# PharmaSUG 2024 - Paper MM-240

# Managing a Single Set of SDTM and ADaM Specifications across All Your Phase 1 Trials

Avani Kaja, Pfizer Inc., South San Francisco, CA

## ABSTRACT

Submitting data in CDISC standards has become mandatory in the drug development industry. The specifications for such standardized data structures and controlled terminology can be managed in metadata collections called data definition tables (DDTs).

While it is common practice for many companies to use a single consolidated master DDT (MDDT) to manage all studies within a single product, this paper will discuss how to standardize and implement an MDDT across multiple products in oncology early-stage development (Phase 1) where studies from different products and indications tends to share common design and many analysis endpoints.

The paper emphasizes implementation, process, and best practices to accommodate different products into one template. It will discuss the process of setting up and maintaining the specifications across various products. In addition, the paper will illustrate effective ways of communicating across different product teams to align specifications within the single MDDT. Lastly, the paper will highlight the process of adapting these specifications when products move to late-stage development (Phase 2/3). Implementing this process in early-stage development has yielded significant reduction in set-up time for SDTM and ADaM specifications, higher compliance to departmental standards, and better utilization of departmental utilities and macros all of these contribute to a much more efficient and streamlined support for study conduct.

## INTRODUCTION

Standardizing SDTM and ADaM specs at product (sometimes referred to as compound or asset) level is very common in the drug development process for a statistical programming function. Towards this goal, we extended this concept and created a set of standard specification files (for SDTM, ADaM, and their corresponding CTs) to accommodate ten different early development products, with capacity to accommodate any future products coming into the pipeline beyond those. The specifications (specs) for all these phase 1 studies were maintained in one standard file contrary to the approach of maintaining one specification file per product.

There are similarities between Phase 1 studies such as dose finding phase, optimization phase, and PK analysis. In addition, in oncology studies, the safety and efficacy analyses in Phase 1 studies are similar to those in late-phase studies. Observing that in these early stages we typically have one early-phase study per product, it is beneficial to use a single MDDT for all early-phase products and establishing the standards early will help carry the consistency to future late-phase studies in the same product.

This paper will describe the approach we took to standardize SDTM, ADaM and CT while describing the best practices and effective ways of communication we implemented to make the standard specifications work for all the studies. The paper will conclude by highlighting the benefits of standardization.

## STANDARDIZING DATA COLLECTION, SDTM, ADAM AND CONTROLLED TERMINOLOGY

### STANDARDIZING RAW DATA COLLECTION

Standardizing the specifications begins with the critical step of creating consistent case report forms (CRFs) and data collection across studies. In our setting, a data standards team worked on standardizing the CRFs

into a departmental library, and all studies are expected to build their respective CRFs based on this standard library.

#### STANDARDIZING SDTM SPECS

Standardizing SDTM specs at the department level is a common practice in Industry. These standard specs are adopted by product teams and, as a good practice, implemented at the product level. We have followed a similar approach for all our products. While this setup delivers myriad benefits and certainly enforces consistency in data structures across all studies within each product, we did observe a few instances where similar variables were named differently from one product to the next. For example, the QNAMs for SUPPQUAL variables may be assigned differently between the products. By following a single early-phase MDDT approach, we can standardize all early development products in the same single specifications file and therefore easily identify and address these issues.

To standardize the specs across all early development products we took the approach of creating a master data definition table (MDDT) in Excel format. The specifications contain a dataset sheet where all data sets are listed. We have the following columns in this dataset sheet:

| Dataset | Description | Comment | InternalComment | DeriveOrder | Class | Structure | Repeating | Study_1 | Study_2 |
|---|---|---|---|---|---|---|---|---|---|
| AE | Adverse Events | AE contains all adverse events for a subject and timepoint | AE is created from the ae raw data set. the variables are mapped to SDTM variables. Additional information like timepoint and questions are mapped to SUPPAE | | EVENTS | One record per adverse event per subject | Yes | x | x |

In the same Excel file, each domain specification and VLM are entered in separate sheets, for example:


The SDTM domain sheet contains the following variables.

| Variable | Label | SourceDerivation | InternalDerivation | Key | Type | Length | SignificantDigits |
|---|---|---|---|---|---|---|---|
| STUDYID | Study Identifier | Protocol number | raw.dm:project | 1 | text | 15 | |

| Origin | Codelist | DisplayFormat | Method | Mandatory | SUPPQUAL | QEVAL | Study_1 | Study_2 |
|---|---|---|---|---|---|---|---|---|
| Protocol | | | | Yes | | | x | x |

The '*SourceDerivation*' column contains the description of the variable that can be used for define.xml, whereas the '*InternalDerivation'* column contains a detailed description referencing raw data variables that are primarily used for production and QC purposes. The SUPPQUAL variables are listed in the same sheet after the main domain variables and the SUPPQUAL column is marked 'Yes' for these variables. When the data sets are actually produced, a departmental SAS macro processes this metadata and automatically splits off and transforms these supplemental qualifier variables into the typical suppqual structure based on this flag, without needing any further manual programming and transformations.

A column is allocated for each study to either acknowledge if the study specification is same as the source derivation and internal derivation by marking "x", or otherwise describing the study-specific derivation in this column.

In our implementation we have seen that studies typically use around ~80% of the standard definitions. When a new study is added to the table, it makes it easy to write the specifications. This approach saves time and resources while ensuring the department standards are met and consistencies are achieved.

#### STANDARDIZING ADAM SPECS

ADaM data sets follow a similar approach as SDTM. The standard specifications are set up following the department-level template. We have the following columns in the ADaM template, as an example.

| Variable | Label | SourceDerivation | InternalDerivation | Key | Type | Length | SignificantDigits | Origin |
|---|---|---|---|---|---|---|---|---|
| STUDYID | Study Identifier | DM.STUDYID | | 1 | text | 15 | | Predecessor |

| Codelist | DisplayFormat | Method | Mandatory | CoreVariable | SC Standards Req or Perm | SC Standards Fixed or Flexible | Study_1 | Study_2 |
|---|---|---|---|---|---|---|---|---|
| | | | Yes | 1 | Req | Fixed | x | x |

The column *"SC Standards Req or Perm"* in the above table specifies if a variable is required according to the standards, and the column *"SC Standards Fixed or Flexible"* suggests if the definition of the variable can differ from the standard definition. Each study is allocated a column to enter the specifications. If the study definition is the same as the source derivation and internal derivation, then the programmer would enter 'x' in the study column; if not, the study specific definition is entered.

We have studies targeting different disease indications and, in some cases, using different response criteria. Despite these variations, we were able to enter the ADaM specs for ten studies in the same template. We observed that ~70% of the ADaM variables are using the standard implementation and specifications.

#### STANDARDIZING CONTROLLED TERMINOLOGY

A single controlled terminology (CT) file is maintained for all the studies, and the CT entries applicable for each study are checked in the study column. This will help ensure all the early development studies will use the same CT and when these studies move on to late phase, the late phase team can continue to use the same CT for all the other new studies in the same product. The consistent terminology will be helpful when working on pooling the data for ISS and ISE. Standardizing CT across studies further helps build macros for critical derivations and automated tools to generate tables, listings, and figures (TLFs).

The following variables are present in the CT file:

| CTName | CTLabel | Code | Decode | Display | CodeType | CRFValue | Extensible |
|---|---|---|---|---|---|---|---|
| PPCAT | Category for Parameter | | DM1 | | | | |

| ValueID | ListID | ListName | StudySpecific | Study_01 | Study_02 |
|---|---|---|---|---|---|
| | | | | | |

## MANAGING THESE SPECIFICATIONS

Many best practices have been created to help early-stage programming teams with managing these specifications.

- The initial template was set up from the department standard. The studies were then added to the template.
- Weekly meetings are set up with the study teams to review and discuss any mapping or derivation related questions.
- We had an issue log where the study lead entered questions to be discussed in a meeting and eventually agree on a solution that would work best for all the studies, resulting in the standard specification to be updated with the change. For study-specific variables, the study leads were asked to follow a consistent approach to add those variables in SUPP or ADaM domains. For example, for any combo drug the first part of the variable name was kept constant and the two-letter abbreviation of the drug name was added to the variable name. For example:

| | |
|---|---|
| AEACNDR | Seagen IP Dose Reduction Flag |
| AACN**AZ**DR | Azacitidine Dose Reduction Flag |

When it comes to keeping up with departmental standard updates, it's helpful to designate one lead to be responsible for updating the early-phase MDDT according to the department standards. The MDDT was cross-checked with the department standards every six months. Any critical updates were implemented in a timely manner and efficiently communicated to study team members in the weekly meeting. We also defined specific timelines by which study teams could implement these changes in their respective studies. These practices helped all product teams stay aligned with the departmental standards.

When a product is moved out of early and into late-phase development, the early-phase MDDT (with all other products removed) will be used as the starting point for the new late-phase product-level MDDT. Additional late-stage studies for that product are then added over time. Once this "break" has occured, if the early-phase study is still ongoing and specs need to be added or updated for it, the early development study lead coordinates with the late development team to ensure the specs for the new studies and the ongoing phase 1 study remain aligned.

## CONCLUSION

Establishing standards from data collection to SDTM, ADaM and TLFs is an important step to drive downstream automation and ensuring efficiency and consistency across studies and products. Having a standard template will help achieve the goal.

In addition to helping automation, the standard specifications save time and resources. For SDTM, ADaM and CT, ~70% of our early-phase specifications used standard definitions, so the lead programmers would just check off the study column to use these existing definitions. As the specifications are standardized, the programming codes can be adapted and reused or even macrotized which saves time and resources.

Going beyond immediate study team efficiencies, if integrated data are requested by a company's drug safety or other department to support pharmacovigilance, exploratory, or new study design efforts, pooling these data will be easy due to the overall data structure consistency across all early-phase products. As another primary advantage, especially when building the early-phase MDDT against departmental data standards, when a product moves to late phase, it will have a good template to start including additional studies towards ultimate regulatory submissions and beyond.

## ACKNOWLEDGMENTS

I would like to acknowledge Michael Hagendoorn, Vice President, Statistical Programming and Wei Qian, Senior Director, Statistical Programming for their review and valuable feedback.

## CONTACT INFORMATION

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

Avani Kaja
Pfizer Inc.
Avani.kaja@pfizer.com