Analysis Data Model (ADaM)

Version 2.1 Final

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
2009-12-07	2.1 Final	Released version reflecting all changes and corrections identified during comment period.
2006-04-11	2.0 Final	Final document

Note: Please see Appendix G for Representations and Warranties; Limitations of Liability, and Disclaimers.

Contents

1 Introduction / Purpose

1 简介/目的

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.
分析数据模型（ADaM）文档规定了创建分析数据集和相关元数据时要遵循的基本原则和标准。元数据是"关于数据的数据"或"关于数据的信息"。分析数据模型支持分析结果高效地生成，再现和审查。

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.
分析数据集的设计通常取决于临床试验的科学性和医学目标。一个基本原则是分析数据集的结构和内容必须支持试验的科学方面和统计方面进行清晰，明确的传达。

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.
ADaM的目的是提供一个能够分析数据的框架，同时让数据的审阅者和其他接收者们清楚地了解数据从收集到分析再到结果的系谱。ADaM经过优化来支持数据的衍生和分析，而CDISC的原始数据标准模型（SDTM）是为了最优化地支持数据制表。

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.
ADaM文档（即本文档）提供了核心要点并定义了ADaM概念和标准的实质和意图。它概述了构建分析数据集和相关元数据时要遵循的基本原则。在该文档中描述了四种类型的ADaM元数据（即，分析数据集元数据，分析变量元数据，分析参数值水平元数据和分析结果元数据），并且提供了范例。

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.
本文档中介绍了受试者水平的分析数据集（ADSL）（参见第4.2节）。即使没有其他分析数据集提交，一个基于CDISC的临床试验数据提交也需要提供ADSL及其相关元数据。

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).
本文档也介绍了ADaM基本数据结构（BDS），此结构被应用于各个治疗领域或者分析类型的大多数ADaM数据集（请参见4.2节）。虽然一般而言BDS支持大多数的统计分析，但是一个研究也会包括特定的标准化结构的分析数据集，用来表达其它的分析信息，比如受试者水平的分析数据集（ADSL）和不良事件分析数据集（ADAE）。

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.
本文档是ADaM实施指南(ADaMIG)的基础，这份指南具体说明这些核心概念的标准化实施，ADaM标准数据集的结构和变量，包括命名规则。它还说明了实施中的问题的标准解决方法。

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.
在采用ADaM的原则和标准构建数据集及其相关元数据时，要想获得分析数据集所能带来的全部益处，审阅者或其他数据接收者与申办者之间尽早而有效的交流是必不可少的，这方面是无论如何强调都不过分的。

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.
在对于ADaM概念的说明中，范例引用了特定的编程语言。在整份ADaM文档中，引用特定供应商的产品只是为了举例，不应理解为是对这些产品的支持。

Note that the examples in this document are only intended as illustrations and should not be viewed as a statement of the standards themselves.
注，本文档中的范例只是用于阐释，不应视为标准本身的说明。

2 Background / Motivation

2 背景/动机

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).
受管制的人类健康产品的上市批准的过程通常包括临床试验数据的提交。在美国，数据是向美国食品药品监督管理局（FDA）提交的必要元素。

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).
在1997年，FDA通过颁布使用电子记录代替纸质记录（21 CFR 第11章）的规章制度建立了电子提交数据的监管基础。在1999年，FDA标准化了临床试验中收集的电子提交数据的文件格式（SAS® 版本5 传输文件[1]）。这在描述临床数据和数据定义文件(define.pdf)提交指南文档系列的第一个文档中进行了解释。

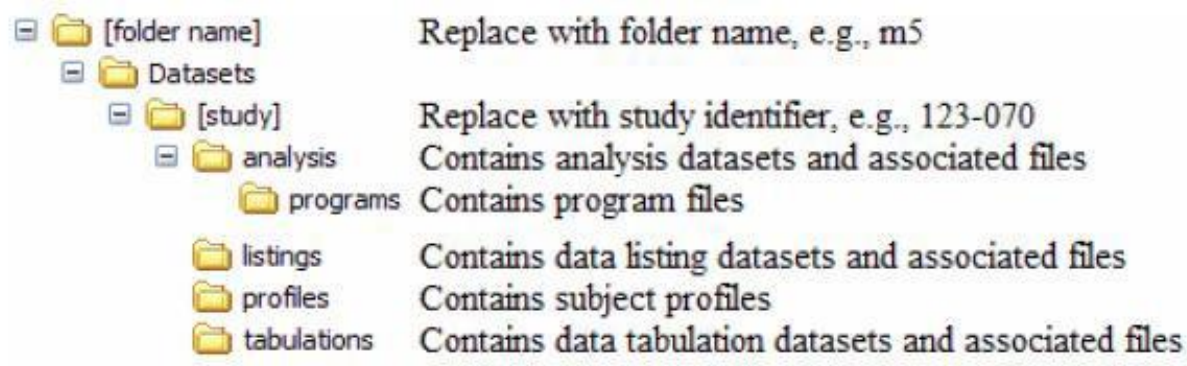
Though the 1999 guidance was withdrawn in 2006, datasets are still submitted using the SAS transport file format, accompanied by a "define file." The define file is a data definition document which provides a list of the datasets included in the submission along with a detailed description of the contents of each dataset (i.e., the metadata for the submitted datasets).
虽然1999年的指南在2006年被撤销，但数据集仍然使用SAS传输文件格式提交，伴有一个定义文件"define file"。这个定义文件是一个数据定义文档，它提供了一个包括在提交中的数据集清单，和对于每个数据集内容的一个详细描述（即提交数据集的元数据）。

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].
从2005年开始，元数据可以使用可扩展的标记语言（XML）格式(define.xml)而不是PDF文档格式(define.pdf)提交，这在FDA的关于研究数据说明书文档中有所描述[7]。更多的关于define.xml的信息可以在CDISC网站找到[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].

与制订临床数据提交指南并行，FDA已经采用人用药物注册技术要求国际协调会议（ICH）作为注册提交的标准，发布了关于电子通用技术文档(eCTD)的指导文档作为医药产品申请的电子提交的框架。这个指南的第2次修订在2008年发布[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].



根据FDA关于eCTD的指导文档，提交数据可分为四类：1）数据制表，2）数据清单，3）分析数据集，4）受试者概况。这些总集在一起称为病例报告制表(CRTs) [6]。提交的组织文件夹中数据集及其相关文件的说明总结于下图，来自"研究数据说明"[7]。

Figure 2.1 Specification for organizing study datasets and their associated files in folders [7].
图2.1 组织文件夹中研究数据集及其相关文件的说明[7]

Data tabulation datasets and analysis datasets are defined as:
数据制表数据集和分析数据集定义为：

- **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].
  研究数据制表（SDTM）- 数据集包含研究中收集的数据，这些数据按临床域来组织。CDISC研究原始数据标准模型[5]和CDISC原始数据标准模型实施指南(SDTMIG) 中对这些数据集进行了描述[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.
1 SAS和所有其他SAS Institute Inc.产品或服务名称是SAS Institute Inc.在美国和其他国家/地区的注册商标或商标。 ®表示美国注册。

- **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].
  分析数据集（ADaM） – 是用于申办者统计分析和报告的数据集， 与SDTM数据集一起提交。ADaM数据集是用于统计分析的所有数据衍生的权威来源，这些数据集在本文档（CDISC分析数据模型文档）和CDISC ADaMIG[1]中描述。

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.
标准化的分析数据集和元数据提供给数据接收者的好处不只是清楚的交流和透明度。一旦接受了关于标准化数据集的原则的培训，审阅者和其它数据接收者能用较少的准备时间、更有效率的使用数据。另外，标准化的结构支持软件工具的开发，从而有益于对分析结果的获取、衍生、分析、重现、审阅。

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.
SDTM不是为支持统计分析而设计的。ADaM数据集把衍生的和收集的数据（来自各种SDTM域、其它ADaM数据集、或任何它们的组合）合并成一个数据集，以便只经过很少的或不用编程就能进行分析。SDTM中不太容易处理的问题的实例是分析窗口、复杂算法和缺失数据填补。

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.
因为ADaM元数据解释了ADaM数据集是如何从SDTM源数据创建的，在ADaM数据集中衍生和填补的变量不应再复制回SDTM源数据，如果这样做反而会在数据流中引入循环依赖，使变量间的重要关系失去关联。

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.
为了简化本文档，分析数据集是在FDA电子提交背景中讨论的。从一开始，CDISC ADaM小组就得到了FDA统计审阅者和医学审阅者们的鼓励和知识分享，FDA统计审阅者和医学审阅者们作为观察员参加ADaM会议，参与CDISC-FDA预试验。ADaM基本原则的起源是与医学和统计审阅者交流的透明性和完整性的需求和医学和统计审阅者的审评的科学上正确性的要求。ADaM标准的制订是为了满足FDA和行业界的需求。除了适用于FDA的注册提交，ADaM还适用于更大范围的药物研究活动。ADaM为申办者和CRO之间以及开发伙伴和独立数据监测委员会之间传输数据提供了标准。当模型的采用变得更加广泛时，申办者间分析数据集及相关元数据的共同模型就会促进引进授权、对外授权、联合投资及合并。

3 Overview of the Analysis Data Model

3 分析数据模型概述

3.1 Fundamental Principles 基本原则

Fundamental Principles 基本原则

Analysis datasets and their associated metadata must:

分析数据集及其关联元数据必须：

facilitate clear and unambiguous communication

有助于清晰明确的交流

provide traceability between the analysis data and its source data (ultimately SDTM)

提供分析数据和源数据（终极源头SDTM）之间的可溯源性

be readily useable by commonly available software tools

可以在常见软件工具上轻松使用

Analysis datasets must:

分析数据集必须：

be accompanied by metadata

有元数据伴随

be analysis-ready

分析就绪的

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.
设计分析数据集和相关元数据的总体原则是，必须清晰明确地传达数据集的内容和来源，这些数据集将支撑临床研究中的统计分析工作。

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.
本原则的本质要求是可溯源性，可溯源性就是让人们理解一个分析值可溯源性（无论分析结果还是分析变量）从哪里来，即数据的系谱或一个分析值和它前身的关系。有关可溯源性的更详细说明，请参见第3.1.1节。

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.
主办方应该努力提交“分析就绪的”的数据集，也就是分析数据集具备的结构和内容，让统计分析得以进行，并且只需编写最少的程序。

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.
一个“分析就绪的”数据集可以在统计分析软件上直接使用，只需最少的额外处理，例如从分析数据集中选择适当记录或对观测值排序。

No complex data manipulations such as transformations or transpositions are required to perform the supported analysis.
无需复杂的数据操作，例如变换和转置，来支持分析。

This approach eliminates or greatly reduces the amount of programming required by analysts such as statistical reviewers.
这种方法消除或大大减少了，统计评审员等分析师所需的编程量。

Appendix D gives an example of applying this principle in SAS, but the concepts apply to all statistical software packages.
附录D给出了在SAS中应用该原理的示例，但这些概念适用于所有统计软件包。

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.
注意：在ADaM的语境中，为了特定统计分析的审查和重现，分析数据集至少包含所需数据。

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.
只是为了支持数据列表或其它非分析的展示，不需要将数据整理成“分析就绪的”的数据集，尽管有些人可能会选择这样做。

Analysis datasets must be readily usable by commonly available software tools, and must be associated with metadata.
分析数据集必须在常见软件工具中，易于使用，也必须关联元数据。

Ideally the metadata are machine-readable.
理想情况下，元数据是机器可读的。

Metadata and other documentation should provide clear and concise communication of the analyses, including statistical methods, assumptions, derivations and imputations performed.
元数据和其他文档应该提供，清晰简明的分析信息，包括所用的统计方法，假设，衍生和填补。

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.
元数据，程序和其他文档服务于系统化分析（描述在统计分析计划（SAP）中的分析，和其他分析）。 这些将在第5节中详细讨论。

3.1.1 Traceability

3.1.1 可溯源性

The concept of traceability is a cornerstone of the Analysis Data Model.

可溯源性的概念是分析数据模型的基石。

This property enables the understanding of the data's lineage or the relationship between an element and its predecessor(s).
此属性让人们理解数据的系谱或一个元素与其前身之间的关系。

Traceability facilitates transparency, which is an essential component in building confidence in a result or conclusion.
可溯源性有助于透明化，在树立对一个结果或结论的信心中，透明化是必不可少的组成成分。

Ultimately, traceability in ADaM permits the understanding of the relationship among the analysis results, the analysis datasets, and the SDTM domains.
最终，ADaM中的可溯源性，让人们理解分析结果，分析数据集和SDTM域之间的关系。
Traceability is built by clearly establishing the path between an element and its immediate predecessor.
可溯源性可溯源性是通过清楚地确定元素与其前身之间的路径而建立的

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.
从一个元素到它前身，再到前身的前身，以此类推，直到SDTM域，最终到收集数据的工具，完整的路径得以追踪。

Note that the CDISC Clinical Data Acquisition Standards Harmonization (CDASH) standard is harmonized with SDTM and therefore assists in assuring end-to-end traceability.
注意，CDISC临床数据采集协调标准（CDASH）与SDTM协调一致，因此有助于确保端到端的可溯源性。

Traceability establishes across-dataset relationships as well as within-dataset relationships.
可溯源性建立数据集之间的关系，也建立数据集内部的关系。

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.
例如，分析数据集中，标帜变量元数据和其他变量的元数据，使用户能够理解派生记录是如何（以及在某种程度上，因何）创建。

There are two levels of traceability:
有两个层面的可溯源性：

- 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).
  元数据的可溯源性，使用户能够理解，分析变量与其源数据集之间的关系，和分析变量与其源变量之间的关系，这也符合ADaM的要求。通过（在元数据中）描述所用算法或步骤，从一个分析值的前身，来衍生或填充这个分析值，可溯源性得以建立。元数据的可溯源性，也用来建立一个分析结果（例如一个p值）和分析数据集之间关系。

- 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.
  数据点的可溯源性，使用户能够直接找到特定的前身记录。如果实际可行，应该实施数据点的可溯

源性。当审阅者试图追踪一个复杂的，数据操作的路径时，此层面的可溯源性非常有用。某个特定数据值，用作一个分析值的输入，通过在数据中提供，到这个特定数据值的，清晰的链接（例如通过••SEQ变量的使用），来建立可溯源性。注意：可能存在这样的情况，数据点的可溯源性非常困难，不切实际，甚至不可行，例如，在多导睡眠图研究中，脑电图记录中的关键结果，可以基于频谱边缘参数，而频谱边缘参数又是通过快速傅里叶变换的绝对功效值，衍生而来。

When traceability is successfully implemented, reviewers are able to identify:
当可溯源性成功实施后，审核人员可以识别：

- information that exists in the submitted SDTM study tabulation data
  哪些信息，在已经提交的，SDTM研究制表数据中
- information that is derived or imputed within the ADaM analysis dataset
  哪些信息，在ADaM分析数据集中，衍生或填补而来
- the method used to create derived or imputed data
  哪些方法，用于创建衍生或填补数据
- information used for analyses, in contrast to information that is not used for analyses yet is included to support traceability or future analysis
  哪些信息用于分析。哪些信息不用于分析，但会用于可溯源性或后续分析

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.
在图3.2.1中，展示了数据从源头到分析结果的典型通用流程概念图。

The schematic only illustrates one reasonable e scenario.
这个示意图只是展示了一个合理场景。

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.
其意图不在于描绘所有可能的关系和所有可能的组件，也不在于指明或暗示这是唯一的方式去运作这个流程。

For example, metadata may actually inform or drive the process rather than be an output of the process.
例如，元数据可以给过程输入信息，或驱动过程，而不是过程的输出。

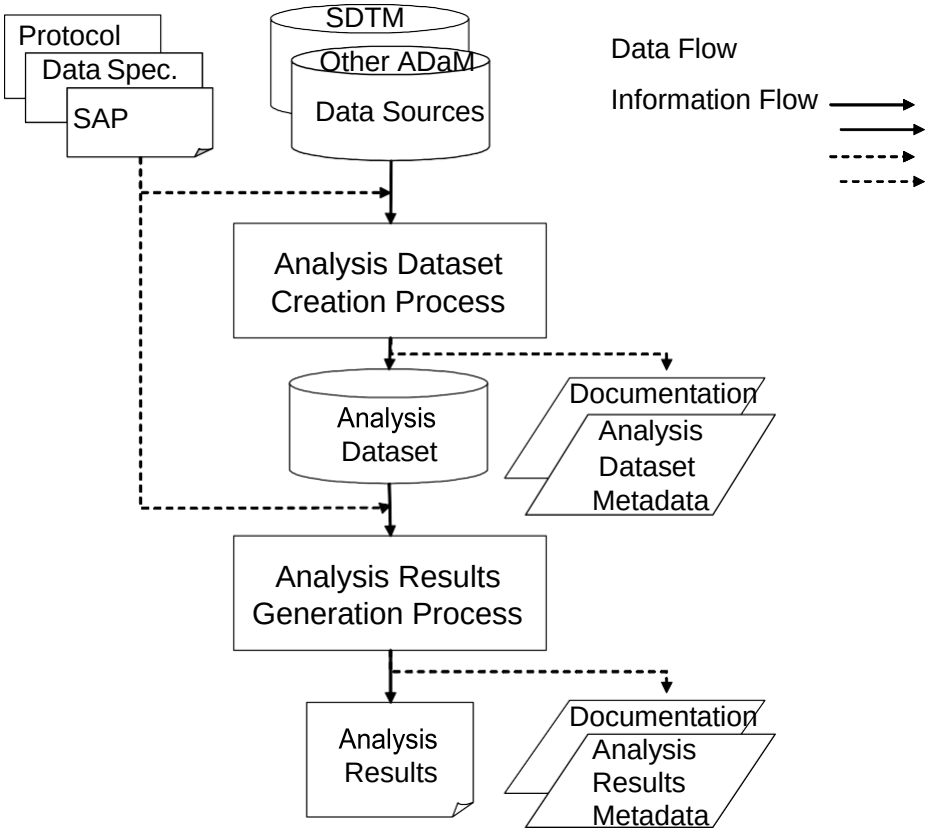


Figure 3.2.1: Analysis Data Flow Diagram Showing One Scenario for the Flow of Data and Information

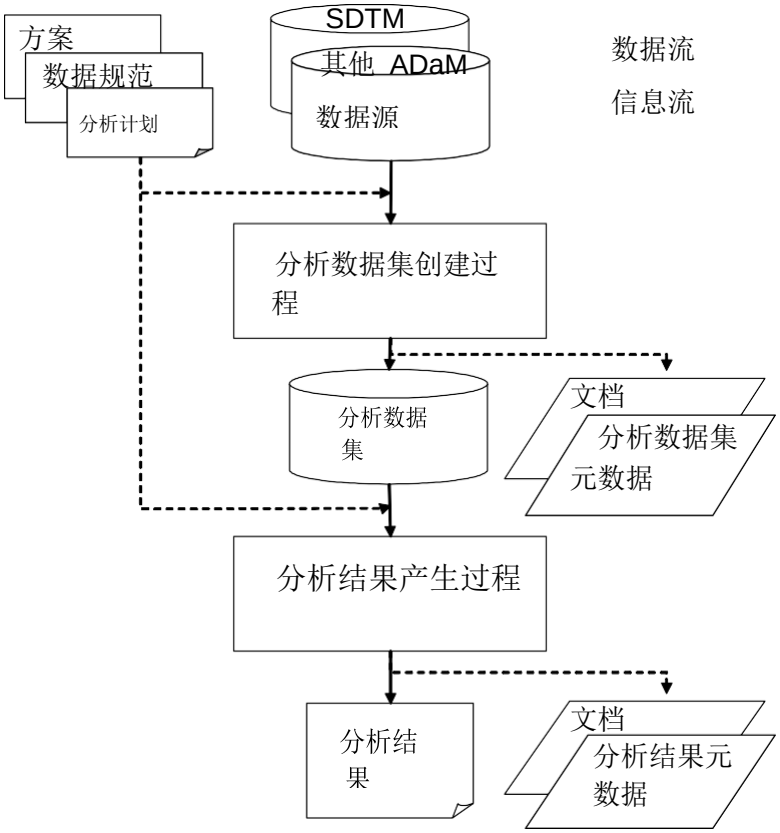


Figure 3.2.1: Analysis Data Flow Diagram Showing One Scenario for the Flow of Data and Information
图 3.2.1 分析数据流程示意图，展示了数据信息流的一个情境

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).
鉴于ADaM标准已经发展成，CDISC标准家族中的一员，假设源数据是SDTM，或其他分析数据集，例如受试者水平的分析数据集（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.
一次符合CDISC的提交包括，SDTM数据集和ADaM数据集；因此，为了便于追溯，元数据需要描述，两个数据集集合之间的关系。

To facilitate clear communication, a distinction is made between the processes of Analysis Dataset Creation and Analysis Results Generation.
为了方便清晰的沟通，创建分析数据集的过程和产生分析结果的过程，做了区分。

These two processes have distinct purposes and consequently require different types of metadata, as outlined in the following section.
这两个过程的目的不同，因此需要不同类型的元数据，如下一节所述。

Analysis Dataset Creation – The processing and programming steps used to create the analysis datasets.
分析数据集的创建 – 用于创建分析数据集的过程步和编程步

As shown in Figure 3.2.1, the analysis dataset creation program is developed based on the analysis plans and dataset specifications.
如图3.2.1所示，创建分析数据集程序的开发，基于分析计划和数据集规范。

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.
进入程序的数据是源数据（例如，SDTM数据集和/或其他分析数据集），而输出是分析数据集。

Analysis Results Generation – The programming steps used to generate analysis results (e.g., summary or inferential statistics presented in tabular or graphical presentations).
分析结果的产生 – 用于产生分析结果的编程步（例如，呈现在表格或图示中的，汇总统计和推断统计）。

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.
如图3.2.1所示，产生分析结果程序的开发，基于统计分析计划，数据衍生规范和数据集规范。

The data going into the program are in the analysis dataset and the output is the analysis result.
进入程序的是分析数据集中的数据，而输出是分析结果。

3.3 Metadata Components

3.3 元数据组件

The analysis datasets and ADaM metadata facilitate the review of the clinical trial data and the analyses performed.
分析数据集和ADaM元数据有助于临床实验数据的审查和分析。

There are four types of metadata described in this document. These include:
本文中，有四种类型的元数据：

- 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.
分析数据集元数据：描述每一个分析数据集，包括内容的简单描述（请参阅第5.1节）所有ADaM数据集都需要此类型的元数据。

- 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.
分析变量元数据：描述数据集中的变量，包括分析变量的来源与创建的信息，例如，所涉及算法的详细描述或创建分析数据集程序的名字。（请参阅第5.2节）所有ADaM数据集都需要此类型的元数据。

- 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.
分析参数值水平元数据：描述分析参数“内”的测量或分析终点（即，对于分析参数的每个唯一值）。就像在ADaM基本数据结构中， 当数据结构允许一个变量包含多种类型的测量或分析终点时，尤其需要这种形式的元数据。（请参阅第5.2.1节）所有ADaM BDS数据集都需要此类型的元数据。

- 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.
分析结果元数据描述了分析的结果 （申办方规定的），包括用了什么分析数据集用来分析以及关于分析的一些信息。（参考5.3章节）分析结果元数据提供了从统计结果的展现到分析数据集的溯源性。分析结果元数据并非是必须的。然而，最佳实践是分析结果元数据可以用来支持审阅人员识别关键的数据分析内容，并提供一个在分析结果、文档、数据集之间的关联关系，以及把分析过程以文档形式记录下来。

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.
前三种元数据描述了分析数据集。它们是在生成分析数据集过程中产生的。分析数据集元数据总体上描述了分析数据集，然而分析变量元数据和分析参数值水平元数据描述了数据集中的变量和记录。

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.
不管是为了记录分析数据产生的过程，还是分析结果的产生过程，可以在元数据文档中包含片段化的程序代码，指向程序、试验方案、统计分析计划或者其他文档的链接。

4 AnalysisDatasets

4. 分析数据集

4.1 Practical Considerations

4.1 实际考虑因素

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

分析数据集必须:

- 包含一个名为"ADSL"的受试者水平分析数据集（参见第6节）
- 由所需的最佳数量的分析数据集组成，并且具有足够的自给自足能力，以满足只需要很少或者不需要额外的编程或数据处理就可以进行分析和审查。
- 使用命名规则"ADxxxxxx" 命名
- 尽可能地使用AdaM标准变量名和命名规则
- 如果复制到分析数据集中而不重新命名，则维护SDTM变量的值和属性，即坚持"同名、同义、同值"协调原则。
- 在给一个特定递交和产品的多个提交的研究中一致地应用数据集和变量的命名规则

4.1.1 The Number and Content of Analysis Datasets

4.1.1 分析数据集的数量和内容

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.
在创建支持临床研究报告或提交的分析结果的分析数据集时，其中一个目标是获得执行各种分析所需的最佳分析数据集数量(最低要求是ADSL)。不需要为不同的分析提供单独的数据集;一个数据集可以支持多个分析。也不需要分析数据集支持每个数据汇总。此外，不要求每个SDTM域都有相应的分析数据集。申办方决定要创建的分析数据集。

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.
创建单个分析数据集可能需要多个数据集(例如SDTM，其他分析数据集)。这是必要的，以便分析数据集包含执行其设计支持的统计分析所需的所有变量。例如，可能需要从ADSL和配置(DS)、人口统计(DM)、受试者特征(SC)、生命体征(VS)、问卷(QS)和暴露(EX)域获取数据，以创建单个分析数据集。

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.
分析数据集被设计成以最少的编程或数据处理来促进分析和审查。分析数据集之间的关联重复数据(即同样的数据在多个数据集中出现)常常是必要的，以便保证数据集是分析就绪的数据集(例如，不良事件分析数据集中的年龄以及功效分析数据集中的年龄)。类似地，也可以包括变量和记录，这些变量和记录实际上不在任何递交的分析中使用，但是对于主办方或评审者来说仍然是感兴趣的（例如，具有临床兴趣事件的受试者的标识标志）或者支持可溯源性。

An example of a composite endpoint requiring complex algorithms and input from multiple datasets is shown in Appendix E.
附录E显示了需要复杂算法和来自多个数据集的输入的复合端点的示例。

4.1.2 Analysis Dataset and Variable Naming Conventions

4.1.2 分析数据集和变量命名规则

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).
分析数据集使用规定“ADxxxxxx”命名。如第6节所述，受试者水平分析数据集被命名为“ADSL”。对于所有其他分析数据集，名称的xxxxxx部分是主办方定义的，在特定的递交或产品的多个递交之间使用公共命名规定。在开发命名规定时，申办方应该考虑eCTD指导文档[6]中提到的要求，以及符合sas传输格式要求的需要(例如名称的总长度不能超过8个字符)。

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.

在分析数据集中创建的变量的命名规则(不要与SDTM所需的任何标准变量混淆)应该遵循ADaMIG中定义的标准变量名称和命名规定。除此之外，分析变量的名称也可以由申办方定义的，并且，尽可能地，在针对某一产品的某一次试验数据递交中或者在不同的试验中的多次数据递交也应该遵循一个通用的命名规则。

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."
任何与SDTM变量同名的ADaM变量都必须是SDTM变量的副本，并且不能修改它的标签、属性和值。应当坚持“同名同义同值”的统一原则。

Refer to the ADaMIG [1] for more general variable naming conventions.
参考ADAMIG[1]以获得常规性的变量命名规则。

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.
理想情况下，分析数据集中变量的顺序遵循了逻辑顺序（不只是简单的字母排序）。有关分析数据集中变量顺序的更多信息，请参见FDA“研究数据规范”[7]。建议申办方为数据集内的变量排序定义一个约定，然后对所有分析数据集一致地应用此排序。数据集中变量的顺序应与数据说明文件中显示的变量顺序匹配

4.2 ADaM Data Structures

4.2 ADaM数据结构

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).
本文和ADaMIG中描述了两个ADaM标准数据结构:受试者水平分析数据集(ADSL)和基本数据结构(BDS)。

4.2.1 The Subject-Level Analysis Dataset (ADSL) Structure

4.2.1 受试者水平分析数据集（ADSL）结构

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.
ADSL数据集结构中每个受试者都有一条记录，并包含一些变量，如受试者水平人群标志、计划和实际治疗变量、人口信息、随机化因素、亚组变量和重要日期。ADSL包含必需的变量(如ADaMIG中指定的)以及其他在描述受试者在试验中的经历时很重要的受试者水平变量。即使没有提交其他分析数据集，基于CDISC的临床试验数据提交也需要ADSL及其相关元数据。有关ADSL的详细描述，请参阅第6节。

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.
虽然从技术上讲，将研究中的每一个数据值作为变量包括在ADSL等受试者水平数据集中是可行的，但这并不是ADSL的目的。关键终点和在研究过程中随时间变化的数据的正确位置包含在BDS数据集中。

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.
BDS包含每个受试者、每个分析参数、每个分析时间点的一个或多个记录。分析时间点是特定条件下必需的，这取决于分析。在没有分析时间点的情况下，该结构是每个受试者每个分析参数一个或多个记录。该结构包含一组描述分析参数的中心变量(例如PARAM和相关变量)，并包含被分析的值(例如AVAL和AVALC和相关变量)。数据集中的其他变量提供了关于正在分析的值(例如，主题标识)的更多信息，或者描述和跟踪它的派生(例如，DTYPE)，或者支持分析(例如，处理变量，协变量)。BDS支持参数化和非参数化分析，如ANCOVA、分类分析、逻辑回归、Cochran-Mantel-Haenszel、Wilcoxon rank-sum、time- to-event分析等。有关BDS标准的详细信息，请参阅ADaMIG[1]。

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.
尽管BDS支持大多数统计分析，但它并不支持所有统计分析。例如，它不支持同时分析多个相关（反应/结果）变量或一系列反应变量的相关性分析。BDS并非设计用于支持不良事件发生率分析或其他发生数据。

4.2.3 Future ADaM Data Structures

4.2.3 未来ADaM数据结构

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.
ADaM团队目前正在为支持不良事件发生率分析的ADAE数据集编写规范文档。ADAE可能是支持分析发病率数据（例如不良事件，伴随药物等）的更加通用的结构的第一个例子，

5 ADaM Metadata

5.ADaM元数据

ADaM Metadata
Analysis dataset metadat 分析数据
集元数据
Analysis variable metadata 分析变
量元数据
Analysis parameter value-level metadata 分析参
数值水平元数据

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.
在临床试验数据分析中基于的基本假设、统计方法、转换、推算和填补等信息或操作，应该是可以很明确的进行交流沟通，并且其数据值和分析结果可以比较容易的被重新生成出来。易于复制的价值和结果的方式清晰地表达。ADaM元数据通过提供介于数据分析的一般性描述(方案中的数据分析章节, 统计分析计划，或者报告的分析方法)、分析结果、分析中使用的数据和SDTM域之间的详细说明和链接来促进这种交流。下面的章节详细地描述了ADaM的元数据的各个组成部分。

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.
所述分析数据集元数据和分析变量元数据的元数据结构是基于病例报告制表数据定义说明标准1.0.0版(CRT-DDS) [2]，更多细节，请参考该文档。

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.
本文档中的元数据示例只是为了解释说明，不是为了指定或者推荐报告的风格、格式或过程。此外，一些图解中的斜体字行只是作为引用，作为对字段定义的提醒，不是为了让申办者把这一行包括在元数据中。

5.1 Analysis Dataset Metadata 分析数据集元数据

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.
分析数据集元数据提供关于分析数据集的信息，包括对数据集内容的描述。最佳实践强烈推荐每个分析数据集都使用表5.1.1中列出的元数据字段描述。ADSL和BDS分析数据集（即遵循ADaM）必须用这些元数据字段描述。实践经验告诉我们，有时候也会用到非这些结构的分析数据集，建议也使用表5.1.1中的元数据字段描述它们。

Table 5.1.1 Analysis Dataset Metadata Fields 表5.1.1 分析数据集元数据字段

Analysis Dataset Metadata Field	Description
DATASET NAME 数据集名	The file name of the dataset, hyperlinked to the corresponding analysis dataset variable descriptions (i.e., the data definition table) within the define file. 数据集的文件名，超链接到定义文件的相应分析数据集变量描述（即，数据定义表）。
DATASET DESCRIPTION 数据集描述	A short descriptive summary of the contents of the dataset 对于数据集内容的简短描述性总结。
DATASET LOCATION 数据集位置	The folder and filename where the dataset can be found, ideally hyperlinked to the actual dataset (i.e., XPT file) 可以找到数据集的文件夹和文件名，最好超链接到实际的数据集（即XPT文件）。
DATASET STRUCTURE 数据集结构	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"). 数据集各记录代表的详细水平（例如，"每个受试者一条记录"，"每个受试者每个访视一条记录"，"每个受试者每个事件一条记录"）。
KEY VARIABLES OF DATASET 数据集的关键变量	A list of variable names that parallels the structure, ideally uniquely identifies and indexes each record in the dataset. 与结构相应的变量名列表，最好能唯一标识和索引数据集中的每条记录。
CLASS OF DATASET 数据集类型	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确定数据集的总的类型，ADaM结构名（即"ADSL," "BDS"）或"其它"（如果不是ADaM说明的结构）。
DOCUMENTATION 文档	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. 关于数据集创建的源数据、处理步骤和分析决定的描述。各种级别的功能性和复杂性的软件代码，可能提供如伪代码、或实际的代码片段。可能使用到外部文档的链接或者引用（如方案、统计分析计划、软件代码）。

5.1.1 Illustration of Analysis Dataset Metadata分析数据集元数据的举例

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.
在本例中，分析的数据是阿尔茨海默病评估量表-认知部分（ADAS-Cog）11项分数的总和。评估的是调查问卷，所以调查问卷（QS）SDTM域用于收集数据。表5.1.1.1阐明了ADAS-Cog分析数据集ADQSADAS的分析数据集元数据。

Table 5.1.1.1 Analysis Dataset Metadata for the ADQSADAS Analysis Dataset[2]
表5.1.1.1 ADQSADAS分析数据集的分析数据集元数据

Dataset Name 数据集名	**Dataset Description 数据集描述**	**Dataset Location 数据集位置**	**Dataset Structure 数据集结构**	**Key Variables of Dataset 数据集关键变量**	**Class of Dataset 数据集的类型**	**Documentation 文档**
filename of the dataset. 数据集的文件名	*short summary of contents of the dataset 数据集内容的简短总结*	*where the dataset can be found 可以找到数据集的地方*	*the level of detail in the dataset 数据集中的细节水平*	*variable names that parallel the structure 和结构平行的变量名称*	*general class of the dataset using controlled terminology 使用受控术语的数据集一般类型*	*links or references to documentation re how the dataset was created 链接或者引用关于数据集如何产生的文档*
ADQSADAS	Data for the ADAS-Cog (11) Analyses ADAS-Cog(11) 分析数据	adqsadas.xpt	one record per subject per parameter per analysis visit 每个受试者，每个参数，每个分析访视一条记录	USUBJID, PARAMCD, AVISIT	BDS 基本数据结构	DSADQSADAS.SAS, Section 14.11 of SAP for detailed ADAS-Cog scoring algorithm DSADQSADAS.SAS，详细的 ADAS-Cog的评分算法见 SAP 的 14.11 节

[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.
元数据的呈现形式应该在申办方和审阅人员之间确定。这个例子只是用来说明内容而不是外观。

5.2 Analysis Variable Metadata分析变量元数据

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.
分析变量元数据描述分析数据集中的每个变量，包括变量的属性和定义。这些描述的元数据字段列在了表5.2.1中。最佳实践强烈推荐每个分析变量都使用这些元数据字段进行描述。遵循ADaM的分析数据集必须用这些分析变量元数据字段进行描述。

Table 5.2.1 Analysis Variable Metadata Fields表5.2.1分析变量元数据字段

Analysis Variable Metadata Field	Description
DATASET NAME 数据集名	The file name of the analysis dataset 分析数据集文件名
VARIABLE NAME 变量名	The name of the variable 变量名
VARIABLE LABEL 变量标签	A brief description of the variable 变量的简短描述
VARIABLE TYPE 变量类型	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") 变量类型。有效值如"病历报告制表数据定义说明标准"所定义（如版本1.0.0，有效值包括"文本"、"整数型"、和"浮点型"）。
DISPLAY FORMAT 显示格式	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). 变量显示信息（即，在结果展示的表格或图形变量的格式）。建议语法和用于分析的软件包内嵌的格式术语相一致（如$16或3.1，如果使用SAS）。
CODELIST / CONTROLLED TERMS 编码列表/受控术语	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. 有效值或允许编码的列表以及对应的变量解码。这个字段可以包括对外部编码列表的引用（由名称和版本识别）或者超链接到定义文件的编码列表/受控术语章节。
SOURCE / DERIVATION 来源/衍生	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. 提供关于变量传承的细节–前身是什么，其中变量来自的源数据（SDTM或其它分析数据集）或者这些变量是怎样衍生的。这个字段用于识别直接前身来源和/或算法的简短的描述、或用于该来源的过程，能够包含超链接的文本，为读者提供更多参考信息。来源/衍生可以简单到只有两个水平的名称(如ADSL.AGEGR)来识别变量来源的数据文件和变量（即拷贝的变量，没有任何改动）；可以是衍生的简单描述和用于衍生的变量（如"ADSL.BMI的归类"）；也可以是复杂的算法，其中元素包含完整的衍生算法的描述和/或到包含它的文档的链接和/或到分析数据集创建程序的链接。

Refer to Section 5.2.2 for an example of analysis variable metadata. 参考5.2.2节的分析变量元数据的例子。

5.2.1 Analysis Parameter Value-Level Metadata

5.2.1 分析参数值水平的元数据

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.
遵循ADaM基本数据结构（BDS）的分析数据集，即BDS类型的一个分析数据集，包含多个分析参数。在BDS分析数据集中，变量PARAM包含对该数据集中每个分析参数的唯一描述。变量PARAMCD包含PARAM中的分析参数的简称，两变量间是一一映射的。每个PARAM值标识数据集中一组的一或多行。
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 variablehave metadata that vary depending on PARAM/PARAMCD, and links the metadata for a variable to the appropriate value of PARAM/PARAMCD.
数据集的列（变量）的元数据通常依赖于PARAM/PARAMCD的值，这个概念与SDTM中的一个单独变量值水平元数据类似，但在BDS中，几个变量的元数据的PARAM/PARAMCD不同很常见。为了描述这些变量的PARAM/PARAMCD不同的情况， BDS分析数据集的变量水平元数据的元素必需有**参数标识符**。参数标识符元数据元素确定随PARAM/PARAMCD不同而异的元数据，把变量元数据链接到PARAM/PARAMCD的相应值。
Controlled terminology reduces the need to enter the same metadata for a variable for multiple values of PARAM/PARAMCD:
受控术语减少了为PARAM/PARAMCD的多个值输入相同变量元数据的需求：

- 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.
  变量的参数标识符为"*ALL*"，说明分析数据集中那个变量的元数据对于 PARAM/PARAMCD 的所有值都是一样的。
- 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.
  变量的参数标识符为"*DEFAULT*"，除非另有说明，那个变量的特定的元数据应视为分析数据集中 PARAM/PARAMCD 所有值的元数据。
- 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.
  变量的参数标识符中 PARAMCD 的一个特定值说明那个变量指定的元数据应视为对特定的 PARAMCD 适用的元数据，如果有的话，优先于指定的*DEFAULT*元数据。

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.
参见5.2.2节中BDS数据集的分析变量元数据的例子，使用了参数标识符元数据元素应该注意的是，这个元数据元素促进了元数据内容的输入和跟踪；在定义文件中如何展示元数据由申办者决定。

Analysis Variable Metadata Field 分析变量元数据字段	**Description 描述**
PARAMETER IDENTIFIER 参数标识符	Contains either 包含 1) the value of PARAMCD that identifies the analysis parameter to which the variable metadata applies; PARAMCD 的值，标识变量元数据适用的分析参数， or 或 2) controlled terminology to indicate groupings of analysis parameters: 受控术语，指出分析参数的归类： *ALL* - Used when the variable metadata applies to all analysis parameters in the dataset. *ALL* - 当变量元数据适用于数据集中所有的分析参数时使用 *DEFAULT* - Used when the variable metadata applies to all analysis parameters in the dataset except for those specifically listed within the metadata. *DEFAULT* - 当变量元数据适用于数据集中除了那些在元数据特别列出的分析参数之外的所有的分析参数时使用。

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.
参考PARAMCD的编码列表，数据集使用者就能够确定数据集中找到的所有分析参数值，能够确定分析参数特有的属性和当PARAMCD为某一特定值的时候变量的衍生算法。
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.

注意对于PARAMCD变量，参数标识符是“PARAMCD”，变量存在的值的列表也作为分析数据集中分析参数和参数标识符的索引。

5.2.1 Illustration of Analysis Variable Metadata, Including Analysis Parameter Value-Level Metadata分析变量元数据举例，包括分析参数数值水平元数据

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.
这个例子是在ADAS-Cog分析举例的扩展，分析的数据是ADAS-Cog (11) 总分，即ADAS-Cog11条目的子量表。评估是调查问卷，所以收集的数据在QS SDTM域中。表5.2.2.1 是ADAS-Cog分析数据集ADQSADAS分析变量元数据的举例，如5.1.1节中所述。

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.
与任何其它BDS分析数据集一样，使用了分析参数值水平元数据。数据集包含各条目的分数，也包括总分，所以某些变量需要有不同的元数据，依赖于PARAM的值。本例用末次观测结转（LOCF）填补总分的缺失值；对单项的分数不进行填补。注意，PARAMCD的所有值必须列在编码列表元素中。

Table 5.2.2.1 Analysis Variable Metadata for the ADQSADAS Dataset[3]
表5.2.2.1 ADQSADAS数据集的分析变量元数据

Dataset Name 数据集名	Parameter Identifier 参数标识符	Variable Name 变量名	Variable Label 变量标签	Variable Type 变量类型	Display Format 显示格式	Codelist / Controlled Terms 编码列表/受控术语	Source / Derivation 来源/衍生
file name of the analysis dataset 分析数据集的文件名	*PARAMCD or *ALL* or *DEFAULT* PARAMCD 或*ALL* 或 *DEFAULT**	*Name 名称*	*Description 描述*	*Type 类型*	*display informa-tion 显示信息*	*valid values or codes and decodes 有效值或编码和解码*	*where the variable came from in the source data or how the variable was derived变量从源数据的何处来或变量是怎样衍生的*
ADQSADAS	*ALL*	STUDYID	Study Identifier 研究标识符	text	$12		ADSL.STUDYID
ADQSADAS	*ALL*	SITEID	Study Site Identifier 研究中心标识符	text	$3		ADSL.SITEID

Dataset Name 数据集名	Parameter Identifier 参数标识符	Variable Name 变量名	Variable Label 变量标签	Variable Type 变量类型	Display Format 显示格式	Codelist / Controlled Terms 编码列表/受控术语	Source / Derivation 来源/衍生
ADQSADAS	*ALL*	SITEGR1	Pooled Site Group 1 合并中心 1	text	$3		ADSL.SITEGR1
ADQSADAS	*ALL*	USUBJID	Unique Subject Identifier 受试者唯一标识符	text	$11		ADSL.USUBJID
ADQSADAS	*ALL*	AVISIT	Analysis Visit 分析访视	text	$19	Baseline, Week 8, Week 16, Week 24基线，第8周，第16周，第24周	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 如果ADQSADAS.ITTRFL='Y'，那么VISIT是分析访视的名称，如果ADQSADAS.ITTRFL=空，那么AVISIT= 空。基于ADQSADAS.ADY确定分析访视的时间窗算法的详细描述，参见SAP的8.2节。
ADQSADAS	*ALL*	VISIT	Visit Name 访视名	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.
元数据的呈现形式应该在申办方和审阅人员之间确定。这个例子只是用来说明内容而不是外观。

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADQSADAS	*ALL*	AVISITN	Analysis Visit (N)分析访视（N）	integer	3.0	3=Baseline, 8=Week 8, 10=Week 16, 12=Week 24 3=基线,8=第8周, 10=第16周,12=第24周	if ADQSADAS.ITTRFL='Y' AVISITN=numeric code for AVISIT, blank if ADQSADAS.ITTRFL=blank 如果ADQSADAS.ITTRFL='Y'，AVISITN= AVISIT的数字编码, 如果ADQSADAS.ITTRFL=空，则为空。
ADQSADAS	*ALL*	ADY	Analysis Relative Day 相对分析日	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 如果ADQSADAS.ADT >=ADSL.TRTSDT，则 ADY=ADQSADAS.ADT - ADSL.TRTSDT + 1；如果ADQSADAS.ADT <ADSL.TRTSDT，则 ADY=ADQSADAS.ADT -ADSL.TRTSDT。
ADQSADAS	*DEFAULT*	PARAM	Parameter 参数	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 当ADQSADAS.PARAMCD指的是单项分数（而不是总分）时，PARAM是 QS.QSTESTCD =ADQSADAS.PARAMCD 时QS.QSTEST的相应值（对于受试者和访视）。
ADQSADAS	ACTOT11	PARAM	Parameter 参数	text	$16	ADAS-Cog11 Total Score	'ADAS-Cog11 Total Score' is assigned to the total score records 'ADAS-Cog11 总分'被分配给总分记录
ADQSADAS	PARAMCD	PARAMCD	Parameter Code 参数编码	text	$8	ACITM01, ACITM02, ACITM03, ACITM04, ACITM05, ACITM06,	Corresponds to PARAM 与 PARAM 对应

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
						ACITM07, ACITM08, ACITM09, ACITM10, ACITM11, ACITM12, ACITM13, ACITM14, ACTOT11	
ADQSADAS	*DEFAULT*	AVAL	Analysis Value 分析值	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 当ADQSADAS.PARAMCD指的是单项分数（而不是总分）时，AVAL是当QS.QSTESTCD = ADQSADAS.PARAMCD时QS.QSSTRESN的相应值（对于受试者和访视）。

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADQSADAS	ACTOT11	AVAL	Analysis Value 分析值	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 ADAS第1, 2, 4, 5, 6, 7, 8, 11, 12, 13 和14项分数之和，校正缺失值的详情请参见统计分析计划的14.2节。
ADQSADAS	*ALL*	BASE	Baseline Value 基线值	float	3.0		ADQSADAS.AVAL when ADQSADAS.ABLFL='Y'
ADQSADAS	*ALL*	CHG	Change from Baseline 相对于基线的变化	float	3.0		ADQSADAS.AVAL - ADQSADAS.BASE
ADQSADAS	*ALL*	ABLFL	Baseline Record Flag 基线记录标识	text	$1	Y	Y if record contains the baseline value, i.e., if AVISITN=3; blank otherwise 如果记录包含基线值，即如果AVISITN=3，时值为Y；否则为空。
ADQSADAS	*ALL*	TRTP	Planned Treatment 计划的治疗	text	$20	Placebo, Xanomeline Low Dose, Xanomeline High Dose安慰剂、咕诺美林低剂量、咕诺美林高剂量	ADSL.TRT01P
ADQSADAS	*ALL*	TRTPN	Planned Treatment (N) 计划的治疗（N）	integer	1.0	0=Placebo, 1=Xanomeline Low Dose, 2=Xanomeline High Dose 0=安慰剂、 1= Xanomeline低剂量、 2= Xanomeline高剂量	ADSL.TRT01PN
ADQSADAS	*ALL*	TRTDOSE	Randomized Daily Dose Strength, mg随机化的每日剂量规格，毫克	integer	2.0	0=placebo, 54=Xanomeline Low Dose, 81=Xanomeline High Dose 0=安慰剂、 54= Xanomeline低剂量、 81= Xanomeline高剂量	ADSL.TRTDOSE

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADQSADAS	*ALL*	AGE	Age 年龄	integer	3.0		ADSL.AGE
ADQSADAS	*ALL*	AGEGR1	Pooled Age Group 1 合并的年龄组 1	text	$5	<65, 65-80, >80	Based on ADSL.AGEGR1, blank if ADSL.AGE is missing 基于ADSL.AGEGR1，如果ADSL.AGE缺失，则为空
ADQSADAS	*ALL*	AGEGR1N	Pooled Age Group 1 (N) 合并的年龄组 1(N)	integer	1.0	1= <65, 2= 65-80, 3= >80	Based on ADSL.AGEGR1N, blank if ADSL.AGE is missing 基于 ADSL.AGEGR1，如果 ADSL.AGE 缺失，则为空
ADQSADAS	*ALL*	SEX	Sex 性别	text	$1	M, F	ADSL.SEX
ADQSADAS	*ALL*	SAFFL	Safety Population Flag 安全人群标识	text	$1	Y., N	ADSL.SAFFL
ADQSADAS	*ALL*	ITTFL	Intent-to-Treat Population Flag 意向性治疗人群标识	text	$1	Y, N	ADSL.ITTFL

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意向性治疗记录水平标识	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.ITTFL='Y' then ITTRFL='Y'; ITTRFL blank otherwise如果观察的数据适合分析（即QS.VISITNUM值在3,8,10,12,201中）且如果QS.VISIT = 包含在ADQSADAS.ADY中的访视窗的名称，且如果ADQSADAS.ITTFL='Y'，则ITTRFL='Y'；否则ITTRFL为空。
ADQSADAS	*DEFAULT*	DTYPE	Derivation Type 衍生类型	text	$4		Not applicable, therefore blank 不适用，所以为空。
ADQSADAS	ACTOT11	DTYPE	Derivation Type 衍生类型	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. 如果ADQSADAS.AVAL（因而整条记录）为LOCF算法填补的值，则DTYPE = 'LOCF'，否则 DTYPE 为空。
ADQSADAS	*ALL*	ONTRTFL	On Treatment Record Flag 治疗期记录标识	text	$1	Y	If ADQSADAS.TRTSDT<= ADQSADAS.ADT<= ADQSADAS.TRTEDT then ONTRTFL='Y'. ONTRTFL blank Otherwise 如果ADQSADAS.TRTSDT<=ADQSADAS.ADT<= ADQSADAS.TRTEDT，则ONTRTFL='Y'，否则ONTRTFL为空。
ADQSADAS	*ALL*	TRTSDT	Date of First Exposure to Treatment第一次暴露于治疗的日期	integer	yymmdd10.		ADSL.TRTSDT
ADQSADAS	*ALL*	TRTEDT	Date of Last Exposure to Treatment最后一次暴露于治疗的日期	integer	yymmdd10.		ADSL.TRTEDT
ADQSADAS	*ALL*	VISITDY	Planned Study Day of Visit计划的研究访视日	integer	3.0		QS.VISITDY

Dataset Name	Parameter Identifier	Variable Name	Variable Label	Variable Type	Display Format	Codelist / Controlled Terms	Source / Derivation
ADQSADAS	*ALL*	VISITNUM	Visit Number 访视编号	float	4.1		QS.VISITNUM
ADQSADAS	*ALL*	ADT	Analysis Date 分析日期	integer	yymmdd10.		QS.QSDTC associated with AVAL, converted to SAS date与AVAL关联的QS.QSDTC，转变为SAS日期

5.3 Analysis Results Metadata分析结果元数据

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.
这些元数据提供从用于统计展示的结果到分析数据集中数据的可溯源性。分析结果元数据不是必需的，但最好的实践是提供它们，通过确定关键分析、提供结果、文档和数据集之间的关联和记录进行的分析以给审阅者提供帮助。

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.
分析结果包括统计展示（如结果的文本、表、图展示）或推断说明，比如p值或者治疗效应的估计。分析结果元数据以标准格式和可预见的位置提供了分析结果与用于产生它的数据之间的链接，这样审阅者可从一个分析结果链接到描述分析的重要信息，如进行分析的原因、用于产生分析的数据集和选择标准。

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.
并不是临床研究报告或递交中的全部分析都需要分析结果元数据，甚至也不建议这么做。由申办者决定哪些分析应该具有分析结果元数据。例如，申办者出于营销宣传的考虑，可能选择只对主要疗效分析和次要疗效分析提供分析结果元数据。

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.

分析结果元数据描述临床研究报告或递交中特定的分析结果的主要属性。用于描述分析结果的元数据字段列在表 5.3.1 中。使用单词"展示"而不是"表"，是因为"展示"更通用，指结果的表或图展示。

Table 5.3.1 Analysis Results Metadata Fields分析结果元数据集字段

Analysis Results Metadata Field 分析结果元数据集字段	Description 描述
DISPLAY IDENTIFIER 展示标识符	A unique identifier for the specific analysis display (such as a table or figure number) 对于特定的分析展示的唯一标识符（如表或图的编号）。
DISPLAY NAME 展示名称	Title of display, including additional information if needed to describe and identify the display (e.g., analysis population) 展示的标题，如果需要描述和确认展示（如分析人群），则包括更多信息。
RESULT IDENTIFIER 结果标识符	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. 确定展示内的特定分析结果。例如，如果一个展示有多个P值，分析结果元数据专门指其中的一个，本字段标识关注的P值。当与展示标识符组合时，它提供了一个特定分析结果的唯一标识。

Analysis Results Metadata Field 分析结果元数据集字段	Description 描述
PARAM 参数	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. BDS分析数据集中的分析参数，是分析结果的焦点。如果结果不是基于BDS分析数据集，则不适用。
PARAMCD	Corresponds to PARAM in the BDS analysis dataset. Does not apply if the result is not based on a BDS analysis dataset. 对应于BDS分析数据集中的PARAM变量。如果结果不是基于BDS分析数据集，则不适用。
ANALYSIS VARIABLE 分析变量	The analysis variable being analyzed 用于分析的分析变量
REASON 原因	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." 进行这个分析的基本原理。它指出分析是何时计划的（例如，"方案中提前规定的"、"统计分析计划中提前规定的"、"数据驱使的"、"监管机构要求的"）和证据体中分析的目的（例如，"主要疗效"、"关键次要疗效"、"安全性"）。原因的一个实例是"方案中规定的主要疗效分析"。
DATASET 数据集	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. 用于产生分析结果的数据集的名称。在大多数的情形下，这是一个单独的数据集。但是如果使用多个数据集，则都会列在这里。
SELECTION CRITERIA 选择标准	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 WHEREclause 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. 特定的和充分的分析子集及/或分子的选择标准-用于确定为分析所选记录的变量及其值的完整列表。虽然语法不是ADaM特有的，期望信息可以很容易的包括在一个WHERE从句或类似的东西中，以保证选择适合分析的记录的精确集合。如果分析不包括分析数据集中的全部记录，则本信息是必需的。

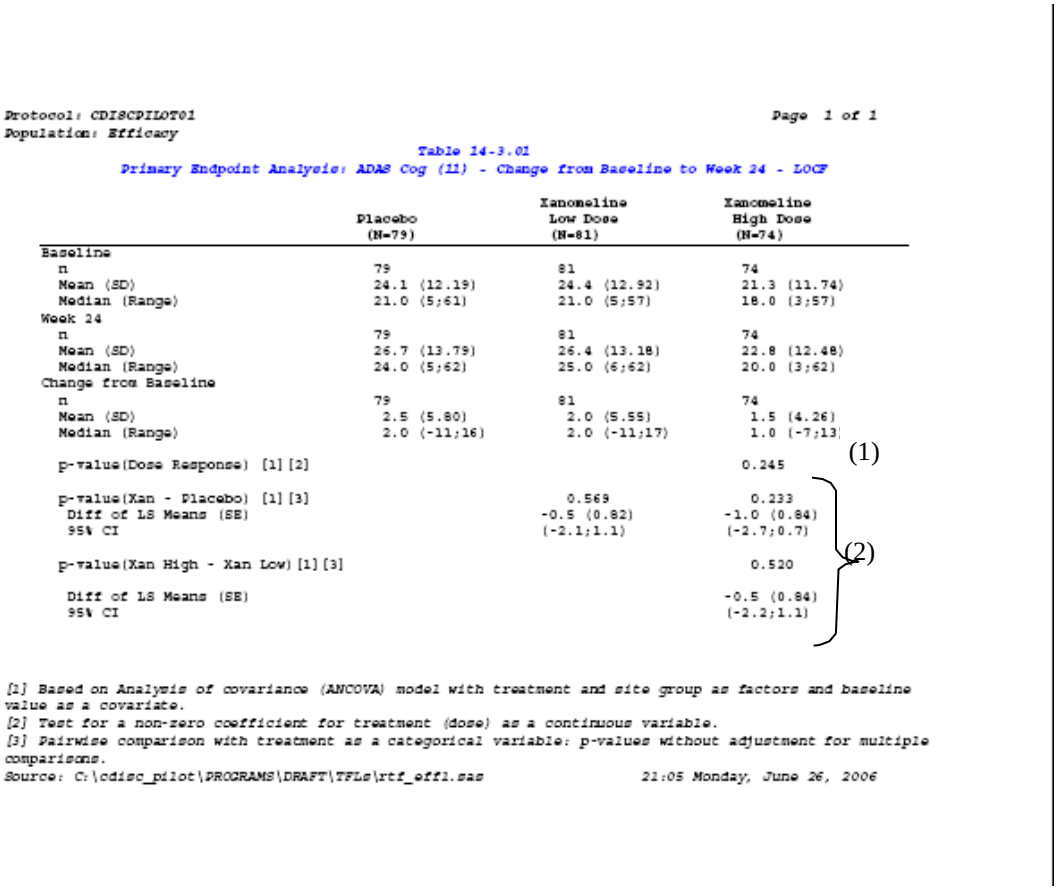
Analysis Results Metadata Field 分析结果元数据集字段	Description 描述
DOCUMENTATION	Textual description of the analysis performed. This information could be a text description, pseudo code, or a link to another document such as the protocol or statistical analysis plan, or a link to an analysis generation program (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. 进行的分析的文字描述。本信息可能是文字描述、伪编码、或到另外文档的链接，如方案或者统计分析计划，或者到分析产生程序的链接（即用于产生分析结果的统计软件程序）。文档元数据元素的内容，取决于用于描述分析本身要求的详细水平，包含申办者是否提供相应的分析产生过程、申办者特有的要求及标准。本文档元数据元素将保持自由形式，意思是它不会遵循严格的结构或者受控术语。
PROGRAMMING STATEMENTS 编程说明	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). 用于进行特定的分析的软件程序代码。这包括，例如模型说明（使用特定的变量名）和再现分析需要的所有的技术说明（例如协方差结构）。名称和适用的软件包的版本应该标明，可以作为此元数据元素的一部分，或者放在另一个文档中，如审阅者指南（关于审阅者指南的更多信息，参见附录B）。

5.3.1 Illustration of Analysis Results Metadata分析结果元数据举例

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.

图5.3.1.1和图5.3.1.2包含5.1.1和5.2.2节中描述的用ADAS-Cog分析数据集进行的分析的数据展示。就像在统计分析计划中描述的，第24周ADAS-Cog(11)总分的主要分析，使用疗效人群，对于24周时的缺失数据使用LOCF方法填补。应用协方差分析(ANCOVA)分析模型，纳入基线分数、中心群和治疗（作为一个连续变量）作为自变量，并且包含一个剂量反应的检验结果。治疗组两两比较使用ANCOVA模型进行，纳入基线分数、中心群和治疗（作为分类变量）作为自变量，提供治疗差异的结果。另外，使用混合效果模型重复测量(MMRM)分析对于ADAS-Cog进行了一次支持性的分析，。在本例中，疗效人群是意向性治疗人群。

Figure 5.3.1.1 Example of Primary Endpoint Analysis Statistical Display[4]
图5.3.1.1 主要终点分析统计展示的举例



Protocol: CDISCPILOT01 Page 1 of 1
Population: Efficacy

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)

(2)

[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.sas 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.
表5.3.1.1和表5.3.1.2阐释了图5.3.1.1中显示的ADAS-Cog的特定元素的分析结果元数据。在图中有下划线的地方最好超链接到临床研究报告中的数据展示、定义文件中其它地方的元数据和SAP中的特定页。

[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.
分析结果的展示类型是由资助者决定的。这个例子只是用来展示内容而不是外观。

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.
表5.3.1.1举例说明了剂量反应分析的分析结果元数据，根据图5.3.1.1（1）确定。它也阐明了所做分析的描述，程序说明中没有提供的模型说明。

Table 5.3.1.1 Analysis Results Metadata for the Dose Response Analysis in the Statistical Display in Figure 5.3.1.1 [5]
图5.3.1.1中统计展示的剂量反应分析的分析结果元数据

Metadata Field 元数据字段	*Definition of field 字段的定义*	Metadata 元数据
DISPLAY IDENTIFIER 展示标识符	*Unique identifier for the specific analysis display 特定分析展示的唯一标识符*	Table 14-3.01
DISPLAY NAME 展示名称	*Title of display 展示标题*	Primary Endpoint Analysis: ADAS Cog (11) - Change from Baseline to Week 24 – LOCF主要终点分析：ADAS Cog (11) – 从基线到第 24 周的变化– LOCF
RESULT IDENTIFIER 结果标识符	*Identifies the specific analysis result within a display 确定展示内的具体分析结果*	Analysis of dose response 剂量反应分析
PARAM	*Parameter 参数*	ADAS-Cog (11) Total Score
PARAMCD	*Parameter code 参数代码*	ACTOT11
ANALYSIS VARIABLE 分析变量	*Analysis variable being analyzed 分析的分析变量*	CHG
REASON 原因	*Rationale for performing this analysis 执行分析的基本原理*	Primary efficacy analysis as pre-specified in protocol 方案中提前规定的主要的疗效分析
DATASET 数据集	*Dataset(s) used in the analysis. 用于分析的数据集*	ADQSADAS
SELECTION CRITERIA 选择标准	*Specific and sufficient selection criteria for analysis subset and / or numerator对于分析子集及/或分子的特定的和充分的选择条件*	ITTFL='Y' and AVISIT='Week 24' and PARAMCD='ACTOT11' IITTFL='Y' and AVISIT='第24周' 和PARAMCD='ACTOT11'
DOCUMENTATION 文档	*Textual description of the analysis performed 执行分析的文本描述*	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). SAP第10.1.1节。第24周ADAS-Cog(11)总分相对基线变化的剂量反应的线性模型分析– 缺失值用LOCF方法填补，疗效人群。使用SAS中的PROC GLM产生P值（来自治疗剂量Type III SS）；模型中自变量是TRTDOSE（随机化剂量：安慰剂是0；54是低剂量；81是高剂量），SITEGR1（中心分组作为类变量）和BASE（基线ADAS-Cog分数）。

Metadata Field 元数据字段	*Definition of field 字段的定义*	Metadata 元数据
PROGRAMMING STATEMENTS 编程说明	*The analysis syntax used to perform the analysis. 用于进行分析的分析语法*	

[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.
表5.3.1.2举例说明了治疗组间二二比较的分析结果元数据，根据图5.3.1.1中的（2）确定。它也阐释了模型说明的纳入条件。

Table 5.3.1.2 Analysis Results Metadata for the Pairwise Treatment Comparisons in the Statistical Display in Figure 5.3.1.1 [6]
图5.3.1.1的统计展示中成对治疗比较的分析结果元数据

Metadata Field 元数据字段	*Definition of field 字段定义*	Metadata 元数据
DISPLAY IDENTIFIER 展示标识符	*Unique identifier for the specific analysis display 特定分析展示的唯一标识符*	Table 14-3.01
DISPLAY NAME 展示名	*Title of display 展示标题*	Primary Endpoint Analysis: ADAS Cog (11) - Change from Baseline to Week 24 - LOCF主要终点分析：ADAS Cog (11) – 第24周相对基线变化– LOCF
RESULT IDENTIFIER 结果标识符	*Identifies the specific analysis result within a display 识别展示中的特定分析结果*	Pairwise treatment comparisons 配对治疗比较
PARAM	*Analysis parameter 分析参数*	ADAS-Cog (11) Total Score
PARAMCD	*Analysis parameter code 分析参数代码*	ACTOT11
ANALYSIS VARIABLE 分析变量	*Analysis variable being analyzed 分析的分析变量*	CHG
REASON 原因	*Rationale for performing this analysis 分析编程的基本原理*	Primary efficacy analysis as pre-specified in protocol 方案中提前规定的主要疗效分析
DATASET 数据集	*Dataset(s) used in the analysis.用于分析的数据集*	ADQSADAS
SELECTION CRITERIA 选择标准	*Specific and sufficient selection criteria for analysis subset and / or numerator分析子集及/或分子的特定的和充分的选择条件*	ITTFL='Y' and AVISIT='Week 24' and PARAMCD='ACTOT11' ITTFL='Y' and AVISIT='第24周' andPARAMCD='ACTOT11'

Metadata Field 元数据字段	*Definition of field 字段定义*	Metadata 元数据
DOCUMENTATION 文档	*Textual description of the analysis performed 执行分析的文字描述*	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 . ADAS-Cog(11)总分线性模型分析：第24周相对基线变化的治疗组配对比较和校正的均值；缺失值填补使用LOCF方法，疗效人群。使用随机化治疗作为分类变量、中心群作为分类变量、基线ADAS-Cog评分纳入模型。
PROGRAMMING STATEMENTS 编程说明	*The analysis syntax used to perform the analysis 用于进行分析的分析语法*	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.
元数据的呈现形式应该在申办方和审阅人员之间确定。这个例子只是用来说明内容而不是外观。

Figure 5.3.1.2 Example of Supportive Analysis Statistical Display[7]
Figure 5.3.1.2 支持性分析统计展示示例

Protocol: CDISCPILOT01 *Page 1 of 1*
Population: Efficacy

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

	Placebo (N=79)	Xanomeline Low Dose (N=81)	Xanomeline 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_mmrm.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.
分析结果的展示类型是由资助者决定的，这次举例只是展示内容而不是外观。

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.
表5.3.1.3举例说明图5.3.1.2中的统计展示的分析结果元数据，阐明了使用元数据来描述一个展示。在示例中有下划线的项，最好超链接到临床研究报告中的数据展示、定义文件中其它地方的元数据和SAP的特定页。

Table 5.3.1.3 Analysis Results Metadata for the Statistical Display in Figure 5.3.1.2 [8]
表5.3.1.3 图5.3.1.2中统计展示的分析结果元数据

Metadata Field 元数据字段	*Definition of field 字段定义*	Metadata 元数据
DISPLAY IDENTIFIER 展示标识符	*Unique identifier for the specific analysis display 特定分析展示的唯一标识符*	Table 14-3.11
DISPLAY NAME 展示名	*Title of display 展示标题*	ADAS Cog (11) - Repeated Measures Analysis of Change from Baseline to Week 24 ADAS Cog (11) – 第24周相对基线变化的重复测量分析
RESULT IDENTIFIER 结果标识符	*Identifies the specific analysis result within a display 识别展示内的特定分析结果*	
PARAM	*Analysis parameter 分析参数*	ADAS-Cog (11) Total Score
PARAMCD	*Analysis parameter code 分析参数代码*	ACTOT11
ANALYSIS VARIABLE 分析变量	*Analysis variable being analyzed 分析的分析变量*	CHG
REASON 原因	*Rationale for performing this analysis 执行分析的基本原理*	Pre-specified in SAP SAP 中提前规定的
DATASET 数据集	*Dataset(s) used in the analysis.用于分析的数据集*	ADQSADAS
SELECTION CRITERIA 选择标准	*Specific and sufficient selection criteria for analysis subset and / or numerator分析子集和/或分子的特定和充分选择条件。*	ITTFL='Y' and AVISITN GT 0 AND DTYPE NE 'LOCF' AND PARAMCD='ACTOT11'
DOCUMENTATION 文档	*Textual description of the analysis performed 进行的分析的文字描述*	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. SAP第10.1.1节：第24周相对基线变化的校正均数和第24周治疗组间配对比较，所用的复测量模型纳入了治疗组（作为分类变量），中心（作为分类变量），时间，治疗*时间的交互作用，基线分数，基线*时间的交互作用，使用的是非结构化协方差矩阵。疗效数据，观察到的病例数据。

Metadata Field 元数据字段	*Definition of field 字段定义*	Metadata 元数据
PROGRAMMING STATEMENTS 编程说明	*The analysis syntax used to perform the analysis 用于进行分析的分析语*	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.
元数据的呈现形式应该在申办方和审阅人员之间确定。这个例子只是用来说明内容而不是外观。

6Subject-Level Analysis Dataset

6 受试者水平分析数据集

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.
无论临床实验设计是何种类型，受试者水平分析数据集 （ADSL）的结构是每个受试者一条记录。ADSL用于提供描述受试者属性的变量。这种结构允许其与其他任何数据集（包括SDTM和分析数据集）进行简单的合并。

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.
监管机构的工作人员表示，ADSL对临床试验的审查非常有帮助。任何基于CDISC的临床试验数据提交都需要ADSL及其相关元数据，即使没有提交其他分析数据集。

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旨在提供有关受试者的描述性信息。它可以用于多种类型的分析，包括描述性分析、分类分析和建模。不应为了试图减少分析数据集的数量而强迫ADSL支持所有分析。虽然提取研究中的每一个数据值并将其作为受试者水平级数据集(如ADSL)中的变量，在技术上是可行的，但这并不是ADSL的意图和目的。关键终点和研究中随时间变化的数据的正确位置包含在BDS数据集中。

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.
ADSL是用于其他分析数据集的受试者水平变量的主要来源，例如人群标记和治疗变量。当合并ADSL到其他分析数据集时，只应包含那些与该分析数据集相关的领域。包含过多无关变量（例如不需要支持分析的变量）会使得用户更难以找到重要变量，并且妨碍清晰简洁的交流。

Table 6.1 provides an example of analysis dataset metadata for ADSL.
表6.1提供了ADSL分析数据集元数据的示例

Table 6.1 Example of Analysis Dataset Metadata for ADSL[9]

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

表6.1 ADSL分析数据集元数据示例

数据集名称	数据集描述	数据集位置	数据集结构	数据集核心变量	数据集类别	*文件记录*
ADSL	受试者特征，人口统计学和基线特征	adsl.xpt	每个受试者一条记录	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):
ADSL中包含的最小变量合集取决于疾病的具体特性和方案，(请参阅ICH E3[8]以获得更详细的清单，并参阅ADaMIG以获得纳入所需变量的进一步描述)。ADSL信息的例子包括（但不仅限于）：

- 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
- 人口统计变量 （例如，年龄，性别，种族和其他相关因素）
- 疾病因素（例如，疾病发病，疾病严重程度）
- 治疗编码/分组
- 其他可能影响治疗反应的预后因素（例如，抽烟，酒精摄入，女性月经状态）
- 重要事件日期（例如，治疗开始和结束时间）
- 研究人群

[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.
元数据的呈现形式应该在申办方和审阅人员之间确定。这个例子只是为了说明内容而不是外观。

ADSL contains variables that describe the subjects in a clinical trial prior to treatment, or group the subjects in some way for analysis purposes.
ADSL包含了描述受试者在临床试验中治疗前的变量，或以某种方式对受试者进行分组以进行分析。

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.
总的来说，ADSL中的变量包括那些描述性的、被认为是重要的基线特征、用作随机化分层、用于将受试者识别为属于特定亚组（例如，种群标志）或用于识别重要事件何时或是否发生的变量（例如，最后给药日期，死亡，中止）。例如，在年龄组内进行的分层随机化中，受试者的年龄类别是研究的重要受试者描述变量，并且包括在ADSL中。

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.
ICH指南（ICH E3, 第11.2节）[8]建议"除了表格和图表给出基线变量的分组数据外，相关的个体受试者人口统计学和基线数据……所有随机化的个体受试者（按照治疗和中心分类或多中心研究）应该在个体水平表格列表中列出。"FDA审查员和申办方通常同意受试者水平数据递交符合此要求。如果是这种情况，ADSL应包括满足此监管规定所需的那些变量。

6.1 Data for Subjects Not Analyzed

6.1 未分析的受试者数据

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.
分析数据集是否包含未分析的受试者数据(如筛选失败)是主办方的决策，应与数据的审阅者或用户进行沟通。如果包含这些数据，则应将它们合并到适当的数据集中，例如ADSL（而不是针对未分析的受试者建立单独的数据集），并且使用适当的标记变量来明确区分这些记录。元数据必须指明包含了这些数据以及怎样区分它们。

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.

附录 B 定义

ADaM – CDISC 分析数据模型

ADaM 基本数据结构 (BDS) –一种便于分析和审阅的数据集结构，由每个受试者的每个分析参数的每个分析时点的一条或多条记录构成。分析时点是根据分析来选择是否需要的。ADaMIG 里的 BDS 能支持大多数的分析。

ADaM 实施指南(ADaMIG) –一份详细说明 ADaM 标准数据集结构和各个变量的文档，包括命名规则。该指南也详细说明了实施问题的标准解决方案。ADaM 文档和 ADaMIG 应配合使用。

分析数据集-用于统计分析和报告的数据集。

分析数据集创建程序-用于创建分析数据集的计算机指令。

分析数据集元数据-描述分析数据集的结构、内容和衍生的信息。

分析生成程序-用于生成分析结果（例如，总结或推断统计的图表结果）的计算机指令

分析参数(PARAM)-用于唯一地区分一组共用通用定义的值的行标识符。例如：主要疗效分析参数为"3 分钟的坐位收缩压(mmHg)"。要注意的是 ADaM 分析参数包含需要唯一定义一组相关分析值的所有信息。而在 SDTM 中，为了定义一组相关值，—TEST 列可能需要和限定语列结合，比如—POS,--LOC,--SPEC 等。本文档中"参数"就是指的"分析参数"。

分析参数值水平元数据-描述一个或一组给定的分析参数内的分析值的信息

分析结果元数据-描述一个特定的包含在临床研究报告或递交物的分析结果的信息。

分析时点-用于把分析参数内的值分类为用于分析的时间组或概念组的行标识符。这些分组可能是观测的、计划的或者衍生的。例如：主要疗效分析是在第 2 周、第 6 周和终点分析时点进行的。

分析值-(1)分析参数描述的字符(AVALC)或数值(AVAL)。分析值可能在输入数据中呈现，输入数据值的分类或者衍生出来的。例如：参数“平均心率”的分析值是从每个访视测量的三个心率值的均值衍生来的。(2)此外，特定含义的值被视为分析值。例如：基线值(BASE)、相对基线的变化(CHG)

分析变量元数据-描述分析数据集中变量的信息。
定义文件-如在病例报告表数据定义规范中所述，1999 年 FDA 电子递交(eSub)指南和电子通用技术文档(eCTD)说明了这是一份描述数据内容和结构的文档，需在递交文件中提供。此文档就是数据定义文档（比如，1999 年指南中的“define.pdf”）。数据定义文档包括递交的数据集的清单以及每个数据集（比如元数据）内容的详细描述。为了提高自动化水平和监管审核过程的有效性，define.xml 可被以机器可读的形式提供数据定义文档。正式名即为病例报告表数据定义(CRT DD)规范。SDTM 和 ADaM 数据集有各自的定义文件。

元数据-关于数据的信息或数据。

记录-数据集中的一行。

审阅者指南-一份可包含在递交文件中的文档，便于审阅者找到递交包中的各个方面。根据参与项目的 FDA 审阅者的建议，审阅者指南包含在 CDISC SDTM/ADaM 试点项目递交包中。（项目报告可在 www.cdisc.org 找到。）此文档有助于提供因太复杂或太冗长而不容易在其他元数据来源中描述的信息。可以描述在变量水平难以沟通的或者适用于多个分析数据集的问题，例如：申办方命名规则、缺失日期的填补规则。不应重复在其他元数据来源中可找到的大量信息。

SDTM-原始数据标准模型-一份由 CDISC 递交数据标准(SDS)团队编写的文档，描述了用于展现递交给法规机构的临床研究数据的一般概念模型。SDTM 提供了一个基本框架，用于描述为临床试验收集的和递交给法规机构的信息组织情况[5]。

SDTM 实施指南(SDTMIG)-一份由 CDISC SDS 团队编写的文档，为了指导递交给法规机构比如 FDA 的标准临床试验制表数据集的组织、结构以及格式。该文档提供了具体的域模型、假设、商业规则和用于准备基于 SDTM 的标准制表数据集的例子[4]。

可溯源性-ADaM 的性质允许分析数据集的用户去理解数据沿袭和/或一个元素和它的前身之间的关系。可溯源性促进了透明度，这对于建立对结果或结论的信心是一个重要的组成部分。根本上来说，ADaM 的溯源性便于理解分析结果、分析数据集和 SDTM 域之间的关系。可溯源性是通过清晰地建立一个元素和它的直接前身之间的路径来建立的。完整的追踪路径是从一个元素到它的前身，然后到它们的前身，以此类推，返回 SDTM 域，最终到达数据收集工具。要注意的是，CDISC 临床数据采集标准 (CDASH)标准是和 SDTM 一致的，因此该标准坚持确保端对端的可溯源性。例如：基于元数据和分析数据集的内容，审阅者可以追溯每个受试者的主要和次要疗效分析是如何从 SDTM 数据衍生来的。

变量-数据集中的一列。

Appendix C Abbreviations and Acronyms
附录 C 缩略语表

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.
下表是在本文档中多次用到的缩略语。不包括各个SDTM域的解释（如：QS,DM）。也不包括引用变量的描述。

Table C.1 Abbreviations and Acronyms Used in the Document
表C.1 本文档中用到的缩略语

缩略语	定义（英文）	定义（英文）
ADAE	ADaM Adverse Event Analysis Dataset	ADaM 不良事件分析数据集
ADaM	CDISC Analysis Data Model	CDISC 分析数据模型
ADaMIG	Analysis Data Model Implementation Guide	分析数据模型实施指南
ADAS-Cog	Alzheimer's Disease Assessment Scale - Cognitive	阿尔茨海默病评定量表-认知亚
ADSL	ADaM Subject-Level Analysis Dataset	ADaM 受试者水平分析数据集
ANCOVA	Analysis of Covariance	协方差分析
BDS	ADaM Basic Data Structure	ADaM 基本数据结构
BMI	Body Mass Index	身体质量指数
CDASH	Clinical Data Acquisition Standards Harmonization	临床数据采集标准协调
CDISC	Clinical Data Interchange Standards Consortium	临床数据交换标准协会
CFR	Code of Federal Regulations	美国联邦法规
CRT	Case Report Tabulation	病例报告表
CRT-DDS	Case Report Tabulation Data Definition Specification	病例报告表数据定义规范
eCTD	electronic Common Technical Document	电子通用技术文档
FDA	United States Food and Drug Administration	美国食品和药物管理局
ICH	International Conference on Harmonisation	国际协调会议
ITT	Intent-to-Treat	意向治疗
LOCF	Last Observation Carried Forward	末次观测结转法
MMRM	Mixed Effects Models Repeated Measures	重复测量混合效应模型
PDF	Portable Document Format	可移植文档格式
SAP	Statistical Analysis Plan	统计分析计划
SDS	Submission Data Standards	递交数据标准
SDTM	Study Data Tabulation Model	原始数据标准模型
SDTMIG	Study Data Tabulation Model Implementation Guide	原始数据标准模型实施指南
XML	Extensible Markup Language	可扩展标记语言
XPT	Filename extension for a SAS transport file	SAS 传输文件的文件扩展名

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' ANDPARAMCD='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;
```

附录 D 分析就绪的的例证

为了阐释"分析就绪的"的概念，再回顾下第 5 部分中的例子。图 5.3.1.1 展示了剂量反应分析和成对剂量比较。准备就绪的分析并不是指这个格式的表格可由单个统计程序生成，而是指表中的每个统计量可用对应的分析数据集作为输入，由一个标准的统计程序（如 SAS PROC、S-PLUS 函数等）被重现。这意味着审阅者能以最小的编程工作来重现和探究这些结果，使得审阅者把注意力集中在结果上而不是编程上。

举个例子，下面的 SAS 代码用包含对应变量的数据集重现了表格 14-3.01（在图 5.3.1.1 中）中的剂量反应分析结果。其中 where 子句为该分析选择了对应的记录。

```
*** 剂量反应分析 ***;
    PROC GLM DATA=A.ADQSADAS(WHERE=(ITTFL='Y' AND AVISIT='Week 24' AND PARAMCD='ACTOT11')); CLASS
    SITEGR1;
    MODEL CHG = TRTDOSE SITEGR1 BASE;
    RUN;
```

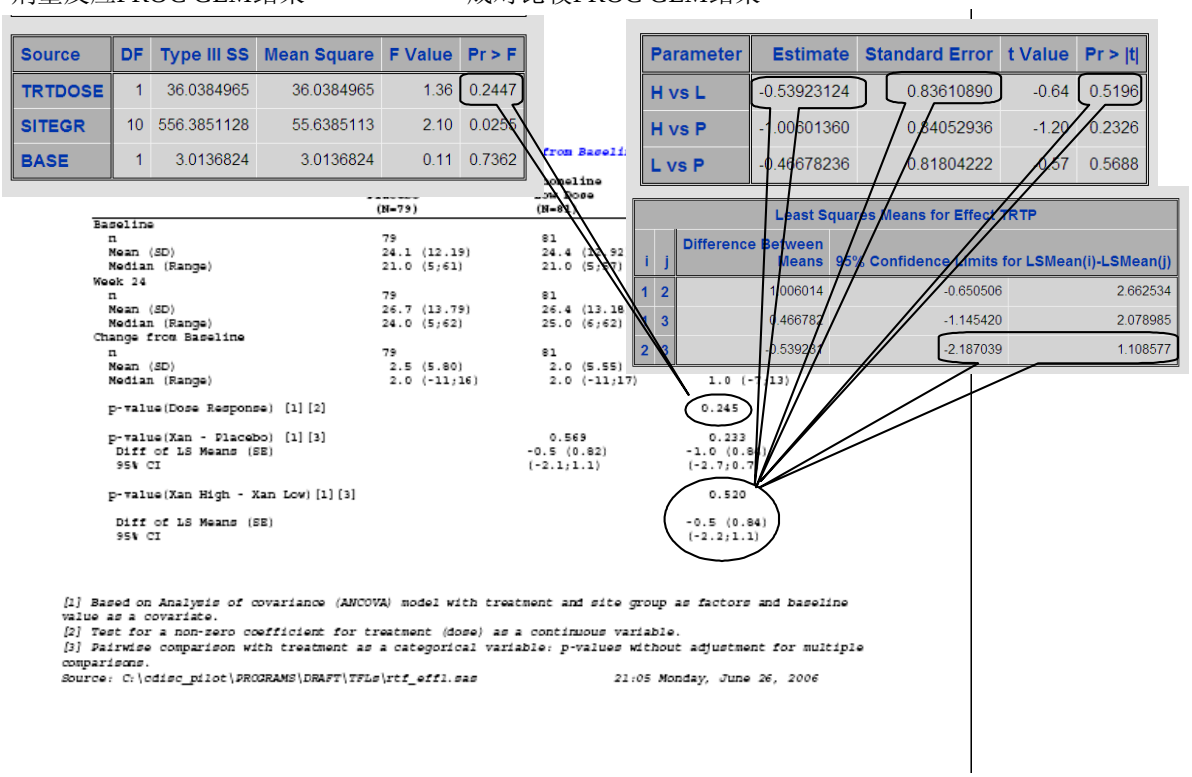
类似的，以下是重现成对剂量比较结果的 SAS 代码。

```
*** 成对剂量比较分析 ***;
    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).
图D.1 阐释以上SAS代码的结果和图5.3.1.1中的表格14-3.01结果中的相应元素的关系。

Result of Dose Response PROC GLM:
剂量反应PROC GLM结果

Results of Pairwise Comparison PROC GLM:
成对比较PROC GLM结果



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	0.0256
BASE	1	3.0136824	3.0136824	0.11	0.7362

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

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	-0.539232	-2.187039	1.108577

Figure D.1 Illustration of SAS Output vs Results Display from Analysis-Ready Dataset[10]
图D.1 SAS分析的结果与"分析就绪的"数据集所产生的结果的联系

[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.

附录 E 复合终点例子

如 4.1.1 节中提到的分析的例子只需要 ADSL，但 SDTM 不会涵盖所有的分析数据，特别是需要考虑疗效分析的时候。甚至全面的安全性分析通常会展现更复杂的例子，但只基于 ADSL 和 SDTM 的分析数据是不能涵盖的。这个例子阐释了一个非常简单的二分类结果变量（一次头痛的治疗结果）具有复杂的生成方法，并需要从不同的源数据集中的数据生成。它描述了一个复合终点，需要来自疗效数据集(不同时间点的头痛严重程度)的数据，以及来自不良事件和伴随药物治疗数据集的数据。此终点是"持续偏头痛疼痛和无症状"。

终点（基于国际头痛协会指南的持续的偏头痛和症状消失）定义如下：

1. 基线时头痛严重程度为中等或严重并且
2. 服药后（即首次服用试验药物后）两小时头痛严重程度为不痛并且
3. 服药后 48 小时内未有头痛复发并且
4. 从首次服药到服药后 48 小时中间未使用止痛或止吐的急救用药并且
5. 从服药后 2 到 28 小时内没有伴随症状（恶心、呕吐、恐光症、恐音症）。

上面描述中的并且表示一个受试者必须发生了所有的情况才被认为是有反应。
举个例子，以下定义和说明适用于：

头痛严重程度

头痛严重程度是由受试者在预先指定的时间点（基线、服药后 0.5h、服药后 1h、服药后 1.5h、服药后 2h、服药后 3h 和服药后 4h）以 0 级（不痛）到 3 级（严重）自主评估。

伴随症状

以下伴随症状在特定时间点（基线、服药后 2h 和服药后 4h）出现或消失的受试者记录：恐光症、恐音症、恶心、呕吐。

此外，受试者需要把以上任何一个症状在日记卡上列为不良症状，如果该症状:（1）在他们服用试验药物后有一个异常的强度增加或者，（2）在他们服用试验药物后，和之前通常的偏头痛症状相比有显著的变化。研究者将会把这些症状记为不良经历。因此一项完整的伴随症状消失的评估会包括不良事件数据集的扫描。

头痛复发

头痛复发是指受试者在服药后 2 小时疼痛缓解（轻度或不痛）而在服药后 48 小时内又变为 2 级或 3 级（中等或严重）的头痛严重程度。受试者记录的是首次服药后 2 到 24 小时和 24 到 28 小时的最大头痛程度。

急救用药

受试者记录的任何在试验剂量、记录日期、时点（上午/下午）药名（如可待因）、药片/胶囊的数量和每片药片/每粒胶囊的剂量之外服用的止痛/止吐药。急救用药也指服药后 48 小时内服用额外剂量的试验药物。急救用药的使用是由合并用药和暴露数据集确定的。

为了确定一个受试者是否符合持续偏头痛和症状消失的标准，五条标准的答案必须都确定。头痛严重程度和伴随症状的数据（一个或多个 SDTM 域）、AE 域、CM 域、EX 域和 ADSL 都需要被导入作为终点的来源

表格 E.1 中的 BDS 分析数据集，假定标准内所有问题的答案都保留在分析数据集中。例子中只列了一些分析数据集变量（PARAMCD,AVAL 和 AVALC），是为了阐释复杂度而不是一个完整的分析数据集。此外，不需要为这个例子去描述具体的 SDTM 域和变量，只需要在来源/衍生的地方写一段简单的文字描述就可以了。在“真实的”元数据中，元数据元素应该指出具体的域和变量，并且包括怎样确定域中的记录是数据的来源（比如当 QSCAT=xxx 是适用于这个 USUBJID 的）。

这个例子说明了来源/衍生可能会比较长和复杂。对于复杂来源的变量，来源/衍生的地方可以提供一个解释各种数据来源和创建这个变量算法的外部文件链接。

Table E.1 Composite Endpoint Example - Illustration of Analysis Variable Metadata for Selected Variables Example[11]

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.

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.

表E.1 复合终点例子-选择变量的分析变量元数据的举例说明

数据集名称	参数标识符	变量名	变量标签	变量类型	显示格式	编码列表/受控术语	来源/衍生
ADSYMFR	PARAMCD	PARAMCD	参数代码	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	分析值	integer	1.0	0=N 1=Y	来源于 ADSYMFR.AVALC, 当 ADSYMFR.AVALC 缺失时为空
ADSYMFR	HASEVBL	AVAL	分析值	integer	1.0	0=无疼痛 1=轻度疼痛 2=中度疼痛 3=重度疼痛	基线的头痛程度来自日记卡数据：给药时记录的头痛严重程度。若缺失则为空。
ADSYMFR	HASEV2	AVAL	分析值	integer	1.0	0=无疼痛 1=轻度疼痛 2=中度疼痛 3=重度疼痛	服药后 2 小时的头痛严重程度来自日记卡数据：记录的服药后两小时的头痛程度，若缺失则为空。
ADSYMFR	*DEFAULT*	AVALC	分析值(C)	text	$1		空
ADSYMFR	HASPNFR	AVALC	分析值(C)	text	$1	N=否 Y=是	服药后 2-48 小时的持续头痛或症状消失是基于这个分析数据集中的其他终点。 'Y'如果 (ADSYMFR.HASEVBL=2 或 3) 且 ADSYMFR.HASEV2=0 且 ADSYMFR.HARECUR=N 且 ADSYMFR.HARESCUE=N 且 (ADSYMFR.HASYMPD=N 且 ADSYMFR.HASYMPAE=N); 否则为'N'

数据集名称	参数标识符	变量名	变量标签	变量类型	显示格式	编码列表/受控术语	来源/衍生
ADSYMFR	HARECUR	AVALC	分析值(C)	text	$1	N=无头痛复发 Y=头痛复发	服药后 48 小时内的头痛复发是基于日记卡数据。 ‘N’，如果受试者记录的他们在首次服药后 2-24 小时和 24-48 小时的最大头痛程度为 0（无疼痛）；’Y’，如果最大头痛程度> 0。 若缺失则为空。
ADSYMFR	HARESCUE	AVALC	分析值(C)	text	$1	N=未服用急救用药 Y=有服用急救用药	从首次服药到服药后 48 小时内服用的急救用药是基于 CM 域、ADSL 和 EX 域。 ’N’，如果从首次服药到服药后 48 小时内（CM 域）未服用止痛或止吐药并且从首次服药到服药后 48 小时内（ADSL 和 EX 域）没有服用额外剂量的研究药物；否则为’Y’。
ADSYMFR	HASYMPD	AVALC	分析值(C)	text	$1	N=无伴随症状 Y=有伴随症状	日记卡上服药后 2-48 小时发生的伴随症状是基于服药后 2 和 4 小时的恐光症、恐音症、恶心或呕吐的出现/消失。 ’N’，如果服药后 2 或 4 小时未产生恐光症、恐音症、恶心或呕吐症状；’Y’，如果服药后 2 或 4 小时产生了任何一种症状。
ADSYMFR	HASYMPAE	AVALC	分析值(C)	text	$1	N=无伴随症状 Y=有伴随症状	AE 数据集中的服药后 2-48 小时发生的伴随症状是基于这些症状在 AE 域中能否找到。 ‘N’，如果服药后 2-48 小时之间未产生恐光症、恐音症、恶心或呕吐被记为不良事件的情况；’Y’，如有有任何症状被记为 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