



Analysis Data Model (ADaM)

Prepared by the
CDISC Analysis Data Model Team

Notes to Readers

This is Version 2.1 of the Analysis Data Model (ADaM) Document. It includes modifications so that it corresponds to Version 1.0 of the Analysis Data Model Implementation Guide (ADaMIG).

Revision History

Date	Version	Summary of Changes
Dec. 17, 2009	2.1 Final	Released version reflecting all changes and corrections identified during comment period.
May 30, 2008	2.1 Draft	Draft for comment.
Aug. 11, 2006	2.0 Final	Final document
31 May 2006	2.0 Draft	Incorporate comments from public review
15 Feb 2006	2.0 Draft	Reformatted from General Considerations v1.0, incorporating Subject-level model, emphasizing requirements and naming and content rules and guidelines.

Note: Please see [Appendix G](#) for Representations and Warranties; Limitations of Liability, and Disclaimers.

Contents

1	Introduction / Purpose	3
2	Background / Motivation	4
3	Overview of the Analysis Data Model	6
3.1	Fundamental Principles	6
3.1.1	Traceability	6
3.2	Analysis Data Flow	7
3.3	Metadata Components	8
4	Analysis Datasets	10
4.1	Practical Considerations	10
4.1.1	The Number and Content of Analysis Datasets	10
4.1.2	Analysis Dataset and Variable Naming Conventions	10
4.1.3	Ordering of Variables	11
4.2	ADaM Data Structures	11
4.2.1	The Subject-Level Analysis Dataset (ADSL) Structure	11
4.2.2	The Basic Data Structure (BDS)	11
4.2.3	Future ADaM Data Structures	12
5	ADaM Metadata	13
5.1	Analysis Dataset Metadata	13
5.1.1	Illustration of Analysis Dataset Metadata	14
5.2	Analysis Variable Metadata	15
5.2.1	Analysis Parameter Value-Level Metadata	15
5.2.2	Illustration of Analysis Variable Metadata, Including Analysis Parameter Value-Level Metadata	17
5.3	Analysis Results Metadata	21
5.3.1	Illustration of Analysis Results Metadata	23
6	Subject-Level Analysis Dataset	28
6.1	Data for Subjects Not Analyzed	29
Appendix A	References	30
Appendix B	Definitions	31
Appendix C	Abbreviations and Acronyms	33
Appendix D	Illustration of Analysis-Ready	34
Appendix E	Composite Endpoint Example	36
Appendix F	Revision History	40
Appendix G	Representations and Warranties; Limitations of Liability, and Disclaimers	41

1 Introduction / Purpose

The Analysis Data Model (ADaM) document specifies the fundamental principles and standards to follow in the creation of analysis datasets and associated metadata. Metadata are “data about the data” or “information about the data.” The Analysis Data Model supports efficient generation, replication, and review of analysis results.

The design of analysis datasets is generally driven by the scientific and medical objectives of the clinical trial. A fundamental principle is that the structure and content of the analysis datasets must support clear, unambiguous communication of the scientific and statistical aspects of the trial.

The purpose of ADaM is to provide a framework that enables analysis of the data, while at the same time allowing reviewers and other recipients of the data to have a clear understanding of the data’s lineage from collection to analysis to results. Whereas ADaM is optimized to support data derivation and analysis, CDISC’s Study Data Tabulation Model (SDTM) is optimized to support data tabulation.

The ADaM document (i.e., this document) provides the core and defines the spirit and intent of the ADaM concepts and standards. It outlines the fundamental principles to follow in constructing analysis datasets and related metadata. Four types of ADaM metadata (i.e., analysis dataset metadata, analysis variable metadata, analysis parameter value-level metadata, and analysis results metadata) are described in this document and examples are provided.

The subject-level analysis dataset (ADSL) is introduced in this document. (Refer to Section 4.2.) ADSL and its related metadata are required in a CDISC-based submission of data from a clinical trial even if no other analysis datasets are submitted.

This document also introduces the ADaM Basic Data Structure (BDS) that is to be used for the majority of ADaM datasets, regardless of the therapeutic area or type of analysis. (Refer to Section 4.2.) Though the BDS generally supports the majority of statistical analyses, a study also includes analysis datasets of specific standardized structures to represent additional analysis information, such as subject-level analysis dataset (ADSL) and ADAE (adverse event analysis dataset).

This document serves as the foundation for the ADaM Implementation Guide (ADaMIG) which specifies the standardized implementation of these core concepts. The ADaMIG specifies ADaM standard dataset structures and variables, including naming conventions. The ADaMIG also specifies standard solutions to implementation issues.

In adopting the principals and standards of ADaM when constructing analysis datasets and their associated metadata, it cannot be emphasized enough that early and effective communication between reviewers or other recipients of the data and sponsors is essential if to achieving the full benefits of analysis datasets are to be achieved.

In an effort to provide illustrations of ADaM concepts, examples are provided that refer to specific programming languages. Throughout ADaM documents, references to specific vendor products are examples only and should not be interpreted as an endorsement of these products.

Note that the examples in this document are only intended as illustrations and should not be viewed as a statement of the standards themselves.

2 Background / Motivation

The marketing approval process for regulated human health products often includes the submission of data from clinical trials. In the United States, data are required elements of a submission to the United States Food and Drug Administration (FDA).

The FDA established the regulatory basis for electronic submission of data in 1997 with the publication of regulations on the use of electronic records in place of paper records (21 CFR Part 11). In 1999, the FDA standardized the file format (SAS® Version 5 Transport Files¹) for electronically submitting data collected in clinical trials. This was explained in the first of a series of guidance documents that described the submission of clinical data and data definition files (define.pdf).

Though the 1999 guidance was withdrawn in 2006, datasets are still submitted using the SAS transport file format, accompanied by a “define file.” The define file is a data definition document which provides a list of the datasets included in the submission along with a detailed description of the contents of each dataset (i.e., the metadata for the submitted datasets).

As of 2005, metadata can be submitted using an extensible markup language (XML) format (define.xml) rather than the portable document format (define.pdf), as described in the FDA document regarding study data specifications [7]. More information about define.xml can be found on the CDISC website [2].

In parallel with the development of clinical data submission guidance, the FDA has adopted the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) standards for regulatory submissions and has issued a guidance document on the electronic Common Technical Document (eCTD) as its framework for electronic submissions of pharmaceutical product applications. Revision 2 of this guidance was posted in 2008 [6].

According to FDA guidance documents on the eCTD, submitted data can be classified into four types: 1) data tabulations, 2) data listings, 3) analysis datasets, and 4) subject profiles. These are collectively referred to as Case Report Tabulations (CRTs) [6]. The specification for organizing datasets and their associated files in folders within the submission is summarized in the following figure, from the “Study Data Specifications” [7].

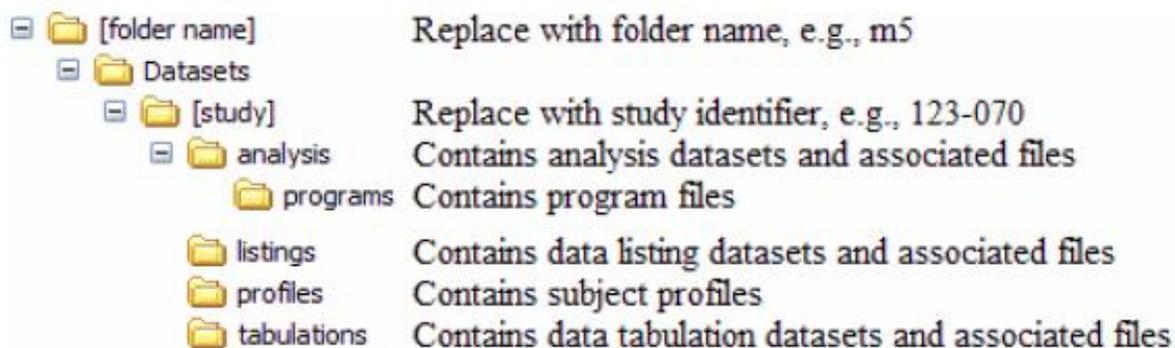


Figure 2.1 Specification for organizing study datasets and their associated files in folders [7].

Data tabulation datasets and analysis datasets are defined as:

- **Study Data Tabulations (SDTM)** – datasets containing data collected during the study and organized by clinical domain. These datasets are described in the CDISC Study Data Tabulation Model [5] and CDISC Study Data Tabulation Model Implementation Guide (SDTMIG) [4].

¹ SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

- **Analysis Datasets (ADaM)** – datasets used for statistical analysis and reporting by the sponsor, submitted in addition to the SDTM domains. ADaM datasets are the authoritative source for all data derivations used in statistical analyses. These datasets are described in this document (the CDISC Analysis Data Model document) and in the CDISC ADaM Implementation Guide [1].

Standardized analysis datasets and metadata provide benefits to recipients of the data beyond clear communication and transparency. Once trained in the principles of standardized datasets, reviewers and other recipients of the data can work with the data more efficiently with less preparation time. In addition, standardized structures allow the development of software tools that facilitate access to, derivations, analyses, and replication and review of the analysis results.

SDTM is not designed to support statistical analysis. ADaM datasets incorporate derived and collected data (from various SDTM domains, other ADaM datasets, or any combination thereof) into one dataset that permits analysis with little or no additional programming. Examples of issues that are not easily handled within SDTM are analysis windows, complicated algorithms, and imputation of missing values.

Since the ADaM metadata explain how the ADaM datasets were created from the SDTM source data, variables that have been derived or imputed in ADaM datasets should not be copied back into the SDTM source data. Attempting to do so would introduce circular dependencies into the data flow and could disassociate important relationships between variables.

For the purposes of simplifying this document, analysis datasets are discussed within the context of electronic submissions to the FDA. Since inception, the CDISC ADaM team has been encouraged and informed by FDA statistical and medical reviewers who participate in ADaM meetings as observers and who have participated in CDISC-FDA pilots. The origin of the fundamental principles of ADaM is the need for transparency and completeness of communication with and scientifically valid review by medical and statistical reviewers. The ADaM standard has been developed to meet the needs of the FDA and industry. ADaM is applicable to a wide range of drug development activities in addition to FDA regulatory submissions. It provides a standard for transferring datasets between sponsors and contract research organizations (CROs), development partners and independent data monitoring committees. As adoption of the model becomes more widespread, in–licensing, out–licensing, joint ventures, and mergers are facilitated by a common model for analysis datasets and associated metadata across sponsors.

3 Overview of the Analysis Data Model

3.1 Fundamental Principles

Fundamental Principles

Analysis datasets and their associated metadata must:

- facilitate clear and unambiguous communication
- provide traceability between the analysis data and its source data (ultimately SDTM)
- be readily useable by commonly available software tools

Analysis datasets must:

- be accompanied by metadata
- be analysis-ready

The overall principle in designing analysis datasets and related metadata is that there must be clear and unambiguous communication of the content and source of the datasets supporting the statistical analyses performed in a clinical study. Inherent in this principle is a need for traceability to allow an understanding of where an analysis value (whether an analysis result or an analysis variable) came from, i.e., the data's lineage or relationship between an analysis value and its predecessor(s). See Section 3.1.1 for a more detailed description of traceability.

Sponsors should strive to submit "analysis-ready" datasets, i.e., analysis datasets that have a structure and content that allows statistical analysis to be performed with minimal programming. An analysis-ready dataset is ready to be used directly by statistical analysis software with only minimal additional processing, for example a sorting of the observations or the selection of the appropriate records from the analysis dataset. No complex data manipulations such as transformations or transpositions are required to perform the supported analysis. This approach eliminates or greatly reduces the amount of programming required by analysts such as statistical reviewers. Appendix D gives an example of applying this principle in SAS, but the concepts apply to all statistical software packages. Note that within the context of ADaM, at a minimum analysis datasets contain the data needed for the review and re-creation of specific statistical analyses. It is not required that the data be collated into analysis-ready datasets solely to support data listings or other non-analytical displays, although some may choose to do so.

Analysis datasets must be readily usable by commonly available software tools, and must be associated with metadata. Ideally the metadata are machine-readable. Metadata and other documentation should provide clear and concise communication of the analyses, including statistical methods, assumptions, derivations and imputations performed. The metadata, programs and other documentation serve to systematize the analyses described in the Statistical Analysis Plan (SAP) as well as other analyses performed. These are discussed in detail in Section 5.

3.1.1 Traceability

The concept of traceability is a cornerstone of the Analysis Data Model. This property enables the understanding of the data's lineage or the relationship between an element and its predecessor(s). Traceability facilitates transparency, which is an essential component in building confidence in a result or conclusion. Ultimately, traceability in ADaM permits the understanding of the relationship among the analysis results, the analysis datasets, and the SDTM domains.

Traceability is built by clearly establishing the path between an element and its immediate predecessor. The full path is traced by going from one element to its predecessors, then on to their predecessors, and so on, back to the SDTM domains, and ultimately to the data collection instrument. Note that the CDISC Clinical Data Acquisition Standards Harmonization (CDASH) standard is harmonized with SDTM and therefore assists in assuring end-to-end traceability. Traceability establishes across-dataset relationships as well as within-dataset relationships. For example, the metadata for flags and other variables within the analysis dataset enables the user to understand how (and, to some extent, why) derived records were created.

There are two levels of traceability:

- Metadata traceability enables the user to understand the relationship of the analysis variable to its source dataset(s) and variable(s) and is required for ADaM compliance. This traceability is established by describing (via metadata) the algorithm used or steps taken to derive or populate an analysis value from its immediate predecessor. Metadata traceability is also used to establish the relationship between an analysis result (e.g., a p-value) and analysis dataset(s).
- Data point traceability enables the user to go directly to the specific predecessor record(s). It should be implemented if practically feasible. This level of traceability can be very helpful when a reviewer is trying to trace the path of a complex data manipulation. This traceability is established by providing clear links in the data (e.g., via use of --SEQ variable) to the specific data values used as input for an analysis value. Note that there may be situations where data point traceability is difficult, impracticable, or even infeasible, e.g., electroencephalographic recordings in polysomnography studies where key outcomes may be based upon spectral edge parameters derived from the absolute power results via Fast Fourier Transform.

When traceability is successfully implemented, reviewers are able to identify:

- information that exists in the submitted SDTM study tabulation data
- information that is derived or imputed within the ADaM analysis dataset
- the method used to create derived or imputed data
- information used for analyses, in contrast to information that is not used for analyses yet is included to support traceability or future analysis

3.2 Analysis Data Flow

A conceptual diagram of a typical general flow of data from its source through the analysis results is shown in [Figure 3.2.1](#). The schematic only illustrates one reasonable scenario. It is not intended to diagram all possible relationships and components, nor to indicate or imply that this is the only way to operationalize the process. For example, metadata may actually inform or drive the process rather than be an output of the process.

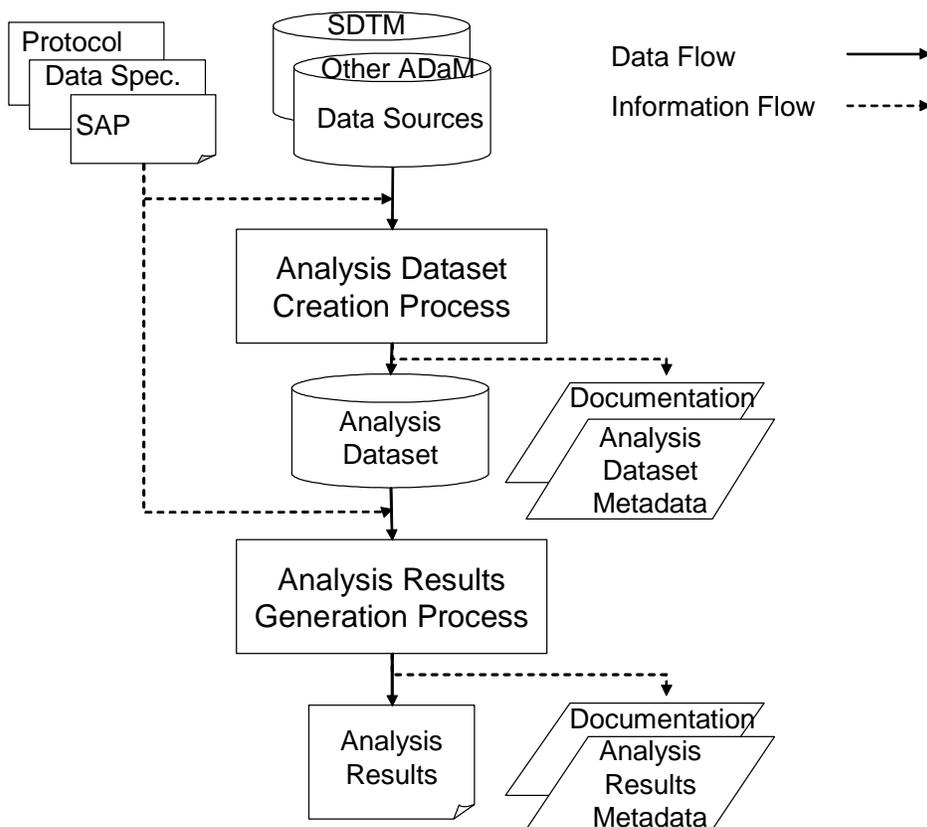


Figure 3.2.1: Analysis Data Flow Diagram Showing One Scenario for the Flow of Data and Information

Given that the ADaM standard has been developed as part of the larger family of CDISC standards, it is assumed that the sources are either SDTM or other analysis datasets such as the Subject-Level Analysis Dataset (ADSL). A CDISC-compliant submission includes both SDTM and ADaM datasets; therefore, to facilitate traceability, the metadata needs to describe the relationship between these two collections of datasets.

To facilitate clear communication, a distinction is made between the processes of Analysis Dataset Creation and Analysis Results Generation. These two processes have distinct purposes and consequently require different types of metadata, as outlined in the following section.

Analysis Dataset Creation – The processing and programming steps used to create the analysis datasets. As shown in Figure 3.2.1, the analysis dataset creation program is developed based on the analysis plans and dataset specifications. The data going into the program are the source data (i.e., SDTM and/or other analysis datasets), and the output is the analysis dataset.

Analysis Results Generation – The programming steps used to generate analysis results (e.g., summary or inferential statistics presented in tabular or graphical presentations). As shown in Figure 3.2.1, the analysis results generation program is developed based on the statistical analysis plan, data derivation and dataset specifications. The data going into the program are in the analysis dataset and the output is the analysis result.

3.3 Metadata Components

The analysis datasets and ADaM metadata facilitate the review of the clinical trial data and the analyses performed. There are four types of metadata described in this document. These include:

- Analysis dataset metadata describe each analysis dataset, including a brief description of the contents. (Refer to Section 5.1) This type of metadata is required for all ADaM datasets.
- Analysis variable metadata describe the variables within the analysis datasets, including information about the source and creation of the analysis variables, e.g., detailed descriptions of algorithms involved and/or references to analysis dataset creation programs. (Refer to Section 5.2) This type of metadata is required for all ADaM datasets.
- Analysis parameter value-level metadata describe the measurements or analysis endpoints “within” an analysis parameter (i.e., for each unique value of the analysis parameter). This form of metadata is particularly needed when the data structure allows a variable to contain multiple types of measurements or analysis endpoints, as in the ADaM Basic Data Structure. (Refer to Section 5.2.1) This type of metadata is required for all ADaM BDS datasets.
- Analysis results metadata describe analysis results (as specified by the sponsor), including which analysis dataset was used and information about the analyses performed. (Refer to Section 5.3) These metadata provide traceability from a result used in a statistical display to the data in the analysis datasets. Analysis results metadata are not required. However, best practice is that they be provided to assist the reviewer by identifying the critical analyses, providing links between results, documentation, and datasets, and documenting the analyses performed.

The first three types of metadata describe the analysis dataset. They are developed during the analysis dataset creation process. The analysis dataset metadata describe the analysis dataset as a whole, whereas the analysis variable metadata and analysis parameter value-level metadata describe the variables and observations within the dataset.

To document either the analysis data creation process or the analysis results generation process, the metadata can include pseudo code, code fragments, links to programs, and/or links to the Protocol, Statistical Analysis Plan, or other documents.

4 Analysis Datasets

4.1 Practical Considerations

Analysis datasets must:

- include a subject-level analysis dataset named “ADSL” (Refer to Section 6)
- consist of the optimum number of analysis datasets needed and have enough self-sufficiency to allow analysis and review with little or no additional programming or data processing
- be named using the convention “ADxxxxxx”
- use ADaM standard variable names and naming conventions when available
- maintain the values and attributes of SDTM variables if copied into analysis datasets without renaming (i.e., adhere to the “same name, same meaning, same values” principle of harmonization)
- apply naming conventions for datasets and variables consistently across studies within a given submission and across multiple submissions for a product

4.1.1 The Number and Content of Analysis Datasets

In creating the analysis datasets supporting the analytic results in a clinical study report or submission, one goal is to have the optimum number of analysis datasets needed to perform the various analyses (with the minimum requirement being ADSL). There is no requirement that there be a separate dataset for different analyses; a single dataset can support multiple analyses. There is also no requirement for every data summary to be supported by an analysis dataset. In addition, there is no requirement that every SDTM domain have a corresponding analysis dataset. The sponsor determines the analysis datasets to be created.

Multiple datasets (e.g., SDTM, other analysis datasets) may be needed for the creation of a single analysis dataset. This is necessary so that the analysis dataset contains all of the variables required for performing the statistical analysis it is designed to support. For example, data may be required from ADSL and the disposition (DS), demographics (DM), subject characteristics (SC), vital signs (VS), questionnaires (QS), and exposure (EX) domains for creating a single analysis dataset.

Analysis datasets are designed to facilitate analysis and review with minimal programming or data processing. Redundancy (i.e., same data appearing in multiple datasets) between analysis datasets is often necessary so that the datasets are analysis-ready (e.g., age in the adverse event analysis data set as well as in an efficacy analysis dataset). Similarly, variables and records can also be included that are not actually used in any of the submitted analyses, but are still of interest to the sponsor or reviewer (e.g., an identification flag for subjects who had an event of clinical interest) or that support traceability.

An example of a composite endpoint requiring complex algorithms and input from multiple datasets is shown in [Appendix E](#).

4.1.2 Analysis Dataset and Variable Naming Conventions

Analysis datasets are named using the convention “ADxxxxxx.” The subject-level analysis dataset is named “ADSL” as described in Section 6. For all other analysis datasets, the xxxxxx portion of the name is sponsor-defined, using a common naming convention across a given submission or multiple submissions for a product. In developing naming conventions, sponsors should consider the requirements noted in the eCTD guidance document

[6], as well as the need to conform to the SAS Transport format requirements (e.g., the total length of the name cannot exceed 8 characters).

Naming conventions for variables created (not to be confused with any standard variables required by SDTM) within the analysis dataset should follow the standard variable names and naming conventions defined in the ADaMIG. Otherwise the analysis variable names are sponsor-defined, and, as much as possible, should also follow a common naming convention across studies within a given submission and across multiple submissions for a product.

Any ADaM variable with the same name as an SDTM variable is required to be a copy of the SDTM variable, and its label, attributes, and values cannot be modified. ADaM adheres to the principle of harmonization known as "same name, same meaning, and same values."

Refer to the ADaMIG [1] for more general variable naming conventions.

4.1.3 Ordering of Variables

Ideally, the ordering of the variables in the analysis dataset follows a logical ordering (not simply alphabetic). Refer to the FDA "Study Data Specifications" [7] for more information regarding the ordering of variables in the analysis dataset. It is recommended that the sponsor define a convention for ordering of variables within a dataset and then apply this ordering consistently for all analysis datasets. The ordering of the variables within a dataset should match the order of the variables as presented in the define file.

4.2 ADaM Data Structures

There are two ADaM standard data structures described within this document and the ADaMIG: the subject-level analysis dataset (ADSL), and the Basic Data Structure (BDS).

4.2.1 The Subject-Level Analysis Dataset (ADSL) Structure

The ADSL dataset structure has one record per subject and contains variables such as subject-level population flags, planned and actual treatment variables, demographic information, randomization factors, subgrouping variables, and important dates. ADSL contains required variables (as specified in the ADaMIG) plus other subject-level variables that are important in describing a subject's experience in the trial. ADSL and its related metadata are required in a CDISC-based submission of data from a clinical trial even if no other analysis datasets are submitted. Refer to Section 6 for a detailed description of ADSL.

Although it would be technically feasible to take every single data value in a study and include them all as variables in a subject-level dataset such as ADSL, that is not the intent or the purpose of ADSL. The correct location for key endpoints and data that vary over time during the course of a study is in a BDS dataset.

4.2.2 The Basic Data Structure (BDS)

A BDS contains one or more records per subject, per analysis parameter, per analysis timepoint. Analysis timepoint is conditionally required, depending on the analysis. In situations where there is no analysis timepoint, the structure is one or more records per subject per analysis parameter. This structure contains a central set of variables that describe the analysis parameter (e.g., PARAM and related variables) and contain the value being analyzed (e.g., AVAL and AVALC and related variables). Other variables in the dataset provide more information about the value being analyzed (e.g., the subject identification) or describe and trace the derivation of it (e.g., DTYPE) or enable the analysis (e.g., treatment variables, covariates). The BDS supports parametric and nonparametric analyses such as ANOVA, ANCOVA, categorical analysis, logistic regression, Cochran-Mantel-Haenszel, Wilcoxon rank-sum, time-to-event analysis, etc. It is often optimal to have more than one BDS analysis dataset. Refer to the ADaMIG [1] for details regarding the BDS standards.

Though the BDS supports the majority of statistical analyses, it does not support all statistical analyses. For example, it does not support simultaneous analysis of multiple dependent (response/outcome) variables or correlation analysis across a range of response variables. The BDS was not designed to support analysis of incidence of adverse events or other occurrence data.

4.2.3 Future ADaM Data Structures

The ADaM team is currently working on a specification document for an ADAE dataset supporting analysis of incidence of adverse events. ADAE may be the first example of a more general structure supporting analysis of incidence data, such as adverse events, concomitant medications, etc.

5 ADaM Metadata

ADaM Metadata

- Analysis dataset metadata
- Analysis variable metadata
- Analysis parameter value-level metadata
- Analysis results metadata

The underlying assumptions, statistical methods, transformations, derivations and imputations performed in the analysis of a clinical trial should be communicated clearly and in such a manner that the values and results can be easily replicated. ADaM metadata facilitates this communication by providing specification of details and links between the general description of the analysis (as found in the protocol's data analysis section, SAP, or the reported analysis methods), the analysis results, the data used in the analysis, and the SDTM domains. The following sections describe in detail the components of the ADaM metadata.

The metadata structures described for the analysis dataset metadata and the analysis variable metadata are based on the Case Report Tabulation Data Definition Specification Standard, version 1.0.0 (CRT-DDS) [2]. Refer to that document for additional details.

The examples of metadata included in this document are for illustration only and are not intended to dictate or recommend presentation style, format or process. In addition, the italicized rows included in some of the illustrations are only for reference, as a reminder of the field definitions. It is not intended that this row be included by sponsors in metadata.

5.1 Analysis Dataset Metadata

Analysis dataset metadata provide information about the analysis dataset, including a description of the contents of the dataset. Best practices strongly recommend that every analysis dataset be described using the metadata fields listed in [Table 5.1.1](#). ADSL and BDS analysis datasets (i.e., ADaM-compliant) must be described by these metadata fields. Practical experience teaches that analysis datasets using structures other than these may occasionally be needed. It is suggested that these also be described by the metadata fields in [Table 5.1.1](#).

Table 5.1.1 Analysis Dataset Metadata Fields

Analysis Dataset Metadata Field	Description
DATASET NAME	The file name of the dataset, hyperlinked to the corresponding analysis dataset variable descriptions (i.e., the data definition table) within the define file.
DATASET DESCRIPTION	A short descriptive summary of the contents of the dataset
DATASET LOCATION	The folder and filename where the dataset can be found, ideally hyperlinked to the actual dataset (i.e., XPT file)
DATASET STRUCTURE	The level of detail represented by individual records in the dataset (e.g., “One record per subject,” “One record per subject per visit,” “One record per subject per event”).
KEY VARIABLES OF DATASET	A list of variable names that parallels the structure, ideally uniquely identifies and indexes each record in the dataset.
CLASS OF DATASET	Identification of the general class of the dataset using the name of the ADaM structure (i.e., “ADSL,” “BDS”) or “OTHER” if not an ADaM-specified structure
DOCUMENTATION	Description of the source data, processing steps, and analysis decisions pertaining to the creation of the dataset. Software code of various levels of functionality and complexity, such as pseudo-code or actual code fragments may be provided. Links or references to external documents (e.g., protocol, statistical analysis plan, software code) may be used.

5.1.1 Illustration of Analysis Dataset Metadata

In this example, the data being analyzed are the total scores of the 11-item Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-Cog). The assessment is a questionnaire, so the Questionnaire (QS) SDTM domain is used for the collected data. Table 5.1.1.1 illustrates the analysis dataset metadata for the ADAS-Cog analysis dataset, ADQSADAS.

Table 5.1.1.1 Analysis Dataset Metadata for the ADQSADAS Analysis Dataset²

Dataset Name	Dataset Description	Dataset Location	Dataset Structure	Key Variables of Dataset	Class of Dataset	Documentation
<i>filename of the dataset.</i>	<i>short summary of contents of the dataset</i>	<i>where the dataset can be found</i>	<i>the level of detail in the dataset</i>	<i>variable names that parallel the structure</i>	<i>general class of the dataset using controlled terminology</i>	<i>links or references to documentation re how the dataset was created</i>
ADQSADAS	Data for the ADAS-Cog (11) Analyses	adqsadas.xpt	one record per subject per parameter per analysis visit	USUBJID, PARAMCD, AVISIT	BDS	DSADQSADAS.SAS, Section 14.11 of SAP for detailed ADAS-Cog scoring algorithm

² The display presentation of the metadata should be determined between the sponsor and the sender. The example is only intended to illustrate content and not appearance.

5.2 Analysis Variable Metadata

The analysis variable metadata describe each variable in the analysis dataset, including the variable attributes and definition. The metadata fields used to provide these descriptions are listed in [Table 5.2.1](#). Best practices strongly recommend that every analysis variable be described using these metadata fields. ADaM-compliant analysis datasets must be described by these analysis variable metadata fields.

Table 5.2.1 Analysis Variable Metadata Fields

Analysis Variable Metadata Field	Description
DATASET NAME	The file name of the analysis dataset
VARIABLE NAME	The name of the variable
VARIABLE LABEL	A brief description of the variable
VARIABLE TYPE	The variable type. Valid values are as defined in the Case Report Tabulation Data Definition Specification Standard (e.g., in version 1.0.0 they include “text,” “integer,” and “float”)
DISPLAY FORMAT	The variable display information (i.e., the format used for the variable in a tabular or graphical presentation of results). It is suggested that the syntax be consistent with the format terminology incorporated in the software package used for analysis (e.g., \$16 or 3.1 if using SAS).
CODELIST / CONTROLLED TERMS	A list of valid values or allowable codes and their corresponding decodes for the variable. The field can include a reference to an external codelist (identified by name and version) or a hyperlink to a list of the values in the codelist/controlled terms section of the define file.
SOURCE / DERIVATION	Provides details about the variable’s lineage – what was the predecessor, where the variable came from in the source data (SDTM or other analysis dataset) or how the variable was derived. This field is used to identify the immediate predecessor source and/or a brief description of the algorithm or process applied to that source and can contain hyperlinked text that refers readers to additional information. The source / derivation can be as simple as a two level name (e.g., ADSL.AGEGR) identifying the data file and variable that is the source of the variable (i.e., a variable copied with no change). It can be a simple description of a derivation and the variable used in the derivation (e.g., “categorization of ADSL.BMI”). It can also be a complex algorithm, where the element contains a complete description of the derivation algorithm and/or a link to a document containing it and/or a link to the analysis dataset creation program.

Refer to [Section 5.2.2](#) for an example of analysis variable metadata.

5.2.1 Analysis Parameter Value-Level Metadata

An analysis dataset that follows the ADaM Basic Data Structure (BDS), i.e., an analysis dataset of the BDS class, can contain multiple analysis parameters. In a BDS analysis dataset, the variable PARAM contains a unique description for every analysis parameter included in that dataset. The variable PARAMCD contains the short name of the analysis parameter in PARAM, with a one-to-one mapping between the two variables. Each value of PARAM identifies a set of one or more rows in the dataset.

The metadata for the columns (variables) in the dataset often depend on the values of PARAM/PARAMCD. This concept is analogous to that of value-level metadata for a single variable in SDTM, but in the BDS it is quite common that the metadata of several variables vary by PARAM/PARAMCD. To describe how variable metadata vary by PARAM/PARAMCD, the metadata element PARAMETER IDENTIFIER is required in variable-level metadata for a BDS analysis dataset. This PARAMETER IDENTIFIER metadata element identifies which variables

have metadata that vary depending on PARAM/PARAMCD, and links the metadata for a variable to the appropriate value of PARAM/PARAMCD.

Controlled terminology reduces the need to enter the same metadata for a variable for multiple values of PARAM/PARAMCD:

- The use of “*ALL*” in the PARAMETER IDENTIFIER for a variable indicates that the metadata for that variable is the same for all values of PARAM/PARAMCD in the analysis dataset.
- The use of “*DEFAULT*” in the PARAMETER IDENTIFIER for a variable indicates that the specified metadata for that variable should be considered the metadata for all values of PARAM/PARAMCD in the analysis dataset unless otherwise specified.
- A particular value of PARAMCD in the PARAMETER IDENTIFIER for a variable indicates that the specified metadata for that variable should be considered the metadata applicable to the particular PARAMCD, overriding the specified *DEFAULT* metadata, if any.

Refer to Section 5.2.2 for an example of analysis variable metadata for a dataset that uses the BDS, including the use of the parameter identifier metadata element. It should be noted that this metadata element facilitates the entry and tracking of the metadata content; how the metadata are displayed in the define file will be determined by the sponsor.

Analysis Variable Metadata Field	Description
PARAMETER IDENTIFIER	Contains either 1) the value of PARAMCD that identifies the analysis parameter to which the variable metadata applies; or 2) controlled terminology to indicate groupings of analysis parameters: *ALL* - Used when the variable metadata applies to all analysis parameters in the dataset. *DEFAULT* - Used when the variable metadata applies to all analysis parameters in the dataset except for those specifically listed within the metadata.

By referencing the codelist for PARAMCD, the user of the dataset can determine the unique analysis parameter values found in the dataset and is able to determine the analysis parameter-specific attributes and derivation algorithms for each variable when PARAMCD is a specific value.

Note that for the PARAMCD variable, the parameter identifier is “PARAMCD.” The list of values that exist for the variable also serves as an index of the analysis parameters and parameter identifiers included in the analysis dataset.

5.2.2 Illustration of Analysis Variable Metadata, Including Analysis Parameter Value-Level Metadata

In this illustration, which expands upon the ADAS-Cog analysis example, the data being analyzed are the ADAS-Cog(11) total scores, an 11-item subscale of the ADAS-Cog. The assessment is a questionnaire, so the QS SDTM domain contains the collected data. [Table 5.2.2.1](#) illustrates the analysis variable metadata for the ADAS-Cog analysis dataset, ADQSADAS, described in [Section 5.1.1](#).

As with any BDS analysis dataset, analysis parameter value-level metadata are used. The dataset contains both the individual item scores as well as the total scores, so there is a need to have different metadata for certain variables, depending on the value of PARAM. In this example, last observation carried forward (LOCF) is used to impute missing values of the total score; no imputation is performed for the individual item scores. Note that all values of PARAMCD must be listed under the codelist element.

Table 5.2.2.1 Analysis Variable Metadata for the ADQSADAS Dataset³

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
<i>file name of the analysis dataset</i>	<i>PARAMCD or *ALL* or *DEFAULT*</i>	<i>name</i>	<i>description</i>	<i>type</i>	<i>display information</i>	<i>valid values or codes and decodes</i>	<i>where the variable came from in the source data or how the variable was derived</i>
ADQSADAS	*ALL*	STUDYID	Study Identifier	text	\$12		ADSL.STUDYID
ADQSADAS	*ALL*	SITEID	Study Site Identifier	text	\$3		ADSL.SITEID
ADQSADAS	*ALL*	SITEGR1	Pooled Site Group 1	text	\$3		ADSL.SITEGR1
ADQSADAS	*ALL*	USUBJID	Unique Subject Identifier	text	\$11		ADSL.USUBJID
ADQSADAS	*ALL*	AVISIT	Analysis Visit	text	\$19	Baseline, Week 8, Week 16, Week 24	If ADQSADAS.ITTRFL='Y' then AVISIT is the name of the analysis visit; if ADQSADAS.ITTRFL=blank then AVISIT=blank. Refer to Section 8.2 of the SAP for a detailed description of the windowing algorithm used to determine the analysis visit based on ADQSADAS.ADY
ADQSADAS	*ALL*	VISIT	Visit Name	text	\$19		QS.VISIT

³ The display presentation of the metadata should be determined between the sponsor and the recipient. The example is only intended to illustrate content and not appearance.

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADQSADAS	*ALL*	AVISITN	Analysis Visit (N)	integer	3.0	3=Baseline, 8=Week 8, 10=Week 16, 12=Week 24	if ADQSADAS.ITTRFL='Y' AVISITN=numeric code for AVISIT, blank if ADQSADAS.ITTRFL=blank
ADQSADAS	*ALL*	ADY	Analysis Relative Day	integer	3.0		if ADQSADAS.ADT >= ADSL.TRTSDT then ADY=ADQSADAS.ADT - ADSL.TRTSDT + 1; if ADQSADAS.ADT < ADSL.TRTSDT then ADY=ADQSADAS.ADT - ADSL.TRTSDT
ADQSADAS	*DEFAULT*	PARAM	Parameter	text	\$16	ADAS-Cog Item 01, ADAS-Cog Item 02, ADAS-Cog Item 03, ADAS-Cog Item 04, ADAS-Cog Item 05, ADAS-Cog Item 06, ADAS-Cog Item 07, ADAS-Cog Item 08, ADAS-Cog Item 09, ADAS-Cog Item 10, ADAS-Cog Item 11, ADAS-Cog Item 12, ADAS-Cog Item 13, ADAS-Cog Item 14	When ADQSADAS.PARAMCD indicates an item score (rather than a total score), PARAM is the corresponding value (for subject and visit) of QS.QSTEST when QS.QSTESTCD = ADQSADAS.PARAMCD
ADQSADAS	ACTOT11	PARAM	Parameter	text	\$16	ADAS-Cog11 Total Score	'ADAS-Cog11 Total Score' is assigned to the total score records
ADQSADAS	PARAMCD	PARAMCD	Parameter Code	text	\$8	ACITM01, ACITM02, ACITM03, ACITM04, ACITM05, ACITM06, ACITM07, ACITM08, ACITM09, ACITM10, ACITM11, ACITM12, ACITM13, ACITM14, ACTOT11	Corresponds to PARAM
ADQSADAS	*DEFAULT*	AVAL	Analysis Value	float	3.0		When ADQSADAS.PARAMCD indicates an item score (rather than a total score), AVAL is the corresponding value (for subject and visit) of QS.QSSTRESN when QS.QSTESTCD = ADQSADAS.PARAMCD

CDISC Analysis Data Model Version 2.1

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADQSADAS	ACTOT11	AVAL	Analysis Value	float	3.0		Sum of ADAS scores for items 1, 2, 4, 5, 6, 7, 8, 11, 12, 13, and 14, see SAP section 14.2 for details on adjusting for missing values
ADQSADAS	*ALL*	BASE	Baseline Value	float	3.0		ADQSADAS.AVAL when ADQSADAS.ABLFL='Y'
ADQSADAS	*ALL*	CHG	Change from Baseline	float	3.0		ADQSADAS.AVAL - ADQSADAS.BASE
ADQSADAS	*ALL*	ABLFL	Baseline Record Flag	text	\$1	Y	Y if record contains the baseline value, i.e., if AVISITN=3; blank otherwise
ADQSADAS	*ALL*	TRTP	Planned Treatment	text	\$20	Placebo, Xanomeline Low Dose, Xanomeline High Dose	ADSL.TRT01P
ADQSADAS	*ALL*	TRTPN	Planned Treatment (N)	integer	1.0	0=Placebo, 1=Xanomeline Low Dose, 2=Xanomeline High Dose	ADSL.TRT01PN
ADQSADAS	*ALL*	TRTDOSE	Randomized Daily Dose Strength, mg	integer	2.0	0=placebo, 54=Xanomeline Low Dose, 81=Xanomeline High Dose	ADSL.TRTDOSE
ADQSADAS	*ALL*	AGE	Age	integer	3.0		ADSL.AGE
ADQSADAS	*ALL*	AGEGR1	Pooled Age Group 1	text	\$5	<65, 65-80, >80	Based on ADSL.AGEGR1, blank if ADSL.AGE is missing
ADQSADAS	*ALL*	AGEGR1N	Pooled Age Group 1 (N)	integer	1.0	1= <65, 2= 65-80, 3= >80	Based on ADSL.AGEGR1N, blank if ADSL.AGE is missing
ADQSADAS	*ALL*	SEX	Sex	text	\$1	M, F	ADSL.SEX
ADQSADAS	*ALL*	SAFFL	Safety Population Flag	text	\$1	Y., N	ADSL.SAFFL
ADQSADAS	*ALL*	ITTFLL	Intent-to-Treat Population Flag	text	\$1	Y, N	ADSL.ITTFLL

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADQSADAS	*ALL*	ITTRFL	Intent-to-Treat Record-Level Flag	text	\$1	Y	If the observed data are eligible for analysis (i.e., QS.VISITNUM in 3,8,10,12,201) and if QS.VISIT = the name of the visit window containing ADQSADAS.ADY and if ADQSADAS.ITTRFL='Y' then ITTRFL='Y'; ITTRFL blank otherwise
ADQSADAS	*DEFAULT*	DTYPE	Derivation Type	text	\$4		Not applicable, therefore blank
ADQSADAS	ACTOT11	DTYPE	Derivation Type	text	\$4	LOCF	DTYPE = 'LOCF' when the value of ADQSADAS.AVAL (and thus the entire record) has been imputed using the LOCF algorithm, blank otherwise.
ADQSADAS	*ALL*	ONTRTFL	On Treatment Record Flag	text	\$1	Y	If ADQSADAS.TRSDT <= ADQSADAS.ADT <= ADQSADAS.TRTEDT then ONTRTFL='Y'. ONTRTFL blank otherwise
ADQSADAS	*ALL*	TRSDT	Date of First Exposure to Treatment	integer	yymmdd10.		ADSL.TRSDT
ADQSADAS	*ALL*	TRTEDT	Date of Last Exposure to Treatment	integer	yymmdd10.		ADSL.TRTEDT
ADQSADAS	*ALL*	VISITDY	Planned Study Day of Visit	integer	3.0		QS.VISITDY
ADQSADAS	*ALL*	VISITNUM	Visit Number	float	4.1		QS.VISITNUM
ADQSADAS	*ALL*	ADT	Analysis Date	integer	yymmdd10.		QS.QSDTC associated with AVAL, converted to SAS date

5.3 Analysis Results Metadata

These metadata provide traceability from a result used in a statistical display to the data in the analysis datasets. Analysis results metadata are not required. However, best practice is that they be provided to assist the reviewer by identifying the critical analyses, providing links between results, documentation, and datasets, and documenting the analyses performed.

Analysis results include statistical displays (e.g., text, tabular or graphical presentation of results) or inferential statements such as p-values or estimates of treatment effect. Analysis results metadata provide a link between analysis results and the data used to generate it in a standard format and a predictable location. This allows reviewers to link from an analysis result to important information describing the analysis such as the reason for performing the analysis, and the dataset and selection criteria used to generate the analysis.

Analysis results metadata are not needed or even advisable for every analysis included in a clinical study report or submission. The sponsor determines which analyses should have analysis results metadata. For example, the sponsor might elect to provide analysis results metadata only for the primary efficacy analysis and the secondary efficacy analyses being considered for a marketing claim.

Analysis results metadata describe the major attributes of a specified analysis result found in a clinical study report or submission. The metadata fields to be used to describe an analysis result are listed in [Table 5.3.1](#). The word “Display” is used instead of “Table” as it is more generic, referring to tabular or graphical presentation of results.

Table 5.3.1 Analysis Results Metadata Fields

Analysis Results Metadata Field	Description
DISPLAY IDENTIFIER	A unique identifier for the specific analysis display (such as a table or figure number)
DISPLAY NAME	Title of display, including additional information if needed to describe and identify the display (e.g., analysis population)
RESULT IDENTIFIER	Identifies the specific analysis result within a display. For example, if there are multiple p-values on a display and the analysis results metadata specifically refers to one of them, this field identifies the p-value of interest. When combined with the display identifier provides a unique identification of a specific analysis result.
PARAM	The analysis parameter in the BDS analysis dataset that is the focus of the analysis result. Does not apply if the result is not based on a BDS analysis dataset.
PARAMCD	Corresponds to PARAM in the BDS analysis dataset. Does not apply if the result is not based on a BDS analysis dataset.
ANALYSIS VARIABLE	The analysis variable being analyzed
REASON	The rationale for performing this analysis. It indicates when the analysis was planned (e.g., “Pre-specified in Protocol,” “Pre-specified in SAP,” “Data Driven,” “Requested by Regulatory Agency”) and the purpose of the analysis within the body of evidence (e.g., “Primary Efficacy,” “Key Secondary Efficacy,” “Safety”). The terminology used is sponsor defined. An example of a reason is “Primary Efficacy Analysis as Pre-specified in Protocol.”
DATASET	The name of the dataset used to generate the analysis result. In most cases, this is a single dataset. However, if multiple datasets are used, they are all listed here.
SELECTION CRITERIA	Specific and sufficient selection criteria for analysis subset and / or numerator – a complete list of the variables and their values used to identify the records selected for the analysis. Though the syntax is not ADaM-specified, the expectation is that the information could easily be included in a WHERE clause or something equivalent to ensure selecting the exact set of records appropriate for an analysis. This information is required if the analysis does not include every record in the analysis dataset.
DOCUMENTATION	Textual description of the analysis performed. This information could be a text description, pseudo code, or a link to another document such as the protocol or statistical analysis plan, or a link to an analysis generation program (i.e., a statistical software program used to generate the analysis result). The contents of the documentation metadata element contains depends on the level of detail required to describe the analysis itself, whether or not the sponsor is providing a corresponding analysis generation program, and sponsor-specific requirements and standards. This documentation metadata element will remain free form, meaning it will not become subject to a rigid structure or controlled terminology.
PROGRAMMING STATEMENTS	The software programming code used to perform the specific analysis. This includes, for example, the model statement (using the specific variable names) and all technical specifications needed for reproducing the analysis (e.g., covariance structure). The name and version of the applicable software package should be specified either as part of this metadata element or in another document, such as a Reviewer’s Guide (see Appendix B for more information about a Reviewer’s Guide).

5.3.1 Illustration of Analysis Results Metadata

Figure 5.3.1.1 and Figure 5.3.1.2 contain data displays illustrating some of the analyses performed using the ADAS-Cog analysis dataset described in Sections 5.1.1 and 5.2.2. As described in the Statistical Analysis Plan, the primary analysis of the ADAS-Cog(11) total score at Week 24 used the efficacy population with LOCF imputation for any missing values at Week 24. An analysis of covariance (ANCOVA) model was used, with baseline score, site group, and treatment (as a continuous variable) included as independent variables, and results for a test of dose response presented. Pairwise treatment comparisons were performed using an ANCOVA model with baseline score, site group, and treatment (as a categorical variable) included as independent variables, and results for the treatment differences presented. In addition, a supportive analysis for the ADAS-Cog was performed, using mixed effects models repeated measures (MMRM) analysis. In this example, the efficacy population is the Intent-to-Treat population.

Figure 5.3.1.1 Example of Primary Endpoint Analysis Statistical Display⁴

Protocol: CDISCPILOT01
Population: Efficacy

Page 1 of 1

Table 14-3.01
Primary Endpoint Analysis: ADAS Cog (11) - Change from Baseline to Week 24 - LOCF

	Placebo (N=79)	Xanomeline Low Dose (N=81)	Xanomeline High Dose (N=74)
Baseline			
n	79	81	74
Mean (SD)	24.1 (12.19)	24.4 (12.92)	21.3 (11.74)
Median (Range)	21.0 (5;61)	21.0 (5;57)	18.0 (3;57)
Week 24			
n	79	81	74
Mean (SD)	26.7 (13.79)	26.4 (13.18)	22.8 (12.48)
Median (Range)	24.0 (5;62)	25.0 (6;62)	20.0 (3;62)
Change from Baseline			
n	79	81	74
Mean (SD)	2.5 (5.80)	2.0 (5.55)	1.5 (4.26)
Median (Range)	2.0 (-11;16)	2.0 (-11;17)	1.0 (-7;13)
p-value(Dose Response) [1] [2]			0.245
p-value(Xan - Placebo) [1] [3]		0.569	0.233
Diff of LS Means (SE)		-0.5 (0.82)	-1.0 (0.84)
95% CI		(-2.1;1.1)	(-2.7;0.7)
p-value(Xan High - Xan Low) [1] [3]			0.520
Diff of LS Means (SE)			-0.5 (0.84)
95% CI			(-2.2;1.1)

[1] Based on Analysis of covariance (ANCOVA) model with treatment and site group as factors and baseline value as a covariate.
[2] Test for a non-zero coefficient for treatment (dose) as a continuous variable.
[3] Pairwise comparison with treatment as a categorical variable: p-values without adjustment for multiple comparisons.

Source: C:\cdisc_pilot\PROGRAMS\DRAFT\TFLs\rtf_eff1.rtf 21:05 Monday, June 26, 2006

Table 5.3.1.1 and Table 5.3.1.2 illustrate the analysis results metadata for specific elements of the ADAS-Cog analyses shown in Figure 5.3.1.1. The items underlined in the illustration would ideally be hyperlinks to the data display in the clinical study report, to metadata elsewhere in the define file, and to specific pages of the SAP.

⁴ The style of the display of the results of an analysis will be determined by the sponsor. The example is intended to illustrate content not appearance.

Table 5.3.1.1 illustrates the analysis results metadata for the analysis of dose response, identified with (1) in Figure 5.3.1.1. It also illustrates the use of a description of the analysis done, with no model statement provided in Programming Statements.

Table 5.3.1.1 Analysis Results Metadata for the Dose Response Analysis in the Statistical Display in Figure 5.3.1.1⁵

Metadata Field	Definition of field	Metadata
DISPLAY IDENTIFIER	Unique identifier for the specific analysis display	Table 14-3.01
DISPLAY NAME	Title of display	Primary Endpoint Analysis: ADAS Cog (11) - Change from Baseline to Week 24 – LOCF
RESULT IDENTIFIER	Identifies the specific analysis result within a display	Analysis of dose response
PARAM	Parameter	ADAS-Cog (11) Total Score
PARAMCD	Parameter code	ACTOT11
ANALYSIS VARIABLE	Analysis variable being analyzed	CHG
REASON	Rationale for performing this analysis	Primary efficacy analysis as pre-specified in protocol
DATASET	Dataset(s) used in the analysis.	ADQSADAS
SELECTION CRITERIA	Specific and sufficient selection criteria for analysis subset and / or numerator	ITTFL='Y' and AVISIT='Week 24' and PARAMCD='ACTOT11'
DOCUMENTATION	Textual description of the analysis performed	SAP Section 10.1.1. Linear model analysis of dose response for the ADAS-Cog(11) total score change from baseline at Week 24 - missing values imputed using LOCF, Efficacy population. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose); Independent terms in model are TRTDOSE (randomized dose: 0 for placebo; 54 for low dose; 81 for high dose) SITEGR1 (site group, as a class variable) and BASE (baseline ADAS-Cog score).
PROGRAMMING STATEMENTS	The analysis syntax used to perform the analysis.	

⁵ The display presentation of the metadata should be determined between the sponsor and the recipient. The example is only intended to illustrate content and not appearance.

Table 5.3.1.2 illustrates the analysis results metadata for the pairwise treatment comparisons, identified with (2) in Figure 5.3.1.1 . It also illustrates the inclusion of a model statement.

Table 5.3.1.2 Analysis Results Metadata for the Pairwise Treatment Comparisons in the Statistical Display in Figure 5.3.1.1⁶

Metadata Field	Definition of field	Metadata
DISPLAY IDENTIFIER	Unique identifier for the specific analysis display	Table 14-3.01
DISPLAY NAME	Title of display	Primary Endpoint Analysis: ADAS Cog (11) - Change from Baseline to Week 24 - LOCF
RESULT IDENTIFIER	Identifies the specific analysis result within a display	Pairwise treatment comparisons
PARAM	Analysis parameter	ADAS-Cog (11) Total Score
PARAMCD	Analysis parameter code	ACTOT11
ANALYSIS VARIABLE	Analysis variable being analyzed	CHG
REASON	Rationale for performing this analysis	Primary efficacy analysis as pre-specified in protocol
DATASET	Dataset(s) used in the analysis.	ADQSADAS
SELECTION CRITERIA	Specific and sufficient selection criteria for analysis subset and / or numerator	ITTFL='Y' and AVISIT='Week 24' and PARAMCD='ACTOT11'
DOCUMENTATION	Textual description of the analysis performed	Linear model analysis of ADAS-Cog(11) total score change from baseline at Week 24 for pairwise treatment comparisons and adjusted means; missing values imputed using LOCF, Efficacy population. Used randomized treatment as class variable; site group as class variable; and baseline ADAS-Cog score in model.
PROGRAMMING STATEMENTS	The analysis syntax used to perform the analysis	PROC GLM; CLASS SITEGR1 TRTP; MODEL CHG = TRTP SITEGR1 BASE; ESTIMATE 'H VS L' TRTP 0 1 -1; ESTIMATE 'H VS P' TRTP -1 1 0; ESTIMATE 'L VS P' TRTP -1 0 1; LSMEANS TRTP / OM STDERR PDIFF CL; RUN;

⁶ The display presentation of the metadata should be determined between the sponsor and the recipient. The example is only intended to illustrate content and not appearance.

Figure 5.3.1.2 Example of Supportive Analysis Statistical Display⁷

Protocol: CDISCPILOT01
 Population: Efficacy

Page 1 of 1

Table 14-3.11
 ADAS-Cog (11) - Repeated Measures Analysis of Change from Baseline to Week 24

	Placebo (N=79)	Exoneline Low Dose (N=81)	Exoneline High Dose (N=74)
LS Means (SE)	1.6 (0.49)	1.5 (0.52)	1.1 (0.56)
p-value(Xan - Placebo)		0.955	0.556
Diff of LS Means (SE)		-0.0 (0.70)	-0.4 (0.72)
95% CI		(-1.4;1.3)	(-1.9;1.0)
p-value(Xan High - Xan Low)			0.606
Diff of LS Means (SE)			-0.4 (0.75)
95% CI			(-1.9;1.1)

Note: The change from baseline is calculated as the post-baseline score minus the baseline score. The covariates included in the MMRM model are treatment, site group, time and treatment by time interaction, baseline ADAS-Cog (11) score, and baseline ADAS-Cog (11) score by time interaction.
 Source: C:\cdisc_pilot\PROGRAMS\DRAFT\TFLs\rtf_eff_mrm.sas 21:06 Monday, June 26, 2006

⁷ The style of the display of the results of an analysis will be determined by the sponsor. The example is intended to illustrate content not appearance.

Table 5.3.1.3 illustrates the analysis results metadata for the statistical display shown in Figure 5.3.1.2, illustrating the use of metadata to describe a single display. The items underlined in the illustration would ideally be hyperlinks to the data display in the clinical study report, to metadata elsewhere in the define file, and to specific pages of the SAP.

Table 5.3.1.3 Analysis Results Metadata for the Statistical Display in Figure 5.3.1.2⁸

Metadata Field	Definition of field	Metadata
DISPLAY IDENTIFIER	<i>Unique identifier for the specific analysis display</i>	<u>Table 14-3.11</u>
DISPLAY NAME	<i>Title of display</i>	ADAS Cog (11) - Repeated Measures Analysis of Change from Baseline to Week 24
RESULT IDENTIFIER	<i>Identifies the specific analysis result within a display</i>	
PARAM	<i>Analysis parameter</i>	ADAS-Cog (11) Total Score
PARAMCD	<i>Analysis parameter code</i>	ACTOT11
ANALYSIS VARIABLE	<i>Analysis variable being analyzed</i>	CHG
REASON	<i>Rationale for performing this analysis</i>	Pre-specified in SAP
DATASET	<i>Dataset(s) used in the analysis.</i>	<u>ADQSADAS</u>
SELECTION CRITERIA	<i>Specific and sufficient selection criteria for analysis subset and / or numerator</i>	ITTFL='Y' and AVISITN GT 0 AND DTYPE NE 'LOCF' AND PARAMCD='ACTOT11'
DOCUMENTATION	<i>Textual description of the analysis performed</i>	SAP Section 10.1.1. Adjusted means for the change from baseline at week 24 and pairwise comparisons between treatment groups at Week 24 using a repeated measures model with treatment group (as class variable); site (as class variable); time; treatment*time interaction; baseline score and baseline*time interaction terms; and an unstructured covariance matrix. Efficacy data, observed cases data.
PROGRAMMING STATEMENTS	<i>The analysis syntax used to perform the analysis</i>	PROC MIXED; CLASS USUBJID SITEGR1 AVISITN TRTP; MODEL CHG = TRTP SITEGR1 AVISITN TRTP*AVISITN BASE BASE*AVISITN / OUTP=PRED DDFM=KR; REPEATED AVISITN / SUBJECT=USUBJID TYPE=UN; LSMEANS TRTP / DIFF CL; RUN;

⁸ The display presentation of the metadata should be determined between the sponsor and the recipient. The example is only intended to illustrate content and not appearance.

6 Subject-Level Analysis Dataset

The structure of the Subject-Level Analysis Dataset (ADSL) is one record per subject, regardless of the type of clinical trial design. ADSL is used to provide the variables that describe attributes of a subject. This structure allows simple merging with any other dataset, including SDTM and analysis datasets.

Regulatory agency staff have stated that ADSL is very helpful in the review of a clinical trial. ADSL and its related metadata are required in any CDISC based submission of data from a clinical trial even if no other analysis datasets are submitted.

ADSL is intended to provide descriptive information about subjects. It can be used in multiple types of analyses, including descriptive, categorical, and modeling. ADSL should not be forced to support all analyses in an attempt to minimize the number of analysis datasets. Although it would be technically feasible to take every single data value in a study and include them all as variables in a subject-level dataset such as ADSL, that is not the intent or the purpose of ADSL. The correct location for key endpoints and data that vary over time during the course of a study is in a BDS dataset.

ADSL is the primary source for subject-level variables included in other analysis datasets, such as population flags and treatment variables. When merging data from ADSL into other analysis datasets, only those fields relevant to these analysis datasets should be included. The inclusion of too many extraneous variables (i.e., variables not needed to support analyses) makes it more difficult for users to find important variables and can impede clear and concise communication.

Table 6.1 provides an example of analysis dataset metadata for ADSL.

Table 6.1 Example of Analysis Dataset Metadata for ADSL⁹

Dataset Name	Dataset Description	Dataset Location	Dataset Structure	Key Variables of Dataset	Class of Dataset	Documentation
ADSL	Subject disposition, demographic, and baseline characteristics	adsl.xpt	One record per subject	USUBJID	ADSL	SAP, DSADSL.SAS

The minimum set of variables to include in ADSL depends on the specific nature of the disease and on the protocol, (refer to ICH E3 [8] for a more detailed listing and to the ADaMIG for further description including required variables). Examples of ADSL information include (but are not limited to):

- Demographic variables (e.g., age, sex, race, other relevant factors)
- Disease factors (e.g., disease onset, disease severity)
- Treatment code/group
- Other possible prognostic factors that might affect response to therapy (e.g., smoking, alcohol intake, menstrual status for women)
- Important event dates (e.g., treatment start and stop dates)
- Study population

⁹ The display presentation of the metadata should be determined between the sponsor and the recipient. The example is only intended to illustrate content and not appearance.

ADSL contains variables that describe the subjects in a clinical trial prior to treatment, or group the subjects in some way for analysis purposes.

In summary, the variables in ADSL include those that are either descriptive, considered an important baseline characteristic, used as strata for randomization, used to identify the subject as belonging to specific subgroups (e.g., population flags) or used to identify when or if important events occurred (e.g., last dose date, death, discontinuation). For example, in a stratified randomization done within age group, a subject's age category is an important subject descriptor variable for the study and is included in ADSL.

ICH Guidance (ICH E3, Section 11.2) [8] recommends that “in addition to tables and graphs giving group data for baseline variables, relevant individual subject demographic and baseline data... for all individual subjects randomized (broken down by treatment and by center or multi-center studies) should be presented in by-subject tabular listings.” Often an FDA reviewer and sponsor agree that submission of subject-level data meets this requirement. If that is the case, ADSL should include those variables needed to meet this regulatory guidance.

6.1 Data for Subjects Not Analyzed

Whether analysis datasets include data for subjects not analyzed (e.g., screen failures) is a sponsor decision and should be communicated with the reviewers or users of the data. If these data are included, they should be incorporated in the appropriate analysis datasets such as ADSL (as opposed to separate datasets for non-analyzed subjects) using appropriate flag variables to clearly differentiate these records. The metadata must specify that these data are included and how to distinguish them.

Appendices

Appendix A References

- [1] CDISC Analysis Data Model (ADaM) Team, 2009, “ADaM Implementation Guide,” available on CDISC website at <<http://www.cdisc.org/standards>>
- [2] CDISC Define.xml Team, 2005, “Case Report Tabulation Data Definition Specification (define.xml),” available on CDISC website at <<http://www.cdisc.org/standards>>
- [3] CDISC SDS Metadata Team, 2007, “Metadata Submission Guidelines, Appendix to the Study Data Tabulation Model Implementation Guide 3.1.1,” available on CDISC website at <<http://www.cdisc.org/standards>>
- [4] CDISC SDS Team, 2008, “Study Data Tabulation Model (SDTM) Implementation Guide Final Version 3.1.2,” available on CDISC website at <<http://www.cdisc.org/standards>>
- [5] CDISC Submission Data Standards (SDS) Team, 2008, “Study Data Tabulation Model (SDTM) Final Version 1.2,” available on CDISC website at <<http://www.cdisc.org/standards>>
- [6] FDA, 2008, “Guidance for Industry: Providing Regulatory Submissions in Electronic Format - Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications,” available on FDA website at <<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm064994.htm>>
- [7] FDA Center for Drug Evaluation and Research (CDER), 2009, “Study Data Specifications, Version 1.5,” available on FDA website at <<http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm>> >
- [8] ICH Expert Working Group, 1995, “ICH Harmonised Tripartite Guideline: Structure And Content of Clinical Study Reports - E3,” available at <<http://www.ich.org/LOB/media/MEDIA479.pdf>>

Appendix B Definitions

ADaM – CDISC Analysis Data Model

ADaM Basic Data Structure (BDS) – A dataset structure designed to facilitate ease of analysis and review, organized as one or more records per subject per analysis parameter per analysis timepoint. Analysis timepoint is conditionally required, depending on the analysis. The BDS, described in the ADaMIG, supports the majority of analyses.

ADaM Implementation Guide (ADaMIG) – A document that specifies ADaM standard dataset structures and variables, including naming conventions. It also specifies standard solutions to implementation issues. The ADaM document and the ADaMIG should be used together.

Analysis Datasets – Datasets used for statistical analysis and reporting.

Analysis Dataset Creation Program – Computer instructions used to create an analysis dataset.

Analysis Dataset Metadata – Information that describes the structure, content, and derivation of an analysis dataset.

Analysis Generation Programs – Computer instructions used to generate analysis results (e.g., summary or inferential statistics presented in tabular or graphical presentations).

Analysis Parameter (PARAM) – A row identifier used to uniquely characterize a group of values that share a common definition. Example: The primary efficacy analysis parameter is “3-Minute Sitting Systolic Blood Pressure (mmHg).” Note that the ADaM analysis parameter contains all of the information needed to uniquely identify a group of related analysis values. In contrast, the SDTM --TEST column may need to be combined with qualifier columns such as --POS, --LOC, --SPEC, etc., in order to identify a group of related values. In this document the word “parameter” is used as a synonym for “analysis parameter.”

Analysis Parameter Value-Level Metadata – Information that describes an analysis value within a given analysis parameter or set of analysis parameters.

Analysis Results Metadata – Information that describes a specified analysis result contained within a clinical study report or submission.

Analysis Timepoint – A row identifier used to classify values within an analysis parameter into temporal or conceptual groups used for analyses. These groupings may be observed, planned or derived. Example: The primary efficacy analysis was performed at the Week 2, Week 6, and Endpoint analysis timepoints.

Analysis Value – (1) The character (AVALC) or numeric (AVAL) value described by the analysis parameter. The analysis value may be present in the input data, a categorization of an input data value, or derived. Example: The analysis value of the parameter “Average Heart Rate (bpm)” was derived as the average of the three heart rate values measured at each visit. (2) In addition, values of certain functions are considered to be analysis values. Examples: baseline value (BASE), change from baseline (CHG).

Analysis Variable Metadata – Information that describes the variables within the analysis dataset.

Define File – As stated in the Case Report Tabulation Data Definition Specification[2], the 1999 FDA electronic submission (eSub) guidance and the electronic Common Technical Document (eCTD) documents specify that a document describing the content and structure of the included data should be provided within a submission. This document is known as the Data Definition Document (e.g., “define.pdf” in the 1999 guidance). The Data Definition Document provides a list of the datasets included in the submission along with a detailed description of the contents of each dataset (i.e., metadata). To increase the level of automation and improve the efficiency of the Regulatory Review process, define.xml can be used to provide the Data Definition Document in a machine-readable format. The formal name for this is the Case Report Tabulation Data Definition (CRT DD) specification. Both SDTM and ADaM datasets have their respective Define files.

Metadata – Information or data about data.

Record – A row in a dataset.

Reviewer's Guide – A document that can be included with a submission to orient reviewers to various aspects of the submission package. A Reviewer's Guide was included in the CDISC SDTM/ADaM Pilot Project submission package at the suggestion of the FDA reviewers participating in the project. (The report for the project is available at www.cdisc.org.) The document is useful for providing information that is either too complex or too lengthy to be described in other sources of metadata. It can describe issues that are difficult to communicate at the variable level or that apply to multiple analysis datasets, e.g., sponsor naming conventions, imputation rules for partial dates. It should not duplicate large amounts of information that can be found in other sources of metadata.

SDTM - Study Data Tabulation Model – A document written by the CDISC Submission Data Standards (SDS) team that describes the general conceptual model for representing clinical study data that are submitted to regulatory authorities. The SDTM provides a general framework for describing the organization of information collected for clinical trials and submitted to regulatory authorities [5].

SDTM Implementation Guide (SDTMIG) – A document written by the CDISC SDS team that is intended to guide the organization, structure, and format of standard clinical trial tabulation datasets submitted to a regulatory authority such as the FDA. It provides specific domain models, assumptions, business rules, and examples for preparing standard tabulation datasets that are based on the SDTM [4].

Traceability – The property in ADaM that permits the user of an analysis dataset to understand the data's lineage and/or the relationship between an element and its predecessor(s). Traceability facilitates transparency, which is an essential component in building confidence in a result or conclusion. Ultimately traceability in ADaM permits the understanding of the relationship between the analysis results, the analysis datasets, and the SDTM domains. Traceability is built by clearly establishing the path between an element and its immediate predecessor. The full path is traced by going from one element to its predecessors, then on to their predecessors, and so on, back to the SDTM domains, and ultimately to the data collection instrument. Note that the CDISC Clinical Data Acquisition Standards Harmonization (CDASH) standard is harmonized with SDTM and therefore assists in assuring end-to-end traceability. Example: Based on the metadata and the content of the analysis dataset, the reviewer can trace how the primary and secondary efficacy analysis values were derived from the SDTM data for each subject.

Variable – A column in a dataset.

Appendix C Abbreviations and Acronyms

The following is a list of abbreviations and acronyms used multiple times in this document. Not included here are explanations of the various SDTM domains (e.g., QS, DM). Also not included is a description of the variables referenced.

Table C.1 Abbreviations and Acronyms Used in the Document

Term	Definition
ADAE	ADaM Adverse Event Analysis Dataset
ADaM	CDISC Analysis Data Model
ADaMIG	Analysis Data Model Implementation Guide
ADAS-Cog	Alzheimer's Disease Assessment Scale - Cognitive Subscale
ADSL	ADaM Subject-Level Analysis Dataset
ANCOVA	Analysis of Covariance
BDS	ADaM Basic Data Structure
BMI	Body Mass Index
CDASH	Clinical Data Acquisition Standards Harmonization
CDISC	Clinical Data Interchange Standards Consortium
CFR	Code of Federal Regulations
CRT	Case Report Tabulation
CRT-DDS	Case Report Tabulation Data Definition Specification
eCTD	electronic Common Technical Document
FDA	United States Food and Drug Administration
ICH	International Conference on Harmonisation
ITT	Intent-to-Treat
LOCF	Last Observation Carried Forward
MMRM	Mixed Effects Models Repeated Measures
PDF	Portable Document Format
SAP	Statistical Analysis Plan
SDS	Submission Data Standards
SDTM	Study Data Tabulation Model
SDTMIG	Study Data Tabulation Model Implementation Guide
XML	Extensible Markup Language
XPT	Filename extension for a SAS transport file

Appendix D Illustration of Analysis-Ready

To illustrate the concept of “analysis-ready,” consider again the example shown in Section 5. In [Figure 5.3.1.1](#), both the dose response analysis and the pairwise dose comparison are included on the display. Analysis-ready does not mean that this formatted table can be generated in a single statistical procedure. Rather it means that each statistic in the table can be replicated by running a standard statistical procedure (e.g., SAS PROC, S-PLUS function, etc.) using the appropriate analysis dataset as input. This means that reviewers can replicate and explore these results with minimal programming effort, allowing reviewers to concentrate on the results, not on programming.

For example, the following SAS code replicates the dose response analysis results of Table 14-3.01 (in [Figure 5.3.1.1](#)) using an analysis dataset containing the appropriate variables. Note that the where clause selects the appropriate records for the analysis.

```
*** DOSE RESPONSE ANALYSIS ***;
PROC GLM DATA=A.ADQSADAS (WHERE=(ITTFL='Y' AND AVISIT='Week 24' AND PARAMCD='ACTOT11')) ;
CLASS SITEGR1;
MODEL CHG = TRTDOSE SITEGR1 BASE;
RUN;
```

Similarly, the following SAS code replicates the pairwise dose comparison results.

```
*** PAIRWISE DOSE COMPARISON ANALYSIS ***;
PROC GLM DATA=A.ADQSADAS (WHERE=(ITTFL='Y' AND AVISIT='Week 24' AND PARAMCD='ACTOT11')) ;
CLASS SITEGR1 TRTP;
MODEL CHG = TRTP SITEGR1 BASE;
ESTIMATE 'H VS L' TRTP 0 1 -1;
ESTIMATE 'H VS P' TRTP -1 1 0 ;
ESTIMATE 'L VS P' TRTP -1 0 1;
LSMEANS TRTP / OM STDERR PDIFF CL;
RUN;
```

[Figure D.1](#) illustrates the relationship of the results of the above SAS code to the corresponding elements of the results display (Table 14-3.01 in [Figure 5.3.1.1](#)).

Result of Dose Response PROC GLM:

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRTDOSE	1	36.0384965	36.0384965	1.36	0.2447
SITEGR	10	556.3851128	55.6385113	2.10	0.0236
BASE	1	3.0136824	3.0136824	0.11	0.7362

Results of Pairwise Comparison PROC GLM:

Parameter	Estimate	Standard Error	t Value	Pr > t
H vs L	-0.53923124	0.83610890	-0.64	0.5196
H vs P	-1.00601360	0.84052936	-1.20	0.2326
L vs P	-0.46678236	0.81804222	-0.57	0.5688

	Baseline (N=79)	Baseline + Dose (N=81)
n	79	81
Mean (SD)	24.1 (12.19)	24.4 (13.92)
Median (Range)	21.0 (5;61)	21.0 (5;57)
Week 24		
n	79	81
Mean (SD)	26.7 (13.79)	26.4 (13.18)
Median (Range)	24.0 (5;62)	25.0 (6;62)
Change from Baseline		
n	79	81
Mean (SD)	2.5 (5.80)	2.0 (5.55)
Median (Range)	2.0 (-11;16)	2.0 (-11;17)
p-value(Dose Response) [1] [2]		
p-value(Xan - Placebo) [1] [3]		0.569
Diff of LS Means (SE)		-0.5 (0.82)
95% CI		(-2.1;1.1)
p-value(Xan High - Xan Low) [1] [3]		0.520
Diff of LS Means (SE)		-0.5 (0.84)
95% CI		(-2.2;1.1)

Least Squares Means for Effect TRTP			
i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)
1	2	1.006014	-0.650506 2.662534
1	3	0.466782	-1.145420 2.078985
2	3	-0.539231	-2.187039 1.108577

[1] Based on Analysis of covariance (ANCOVA) model with treatment and site group as factors and baseline value as a covariate.
 [2] Test for a non-zero coefficient for treatment (dose) as a continuous variable.
 [3] Pairwise comparison with treatment as a categorical variable; p-values without adjustment for multiple comparisons.
 Source: C:\cdisc_pilot\PROGRAMS\DRIFT\FFLs\rtf_eff1.sas 21:05 Monday, June 26, 2006

Figure D.1 Illustration of SAS Output vs Results Display from Analysis-Ready Dataset¹⁰

¹⁰ The style of the display of the results of an analysis will be determined by the sponsor. The example is intended to illustrate content not appearance.

Appendix E Composite Endpoint Example

As mentioned in Section 4.1.1, examples of analyses that would require only ADSL and SDTM do not cover the full range of analyses, especially when considering efficacy analyses. Even the full scope of safety analyses commonly presents more complex examples that cannot be covered by analysis data based solely on ADSL and SDTM. This example illustrates how an apparently simple binary outcome variable (outcome of the treatment of a single headache episode) has complex underpinnings and draws from data elements from different source datasets. It describes a composite endpoint that requires data from an efficacy dataset (headache severity at different time points), as well as from adverse experiences and concomitant medications datasets. The endpoint is “Sustained migraine pain and symptom free.”

The endpoint (sustained migraine pain and symptom free, based on the International Headache Society Guidelines) is defined as:

1. Headache severity of either Moderate or Severe at Baseline AND
2. Headache severity of No Pain by 2 hours post dose (i.e., after initial dose of test medication) AND
3. No headache recurrence within 48 hours post dose AND
4. No rescue medications for analgesia or anti-emetic from time of initial dose through 48 hours post dose AND
5. No associated symptoms (nausea, vomiting, photophobia, phonophobia) from two through 48 hours post dose.

The AND's in the above text indicate that all conditions must be met for a subject to be considered to have experienced the response.

For this example, the following definitions and specifications apply:

Headache severity

Headache severity is subjectively rated by subjects at pre-specified time points (baseline, 0.5, 1, 1.5, 2, 3, and 4 hours post dose) on a scale from zero (no pain) to 3 (severe pain).

Associated Symptoms

The subject records whether the following associated symptoms were present or absent at regular time points (baseline, 2, and 4 hours post dose): photophobia, phonophobia, nausea, vomiting.

In addition, subjects are instructed to list any of the above symptoms as an “Adverse Symptom” on the diary card if it: (1) shows an unusual increase in intensity after they have taken their test medication or, (2) otherwise shows an important change in character after they have taken their test medication, as compared with their usual migraine symptoms. The investigator is to record all such symptoms as adverse experiences. Therefore, a full assessment of the absence of associated symptoms will include a scan of the adverse event dataset.

Headache Recurrence

Headache recurrence is defined as the return of headache to a severity of two or three (moderate or severe) within 48 hours post dose in subjects who report pain relief (mild or no pain) at 2 hours post dose. Subjects record the maximum headache severity between 2 and 24 hours post-initial dose and between 24 and 48 hours post-initial dose.

Rescue Medications

The subject records any additional analgesics/anti-emetics taken after any test dose, documenting date, clock time (AM/PM), name of drug (e.g., codeine), the number of tablets/capsules, and the dose per tablet/capsule. Rescue medication is also defined as taking any additional doses of test medication within 48 hours post dose. The use of rescue medications is determined using the concomitant medication and exposure datasets.

To determine whether a subject meets the criteria for sustained migraine pain and symptom free, the answers to each of the five criteria must be determined. The headache severity and associated symptoms data (one or more SDTM domains), the AE domain, the CM domain, the EX domain, and ADSL all need to be input into the derivation for the endpoint.

In the BDS analysis dataset illustrated in [Table E.1](#) it is assumed that the answers to all of the questions inherent in the criteria are retained in the analysis dataset. Only a few of the analysis dataset variables (PARAMCD, AVAL, and AVALC) are listed in the illustration, since the purpose is to illustrate the complexity and not a full analysis dataset. In addition, rather than attempt to describe specific SDTM domains and variables for this example, a simple text description is provided for the source / derivation field. In “real” metadata, this metadata element should point to the specific domain and variable, and should include how to identify which record in the domain is the source of the data. (e.g., when QSCAT=xxx for this USUBJID).

This example illustrates that the source / derivation could be quite lengthy and complicated. For complex derived variables, the source / derivation field could provide a link to external documentation that explains the various sources of data and the algorithms involved in creating the variable.

Table E.1 Composite Endpoint Example - Illustration of Analysis Variable Metadata for Selected Variables Example¹¹

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADSYMFR	PARAMCD	PARAMCD	Parameter Code	text	\$8	HASPNFR HASEVBL HASEV2 HARECUR HARESCUE HASYMPD HASYMPAE	HASPNFR when ADSYMFR.PARAM= Sustained migraine pain and symptom free from 2-48 hours post-dose HASEVBL when ADSYMFR.PARAM = Headache severity at baseline HASEV2 when ADSYMFR.PARAM= Headache severity at 2 hours post-dose HARECUR when ADSYMFR.PARAM = Headache Recurrence within 48 hours post-dose HARESCUE when ADSYMFR.PARAM = Rescue medications taken from initial dose through 48 hours post-dose HASYMPD when ADSYMFR.PARAM = Associated symptoms as indicated on diary card from 2-48 hours post-dose HASYMPAE when ADSYMFR.PARAM = Associated symptoms as indicated in AE datasets from 2-48 hours post-dose
ADSYMFR	*DEFAULT*	AVAL	Analysis Value	integer	1.0	0=N 1=Y	Derived based on ADSYMFR.AVALC, null if ADSYMFR.AVALC missing
ADSYMFR	HASEVBL	AVAL	Analysis Value	integer	1.0	0=No pain 1=Mild pain 2=Moderate pain 3=Severe pain	Headache severity at baseline is from the diary card data: the recorded headache severity at the time of dosing. Null if missing.
ADSYMFR	HASEV2	AVAL	Analysis Value	integer	1.0	0=No pain 1=Mild pain 2=Moderate pain 3=Severe pain	Headache severity at 2 hours post-dose is from the diary card data: the recorded 2-hour post-dose headache severity . Null if missing.
ADSYMFR	*DEFAULT*	AVALC	Analysis Value (C)	text	\$1		Blank

¹¹ The display presentation of the metadata should be determined between the sponsor and the recipient. The example is only intended to illustrate content and not appearance.

CDISC Analysis Data Model Version 2.1

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADSYMFR	HASPNFR	AVALC	Analysis Value (C)	text	\$1	N=No Y=Yes	Sustained migraine pain and symptom free from 2-48 hours post-dose is based on other endpoints in this analysis dataset. It is Y if (ADSYMFR.HASEVBL=2 or 3) AND ADSYMFR.HASEV2=0 AND ADSYMFR.HARECUR=N AND ADSYMFR.HARESCUE=N AND (ADSYMFR.HASYMPD=N and ADSYMFR.HASYMPAE=N); N otherwise
ADSYMFR	HARECUR	AVALC	Analysis Value (C)	text	\$1	N=No headache recurrence Y=Headache did recur	Headache Recurrence within 48 hours post-dose is based on diary card data. It is N if the subject recorded their maximum headache severity as 0 (no pain) for the time periods between 2 and 24 hours post-initial dose and between 24 and 48 hours post-initial dose; Y if either maximum headache severity > 0. Blank if missing.
ADSYMFR	HARESCUE	AVALC	Analysis Value (C)	text	\$1	N=No rescue medication taken Y=Rescue medication taken	Rescue medications taken from initial dose through 48 hours post-dose is based on the CM domain, ADSL, and the EX domain. It is N if no analgesics or anti-emetics were taken from time of initial dose through 48 hours post-dose (CM domain), and if no additional doses of study medication were taken from time of initial dose through 48 hours post-dose (ADSL and EX domain); Y otherwise.
ADSYMFR	HASYMPD	AVALC	Analysis Value (C)	text	\$1	N=No associated symptoms present Y=Associated symptoms are present	Associated symptoms as indicated on diary card from 2-48 hours post-dose is based on the presence/absence of photophobia, phonophobia, nausea or vomiting at 2 and 4 hours post dose. It is N if no photophobia, phonophobia, nausea or vomiting at 2 or 4 hours post dose; Y if any were present at 2 or 4 hours post dose.
ADSYMFR	HASYMPAE	AVALC	Analysis Value (C)	text	\$1	N=No associated symptoms present Y=Associated symptoms are present	Associated symptoms as indicated in AE datasets from 2-48 hours post-dose is based on whether or not these symptoms are found in the AE domain. It is N if no photophobia, phonophobia, nausea or vomiting were noted as AEs from 2-48 hours post-dose; Y if any of these were reported as AE.

Appendix F Revision History

Version 2.1 represents the second formal release of the Analysis Data Model. The original version was released as the Analysis Data Model V2.0 in August 2006. Version 2.1 includes modifications so that the document corresponds to Version 1.0 of the Analysis Data Model Implementation Guide (ADaMIG).

Not all changes to the document can be listed here, as significant revising and reformatting was performed. The examples in the document have been substantially revised. The ADaM Basic Data Structure is introduced. The concept of traceability is now in the document. Significant portions of the document (e.g., analysis dataset variables) have been moved to the ADaMIG.

One significant change is the clarification that SDTM is expected to be the input source for ADaM.

The ADaM metadata has been clarified and expanded as needed for clarification:

Analysis dataset metadata: Removed “Purpose” as a metadata field. Removed the requirement that an analysis dataset should have “analysis” or “statistics” in the dataset label.

Analysis variable metadata: Clarification of the source field. Definition of parameter value-level metadata, replacing value-level metadata.

Analysis results metadata: Significant expansion of the fields used in analysis results metadata, to facilitate clarification of the components.

Significant expansion and modifications to analysis dataset variables were made when the text was moved to the ADaMIG.

Appendix G Representations and Warranties; Limitations of Liability, and Disclaimers

CDISC Patent Disclaimers

It is possible that implementation of and compliance with this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken with respect to the existence or validity of any claim or of any patent rights in connection therewith. CDISC, including the CDISC Board of Directors, shall not be responsible for identifying patent claims for which a license may be required in order to implement this standard or for conducting inquiries into the legal validity or scope of those patents or patent claims that are brought to its attention.

Representations and Warranties

Each Participant in the development of this standard shall be deemed to represent, warrant, and covenant, at the time of a Contribution by such Participant (or by its Representative), that to the best of its knowledge and ability: (a) it holds or has the right to grant all relevant licenses to any of its Contributions in all jurisdictions or territories in which it holds relevant intellectual property rights; (b) there are no limits to the Participant's ability to make the grants, acknowledgments, and agreements herein; and (c) the Contribution does not subject any Contribution, Draft Standard, Final Standard, or implementations thereof, in whole or in part, to licensing obligations with additional restrictions or requirements inconsistent with those set forth in this Policy, or that would require any such Contribution, Final Standard, or implementation, in whole or in part, to be either: (i) disclosed or distributed in source code form; (ii) licensed for the purpose of making derivative works (other than as set forth in Section 4.2 of the CDISC Intellectual Property Policy ("the Policy")); or (iii) distributed at no charge, except as set forth in Sections 3, 5.1, and 4.2 of the Policy. If a Participant has knowledge that a Contribution made by any Participant or any other party may subject any Contribution, Draft Standard, Final Standard, or implementation, in whole or in part, to one or more of the licensing obligations listed in Section 9.3, such Participant shall give prompt notice of the same to the CDISC President who shall promptly notify all Participants.

No Other Warranties/Disclaimers. ALL PARTICIPANTS ACKNOWLEDGE THAT, EXCEPT AS PROVIDED UNDER SECTION 9.3 OF THE CDISC INTELLECTUAL PROPERTY POLICY, ALL DRAFT STANDARDS AND FINAL STANDARDS, AND ALL CONTRIBUTIONS TO FINAL STANDARDS AND DRAFT STANDARDS, ARE PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, WHETHER EXPRESS, IMPLIED, STATUTORY, OR OTHERWISE, AND THE PARTICIPANTS, REPRESENTATIVES, THE CDISC PRESIDENT, THE CDISC BOARD OF DIRECTORS, AND CDISC EXPRESSLY DISCLAIM ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR ANY PARTICULAR OR INTENDED PURPOSE, OR ANY OTHER WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, FINAL STANDARDS OR DRAFT STANDARDS, OR CONTRIBUTION.

Limitation of Liability

IN NO EVENT WILL CDISC OR ANY OF ITS CONSTITUENT PARTS (INCLUDING, BUT NOT LIMITED TO, THE CDISC BOARD OF DIRECTORS, THE CDISC PRESIDENT, CDISC STAFF, AND CDISC MEMBERS) BE LIABLE TO ANY OTHER PERSON OR ENTITY FOR ANY LOSS OF PROFITS, LOSS OF USE, DIRECT, INDIRECT, INCIDENTAL, CONSEQUENTIAL, OR SPECIAL DAMAGES, WHETHER UNDER CONTRACT, TORT, WARRANTY, OR OTHERWISE, ARISING IN ANY WAY OUT OF THIS POLICY OR ANY RELATED AGREEMENT, WHETHER OR NOT SUCH PARTY HAD ADVANCE NOTICE OF THE POSSIBILITY OF SUCH DAMAGES.

Note: The CDISC Intellectual Property Policy can be found at http://www.cdisc.org/about/bylaws_pdfs/CDISCIPPolicy-FINAL.pdf.