PharmaSUG 2019 - Paper DS-082

Incremental Changes: ADaMIG v1.2 Update

Nancy Brucken, Syneos Health™; Deb Goodfellow, Covance; Brian Harris, MedImmune;
Terek Peterson, Covance; Alyssa Wittle, Covance

ABSTRACT

The ADaM Implementation Guide (ADaMIG) has now been available to industry since 2009, providing a standardized way to communicate and analyze study data. Improvements and clarifications were added in 2016 with the release of v1.1. Since that release, the ADaM team has been working on some items that were not yet ready for v1.1 but are now ready for the next release, v1.2. These items include important clarifications to existing text, standard nomenclature for stratification variables within ADSL, and a recommended approach for bi-directional toxicity grades. In addition, an update on the removal of the new suggested permissible variable within the Basic Data Structure (BDS) called PARQUAL will be provided. The ADaMIG v1.2 will be discussed from both the perspective of changes from v1.1 as well as changes made since the public review of v1.2.

INTRODUCTION

The ADaM team is proposing a minor version update of the ADaMIG with two new additions as well as many small clarifications:

- Addition of stratification variables in Subject-level Analysis Dataset (ADSL)
- Addition of bi-directional toxicity variables in BDS datasets
- Clarifications on ADSL timing variables, BDS indicator flags, the pre-ADSL dataset concept, depreciation of the PARAMTYP variable, scope of BASETYPE, and relationship between primary & secondary variables

STRATIFICATION VARIABLES IN ADSL

Stratified randomization is used to ensure balance of treatment assignments across one or more prognostic factors. A prognostic factor is an aspect of the disease or a characteristic of the subject that may influence treatment response. The analysis may need (1) As Randomized – the value used to randomize the subject and (2) As Verified – the actual value for the subject verified from the source documents at the site. As seen in Table 1 below, ADaM IG v1.2 provides a set of variables to allow maximum flexibility in representing the description of the prognostic factors as well as the values used for randomization and the values that were verified. To illustrate the interrelationships of the variables, examples are provided based on the combination of three stratification factors: Age Group (“<50” or “>=50”), Prior Treatment Status (“Treatment naïve”, “Treatment experienced”), and Hypertension (“Y” or “N”). ‘w’ in the variable names represents the ‘nth’ stratification factor. In the example in Table 1 below, w=3, so the variable STRATR3 represents the third stratification factor, which is Hypertension.

Variable Name	Variable Label	Type	Core	Example
STRATAR	Strata from Randomization	Char	Perm	STRATAR=">=50, Treatment experienced, N"
STRATARN	Strata from Randomization (N)	Num	Perm	STRATARN=3 when STRATAR=">=50, Treatment experienced, N"
STRATAV	Strata from Verification Source	Char	Perm	STRATAV=">=50, Treatment experienced, Y"
STRATAVN	Strata from Verification Source (N)	Num	Perm	STRATAVN=4 when STRATAV=">=50, Treatment experienced, Y"
STRATRw	Strat Factor w Value from Rand	Char	Perm	STRATR3="N"
STRATRwN	Strat Factor w Value from Rand (N)	Num	Perm	STRATR3N=0 when STRATR3="N"
STRATDw	Description of Stratification Factor w	Char	Perm	STRATD3="Hypertension"

Variable Name	Variable Label	Type	Core	Example
STRATVw	Strat Factor w Value from Verif Source	Char	Perm	STRATV3="Y"
STRATVwN	Strat Fact w Val from Verif Source (N)	Num	Perm	STRATV3N=1 when STRATV3="Y"

Table 1. Stratification Variables in ADSL

BI-DIRECTIONAL TOXICITY VARIABLES IN BDS

The ADaM team discussed many options for handling lab limits that need to be assessed in more than one direction with different meaning. Rather than creating additional records, the ADaM team agreed to add guidance around the variables proposed to handle bi-directional information. The examples in this paper are based on NCI CTCAE v4.03 Toxicity Grades; however, the bi-directional variables can also be used for any sponsor-specific toxicity grading.

Variable Name	Variable Label	Type	Core	Comments
ATOXGRL	Analysis Toxicity Grade Low	Char	Perm	Low toxicity grade of AVAL or AVALC for analysis
ATOXGRLN	Analysis Toxicity Grade Low (N)	Num	Perm	Numeric representation of ATOXGRL.
ATOXGRH	Analysis Toxicity Grade High	Char	Perm	High toxicity grad of AVAL or AVALC for analysis
ATOXGRHN	Analysis Toxicity Grade High (N)	Num	Perm	Numeric representation of ATOXGRH.
BTOXGRL	Baseline Toxicity Grade Low	Char	Perm	ATOXGRL of the baseline record identified by ABLFL
BTOXGRLN	Baseline Toxicity Grade Low (N)	Num	Perm	Numeric representation of BTOXGRL.
BTOXGRH	Baseline Toxicity Grade High	Char	Perm	ATOXGRH of the baseline record identified by ABLFL
BTOXGRHN	Baseline Toxicity Grade High (N)	Num	Perm	Numeric representation of BTOXGRH.
ATOXDSCL	Analysis Toxicity Description Low	Char	Perm	The analysis toxicity term used to describe toxicity in the low direction.
ATOXDSCH	Analysis Toxicity Description High	Char	Perm	The analysis toxicity term used to describe toxicity in the high direction

Table 2. Bi-directional Toxicity Variables

Table 2 displays the bi-directional toxicity grading proposed in ADaM IG v1.2. The toxicity descriptions (ATOXDSCL and ATOXDSCH) can help to determine denominators. For example, if ATOXDSCL is populated then that record is counted for low toxicity grading tables, and if ATOXDSCH is populated then that record is counted in high toxicity grading. In some cases, a record can be counted for both since the record would be assessed in both directions. The population of ATOXGRL and ATOXGRH is sponsor-defined. For example, in Table 3, ATOXGRL and ATOXGRH are derived using SDTM LBTOXGR, adding a ‘Grade 0’ for assessments that did not meet any of the grading criteria in a given direction:

Row	USUBJID	PARAMCD	VISITNUM	AVAL	BASE	ABLFL	ANRLO	ANRHI
1	001-0001	HGB	1	7.4	7.4	Y	11	16.1
2	001-0001	HGB	2	20.5	7.4		11	16.1
3	001-0001	SGOT	1	33	33	Y	5	25
4	001-0001	SGOT	2	55	33		5	25
5	001-0001	SGOT	3	60	33		5	25
6	001-0001	SGOT	4	77	33		5	25
7	001-0001	PLAT	1	250	250	Y	150	450
8	001-0001	PLAT	2	100	250		150	450
9	001-0001	PLAT	3	99	250		150	450
10	001-0001	PLAT	4	75	250		150	450
11	001-0001	PLAT	5	49	250		150	450
12	001-0002	HGB	1	21.1	21.1	Y	11	16.1

Row	ATOXDSCL	ATOXGRL	BTOXGRL	ATOXDSCH	ATOXGRH	BTOXGRH
1	Anemia	Grade 3	Grade 3	Hemoglobin increased	Grade 0	Grade 0
2	Anemia	Grade 0	Grade 3	Hemoglobin increased	Grade 3	Grade 0
3				Aspartate aminotransferase increased	Grade 1	Grade 1
4				Aspartate aminotransferase increased	Grade 1	Grade 1
5				Aspartate aminotransferase increased	Grade 1	Grade 1
6				Aspartate aminotransferase increased	Grade 2	Grade 1
7	Platelet count decreased	Grade 0	Grade 0			
8	Platelet count decreased	Grade 1	Grade 0			
9	Platelet count decreased	Grade 1	Grade 0			
10	Platelet count decreased	Grade 1	Grade 0			
11	Platelet count decreased	Grade 3	Grade 0			
12	Anemia	Grade 0	Grade 0	Hemoglobin increased	Grade 3	Grade 3

Table 3. Bi-directional Toxicity Grades Example

Table 3 contains an example of variables that support the analysis of parameters that have toxicity in two directions. Here the sponsor decided to include the word “Grade” in ATOXGRL, BTOXGRL, ATOXGRH, and BTOXGRH; however, the inclusion of the word “Grade” is not part of the ADaM standard and is neither encouraged nor discouraged. ATOXDSCL and ATOXDSCH are descriptions of the toxicity being assessed; in Table 3, they are based on the LBTOX values in SDTM, but in other cases could be sponsor-defined.

As can be seen in the highlighted red box in Table 3, PARAMCD=PLAT has toxicity grading only in the low direction, so only BTOXGRL, ATOXGRL, and other toxicity variables in the low direction are populated. The high direction toxicity variables for PARAMCD=PLAT are never populated, even if the value is out of range in the high direction (ANRIND=HIGH).

Note the highlighted orange box in Table 3; this demonstrates an example of bi-directional toxicity, where AVAL is graded in both the low and high directions, so both ATOXDSCL and ATOXSDCH are populated.

ATOXGRL, ATOXGRH, and the corresponding baseline toxicity variables were used to derive shift in toxicity grade, as is illustrated in the example in Table 4 below. In this case, SHIFT1 represents the shift from baseline in the low direction and is derived from BTOXGRL and ATOXGRL, whereas SHIFT2 represents the shift from baseline in the high direction and is derived from BTOXGRH and ATOXGRH.

USUBJID	AVAL	BASE	ABLFL	ATOXGRL	BTOXGRL	ATOXGRH	BTOXGRH	SHIFT1	SHIFT2
001-0001	7.4	7.4	Y	Grade 3	Grade 3	Grade 0	Grade 0		
001-0001	20.5	7.4		Grade 0	Grade 3	Grade 3	Grade 0	Grade 3 to Grade 0	Grade 0 to Grade 3
001-0001	33	33	Y			Grade 1	Grade 1		
001-0001	55	33				Grade 1	Grade 1		Grade 1 to Grade 1
001-0001	60	33				Grade 1	Grade 1		Grade 1 to Grade 1
001-0001	77	33				Grade 2	Grade 1		Grade 1 to Grade 2
001-0001	250	250	Y	Grade 0	Grade 0				
001-0001	100	250		Grade 1	Grade 0			Grade 0 to Grade 1	
001-0001	99	250		Grade 1	Grade 0			Grade 0 to Grade 1	
001-0001	75	250		Grade 1	Grade 0			Grade 0 to Grade 1	
001-0001	49	250		Grade 3	Grade 0			Grade 0 to Grade 3	
001-0002	21.1	21.1	Y	Grade 0	Grade 0	Grade 3	Grade 3		

Table 4. Use of Shift Variables Example

OTHER CHANGES

REMOVAL OF PARAMETER QUALIFYING VARIABLE – PARQUAL

In the draft version of ADaMIG v1.2 that went through public review, PARQUAL was included as a new permissible variable in BDS. Due to confusion discovered during public review on when to use PARQUAL, the ADaM team determined that PARQUAL needs more clarification and may be considered for a future release. As in prior releases, there are no qualifiers allowed for PARAM.

ADSL TIMING VARIABLES

The set of ADSL timing variables includes all of the start and end dates within each of phase, period or sub-period. The set of phase timing variables is independent of period and sub-period, while sub-period is dependent on period timing variables.

The ADaMIG was clarified to state that each set of timing variables may be defined at the subject-level or dataset level. If the set is defined at the subject-level, then the dataset level assignment must be consistent with the subject-level and the set of variables can be defined with in ADSL. However, if the timing variables are defined at the dataset level, and therefore vary across datasets, then they should not be included within ADSL. An example of when the period timing variables would not be allowed in ADSL would be where periods have the same definition across all datasets except Adverse Events, in this example, for Adverse Events, the last period start date is the same, but the end date extends to all observations up to 30 days after last dose, while in other datasets the end date is date of last treatment. In this case, separate timing variables need to be defined at the dataset level.

Sub-period timing variables may only be included in ADSL if the period timing variables are included. The same rule would apply to sub-period as above. If the period timing variables remain consistent but the sub-periods do not, then ADSL would contain the period timing variables, but not the sub-period timing variables.

BDS INDICATOR FLAGS

In the previous ADaMIG, the controlled terminology for these flags was Y for character and 1 for numeric. The terminology for these variables has now been extended to allow either Y or Y/N values. The use of Y/N should be applied when null and N values are treated differently in an analysis.

PRE-ADSL DATASET CONCEPT

There are cases where the derivations for some of the variables in ADSL are extremely complex or rely on imputed or derived values not available in SDTM datasets. In such cases, it may be helpful to create an intermediate dataset which contains a partial set of ADSL variables on the way to a complete ADSL. This will then be followed by additional processing to derive the remaining variables, and a final step to merge those variables back into pre-ADSL in order to obtain a complete ADSL. This type of data flow was not prohibited by previous versions of the ADaMIG since the ADaMIG does not dictate process. In ADaMIG v1.2, the text was closely reviewed to make sure that none of it implied that a pre-ADSL could not be used. For example, any text stating that variables were "…copied from ADSL" was re-written to be data flow neutral.

Here is an example of using a pre-ADSL, as described in the Rheumatoid Arthritis Therapeutic Area Users Guide:

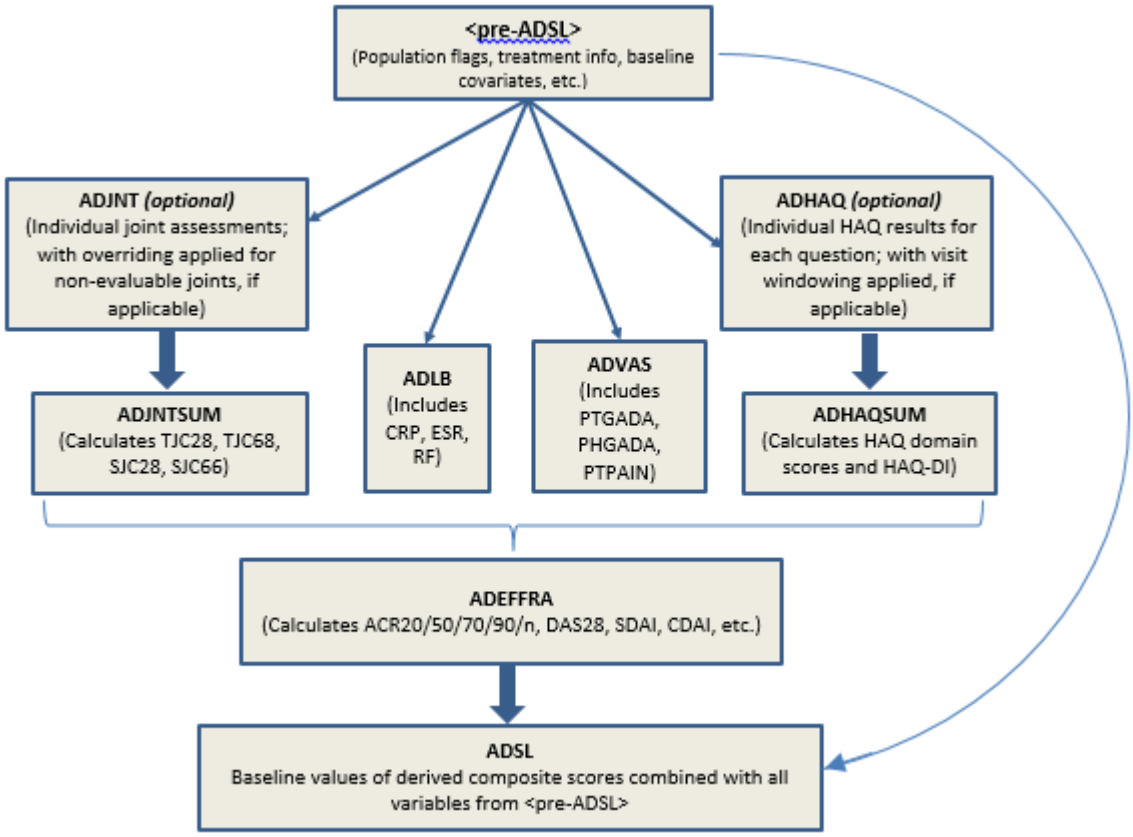

Figure 1. <pre-ADSL> processing example

The flow chart presented in Figure 1 above describes a possible approach for the order of creation of the analysis datasets typically used in rheumatoid arthritis (RA) studies. The dataset names used in this flow chart are meaningful examples of ADaM-compliant names but are not being proposed as a standard naming convention. The intermediate dataset ADEFFRA is created to hold several highly derived baseline values typically found in RA studies which are used for subgroup analyses, and thus need to ultimately be stored in ADSL.

As an initial step in this example, <pre-ADSL> is created. This dataset contains all the usual variables found in ADSL, including population flags and treatment information, but without the RA-specific baseline values to be calculated later, and will be used as the source for ADSL variables in all other ADaM datasets created for the study. Since <pre-ADSL> contains only a subset of variables in ADSL, it is the sponsor's decision whether to include <pre-ADSL> in a regulatory submission.

After <pre-ADSL>, there are two datasets that are optional; in many cases they would be a simple transformation of the source SDTM data into an ADaM BDS dataset and may be unnecessary. These are the analysis datasets for tender/swollen joints (ADJNT) and the Health Assessment Questionnaire (HAQ) (ADHAQ). The dataset ADJNT, if present, would be used to create ADJNTSUM, which contains the derived summary joint scores. Similarly, ADHAQ would be used to create ADHAQSUM, which would contain the eight sub-domain scores and the overall HAQ-DI score. Reasons why it may be advantageous to create these intermediate datasets include imputation or the need to perform significant visit reassignment.

The ADaM dataset (ADEFFRA) is created from joint count calculations (ADJNTSUM), HAQ-DI score (ADHAQSUM), CRP and/or ESR values (ADLB), and VAS scores (ADVAS). Composite endpoints, such as ACRx scores and the DAS28 scores, can be derived from these data. It is recommended that as much as possible is harmonized across those four predecessor datasets to simplify the merge of the four datasets. For example, applying the same visit windowing algorithms and consistent use of analysis flags within each of those four predecessor ADaM datasets will facilitate the merge by USUBJID and AVISIT as the first step towards the derivation of the composite scores for each visit.

DEPRECATION OF THE PARAMTYP VARIABLE

As promised in ADaMIG v1.1, PARAMTYP has been deprecated in ADaMIG v1.2. The ADaM sub-team feels that the there was a considerable amount of confusion regarding the differences between DTYPE and PARAMTYP, and since information describing variable derivations is contained in the metadata, there really was no need to explicitly flag derived parameters with a separate variable whose only possible values were DERIVED or null. However, PARAMTYP may still be used if required by analysis needs.

BASETYPE SHOULD BE POPULATED FOR A PARAM IF USED

ADaMIG v1.1 indicated for BASETYPE that, if used for any PARAM within a dataset, it should be non-null for all records of that dataset. However, it is often the case that BASETYPE is only defined for some of the parameters within a dataset. As a result, this has been clarified to state that if BASETYPE is used for any PARAM within a dataset, it should be non-null for all records for that PARAM within that dataset. In Table 5 below is an example showing BASETYPE defined for only some of the parameters in the dataset. In the example, when PARAMCD equals ALT all rows for BASETYPE are populated and when PARAMCD equals ALP they are not.

USUBJID	PARAMCD	AVISIT	AVAL	BASE	ABLFL	BASETYPE
001-0001	ALT	Screening 1	20	20	Y	MIN
001-0001	ALT	Screening 2	25	25	Y	MAX
001-0001	ALT	Week 1	19	20		MIN
001-0001	ALT	Week 1	19	25		MAX
001-0001	ALT	Week 2	21	20		MIN
001-0001	ALT	Week 2	21	25		MAX
001-0001	ALP	Screening 1	25	27		
001-0001	ALP	Screening 2	27	27	Y	
001-0001	ALP	Week 1	26	27		
001-0001	ALP	Week 2	24	27		

Table 5. Populating BASETYPE example

RELATIONSHIP BETWEEN PRIMARY AND SECONDARY VARIABLES

The text within the CDISC notes for secondary variables was expanded to clarify its relationship to the primary variable. Below is a typical example of this change.

Primary Variable CDISC Notes (v1.1 & v1.2) for AGEGRy	Secondary Variable CDISC Notes (v1.1) for AGEGRyN	Secondary Variable CDISC Notes (v1.2) for AGEGRyN
Character description of a grouping or pooling of the subject's age for analysis purposes. For example, AGEGR1 might have values of "<18", "18-65", and ">65"; AGEGR2 might have values of "Less than 35 y old" and "At least 35 y old".	The numeric code for AGEGRy. Orders the grouping or pooling of subject age for analysis and reporting. One-to-one mapping to AGEGRy within a study.	Numeric representation of AGEGRy. Orders the grouping or pooling of subject age for analysis and reporting. There must be a one-to-one relationship between AGEGRyN and AGEGRy within a study. AGEGRyN cannot be present unless AGEGRy is also present. When AGEGRy and AGEGRyN are present, then on a given record, either both must be populated or both must be null.

Table 6. Relationship between Primary and Secondary Variables

CONCLUSION

Since 2009, the ADaM Implementation Guide has provided the industry with a standardized way to communicate and analyze study data. The ADaM team is proposing the release of ADaMIG v1.2 as part of an ongoing commitment to improving the ADaMIG. Version 1.2 will include the addition of stratification variables within ADSL, as well as guidance for bi-directional toxicity grades and additional clarifications to text. These updates will make the implementation of ADaM standards more efficient while also improving the overall quality and legibility of ADaM datasets.

ACKNOWLEDGMENTS

This paper was made possible by the dedicated work of the ADaM v1.2 sub-team and ADaM team mentors.

RECOMMENDED READING

CDISC ADaM Team. 2019. “ADaM Implementation Guide v1.2”. Accessed April 20, 2019. https://www.cdisc.org/standards/foundational/adam

U.S. Food and Drug Administration. 2019. “Study Data Technical Conformance Guide v4.3”. Accessed April 20, 2019. https://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm

Japan Pharmaceuticals and Medical Devices Agency. 2015. “Technical Conformance Guide on Elecgtronic Data Submissions”. Accessed April 20, 2019. https://www.pmda.go.jp/english/review-services/reviews/advanced-efforts/0002.html

CONTACT INFORMATION

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

Nancy Brucken
Syneos Health
nancy.brucken@syneoshealth.com

Deb Goodfellow
Covance
Deborah.Goodfellow@Covance.com

Brian Harris
MedImmune (AstraZeneca group)
harrisb@medimmune.com

Terek Peterson
Covance
Terek.Peterson@Covance.com

Alyssa Wittle
Covance
Alyssa.Wittle@Covance.com