDTC	TRORRES	TRORRESU	TRBLFL
101	RCC	LDIAM	Longest Diameter	TARGET	2024-01-01	50	mm	Y
101	RCC	LDIAM	Longest Diameter	TARGET	2024-01-15	45	mm	
101	PPGL	LDIAM	Longest Diameter	TARGET	2024-01-01	40	mm	Y
101	PPGL	LDIAM	Longest Diameter	TARGET	2024-01-15	35	mm	
101	pNET	LDIAM	Longest Diameter	TARGET	2024-01-01	30	mm	Y
101	pNET	LDIAM	Longest Diameter	TARGET	2024-01-15	28	mm	

Example for RS:

USUBJID	TUMORTYP	RSTESTCD	RSTEST	RSORRES	RSDTC	RSSEQ	RSBLFL
101	RCC	OVRLRESP	Overall Response	PR	2024-01-15	1	
101	PPGL	OVRLRESP	Overall Response	CR	2024-01-15	2	
101	pNET	OVRLRESP	Overall Response	SD	2024-01-15	3	
101	RCC	OVRLRESP	Overall Response	PD	2024-02-01	4	
101	pNET	OVRLRESP	Overall Response	PR	2024-02-01	5	

ANALYSIS DATASET FOR TARGET LESIONS (ADTL)

The ADTL dataset is a key efficacy dataset used in oncology trials to capture data related to target lesions, selected at baseline for assessment of treatment efficacy. Target lesions are measured over time to evaluate tumor response, such as shrinkage or progression, as part of endpoints like objective response rate (ORR) and progression-free survival (PFS). ADTL is structured following the BDS format, which organizes data with one record per subject, per parameter, and per analysis time point. To accommodate multiple tumor types in a trial, there are two common approaches to structuring the ADTL dataset, each with its pros and cons:

OPTION 1: ONE PARAMCD FOR EACH TUMOR TYPE

In this approach, a unique PARAMCD is created for each tumor type (e.g., RCC, PPGL, pNET, CNS Hemangioblastomas). This results in multiple records per subject, with each record corresponding to a specific tumor

type. For each tumor, a separate PARAMCD is derived, such as PARAMCD='RCC_SLD', PARAMCD='PPGL_SLD', etc. The derivation is included in Appendix 1 (Option 1)

Advantages:

- Each tumor type has its own distinct set of records.
- Allows for clear distinction and easier analysis by tumor type.

Disadvantages:

- Requires careful management of multiple tumor specific PARAMCDs.

USUBJID	PARAMCD	PARAM	AVISIT	AVAL	ADT	ABLFL
101	RCC_SLD	SLD (RCC)	Baseline	50.0	2024-01-01	Y
101	RCC_SLD	SLD (RCC)	Week 2	45.0	2024-01-15	
101	PPGL_SLD	SLD (PPGL)	Baseline	40.0	2024-01-01	Y
101	PPGL_SLD	SLD (PPGL)	Week 2	35.0	2024-01-15	
101	pNET_SLD	SLD (pNET)	Baseline	30.0	2024-01-01	Y
101	pNET_SLD	SLD (pNET)	Week 2	28.0	2024-01-15	

OPTION 2: ONE PARAMCD FOR ALL TUMOR TYPES, WITH A TUMORTYP VARIABLE

In this approach, a single PARAMCD is used for all tumors (e.g., SLD), and an additional variable, TUMORTYP, is used to differentiate between tumor types. This approach minimizes the number of records, but there are implications for baseline records. The derivation is included in Appendix 1 (Option 2).

Advantages:

- Simplifies the structure by using one PARAMCD across all tumors.

Disadvantages:

- Can lead to P21 validation issues (e.g., AD0154: "Multiple baseline records exist for a unique USUBJID, PARAMCD, BASETYPE) because multiple baseline flags (ABLFL='Y') are created for the same subject with a single PARAMCD.

SOLUTION OPTION 2: USE OF BASETYPE VARIABLE

To resolve the P21 validation issue, you can introduce the BASETYPE variable with values matching the tumor type (TUMORTYP). This distinguishes baseline records for each tumor type under the same PARAMCD. The BASETYPE variable essentially serves to differentiate the baseline record for each tumor.

USUBJID	PARAMCD	PARAM	AVISIT	AVAL	ADT	ABLFL	TUMORTYP	BASETYPE
101	SLD	SLD	Baseline	50.0	2024-01-01	Y	RCC	RCC
101	SLD	SLD	Week 2	45.0	2024-01-15		RCC	RCC
101	SLD	SLD	Baseline	40.0	2024-01-01	Y	PPGL	PPGL
101	SLD	SLD	Week 2	38.0	2024-01-15		PPGL	PPGL
101	SLD	SLD	Baseline	30.0	2024-01-01	Y	pNET	pNET
101	SLD	SLD	Week 2	28.0	2024-01-15		pNET	pNET

After exploring these two options, we have implemented Option 2 in our study.

ANALYSIS DATASET OF RESPONSE (ADRS)

The Analysis Dataset of Response (ADRS) is a key component in oncology trails, particularly for tracking and analyzing tumor response to treatment over time. The dataset typically follows the BDS format, where each subject

has multiple records, representing different parameters (Investigator Timepoint Assessment, Best Overall Response with Confirmation) at various analysis time points.

In Oncology, response assessment is often based on criteria such as RECIST 1.1, which classifies the response categories like Complete Response (CR), Partial Response (PR), Stable Disease (SD), and Progressive Disease (PD). The ADRS dataset is used to calculate response rates (e.g., Objective Response Rate (ORR), Disease Control Rate (DCR)), and track tumor progression over time.

DERIVATION:
In this dataset, the Best Overall Response (BOR) should be derived for each subject and each tumor individually, rather than at the subject level alone. This can be accomplished by sorting the dataset by subject, tumor type, and analysis time point, then selecting the best response within each subject-tumor combination. To distinguish the BOR for different tumors within the same subject, the variable TUMORTYP is included. The derivation is included in Appendix 2

USUBJID	PARAMCD	AVISIT	ADT	AVAL	AVALC	TUMORTYP	BOR
101	OVRRESP	Baseline	2024-01-01	2	PR	RCC	PR
101	OVRRESP	Week 4	2024-02-01	1	CR	RCC	CR
101	OVRRESP	Week 8	2024-03-01	3	SD	RCC	CR
101	OVRRESP	Baseline	2024-01-01	3	SD	PPGL	SD
101	OVRRESP	Week 4	2024-02-01	2	PR	PPGL	PR
101	OVRRESP	Week 8	2024-03-01	4	PD	PPGL	PR

INTERMEDIATE DATES DATASET FOR ADTTE (ADINTDT)

ADINTDT is an intermediate dataset designed to capture key dates that are essential for creating the ADTTE (Time-to-Event) analysis dataset. Each parameter in ADINTDT represents a significant date that contributes to the analysis of efficacy endpoints such as Duration of Response (DOR), Progression-Free Survival (PFS), and Overall Survival (OS).

The records in ADINTDT are derived from multiple sources including the RS, CM, DS, EX, ADSL, and ADRS. Key parameters include - Date of Progression, Date of First Confirmed Response, Stable Disease Start Date etc. In these parameters, the relevant dates are derived for each subject and tumor type to support the analysis of multi-tumor basket trials. The dataset ensures that each time point, critical for DOR, PFS is appropriately recorded for each subject-tumor combination. However, certain parameters are subject-level and are not tumor-specific. For example: Date of Death (DTHDT), and Date of Treatment Discontinuation (DISCDT). The derivation is included in Appendix 3.

USUBJID	PARAMCD	ADT	TUMORTYP
101	DISCDT	2024-05-15	
101	DTHDT	2024-06-01	
101	IRCPDDT	2024-02-15	RCC
101	IRCPDDT	2024-03-10	PPGL
101	IRCPDDT	2024-03-20	pNET
101	IRCPDDT	2024-04-05	CNS
102	DTHDT	2024-04-15	
102	IRCPDDT	2024-01-25	CNS

TIME TO EVENT ANALYSIS DATASET (ADTTE)

The ADTTE dataset is a key analysis dataset used to capture time-to-event data, such as survival times, progression-free survival (PFS), and duration of response (DOR). It follows the BDS (Basic Data Structure) format, where each

record corresponds to a specific subject, tumor type, and parameter, allowing for detailed time-to-event analysis across multiple tumor types.

As discussed in the above datasets, the endpoints in the ADTTE datasets, such as progression-free survival, duration of response, and time to response, must be derived for each subject and each tumor type, rather than for each subject alone. To achieve this, date variables (e.g., progression dates or response dates) are first extracted from the ADINTDT dataset. The ADINTDT dataset, described earlier, can be transposed to create a structure that includes USUBJID, TUMORTYP (tumor type), and relevant PARAMCD values as column names. This transposed structure serves as the foundation for deriving time-to-event endpoints, ensuring that progression or response is calculated independently for each subject and tumor type.

USUBJID	TUMORTYP	DISCDT	DTHDT	IRCPDDT
101	RCC	2024-05-15	2024-06-01	2024-02-15
101	PPGL	2024-05-15	2024-06-01	2024-03-10
101	pNET	2024-05-15	2024-06-01	2024-03-20
101	CNS	2024-05-15	2024-06-01	2024-04-05
102	CNS		2024-04-15	2024-01-25

The final ADTTE dataset is structured to include one record per subject per tumor type, allowing the analysis to account for subjects with multiple tumors in a single trial. The derivation is included in Appendix 4.

USUBJID	TUMORTYP	EVENTDESC	CNSDTDSC	PARAMCD	ADT	CNSR
101	RCC	Documented Progression		PFS	2024-02-15	0
101	PPGL	Documented Progression		PFS	2024-03-10	0
101	pNET	No adequate post-baseline disease assessment	First dose date	PFS	2024-03-25	1
101	CNS	No adequate post-baseline disease assessment	First dose date	PFS	2024-04-01	1
102	CNS	Documented Progression		PFS	2024-01-25	0
102	CNS	No adequate post-baseline disease assessment	First dose date	PFS	2024-02-15	1

CONCLUSION

In oncology basket trails, analyzing efficacy data across multiple tumors within the same subject represents a unique and complex challenge. This paper highlights the application of CDISC ADaM standards in structuring ADaM datasets like, ADTL, ADRS, ADINTDT, and ADTTE to address these challenges, ensuring the regulatory compliance for the submission of datasets.

ACKNOWLEDGMENTS

The author would like to thank Amy, Sai Govind Chenna, Himanshu Patel, and Dileep Penmetsa for reviewing the paper and providing great suggestions.

CONTACT INFORMATION

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

Venkatesh Burla
Merck & Co., Inc., Rahway, NJ, USA
215-631-5295
Email: venkatesh.burla@merck.com
Manasa Ravi
Merck & Co., Inc., Rahway, NJ, USA
267-305-6745
Email: manasa.ravi@merck.com

APPENDIX 1: ADTL.SAS (OPTION 1)

SAS SCRIPT	FUNCTION
`select usubjid, tumortyp, avisitn, avisit,` `sum(trstresn) as sumdiam,` `'SUMDIAM_' || strip(tumortyp) as paramcd length=8,` `'Sum of Diameters for ' || strip(tumortyp) as param length=40` `from tr` `group by usubjid, tumortyp, avisitn, avisit;` `by usubjid tumortyp avisitn;` `if avisitn = 1 then ablfl = 'Y';` `else call missing(ablfl);` `if first.tumortyp then base_sumdiam = .;`	Assuming the TR dataset has individual lesion measurements, create the dataset with paramcd as "SUMDIAM_" || tumor type (tumortyp) variable and sort the dataset by USUBJID, TUMORTYP, AVISITN, and AVISIT. Using BY statement with TUMORTYP variable in SET statement to derive

ADTL.SAS (OPTION 2)

SAS SCRIPT	FUNCTION
`paramcd = "SLD";` `param = "SLD";` `/* Deriving BASETYPE only for the first record per USUBJID */` `by usubjid;` `retain basetype;` `if first.usubjid then basetype = tumortyp;` `else basetype = "";` `/* Initialize ABLFL as missing */` `ablfl = "";` `if avisit = 'Baseline' then ablfl = 'Y';` `by usubjid tumortyp avisitn;` `if avisitn = 1 then ablfl = 'Y';` `else call missing(ablfl);` `if first.tumortyp then base_sumdiam = .;`	Assuming the TR dataset has individual lesion measurements, create the dataset with PARAMCD of each tumor type (tumortyp). Create BASETYPE record from tumortyp to avoid P21 message.

APPENDIX 2: BOR.SAS

SAS SCRIPT	FUNCTION
`by usubjid tumortyp;` `length bor $2; /* BOR variable to store the best overall response */` `retain bor;` `/* Deriving BOR based on response hierarchy (CR > PR > SD > PD) */` `if first.tumortyp then bor = ''; /* Reset BOR at the start of each tumor type */` `select (avalc);`	Assuming the RS dataset has individual response measurements, sort the dataset by USUBJID, TUMORTYP and pick the last record for each tumor type after

`` when ('CR') bor = 'CR'; when ('PR') if bor ne 'CR' then bor = 'PR'; when ('SD') if bor not in ('CR', 'PR') then bor = 'SD'; when ('PD') if bor not in ('CR', 'PR', 'SD') then bor = 'PD'; otherwise; /* Keep bor unchanged for any other values */ end; /* Output only the final BOR for each subject-tumor combination */ if last.tumortyp then output; ```	deriving BOR for each subject and tumor type.

APPENDIX 3: ADINTDT.SAS

SAS SCRIPT	FUNCTION
`` set sorted_ircpddt; by usubjid tumortyp adt srcdtc; if first.tumortyp then output; adt = rsdt; srcdom = "RS"; srcvar = "RSDTC"; srcseq = rsseq; paramcd = "IRCPDDT"; parcat1 = "BICR"; parcat2 = "RECIST 1.1"; srcdtc = rsdtc; ```	Subset the input dataset for required parameter (e.g., RSGRPID=”CIMG“ and RSORRES = “PD” for RSTESTCD=”OVRLRESP” for deriving the Date of progression for each subject and tumor type.

APPENDIX 4: ADTTE.SAS

SAS SCRIPT	FUNCTION
`` proc transpose data = adintdt out = adintdt_1(drop = _name_ _label_); by usubjid tumortyp; id paramcd; var adt; where tumortyp ne " "; run; proc transpose data = adintdt out = adintdt1_1(drop = _name_ _label_); by usubjid; id paramcd; var adt; where tumortyp eq " "; run; data adintdt2; merge adintdt_1(in=b) adintdt1_1(in=a) ; by usubjid; run; if IRCPDDT ne . then do; paramcd = 'PFS'; adt = IRCPDDT; evntdesc = 'Documented Progression'; cnsr = 0; /* Event occurred */ output; end; else if DISCDT ne . then do; paramcd = 'PFS'; adt = DISCONT; evntdesc = 'Discontinued Treatment'; cnsr = 1; /* Censored due to discontinuation */ output; end; ```	Transposing the ADINTDT data by USUBJID and TUMORTYP. Transposing the ADINTDT data by USUBJID ONLY. Merge the datasets and derive ADT, EVNTDESC as required.