Analysis Data Reviewer's Guide Completion Guidance

Version 1.0

Analysis Data Reviewer's Guide – Completion Guidance

Contents

Analysis Data Reviewer's Guide Completion Guidance Overview

Analysis Data Reviewer's Guide Purpose

The Analysis Data Reviewer's Guide (ADRG) provides FDA Reviewers with additional context for analysis datasets (AD) received as part of a regulatory submission. The ADRG is intended to describe analysis data submitted for an individual study in the Module 5 clinical section of the eCTD. The ADRG purposefully duplicates limited information found in other submission documentation (e.g., the protocol, statistical analysis plan, clinical study report, define.xml) in order to provide FDA Reviewers with a single point of orientation to the analysis datasets. The submission of a reviewer guide does not obviate the requirement to submit a complete and informative define.xml document to accompany the analysis datasets.

ADRG Overview

The ADRG has seven sections - Introduction, Protocol Description, Analysis Considerations Related to Multiple Analysis Datasets, Analysis Data Creation and Processing Issues, Analysis Dataset Descriptions, Data Conformance Summary, Submissions of Programs, and an optional Appendix. The Introduction provides an overview, an inventory of standards used on the study, and describes the source data used to create the analysis datasets. The Protocol Description provides a brief orientation to the study and describes how planned treatment and timing variables relate to the study design. The Analysis Considerations Related to Multiple Analysis Datasets section provides an overview of topics relevant to multiple datasets such as a description of core variables appearing on most datasets, a comparison of data appearing in source (SDTM) versus analysis datasets, subjects requiring special analysis rules, windowing rules, and imputation/derivation methods. The Analysis Data Creation and Processing Issues section describes data dependencies, variable conventions, and, if submitted, split datasets, and intermediate datasets. The Analysis Dataset Descriptions section provides an overview of the analysis datasets with additional detail beyond that found in the define.xml where warranted. The Data Conformance Summary describes how ADaM conformance was assessed and summarizes conformance findings. The Submission of Programs section itemizes the programs that are included in the submission. An optional Appendix section may be included if needed.

The ADRG assumes that SDTM is used as input to the creation of analysis datasets and that the analysis datasets adhere to the ADaM standard to the largest extent possible. If a sponsor has created analysis datasets that are not based on SDTM and/or do not adhere to the ADaM model, then it is incumbent upon the sponsor to determine what sections, if any, of this ADRG are pertinent and to edit as necessary. The completion guidelines will not present examples of an ADRG for legacy data.

ADRG Completion Guideline Purpose

The purpose of this document is to provide sponsors with a clear, concise set of instructions that facilitates the consistent development of the ADRG from the Analysis Data Reviewer's Guide Template. In addition to the ADRG Completion Guideline, ADRG examples are available as an additional reference.

Organization of This Document

This document has three sections: this guideline overview, ADRG Template Completion Instructions, and ADRG Finalization Instructions. The section number in the ADRG Template Completion Instructions

corresponds directly to the ADRG Template. The ADRG Finalization Instructions describe how to format the document for submission after completing the ADRG Template.

Analysis Data Reviewer's Guide Template Completion Instructions

This section provides companion instructions for the ADRG Template. The section numbering corresponds directly to the ADRG Template. **Note: Certain ADRG Sections include a series of bulleted questions intended to aid FDA Reviewers. Provide complete answers to all bulleted questions. Do not delete the primary questions from the final document.**

1. Introduction

1.1 Purpose

This required section states the purpose of the ADRG. Please refer to the ADRG Template for standard text.

1.2 Acronyms

This optional section documents any sponsor-specific or non-industry standard acronyms used in the ADRG. Standard industry acronyms (e.g., MedDRA, LOINC, CDISC, SDTM, ADaM) do not need to be documented.

Acronym	Translation

1.3 Study Data Standards and Dictionary Inventory

This required section documents the ADaM, SDTM, and Define version(s) used in the study. The version specified for SDTM in this ADRG should match exactly with the similar information found in the Study Data Reviewer's Guide (SDRG). Version(s) of conformance checks are documented in Section 6. It is not necessary to repeat versions of controlled terminology, coding dictionaries or any other standards information used for SDTM.

Versions of standard published questionnaires, scoring algorithms, or other published standards used for analysis should be mentioned within the section pertaining to the analysis datasets in which the standard occurs.

Example:

Database Model	Version
SDTM	SDTM v1.3/SDTM IG v3.1.3

ADaM	ADaM Model Document 2.1 ADaM Implementation Guide v1.0 ADaM Data Structure for Adverse Event Analysis v1.0 ADAM Basic Data Structure for Time-to-Event Analysis v1.0
Data Definitions	define.xml v2.0

End Example

1.4 Source Data Used for Analysis Dataset Creation

This required section is used to describe the type of data sources used to create the analysis datasets. Whereas this ADRG was developed with the assumption that SDTM was the source of data for analysis dataset creation, it is recognized that submissions continue to be made where the source may be SDTM, a non-SDTM clinical database or a combination of these. The purpose of this section is to provide a high level introduction to the types of data used for analysis dataset creation. If SDTM was used as the sole source of data for the development of analysis dataset, then a simple sentence stating this fact is sufficient. Otherwise it is beneficial to highlight other sources and reference sections below where more information can be found or describe in here if no other section provides the necessary content. Content may include but is not limited to the following:

- ✓ In the case of a study which is ongoing or has an ongoing follow-up component, the data cutoff rules may be described.
- ✓ If there are any special cases of data supplied, they should be described here. For example, in some cases sponsors may create customized lookup tables in order to classify certain data, such as adverse events of special interest.
- ✓ There could also be cases where adjudication information was supplied, such as by a clinical review panel. This section can describe the data handling methods used for the adjudicated data. It is not necessary to restate any discussion of the adjudication process that may be contained in the statistical analysis plan (SAP). If there is a Section 5.2.x that specifically addresses the dataset in which adjudicated results are found, then this could be referenced here.

Following are examples of the type of statements that might be included in this section.

Example 1:

"The source data for the ADaM datasets were SDTM version 3.1.3. The protocol for this study consisted of a double blind phase, an open label follow-up phase, and an extended follow-up phase which was used to gather additional survival information. The source data includes all data for the double blind and open label phase, as well as any extended follow-up information that was available as of 31JAN2013. "

Example 2:

"The source data includes all data that were available as of 31JAN2013. However, the sponsor was notified of 3 deaths that occurred after this date. Due to the importance of death information to this analysis, death information only had a separate cutoff date of 30APR2013. "

Example 3:

"In addition to the clinical database, the source data contains file ADJUCRES which contains the results of the clinical outcome review committee meeting held on 14Jul2012. The data supplied to the committee and the review methodologies are described in SAP section 10.5.2. The source files for the review included AE and CE. The adjudication results were entered manually by the committee chair and the results were reviewed and signed off by committee members as described in the protocol. The adjudication records may be linked to the source records using key variables USUBJID, AESPID, AESEQ. They were used to derive the efficacy dataset ADEFF."

Example 4:

"In addition to the clinical database, the source data contains file AEEOI which is used as a dictionary of adverse events of special interest. Since this study has a particular concern regarding specific cardiac events, those specific events are flagged for analysis in the adverse events file. These events were identified after the database was locked and MedDRA coding was applied, and before unblinding. Events were identified by generating a spreadsheet of all unique AEDECOD values. This list (which contained no subject or treatment identifying information) was reviewed by two clinical investigators and each term was assigned a flag value for the special interest category (Y or N). The spreadsheet was converted to a dataset and used to apply the flag value to all adverse event records."

2. Protocol Description

2.1 Protocol Number and Title

This required section provides the protocol number or identifier, title, and versions included in the submission. For protocol amendments, note changes that significantly affected data interpretation or analysis, if any. If an amendment did not significantly affect data analysis, it is not necessary to note.

2.2 Protocol Design in Relation to ADaM Concepts

This required section describes how standard ADaM analysis variables relate to the protocol design. For example, variables such as planned treatment assignments (TRTxxP, TRTSEQP), analysis phase (APHASE), analysis period (APERIOD), subperiod (ASPER), cycle (ACYCLE), etc. are defined in ADSL and other analysis datasets and help define how a particular observation relates to treatment and timing in a protocol. The manner in which these variables are defined for a given study aid the understanding of how the protocol design relates to key analysis concepts used in ADaM. Note that the ADaM model does not regulate how these variables are defined and used to produce a given analysis. Because the terms 'phase' and 'period' are not used in a standard fashion across the industry within the text of a protocol or statistical analysis plan, it is useful to describe how the standard ADaM variables relate to key analysis concepts.

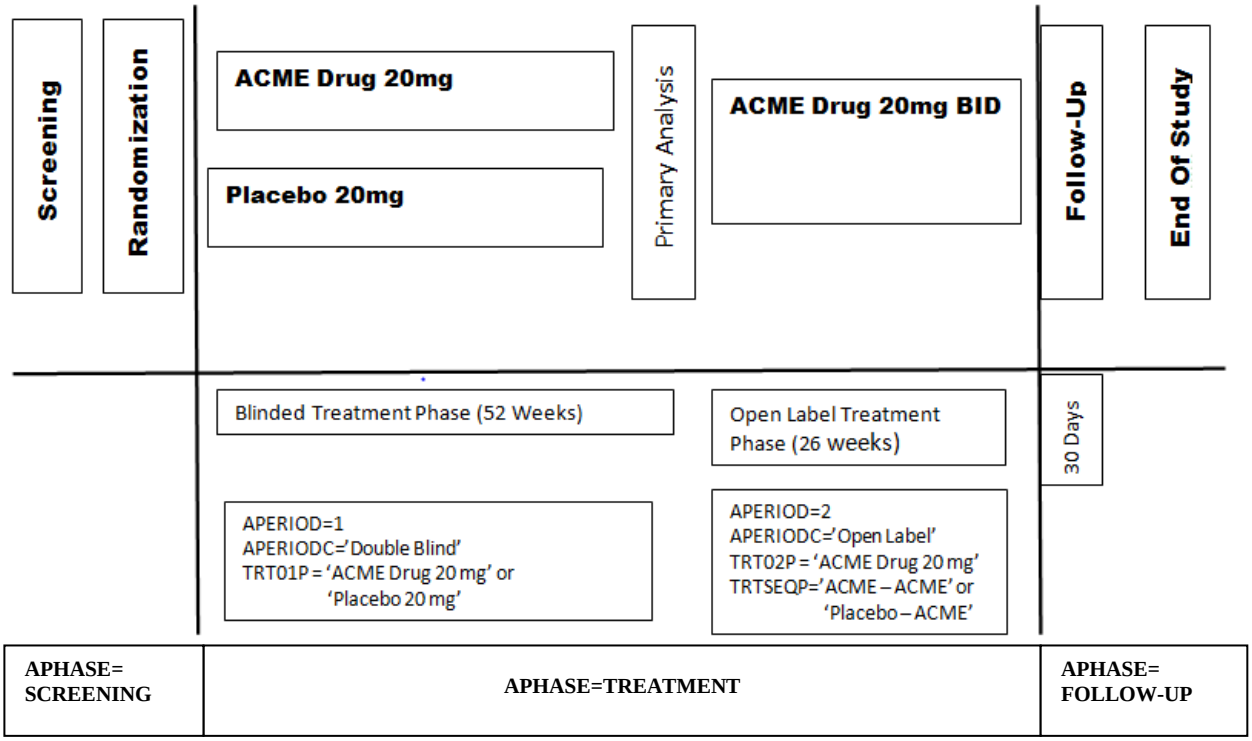
These variables can be described textually and/or via annotation onto a protocol schema.

The textual and the pictorial examples below are for illustrative purposes only:

Text Example:

This is a two arm randomized double-blind to open-label study. APERIOD is used to describe the double-blind period (APERIOD=1) and the open label period (APERIOD=2). TRT01P represents the treatment to which a subject was randomized at the start of the double-blind period and TRT02P represents the open label treatment. The variable TRTSEQP provides a description of the sequence of planned treatments from double-blind to open-label. Records collected prior to randomization are considered to be APHASE=Screening, all records collected during double-blind or open label have APHASE='Treatment' and records collected during the 30 day follow-up have APHASE='Follow-up'

Pictorial Example:



3. Analysis Considerations Related to Multiple Analysis Datasets

3.1 Comparison of SDTM and ADaM Content

The following questions must be answered:

- Are data for screen failures, including data for run-in screening (for example, SDTM values of ARMCD='SCRNFAIL', or 'NOTASSGN') included in ADaM datasets?

  If yes, then refer reader to the appropriate Section 5.2.x below for individual datasets that contain screen failure and/or run-in failure data.

  If no, if screen failure/run-in data are in SDTM but not included in ADaM, then briefly explain why these data are not needed for ADaM.

If screen failure/run-in data are not in SDTM either, then state this.

- Are data taken from an ongoing study?

  If yes then describe the data cut and any issues that might influence record selection from SDTM to ADaM and refer to Section 4.1 for more information as appropriate.

Additional content may include, but is not limited to the following:

Discussion of any differences between SDTM and ADaM in the definitions of derived variable concepts (for example, baseline (xxBLFL versus ABLFL), actual study day (xxDY versus ADY), population flags (SAFETY versus SAFFL))? If yes, refer reader to the appropriate section 5.2.x below as necessary.

3.2 Core Variables

Core variables are those that are represented across all/most analysis datasets. The designation of 'core' is given to a variable that is useful for nearly all analyses (such as age group, sex, race, treatment arm) and/or serves as an important reference variable (such as studyid). If core variables are defined, then a table with the core variable name and a brief description is required.

Since both USUBJID and STUDYID are required by the ADaM model, then at a minimum, this table would contain these two variables.

Example:

Variable Name	Variable Description
USUBJID	Unique subject identifier
STUDYID	Study identifier used for this protocol
SITEID	Unique site identifier for the investigator site
COUNTRY	Country code using ISO
ARM	Planned treatment arm from SDTM
TRT01P	Randomized treatment description
SEX	Sex
AGEGRP1	Age group (<65 and >=65)
RACE	Race description
ITTFL	Flag to indicate inclusion ('Y') or exclusion ('N') from the intent to treat population
HBA1CBL	Baseline value of HbA1C which is used as a covariate in all efficacy analyses

3.3 Treatment Variables

This required section provides information specific to the comparison of the values of SDTM and ADaM treatment variables and the use of planned and actual treatment variables in the analyses. The following questions must be answered. Italicized text is included for guidance. Additional information may be added below these required questions as needed.

ARM versus TRTxxP

<<The purpose of this section is to describe / contrast values of ARM vs. TRTxxP.>>

- Are the values of ARM equivalent in meaning to values of TRTxxP?

  If yes, state this here.

  If no, explain the relationship in text or tabular form.

ACTARM versus TRTxxA

<<The purpose of this section is to describe / contrast values of ACTARM vs. TRTxxA. >>

- If TRTxxA is used, then are the values of ACTARM equivalent in meaning to values of TRTxxA?

  If yes, state this here.

  If no, explain the relationship in text or tabular form.

- If TRTxxA was not used, then state this here.

Use of ADaM Treatment Variables in Analysis

<<The purpose of this section is to describe the use of planned and actual treatment variables in the analysis >>

- Are both planned and actual treatment variables used in analyses?

  If no, state this here.

  If yes, explain at a higher level (e.g., across safety, efficacy) planned versus actual treatment for each type of analysis.

Refer to the appropriate part(s) of Section 5 for use of ADaM treatment variables within individual datasets.

3.4 Subject Issues that Require Special Analysis Rules

This required section provides a description of any situation that occurred which affects the analysis of an individual subject. If there are no issues, then state this. Content may include but is not limited to the following:

- ✓ Did subjects receive the wrong treatment entirely compared to assigned randomization? If yes, how many and explain the deviation(s).
- ✓ Did subjects receive the wrong treatment and/or wrong dose at least once, but not entirely from what was expected per the assigned randomization? If yes, indicate how many subjects received

wrong treatment and/or wrong dose and how to identify these subjects and how it affected the analysis.

- ✓ Did subjects have incorrectly defined randomization strata? If yes, how many and how it affected the analysis.
- ✓ Describe whether any subjects were excluded from analysis datasets (other than screen failure exclusions describe in Section 3.1) and the rationale for the exclusion.
- ✓ Did subjects switch sites? If yes, describe how it was handled in the analysis.
- ✓ Were subjects randomized multiple times at different sites? If yes, give details on how this was handled in the analysis. Note that if this occurs, then the same USUBJID should be used all records for a given subject. If this is not the case, it would be important to note.
- ✓ Were there any protocol deviators that were handled differently in the analysis than what was expected per the definitions in the SAP? If yes, provide details.

3.5 Use of Visit Windowing, Unscheduled Visits, and Record Selection

This required section provides an overview of how the observed visit records from SDTM were used in the analysis. Content should include but is not limited to the following:

- Was windowing used in one or more analysis datasets?

  If no, then state this here.

  If yes, then were the same rules applied to all analysis datasets?

  If yes, then describe how to determine which records were used for analysis.

  If no, refer to Section 5.2.x as appropriate for each individual analysis dataset.

- Were unscheduled visits used in any analyses?

  If no, then state this here.

  If yes, then refer to Section 5.2.x as appropriate.

Additional content may include but is not limited to the following:

Were there records which are included in one or more analysis datasets that were never used for any analysis (such as after follow-up period, screening, etc.)?

3.6 Imputation/Derivation Methods

This required section provides an orientation to the use of record level imputation or derivation and the use of associated ADaM variables.

- If date imputation was performed, were there rules that were used in multiple analysis datasets?

  If yes, then either point the reviewer to the location of the description of these common rules in the specific section of the SAP (for example 'see Section 9.3 in SAP') or describe the rules here. Include in which analysis datasets these common rules were applied.

If common date imputations were not done but imputations were specific to individual analysis datasets, then refer reader to the appropriate part(s) of Section 5 for more information regarding specific analysis datasets where these imputations occurred.

Additional content may include but is not limited to the following:

- ✓ Was DTYPE used in one or more analysis datasets?

  If yes, describe the controlled terminology and associated definitions. Consider referencing SAP if appropriate.

- ✓ Was BASETYPE used in one or more analysis datasets?

  If yes, describe the use of BASETYPE and provide controlled terminology and definitions. Consider referencing SAP if appropriate.

4. Analysis Data Creation and Processing Issues

4.1 Split Datasets

This conditionally required section is intended for use when the sponsor must split an analysis dataset for submission due to size constraints. It is required if any analysis data was split for submission but is optional otherwise. The sponsor should clearly describe the method by which the dataset was split (e.g., by parameter) and notify reviewers of the need to reassemble the analysis dataset prior to any analysis.

Example:

"The Laboratory Chemistry analysis dataset (ADCHM) size exceeded 1 GB so it was split into two datasets for submission (ADCHM1 and ADCHM2). The dataset was split based on the value of PARCAT1. ADCHM2 includes parameters for hepatic function tests (PARCAT1='LFT'); all other lab chemistry parameters can be found in ADCHM1. Reviewers who wish to execute the SAS programs provided for safety laboratory analysis (see Program Inventory in section 7) should first reassemble the two datasets into a single dataset named ADCHM. The metadata describing laboratory chemistry results is described under dataset ADCHM in the define.xml."

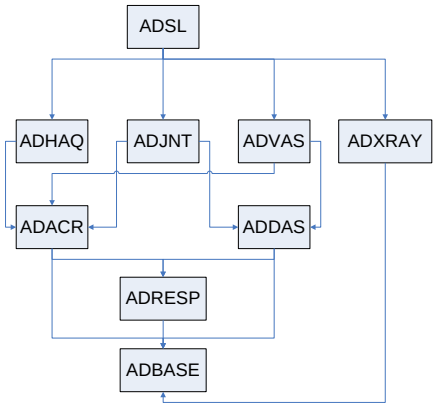
Note that descriptions of decisions regarding how to organize source data for analysis are out of scope for this section. This type of information may be presented in Section 5. For example, source data for laboratory results may be submitted in a single LB dataset, but for analysis, the data may be organized into separate analysis datasets by hematology, chemistry, etc. and in so doing may avoid the need to split the ADaM dataset.

4.2 Data Dependencies

This required section is used to describe any dependencies between analysis datasets. A flowchart is recommended when there are dependencies between analysis datasets beyond a dependency on ADSL. In the case of very minimal analysis dataset dependencies, the user may opt for creating a table to explain the dataset dependencies as an alternative to a flow chart. Where no dependencies exist between analysis datasets beyond a dependency on ADSL, then a simple statement asserting that fact is recommended. Dataset dependencies involving the creation of intermediate analysis datasets should be noted here and further described in Section 4.3, Intermediate Datasets as appropriate.

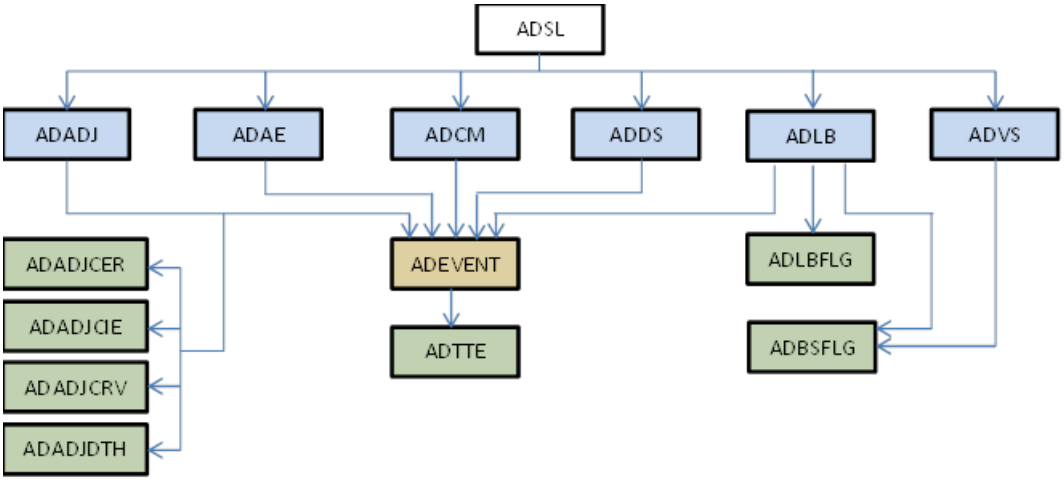
Following are examples of the type of information that might be included in this section

***Example 1*:**



Example 2

In this diagram, blue is used to indicate datasets that have dependency only on ADSL, green indicates dependency on other analysis datasets, and yellow indicates intermediate datasets that were not used for any analysis.



Example 3:

Dataset	**Input Datasets**
ADTTE	ADAE, ADCE, ADSL

Example 4:

There are no analysis dataset dependencies other than ADSL.

4.3 Intermediate Datasets

This optional section is used to describe the existence of intermediate analysis dataset(s) and the resultant analysis dataset(s). Intermediate datasets may have been created during the trial to handles cases when working with complex derivations and/or when a smaller dataset was created from the larger parent analysis parent for reporting purposes and internal review. If applicable, describe any naming convention used for interim datasets.

Following are examples of the type of information that might be included in this section.

Example 1:

Intermediate Dataset	**Output Dataset**
ADEVENT	ADTTE

Example 2:

"No intermediate analysis datasets were created in this trial."

Example 3:

Intermediate Dataset	**Output Dataset(s)**
ADEX	ADEXCYCL, ADEXTOT

"Dataset ADEX is not used in analyses, but is supplied to provide traceability for ADEXCYCL and ADEXTOT and used for a listing. The source data were collected using a per-dose case report form page, which recorded the actual amount infused. The ADEX intermediate file was used to convert actual amounts infused to actual amounts in mg/kg using the last available body weight. This file was then used to create ADEXCYCL which summarizes the total amount received per treatment cycle, and to account for interruptions and changes in dosing regimens. ADEXCYCL was then used to derive summary variables in a one-record-per-subject structure, stored in ADEXTOT."

4.4 Variable Conventions

This optional section provides an orientation to important variable conventions used by a sponsor that cannot be easily established in the define.xml. The ADaM standards allow a good deal of flexibility to choose from standard variables and in some cases to add variables to the standard ones. The definition of individual variables in specific datasets is usually adequately handled in the define.xml. However, it may be useful to explain at a higher level the rationale for using certain standard or additional variables, particularly if a set of conventions applied to multiple datasets. The conventions described here should be those that are over and above the conventions specified in any of the CDISC ADaM documentation. For example, if a sponsor has used conventions for particular variables, such as ANLzzFL or other common used flag variables, AVISIT:AVISITN, PARAM:PARAMCD, etc, these can be described here. It may also be useful to discuss how the setup of certain variables supported analysis.

Following are examples of the type of statements that might be included in this section.

Example 1:

"Study ABC1234 included one subject (USUBJID='abc-7023') who had dosing errors in the first cycle of treatment. This subject was randomized to active treatment but actually received placebo for the first cycle. For this reason, all safety-related analysis datasets included the record-level treatment variables TRTA and TRTAN. In safety analysis tables, subjects are categorized by the actual treatment at the time of the observation. Tables that summarize data by cycle will show a change in subject count from cycle to cycle and are footnoted accordingly."

Example 2:

"The analysis plan calls for change from baseline in efficacy parameters to be calculated from the start of study drug dosing, and also from the start of a given cycle of treatment. This was implemented in datasets ADLB, ADEFF, using the ADaM convention of creating a separate row for each definition of baseline. The type of baseline was distinguished using the variable BASETYPE. In analysis tables, the table title describes the baseline type and a footnote describes the selection criteria using the BASETYPE variable. The BASETYPE variable was not used for datasets that are designed only for safety analysis. Safety datasets (ADVS, ADLBSAF, and ADECG) calculated change from baseline only from the start of study drug dosing. Therefore the BASETYPE variable was not included in these datasets."

5. Analysis Dataset Descriptions

5.1 Overview

This required section provides a summary orientation to the analysis datasets.

Answers to the following question must be provided:

- Do the analysis datasets support all protocol and SAP specified objectives?

  If no, include all objectives listed in the protocol and SAP which are not supported in the analysis datasets and the reason for their absence.

Additional content may include, but is not limited to:

- ✓ Describe any flagging variables that are used in the specific analysis dataset that is not otherwise described above in Section 4: Variable Conventions in particular those which are not used routinely or require additional information to aid interpretation.
- ✓ Location of key safety, efficacy, or other data of special interest.
- ✓ Document the location of adjudication data and the method used to differentiate and to relate this data to data collected at the investigational site.
- ✓ Description of the algorithms followed to calculate timing variables used across datasets (e.g., ADY). These should align with the definitions in define.xml.
- ✓ Document any analysis datasets which are included for supportive purposes but not utilized for submitted analyses.

5.2 Analysis Datasets

This required section provides an inventory of the analysis datasets. The content below is provided to describe standard practice for how to reference the analysis datasets. This may be done in a table, as shown below, or in textual format.

List all analysis datasets included in the submission starting with ADSL followed by all others alphabetically by dataset name.

Include a separate row for each split analysis dataset.

Provide a hyperlink to the sections below from the value in the Dataset-Dataset Label to any analysis dataset that requires additional explanation within the context of the study.

Specify the ADaM class.

Specify the functional category or categories for each analysis dataset.

Include categories of Efficacy, Safety, Baseline or Other Subject Characteristics, and PK/PD (if these data are present).

Additional categories may be defined at the discretion of the sponsor.

Indicate if the analysis dataset is used for the primary analysis.

Optionally, describe the structure of the analysis dataset. If included in the table, the structure should align with define.xml.

Example 1:

Dataset – Dataset Label	Class	Efficacy	Safety	Baseline or other subject characteristics	PK/PD	Primary Objective	Structure
ADSL Subject Level Analysis Dataset	ADSL			x			One observation per subject
ADAE Adverse Event Analysis Dataset	ODS		x				One observation per subject per event
ADEFF Primary Efficacy Analysis Dataset	BDS	x				x	One observation per subject per parameter per visit
ADEX Exposure Analysis Dataset	OTHER		x				One observation per subject per intervention
…							
ADPK PK Parameters Analysis Dataset	BDS				x		One observation per subject per parameter per visit
ADTTE Time to Cardiac Events Analysis Dataset	BDS	x	x			x	One observation per subject per endpoint

End of Example

Example 2:

Dataset – Dataset Label	Class	Efficacy	Safety	Baseline or other subject characteristics	Primary Objective
ADSL Subject Level Analysis Dataset	ADSL			x	
ADBASE	OTHER			x	

Baseline Characteristic Analysis Dataset					
ADAE Adverse Event Analysis Dataset	ODS		x		
ADEFF Primary Efficacy Analysis Dataset	BDS	x			x
ADEX Exposure Analysis Dataset	OTHER		x		

End of Example

5.2.1 ADSL – Subject Level Analysis Dataset

This section is required for the subject level analysis dataset. Provide explanation beyond which is documented in define.xml or the ADaM Implementation Guide and its supplements.

Content may include, but is not limited to the following:

- ✓ Describe breadth of coverage of ADSL.
- ✓ Does ADSL have the same number of subjects (records) as in the SDTM DM domain? If no, then describe any difference.
- ✓ Document which analysis populations are defined in ADSL using their variable names.
- ✓ Are there other analysis datasets that contain other subject level information pertaining to baseline characteristics, disposition, etc? If yes, then list the name of the other subject level datasets.
- ✓ List the variable names for the covariates used for inferential statistical analysis relating to the primary or secondary objectives.
- ✓ Are all covariates used for inferential statistical analysis relating to the primary or secondary endpoint included in ADSL? If no, then indicate where the other covariates can be found.

5.2.x Dataset – Dataset Label

This section is required for each analysis datasets (AD) with hyperlinks that have been provided in Section 5.2 for analysis datasets that benefit from additional description. At a minimum, the dataset containing the primary objective must be described and hyperlinked to the table in section 5.2. Provide explanation beyond which is documented in define.xml or the ADaM Implementation Guide and its supplements. For the dataset(s) that contain the primary efficacy measures, it is advisable to identify the variables that are used for the analyses of these primary endpoints. This is especially important if a non-BDS structure is used since this implies non-standard variable names that may not be obviously related to

efficacy endpoints. When using BDS, it is advisable to indicate the relevant parameter(s), variable(s) analyzed (AVAL, AVALC, CHG, etc), and flags or timing variables as appropriate.

Do not duplicate information that pertains to multiple datasets that may be discussed in Section 3 above. Provide a section number for each AD requiring additional explanation (e.g., 5.2.2, 5.2.3, 5.2.4).

Note that this section header is NOT a Word Header Style. It must be manually edited. This avoids problems with automatic 3- level section numbering that sometimes occurs with Word Header Styles.

Specify key parameters and/or variables of interest. At a minimum, those related to the primary objective should be indicated. Note that it is not necessary to describe the derivation of the primary objective as this would be in the SAP and define.xml. An example is included below for dataset ADEFF. The table is an example and other formats are acceptable.

Content may include, but is not limited to the following:

- ✓ Describe the purpose and breadth of coverage of the AD.
- ✓ Are there substantial number of records in this AD that are found in other AD's, for example a 'parent' and 'child' relationship with another AD? If yes, then indicate the name of the other AD's.
- ✓ Document if there are separate analysis datasets that contain similar content and the purpose for separating the data into multiple analysis datasets. For example, suppose you separate different sensitivity analyses for time to event in different datasets or create different ADs for different baselines instead of using BASETYPE.
- ✓ Are the same number of subjects included in this AD as in the source SDTM domain? If no, then describe the reason for the difference. If this difference is due to screen failures, then this should already have been noted in Section 3.1.
- ✓ Are there derived variables in this AD that are also represented in SDTM but the derivation differs? If yes, then itemize variables and/or describe differences that are not easily understood from the define file.
- ✓ If there are multiple treatment variables in the AD, then describe which treatment variables are used for the analyses generated from this AD.
- ✓ Is BASETYPE used in this AD? If yes, then briefly describe why BASETYPE is needed.
- ✓ Is DTYPE used in this AD? If yes, then briefly describe why DTYPE is needed.
- ✓ Were any external reference data or look up tables used for derivations in this AD? If yes, then indicate whether they are included in the submission and the location.
- ✓ Are there specific flag variables (excluding population flags) that are important for the analyses? If yes, then describe.
- ✓ Describe any derived variables which were created to mitigate issues relating to data that was demonstrably incorrect.
- ✓ If special windowing rules were used in this AD, then describe.

Example:

5.2.1 ADSL – Subject Level Analysis Dataset

In addition to supporting all analyses, ADSL contains variables to also support baseline characteristics and disposition analyses. The population indicator variables, treatment variables and variables used as covariates for statistical analyses (Age Group <65, >=65 - AGEGR1 and Country - COUNTRY) are copied onto all analysis datasets. All subjects in DM, with the exception of screen failures, were included in ADSL.

5.2.2 ADEFF – Efficacy Analysis Dataset

ADEFF is a sponsor-defined analysis dataset following the ADaM Basic Data Structure (BDS) supporting the primary efficacy endpoint of ACR 20 and the secondary efficacy endpoint of ACR. Only one type of baseline was defined in the SAP but BASETYPE = "SCREENING" was included to indicate that baseline is defined as screening for the included parameters. For all analyses of these parameters, AVAL is the variable analyzed.

PARAMCD Value	PARAM Value	Description	Usage
TSJC	Total Swollen Joint Count	Number of swollen joint counts identified by the investigator	Included for traceability for the derivation of the ACR score
TTJC	Total Tender Joint Count	Number of swollen joint counts identified by the investigator	Included for traceability for the derivation of the ACR score
ACR	ACR	The ACRn score is defined as each patient's lowest percentage improvement from baseline of the contributing parameters. A positive value indicates improvement	Utilized in summary statistics and included for traceability in the derivation of ACR 20
ACR20	ACR 20	Binary response as to whether the ACR 20 criteria is met or not	Primary Analysis endpoint analyzed using a CMH test

End of Example

6. Data Conformance Summary

This section describes the validation checks and inputs used to evaluate conformance.

6.1 Conformance Inputs

This required section summarizes how ADaM conformance was established. Answers to the following questions must be provided:

- Were the analysis datasets evaluated for conformance with CDISC ADaM Validation Checks?

  If yes:

  - Version of CDISC ADaM Validation Checks:
  - Specify software used:
    - OpenCDISC
    - Sponsor-defined
    - Other (describe)

- Were the ADaM datasets evaluated in relation to define.xml?

  <<The structure and content of the study analysis datasets should be checked for consistency with the study metadata as described in the define.xml. >>

- Was define.xml evaluated?

  <<The define.xml should be checked for compliance with the appropriate version of the Case Report Tabulations Data Definition Specification (CRT-DDS). It should also be validated against the define.xml XML schemas. For further information, see the "XML Schema Validation for Define.xml" whitepaper on the CDISC website. >>

Because ADaM conformance is not solely established by computerized checks, sponsors may use other methods to assess conformance, such as manual review of the data or internal testing of the clarity of variable metadata. If such methods are used and are worthy of noting, then add additional text as desired.

6.2 Issues Summary

This required section summarizes compliance findings.

- ✓ Summarize findings from an ADaM conformance report (e.g., the OpenCDISC report's Issues Summary tab or similar) in table form. The table below may be used and sponsors may include additional columns if desired.
- ✓ List only those findings that appear in the submission.
- ✓ Annotate issues with a brief, non-technical explanation of the findings.
- ✓ Do not include skipped validation checks or validation checks for which datasets do not exist.
- ✓ If your conformance diagnostics do not include severity, leave that column blank.
- ✓ If non-automated issues were detected, these should be explained as well.

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Count and/or Issue Rate	Explanation

Dataset(s)	Diagnostic Message and/or Check ID	Severity	Count and/or Issue Rate	Explanation

7. Submission of Programs

This section is required if programs are included in the submission. It is advisable for sponsors to discuss the submission of programs with the specific FDA Division review team before preparing a submission. Sponsors should be prepared for FDA reviewers to conduct independent quality validation to verify results in submitted clinical studies. The sponsor should try to understand as clearly as possible if and how their reviewer will use the submitted programs and what type of 'packaging' will best support the review. Some points to consider might include:

- ✓ Which programs should be submitted?
  It is common to include programs that create analysis datasets, as well as programs that generate primary and secondary efficacy analyses and important safety analyses, and to omit programs that are either repetitive (for example, closely related adverse event summaries) or unlikely to affect the statistical conclusions of the study (for example, listing programs). This section can be used to discuss the rationale for including/excluding programs. If analysis results metadata and/or annotated table mockups are included in the submission package, this may be noted here and may obviate the need to send table programs.
- ✓ Do the programs need to be executable?
  The FDA has no policy that requires program code to be executable. If there is a question about whether to do any extra work to make programs executable, it is best to discuss with the reviewer.
- ✓ How should programs that include macro code be handled?
  Although there is a common perception that macro code is unacceptable to submit, this is not necessarily true; reviewers may consider macro code usable. Macro code may nevertheless be difficult for the sponsor to package. If submitting macro code is problematic, the sponsor may discuss with the reviewer whether an alternate approach is acceptable. Some possible alternate approaches might include:

  1) Submitting validation programs that do not contain macro code
  2) Submitting resolved macro code.
     If the results were programmed using SAS, the MFILE/MPRINT options can be used to write the resolved macro code to a separate file. The resolved macro code could be submitted in place of the original program.

Program files that are submitted should have documentation that identifies the inputs and outputs to the program and allows the reviewer to connect the program with the results that it supports. Industry best

practice advises that the inputs/outputs be specified in the program header block. If this practice is adhered to it may be worthwhile noting this. Internal comments to explain important sections of logic are highly recommended. It is worthwhile to accompany the programs with a statement about the reasoning for selecting the programs that were submitted. If there are any special methods for preparing the code (such as the alternatives described above), then the relationship between the submitted code and the outputs should be explained.

Example 1:

"The submitted programs include all programs for producing analysis datasets, including a macro that was used to standardize units of study drug dosing, which is referenced in several datasets. The programs also include all programs for primary and secondary efficacy.

Program inventory:

Program name	Output	Inputs	Macros used
adsl.sas	adsl	dm,ex,ds,ie	dsunit.sas
adtte.sas	adtte	Dm, ds, adsl	
teff.sas	Table 14.2.1.1	adtte	
tsurv.sas	Table 14.2.2.1	adsl	

End of Example

E***xample 2:***

"The submitted programs include the programs that produce each analysis datasets. The name of the creation program is identical to the name of the analysis dataset. For example, ADSL.SAS produces the dataset ADSL. All inputs, such as SDTM domains or other analysis datasets, are specified in the program header. No macros were used in the production of analysis datasets.

The submitted programs also include programs that produce primary efficacy and safety tables. The table below associates the table number with the program that produced it. The second table lists the macros that were used in table production. Note that one macro may be used to produce multiple tables. The header block in each table program clearly specifies this information as well.

Table Number	**Abbreviated Table Title**	**Program Name**
14.1.1	Subject Disposition	Subjdisp.sas
14.1.2	Demographics and Baseline Characteristics	Demog.sas
14.2.2.1	Incidence of Clinical Success	Clinimprove.sas
14.2.2.2	Incidence of Clinical Improvement	Clinimprove.sas
14.3.1.1	Overall Summary of Treatment Emergent Adverse Events	Teaesum.sas

14.3.1.3	Treatment Emergent Adverse Events by Body System and Preferred Term	Teaesum.sas
Table 14.3.4.1	Change from Baseline for Hematology Laboratory	Labcfb.sas
Table 14.3.4.2	Change from Baseline for Chemistry Laboratory	Labcfb.sas

Macro Name	**Macro Description**
statdesc.sas	Prepares descriptive statistics from the specified data set and variable
statcnt.sas	Prepares count and percent statistics from the specified data set and variable
statcnp.sas	Prepares count/percentage and p-value from the specified data set and variable
aeout.sas	Counts Adverse Events by body system and preferred term
auxlbout.sas	Summarizes change from baseline values of lab and vital signs parameters

End of Example

8. Appendix

This is an optional section that can be used if needed. If it is not needed, then delete this section entirely.

Analysis Data Reviewer's Guide Finalization Instructions

This section describes how to format the document for submission after completing the ADRG Template.

1. Create hyperlinks from dataset names in section 5.2 to descriptions in 5.2.x

Select the text in the first column in the table in section 5.2 that needs a hyperlink. **Right click the selected text** and choose "**Hyperlink**" from the menu. In the left panel of the Hyperlink window, make sure that "**Place in this document**" is selected. Then, in the list of document places select the dataset name's header under ADaM Domains (e.g. ADAE – Adverse Events Analysis Dataset) and click **OK**. Ctrl+click the hyperlink to test it.

2. Do not remove unused sections from the document

For consistency, it is best to leave unused sections in the document and indicate 'N/A', or 'This section does not apply'. It is acceptable to remove the optional Appendix 8 section.

3. Update the Table of Contents, document header and version date

After all edits have been completed, update the table of contents at the top if the document. **Right click on any line in the table** and select "**Update Field**." In the dialog window, select "**update entire table**," then click **OK**.

Do not edit the document header or footer. The study number in the header references the study number on the title page. When you edit the study number on the title page, the study number in the header is updated automatically. To update the version date on the title page and the PDF creation date in the document footer, **save and close the document**, then **re-open it**. All necessary fields will be updated.

4. Convert the document to PDF format

These instructions are for Microsoft Word 2003 or newer, using either the Adobe Acrobat plug-in or the MS Office PDF creation feature.

4.1 Using the Adobe Acrobat plug-in for Microsoft Office:

Click the **Acrobat tab** in the Word menu at the top of the screen. Select "**Create PDF**." If a dialog window pops up asking you to save and continue, click **Yes**. In the second dialog window, **navigate to the directory** in which you want to save the PDF, **name the file** "**analysis-data-reviewers-guide.pdf"**, and click **Save**.

4.2 Conversion without Adobe Acrobat plug-in:

Click the **Office button** at the top left of your screen. Select "**Save As**," then "**PDF or XPS**". **Navigate to the directory** in which you want to save the PDF, name **the file** "**analysis-data-reviewers-guide.pdf"**, and click **Save**.

4.3 Formatting and verifying the PDF

Open the PDF. Go to the **File menu** and select **"Properties**." Navigate to the **Initial View tab.** In the drop-down menu for **Navigation tab**, select "**Bookmarks Panel and Page**." In the drop-down menus for both **Page Layout** and **Magnification**, select "**Default**." Click **OK.**

Go to the **Document menu** and select "**Reduce File Size**." In the drop-down menu, select "**Acrobat 5.0 and later**." Click **OK**, then **navigate to the directory** in which you want to save the PDF, **name the file** "**analysis-data-reviewers-guide.pdf"**, and click **Save**.

Go to **File**, and select "**Properties**." Verify at the bottom of the dialog window that the **PDF version is 1.4.**