PharmaSUG 2022 - Paper SS-087

Including Population Information in ADaM define.xml for Better Understanding of Datasets

Hong Qi, Majdoub Haloui, Guowei Wu, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

To create analysis datasets for clinical trials, an important element is population selection criteria in the ADaM (Analysis Dataset Model) dataset specification to define study populations included in each dataset based on the Statistical Analysis Plan (SAP). Per current industry guideline/practice, population selection criteria or population information are not required to be listed directly in the submission document for each ADaM dataset, such as the analysis data definition document, ADaM define.xml, and Analysis Reviewer's Guide (ADRG). Therefore, this information is not described in ADaM define.xml and often missed in ADRG at the dataset level. Meanwhile, no software is available to automatically include the population selection criteria in define.xml. To help regulatory agencies for an easier review, effort can be made to communicate the study population enclosed in each submitted analysis dataset. This paper discusses the importance of including population information in a submission document and the current gaps. It explains why ADaM define.xml/Dataset-level metadata is the ideal document/section to include this information, and it provides implementation approaches.

INTRODUCTION

Clinical trial datasets contain data and variable information pertinent to the analysis and presentation of the study data. Within the same trial, each analysis dataset has its specific objective and may have a different study population from other datasets. For example, the subject-level analysis dataset (ADSL) usually contains all screened subjects; the concomitant medications analysis dataset (ADCM) usually contains all randomized or allocated subjects with an available concomitant medication record.

The generation of analysis datasets follows an ADaM dataset specification which defines each dataset and variable according to the SAP. As a critical ADaM document, the ADaM dataset specification also provides data/variable/parameter-level traceability and serves as a key input to ADaM define.xml required for an ADaM compliant regulatory submission. Many pharmaceutical companies/organizations have their own user-friendly ADaM dataset specification template, such as shown in Display 1. The study population information for each analysis dataset is provided by the data "Selection Criteria" column. However, it is not automatically transferred to ADaM define.xml by Pinnacle 21 sponsor adaptor (P21 adaptor) along with other dataset-level descriptions, such as dataset label, class, structure, key variables, and location (Display 2). Note: The P21 adaptor is a custom software that is developed on top of the base Pinnacle 21 Enterprise software. It is used to create ADaM define.xml from the sponsor's ADaM dataset specification in Excel format (Display 3).

ADaM Dataset	ADaM Dataset Description	Class of Dataset	Subclass	ADaM Dataset Structure	Key Variables	Dataset Location	Selection Criteria
ADSL	Subject-Level Analysis Dataset	ADSL		One record per subject	USUBJID	adsl.xpt	All screened subjects.
ADDILI	DILI Reporting Analysis Dataset	BDS		One observation per subject, parameter, and analysis time point.	USUBJID, PARAMCD, AEPOCH, ADT	addili.xpt	All randomized subjects with available measurements.
ADEXSUM	Exposure Summary Analysis Dataset	BDS		One record per subject, per analysis period, per study medication, per dosage level, per analysis parameter	USUBJID, AEXTRT, PARAMCD	adexsum.xpt	All randomized subjects who took at least one dose.
ADINTDT	Interim Dataset of Dates Used for ADTTE	BDS		One record per subject per parameter	USUBJID, PARCAT1, PARCAT2, PARAMCD, ADT	adintdt.xpt	All randomized subjects

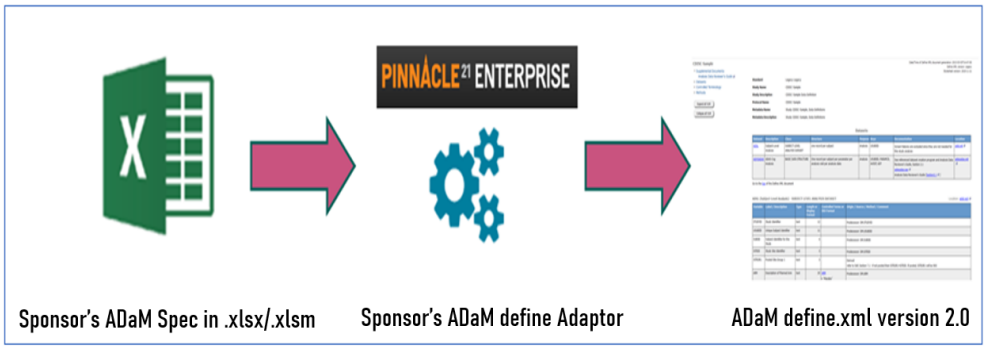
Display 1 Study population information defined by "Selection Criteria" in the Sponsor's ADaM dataset specification

Metadata Name Study P000V01MK9999 Data Definition

Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
ADSL	Subject-Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	One record per subject	Analysis	USUBJID		adsl.xpt
ADDILI	DILI Reporting Analysis Dataset	BASIC DATA STRUCTURE	One or more record per subject, parameter, and analysis time point	Analysis	USUBJID, PARAMCD, AEPOCH, ADT		addili.xpt
ADEXSUM	Exposure Summary Analysis Dataset	BASIC DATA STRUCTURE	One record per subject, study medication and analysis parameter	Analysis	USUBJID, AEXTRT, PARAMCD		adexsum.xpt
ADINTDT	Intermediate Dates Dataset for ADTTE	BASIC DATA STRUCTURE	One record per subject, and parameter	Analysis	USUBJID, PARCAT1, PARCAT2, PARAMCD, ADT		adintdt.xpt

Display 2 ADaM define.xml without population information for the datasets



Display 3 Process of ADaM define.xml Generation by Pinnacle 21 Sponsor Adaptor

Current industry guideline/practice does not require population selection criteria, nor population information, to be listed directly in ADaM define.xml or ADRG for each analysis dataset. Therefore, this information is not described in the ADaM define.xml and is often missed in the ADRG at the dataset level. Meanwhile, no software is available to automatically include the population selection criteria in the define.xml. Including the description of study population in each submitted dataset in the define.xml can make it easier for a regulatory agency to review or duplicate the analysis dataset.

The objective for this paper is to describe why, where, and how to include population information in the ADaM define.xml to improve understandability of the datasets, which could potentially increase the efficiency of regulatory review. In the following sections, we 1) explain in detail why ADaM define.xml/Dataset-level metadata is the ideal document/section in which to include this information and 2) describe the implementation approaches.

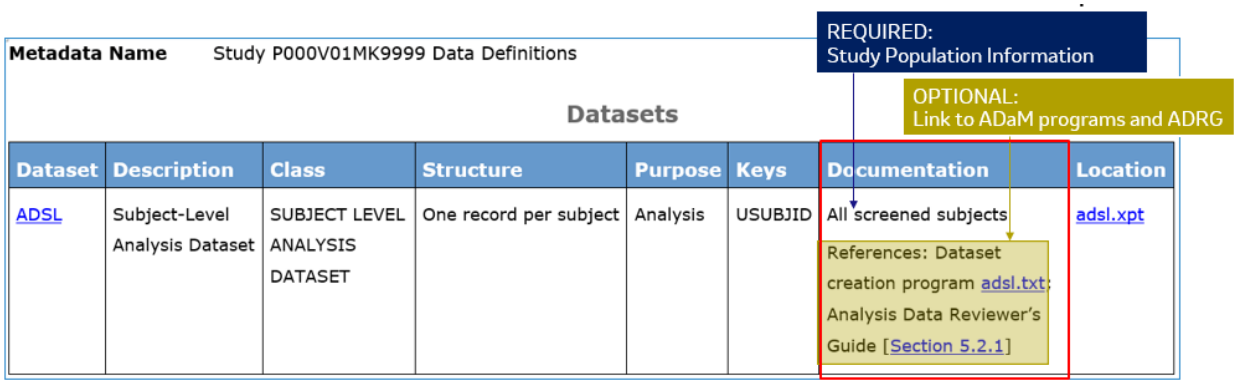
INCLUSION OF STUDY POPULATION INFORMATION IN ADaM DEFINE.XML/ DATASET-LEVEL METADATA

In the analysis package submitted to regulatory authorities, ADaM define.xml and ADRG are the only components that contain description of the analysis datasets. While ADRG provides an overview storyline about the study, **DEFINE.XML** is a required document primarily used to define the analysis datasets and thus is identified as the ideal ADaM submission document in which the study population information can be added at the dataset level. Table 1 presents the advantages of using define.xml to include this information over ADRG.

Document	ADaM define.xml	ADRG
Purpose	Describes the structures and contents of the submitted ADaM datasets.	Provides additional context beyond ADaM define.xml 1. Identifies the standards/versions implemented beyond ADaM 2. Describes the trial design 3. Notes any special considerations for the datasets and how they relate to or impact the study results.
Requirement in FDA Submission	**Required**	**Recommended**
Important Consideration	Describes dataset/variable as detailed as possible allowing easy locate/replicate of information	Complements ADaM define.xml: **Conveys information that does not have a place in/is not easily expressed in define.xml.**
Key Distinction	Machine-readable file which can be used by tools to help translate or verify the data. **Focuses more on ADaM data**	Provides additional information to regulatory reviewers in a human readable format **Provides an overview storyline for the study or additional information besides define.xml.**

Table 1 Comparison of ADaM define.xml and ADRG

Within the ADaM define.xml, **dataset-level metadata** provides information for each dataset and therefore is considered the appropriate location for study population information (Display 3). In addition, this could facilitate the future automation of reading in the same information from the ADaM dataset specification when a P21 adaptor is updated.



Display 3 Dataset-Level Metadata Section of ADaM define.xml as the Location to Add Study Population Information

IMPLEMENTATION APPROACHES

Manual add-in and automatic solutions are planned to accommodate the immediate and longstanding needs to include population information in ADaM define.xml.

MANUAL ADD-IN SOLUTION

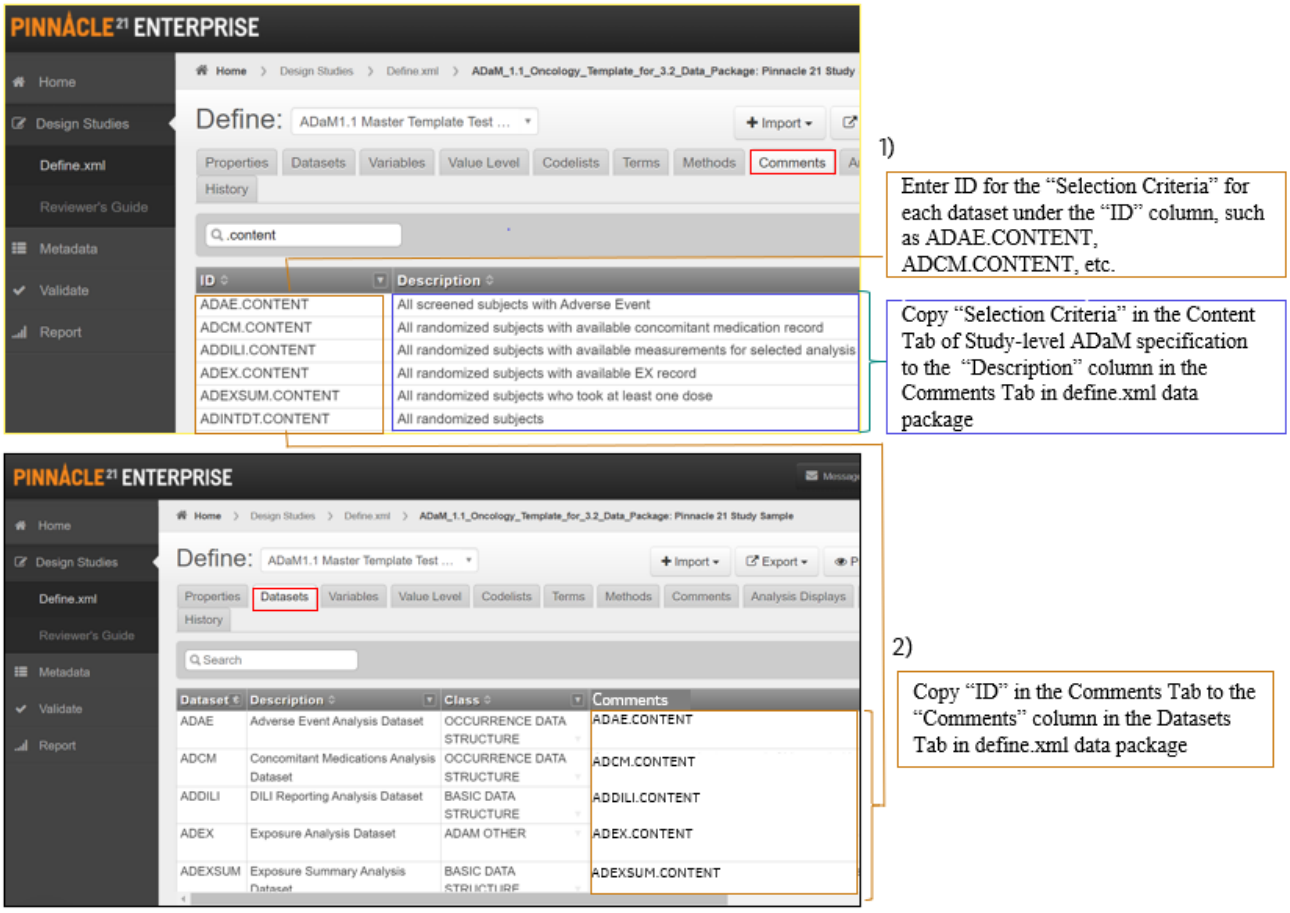
Due to the limitation of the current P21 adaptor and unavailability of an automatic tool, a manual add-in approach is considered to address immediate needs. The following steps outline how to add the data to P21 adaptor and generate ADaM define.xml with desired information.

Step 1. Generate the draft ADaM define.xml using the sponsor-defined study-level ADaM dataset specification as the input file

Step 2. Before finalizing ADaM define.xml, manually add "Selection Criteria" from the study-level ADaM dataset specification (Content Tab) to the P21 adaptor: As shown in Display 4, in the P21 adaptor, select Design Studies → Define.xml → Data Package,

1) In the Comments Tab of define.xml data package:
   - enter an ID for the "Selection Criteria" for each dataset under the "ID" column, such as ADAE.CONTENT, ADCM.CONTENT, etc.
   - copy "Selection Criteria" from the Content Tab of study-level ADaM dataset specification to the "Description" column

2) In the Datasets Tab of define.xml data package:
   - copy "ID" from the Comments Tab to the "Comments" column



Display 4 Add "Selection Criteria" in Pinnacle 21 Sponsor Adaptor

Step 3 Generate final ADaM define.xml
As shown in Display 5, export define.xml from the P21 adaptor. "Selection Criteria" is then populated in the "Documentation" column of the Datasets section.

Datasets

Dataset	Description	Class	Structure	Purpose	Keys	Documentation	Location
ADSL	Subject-Level Analysis Dataset	SUBJECT LEVEL ANALYSIS DATASET	One record per subject	Analysis	USUBJID	All screened subjects	adsl.xpt
ADDILI	DILI Reporting Analysis Dataset	BASIC DATA STRUCTURE	One or more record per subject, parameter, and analysis time point	Analysis	USUBJID, PARAMN, AEPOCH, ADT	All randomized subjects with available measurements for selected analysis	addili.xpt
ADEXSUM	Exposure Summary Analysis Dataset	BASIC DATA STRUCTURE	One record per subject, analysis period, study medication, dosage level, and analysis parameter	Analysis	USUBJID, AEXTRT, PARAMCD	All randomized subjects who took at least one dose	adexsum.xpt
ADINTDT	Interim Dataset of Dates Used for ADTTE	BASIC DATA STRUCTURE	One record per subject per parameter	Analysis	USUBJID, PARCAT1, PARCAT2, PARAMCD, ADT	All randomized subjects	adintdt.xpt

Display 5 "Selection Criteria" included under the "Documentation" column of the Dataset section in ADaM define.xml

For companies and organizations without a customized P21 adaptor, the manual add-in solution can be done as described below.

Step 1. Add the population selection criteria in the Pinnacle 21 Excel specification as demonstrated in Display 6.

1) In the Comments Tab:
   - enter an ID for the "Selection Criteria" for each dataset under the "ID" column
   - fill in study population information under the "Description" column

2) In the Datasets Tab:
   - copy "ID" from the Comments Tab to the "Comments" column

Step 2. Generate final ADaM define.xml
Export define.xml as shown in Display 5.

Pinnacle 21 Excel Specification:

1) Comments Tab

ID	Description
ADAE.CONTENT	All screened subjects with Adverse Event
ADCM.CONTENT	All randomized subjects with available concomitant medication record
ADDILI.CONTENT	All randomized subjects with available measurements for selected analysis
ADEX.CONTENT	All randomized subjects with available EX record

2) Dataset Tab

Datas	Description	Class	Comment
ADAE	Adverse Event Analysis Dataset	OCCURRENCE DATA STRUCTURE	ADAE.CONTENT
ADCM	Concomitant Medications Analysis	OCCURRENCE DATA STRUCTURE	ADCM.CONTENT
ADDILI	DILI Reporting Analysis Dataset	BASIC DATA STRUCTURE	ADDILI.CONTENT
ADEX	Exposure Analysis Dataset	ADAM OTHER	ADEX.CONTENT

Display 6 Add Population Information in Pinnacle 21 Excel Specification

AUTOMATED SOLUTION

The current thought on a long-term solution is to rely on P21 adaptor's update to automate the process. The desired outcome is to generate the ADaM define.xml with the study population information reading directly from the ADaM dataset specification to the dataset-level metadata.

CONCLUSION

Study population information is critical in defining and communicating the content of analysis datasets in clinical trials and regulatory review. Including this information in ADaM define.xml/Datasets-level metadata provides a better picture of the datasets at one snapshot and allows easy duplication of datasets through machine reading. Currently, we are implementing this in pilot studies, and anticipate a smooth agency review with this enhancement in the ADaM define.xml.

ACKNOWLEDGEMENTS

The authors would like to thank the support from BARDS ADaM Center of Excellence team and Oncology Statistical Programming Leadership Team at Merck & Co., Inc., and thank Ms. Mary N. Varughese, Ms. Ellen Asam, Ms. Susan Kramlik and Ms. Amy Gillespie for reviewing the paper and their great suggestions.

REFERENCES

1. CDISC Define-XML Specification Version 2.1 (Final), Define-XML v2.1 | CDISC
2. CDISC Analysis Data Model Version 2.1, ADaM | CDISC
3. Wang, X., Huang, D. "Design ADaM Specification Template that Simplifies ADaM Programming and Creation of Define XML in CDISC Era". 2021 PharmaSUG, SS-016
4. Nguyen, S. V., Asam, E., Dong, W., Travalent, A. "Sorting Out the Paperwork – Define.xml versus Reviewer's Guide and other Submission Documents". 2018 FDA/PhUSE CSS, PP19

CONTACT INFORMATION

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

Hong Qi
Principal Scientist, Statistical Programming
Merck & Co., Inc.
hong_qi@merck.com

Majdoub Haloui
Principal Scientist, Statistical Programming
Merck & Co., Inc.
Majdoub.haloui@merck.com

Guowei Wu
Sr. Principal Scientist, Statistical Programming
Merck & Co., Inc.
Guowei_wu@merck.com