



# 分析数据模型实施指南（ADaMIG）

## CDISC 分析数据模型团队 （ADaM Team）开发

### 中文版读者说明

本文档为分析数据模型实施指南（ADaMIG）1.0 版，它对应于分析数据模型（ADaM）2.1 版。本中文版初稿由 Absolute Clinical Data Systems Co., Ltd 提供，C-STAR ADaMIG 团队负责修改、并校对和审订。翻译和审核流程详见第 2 页。

### 中文版修订历史

日期	版本	修订小结
2015年2月1日	征求意见稿	在CDISC中国官网发布，意见征集截止时间为2015年5月31日
2013年12月2日	终版	发布的版本，反映了在评论时期确定的所有改动和更正。
2013年5月5日	草案	待评论草案。

### 英文版修订历史

修订日期	版本	修订小结
2009年12月17日	1.0终版	最终发布版本，综合了所有读者意见与建议。
2008年5月30日	1.0草案	征求意见稿。

注：有关陈述和保证、责任限制和免责声明，参见\_附录 B\_陈述和保证、责任限制和免责声明。

## CDISC 分析数据模型 实施指南 (ADaMIG) V1.0 翻译和审核 流程及致谢

### 翻译

由 Absolute Clinical Data Systems Co., Ltd (AbsCD) 公司提供 V1.0 翻译初稿。

### 审核

中国 CDISC 协调委员会 (C3C) 于 2011 年 5 月成立 CDISC 数据标准翻译和审阅 (C-STAR) 项目组, 负责审阅和校对 AbsCD 提供的初稿, 本中文版基于英文 V1.0 正式版, 该版本于 2009-12-17 正式发布。

校对审核遵循 C-STAR 项目组批准的流程。尽管所有校对人员都是有经验的志愿者, 但时间有限。ADaMIG 按照内容分成 3 大部分, 分别由 3 个小组来承担, 每个小组内任何部分内容均由两个独立的翻译校对。小组负责人、ADaMIG 审阅团队负责人和 C-STAR 项目组负责人分别依次进行核对和讨论决定。具体流程如下:

1. 第一位翻译校对员独立审阅, 并用 MS WORD Track Changes 记录
2. 第二位翻译校对员独立审阅, 也用 MS WORD Track Changes 记录
3. 其中一位翻译校对员负责合并, 校对并与另外一位讨论, 直至双方意见一致
4. 如果两个校对员意见不一致, 递交给 ADaMIG 审阅团队负责人讨论判定; 如果还不能定夺, 交由 C-STAR 项目组负责人会同 CDISC 专家评定
5. 所有各部分校对完成, 再进行合并后的全文一致性 QC 及和与 CDISC 其它标准 (SDTM, ADaM) 的术语一致性 QC 检查
6. C-STAR 负责人进行最后审阅

### 翻译征求意见稿

经过 C-STAR 所有流程审阅后的终稿即发布在 CDISC 网站征求意见 (即本文稿)。ADaMIG 审阅团队将根据收集到意见会同 AbsCD 进行中文修改。最终定稿将发布在 CDISC 网站。

由于这是第一次尝试对 ADaMIG 进行中文翻译和核对, 不足之处在所难免。请将您的反馈意见经由 CDISC 发给我们, 以便在终稿时更正。

### 致谢!

没有 AbsCD 和全体 ADaMIG 翻译审阅团队的努力, ADaMIG V1.0 中文稿的面世是难以想象的。核对的工作有时是很枯燥乏味的。对一些不易理解的句子, 难以翻译的术语和多个不同中文术语同时存在的, 翻译校对员需要经过多次反复推敲才能定夺。谢谢所有志愿者的奉献和 (审核时) 所属公司的支持!

C-STAR ADaMIG 翻译审阅组

## 目录

68		
69	1 介绍.....	4
70	1.1 目的 .....	4
71	1.2 背景 .....	4
72	1.3 ADaMIG 的内容概括 .....	4
73	1.4 本文档的组织 .....	5
74	1.5 定义 .....	5
75	1.5.1 ADaM 一般定义.....	5
76	1.5.2 基本数据结构的定义.....	5
77	2 ADaM 标准的基础.....	7
78	2.1 基本原则 .....	7
79	2.2 可溯源性 .....	7
80	2.3 ADaM 数据结构.....	8
81	2.3.1 ADaM 受试者水平分析数据集 ADSL .....	8
82	2.3.2 ADaM 基本数据结构(BDS).....	8
83	3 标准 ADaM 变量.....	9
84	3.1 ADSL 变量 .....	11
85	3.2 ADaM 基本数据结构 (BDS) 变量.....	16
86	3.2.1 BDS 数据集的受试者标识变量 .....	16
87	3.2.2 BDS 数据集的治疗变量 .....	17
88	3.2.3 BDS 数据集集中的时间变量 .....	18
89	3.2.4 BDS 数据集的分析参数变量.....	23
90	3.2.5 BDS 数据集的分析描述符变量 .....	26
91	3.2.6 BDS 数据集的标识变量 .....	28
92	3.2.7 SDTM 及 ADaM 人群和基线标志之间的区别 .....	32
93	3.2.8 其它变量 .....	32
94	4 执行过程中的问题与解决方案及示例 .....	34
95	4.1 常用试验设计治疗变量的例子 .....	34
96	4.2 衍生列 vs 衍生行的创建 .....	36
97	4.2.1 创建行和列的规则 .....	36
98	4.3 分析包含一个参数所有观察的和衍生记录 vs 记录的子集 .....	50
99	4.3.1 ADaM 方法和实例 .....	50
100	4.4 在分析数据集中包含不用于分析但与衍生数据有关的输入数据 .....	52
101	4.4.1 ADaM 方法和实例 .....	52
102	4.5 分析所用行的标识 .....	54
103	4.5.1 用于一个时间点填补分析的行的识别 .....	54
104	4.5.2 标识基线行 .....	56
105	4.5.3 基线后概念性时间点行的标识 .....	57
106	4.5.4 标识用于分析的行 - 一般情况 .....	58
107	4.6 特定人群分析行的标识 .....	60
108	4.6.1 ADaM 方法和实例 .....	60
109	4.7 标识符合预定分析标准的行 .....	61
110	4.7.1 ADaM 方法和实例 .....	61
111	4.8 其它要考虑的问题 .....	63
112	4.8.1 添加记录, 为每个受试者创建所有分析时间点.....	63
113	4.8.2 创建多个数据集, 支持相同类型数据的分析 .....	63
114	附录.....	64
115	附录 A 缩略词.....	64
116	附录 B 免责声明 .....	65
117		

# 1 介绍

## 1.1 目的

CDISC分析数据模型实施指南（Analysis Data Model Implementation Guide, 以下简称ADaMIG, 版本1.0）由CDISC分析数据模型小组负责起草。ADaMIG具体阐述了分析数据模型（Analysis Data Model 简称 ADaM）标准数据集应具有的结构与变量, 包括命名惯例与规则等; 同时也具体指出了在执行过程中有关问题的标准解决办法。

ADaMIG必须与相对应版本的CDISC ADaM文档结合使用, ADaM文档可以在CDISC官方网址<http://www.cdisc.org/adam>进行下载。ADaM文档解释了分析数据模型的目的, 并描述了应用于所有分析数据集的基本原则, 其主要目的旨在促使分析数据集及相关元数据的设计能够清楚明白的表达所提交分析数据集的内容、数据来源及其目的等。分析数据模型能帮助更有效的产生, 重现及评审所分析或提交的结果或数据。

## 1.2 背景

考虑到CDISC研究数据表格模型（Study Data Tabulation Model 简称 SDTM）是ADaM数据的来源, ADaMIG的用户必须熟悉SDTM及其实施指南（Study Data Tabulation Model Implementation Guide 简称 SDTMIG）, 二者都可以在CDISC官方网址<http://www.cdisc.org/sdtm>进行下载。

SDTM和ADaM标准的设计宗旨是为了帮助申办者向监管机构, 例如美国食品药品监督管理局(FDA)进行数据提交。起初, CDISC ADaM小组得到了来自FDA统计学家和临床评审者们的鼓励 and 影响, 他们作为观察员参加ADaM的会议, 还参与了CDISC-FDA最初的试行项目。创建ADaM的基本原则基于监管机构对数据透明度并能满足其科学有据的评审需求。ADaM标准的开发不仅方便向FDA监管提交数据, 同时亦是满足制药行业自身发展的需求, 因其可广泛的应用于药品研发活动。它提供了一个在申办者与合同研究组织（CROs）、研究伙伴及独立数据监测委员会之间进行数据交换的标准。由于ADaM模型被业界越来越广泛的采纳和应用, 不同申办者运用相同的分析数据模型和元数据, 使未来获得相关的使用授权许可与兼并等变得更加容易。

## 1.3 ADaMIG 的内容概括

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

ADSL数据集中, 每个受试者只允许拥有一条记录, 包含受试者水平人群标志、试验每个阶段计划/实际的治疗信息、人口特征信息、分层和亚群信息及重要的日期等。ADSL除包含有必要的常规变量外(如在本文档中后面所指出的), 还包括试验中其他用来描述受试者试验信息的重要变量。基于CDISC的临床试验数据申报中, 即使没有其他可用分析数据提交, ADSL和它相关的元数据也要求必须提供。

BDS数据集中, 每个受试者的每个分析参数, 在每个分析时间点可包含一条或多条记录。分析时间点的有无取决于分析, 并不是必须的。在没有分析时间点的情况下, 每个受试者的每个分析参数可包含一条或多条记录。这种结构包含了一系列代表要被分析的实际数据的变量集合。BDS数据结构支持参数和非参数分析, 比如方差分析(ANOVA)、协方差分析(ANCOVA)、类别分析、Logistic回归、CMH检验、Wilcoxon秩和检验、时间事件分析等。

BDS 数据结构虽能支持绝大多数的统计分析, 但并不能涵盖所有的统计分析。例如, 它不支持多个因变量(响应/结果)的多组分分析或者对一系列因变量的相关性分析。BDS 数据结构设计亦不支持对不良事件发生率或者其它类似发生事件数据的分析。

当前版本的实施指南亦不完全适用于剂量递增试验分析或者是多个研究的整合分析等。

## 未来发展

ADaM 小组正在准备起草的其它分析数据结构及说明文档包括：

- 一个用来支持不良事件发生率分析的 **ADAE** 数据集说明文档。**ADAE** 可以说是第一个更通用的用来支持事件发生率数据分析的数据结构，比如对合并用药、病史等的分析。
- 一个在时间事件分析中应用 **BDS** 数据结构的详细说明文档和示例。
- 一个包含应用 **BDS** 数据和元数据进行分析的示例文档，比如协方差分析。
- 一个提供 **ADaM** 元数据模型及其执行的的详细说明文档。
- 一个定义 **ADaM** 依从性的文档。

预期大多数、亦或所有上述文档将最终整合到未来发布的 **ADaM** 文档及其实施指南，多个研究的整合分析亦会在将来发布的文档中加以解决。

## 1.4 本文档的组织

本文档包括如下章节：

- 第1节提供了对**ADaM**标准的重要性的一个全面介绍以及它们如何关联到其它**CDISC**数据标准。
- 第2节介绍了应用**ADaM**数据集的基本原则，并且介绍了两个足够灵活而且能够应用于大多数数据分析的标准数据结构。同时定义了分析变量的类别及那些对分析数据集用户来说重要的标准。
- 第3节定义了通常用于**ADaM**标准数据结构分析变量的标准变量。
- 第4节通过示例的方式，提供了对**BDS**执行中问题的标准解决办法。

## 1.5 定义

### 1.5.1 ADaM 一般定义

**即可分析性** – 实施分析所必需具有的列或行数据。例如：高血压类型列被加入到分析数据集是为了能够实施亚组分析，添加协变量年龄列是为了能够实施对年龄调整的分析，而在多中心研究中通常以试验中心进行分层分析等。

**可溯源性** – 一种可以帮助理解数据体系及/或一个元素和它的前身之间关系的特性。可溯源性提高了数据的透明度，这是构成分析结果或结论可信度的一个非常重要的部分。**ADaM**中的可溯源性从根本上可以帮助对于分析结果、分析数据集、和**SDTM**域之间关系的理解。可溯源性通过建立一个元素和它最近前身之间清晰的路径而获得，完全的路径可以通过从一个元素到它的前身然后再到前身的前身等等来溯源，可以溯源到**SDTM**域，最终到达源数据采集端。同时应该注意到**CDISC**临床数据采集协调标准（**CDASH**）与**SDTM**应相互协调，这样有助于保证数据点到点的可溯源性。

**可支持性** – 使可溯源性可行。当一列或一行在分析中并不必须，但为了加强可溯源性而被包含进来的时候，这一列或一行可以被看做支持性的数据。例如，从**SDTM**转入的**LBSEQ**和**VISIT**列是为了促进理解分析数据集是如何与研究表格数据集相联系的。

**记录** – 数据集中的一行。

**变量** – 数据集中的一列。

### 1.5.2 基本数据结构的定义

**分析参数** – 一个行标识符，用来唯一的描绘一组有相同定义的数值。注意，**ADaM**分析参数是包含能唯一确定一组分析相关的值所需要的信息。而与此相对应，在**SDTM**中，列 **-TEST**可能需要和其它限定修饰语列，如**-POS**，**-LOC**，**--SPEC**等相结合，才来确定一组相关的数值。例如，主要的疗效分析参数为“3分钟坐姿收缩血压（毫米汞柱）”。在本文档中，“参数”这个词用于指代“分析参数”。

**分析时间点** – 一个行标识符，用于把一个分析参数的值细分到具体的分析时间或概念组。这些组别可能是实际观测到的、预先计划的、或后期推导的。例如：主要的疗效分析在第2周、第6周和终点分析时间进行。

217 **分析值** - (1) 由分析参数描述的字符型值(AVALC)或数字型(AVAL)值。分析值可来自输入数据、输入数据  
218 值的分类或其推导值。例如，参数“平均心率（次/分钟）”的分析值由每次访视测量的三次心率值的平均衍生。  
219 (2) 另外，特定函数的值可被认定为分析值。例如：基线值（BASE）、相对于基线的变化值（CHG）  
220 等。  
221  
222 **参数不变性** – 衍生列是不随参数变化的，即在分析数据集内，无论何时被填补，它总是以相同的算法加  
223 以计算。例如列 CHG，无论何时填补，亦无论什么参数，它总是被计算为 AVAL – BASE。然而在某些不  
224 适用的情形下，如对于一个时间事件参数，或者基线前数据不需要计算的时候，CHG 可被置为空值 NULL。  
225 参数不变性只可应用于基于 AVAL 值的函数衍生分析变量（列）。定义参数不变性的目的是为了在 4.2 节  
226 中帮助定义 DBS 数据结构的应用规则。  
227  
228

## 2 AdaM 标准的基础

### 2.1 基本原则

分析数据集必须遵循特定的基本原则，在分析数据集模型文档中这些原则已经得到详细的描述：

- 分析数据集和相关的元数据必须清楚的表达在临床研究中进行统计分析支持的数据集的内容和来源。
- 分析数据集和相关的元数据必须具备可溯源性，以帮助理解一个分析值（分析结果或者分析变量）的来源，即分析值和它前身之间数据体系或关系。元数据也必须提供在什么时间分析值被推导或者输入的相应的标识。分析数据集必须适用于通常的软件工具。
- 分析数据集应该和它的元数据一起来帮助清楚无歧义的表达。理想的元数据应该是机器可读的。
- 分析数据集应该保证它的结构和内容能够允许通过最少的编程工作来进行统计分析。这样的数据集被描述为“可分析的”。应该注意到，在 ADaM 的背景下，分析数据集包含需要审阅的和能重现具体的统计分析结果所需要的数据。如果只是为了支持数据列表或其它非分析相关的展示，不必把数据整理成“可分析的”的数据集。

请参考 [www.cdisc.org](http://www.cdisc.org) 网站的 ADaM 文档提供的更多细节。

### 2.2 可溯源性

为了有助于审阅，分析数据集和元数据必须清楚的交代分析数据集是如何创建的。这个要求表明在使用分析数据集的时候用户必须同时拥有用来创建分析数据集的输入数据以便验证可能的偏差。一个遵从 CDISC 标准的数据提交包括 SDTM 和 ADaM 数据集，因此 SDTM 和 ADaM 之间的关系必须清晰，这凸显了输入数据（SDTM）和分析数据（ADaM）之间的可溯源性的重要性。

可溯源性通过建立一个元素和它的最近前身之间的清晰路径来加以实现。完整的路径跟踪是通过从一个元素到它的前身，再到前身的前身，然后溯源到 SDTM 域，并最终到达最原始数据采集工具端来完成的。值得注意的是，CDISC 临床数据采集统一标准（CDASH）与 SDTM 是内在协调统一的，因此也保证了数据端到端的可溯源性。可溯源性建立了数据集间的关系，也明确了数据集内的关系。例如，分析数据集中的元数据标志和其它支持性变量能够帮助用户理解数据记录是如何（或为什么）被创建的。

可溯源性的两个层次：

- 元数据可溯源性使得用户能够理解分析变量和它的源数据集及其变量之间的关系，这同时也是 ADaM 依从性所要求的。这种可溯源性通过描述（通过元数据）使用的算法或者如何从它的最近前身衍生或填补一个分析值的步骤来建立的。元数据可溯源性也用于建立分析结果和分析数据集之间的关系。
- 数据点的可溯源性使得用户可以溯源到具体的前身记录。如果这种可溯源性可以实现，就应该执行。当审阅者想要溯源一个复杂数据处理的路径时，这个层次的可溯源性是非常有用的。这种可溯源性通过提供（例如，使用 --SEQ 变量）在分析值和用于分析值的输入数据之间的清楚的联系来建立的。

通过来自源数据集的记录标识符变量提供数据点的可溯源性可能不总是实际可行的，但是不管有没有提供数据点的可溯源性，元数据可溯源性必须总是清楚的解释一个分析值是如何填补的。

非常复杂的数据衍生可能需要创建中间分析数据集，对于这样的情形，可溯源性可通过提交那些中间分析数据集和它们的相关元数据来实现。这样可溯源性就会涉及几个步骤，分析结果会通过适当的元数据链接到用来支持分析过程的数据，利用这些数据可以链接到中间分析数据集，然后可通过中间数据溯源链接到 SDTM 源数据。

对于一个成功的可溯源性创建例子，审阅者能够确定：

- 提交的 SDTM 研究表格数据中存在的信息
- 在 ADaM 分析数据集中衍生或填补的信息
- 用于创建衍生或填补数据的方法
- 用于分析的信息及那些不用来分析，但用来支持数据可溯源性或将来分析用的信息。

## 2.3 ADaM 数据结构

清晰、易懂是分析数据集的基本原则。当分析数据集包含源数据和衍生数据时，中心问题就变为让用户理解衍生变量及其观测值是如何产生的，又如何运用这些变量及其观测值产生所需要的分析结果。分析数据集的用户必须能够清楚的确认输入数据和用于产生衍生数据的算法。如果这些信息是通过一种可预测的、并且运用标准数据结构和元数据来传递的话，分析数据集的用户就能够据此而适当的使用该分析数据集来复制分析结果或者实施另外的分析。

许多类型的统计分析不需要一个特定的数据结构，换句话说，分析数据集的结构不一定会限制要进行的分析类型，也不一定会限制数据集本身之间的交流。然而，如果大多数的分析数据集能够使用一个可预测的数据结构，这种交流会得到增强。

除了支持分析数据集用户间的清晰交流，一个可预测的数据结构同时具有其它好处。首先，由于观测值和变量类型的可变性不大，一个可预测的数据结构减轻了管理数据集元数据的负担；其次，也能通过开发软件工具来帮助元数据管理和数据审阅，包括基于已知的关键变量数据重组的工具（例如，转置）；最后，一个可预测的数据结构允许使用一组可被验证的已知惯例来检查分析数据集是否符合ADaM标准。

如在1节指出的，ADaMIG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）。标准的ADaM变量将在第3节描述，实施争议、解决办法及示实例将在第4节加以介绍，第3节和第4节将一起用来完整的描述标准数据结构。

### 2.3.1 ADaM 受试者水平分析数据集 ADSL

不论临床试验设计的类型，ADSL中每个受试者只有一条记录。ADSL提供了描述关于受试者属性的变量，能够允许与其它数据集方便的合并，例如SDTM域和分析数据集。ADSL亦用于其它分析数据集中的受试者级别变量的来源，例如，人群标志和治疗变量。一个研究只有一个ADSL，在一个基于CDISC的数据提交中，在没有其它分析数据集提交的情况下，ADSL和它的相关元数据也是必须的。

临床研究中每个单独的数据值均可以通过创建变量的形式，存放在一个受试者水平的数据集中，比如ADSL，尽管这种方法技术上是可行的，但这并不是创建ADSL的目的。对于那些研究过程中随时间变化的关键终点和数据，BDS数据集才是它们应该存放的地方。

### 2.3.2 ADaM 基本数据结构(BDS)

在一个BDS数据集中，每个受试者、每个分析参数、每个分析时间点包含一条或多条记录，分析时间点的有无要看分析的需要。在没有分析时间点的情况时，数据结构是每个受试者的每个分析参数拥有一条或多条记录。这个结构包含一组中心变量，用来表达被分析的数据。这些变量包含被分析的值（例如AVAL）和对于被分析的值描述（例如PARAM）。数据集中的其它变量提供关于被分析的值更多信息（例如，受试者识别）或者描述衍生此值的轨迹（例如DTYPE）或者是支持对此值的分析（例如治疗变量，协变量）。

读者需要注意的是，在SDTM中并不具备ADaM数据集相对应的结构，因为BDS倾向于纵向设计，有些人可能会认为它和SDTM发现类域相似，然而实际上，BDS数据集可以从SDTM中的发现类域、事件类域、特殊用途类域，或其它ADaM数据集中衍生，亦或者是它们的任何组合。而且和SDTM发现类域数据集不同的是，BDS数据集结构对于实施稳健而又灵活的的大多数统计分析及其审阅提供了支持。

在分析数据集中，一条记录代表一条需要进行分析的，观测到的、衍生的、或填补的数值。例如，它可能是一个时间到事件的记录，比如当一个得分增加到一个阈值的时间，亦或是中止时间，或者是一个高度衍生值，比如对实验室数据进行回归分析衍生的肿瘤生长率。一个数据值可以从任何SDTM及/或ADaM数据集的组合中衍生。

根据4.2节描述的规则，为了支持分析和提供数据的可溯源性，BDS可以灵活的添加行和列。但是应该说明的是，在一个研究中经常有多于一个的BDS分析数据集。添加行和列的能力并不意味着需要把所有数据都放到一个单独的数据集中。在一个研究中，设计最佳的分析数据集个数可以参照在ADaM文档中的相关讨论部分。



### 3 标准 ADaM 变量

在本节定义了一些在分析数据集中常用的标准变量特征。ADaM标准要求，如果一个分析数据集中的变量已包含在本书第3节相关变量定义时，则应该使用这些标准变量名称。

3.1节描述ADSL中使用的变量。3.2节将描述BDS中使用的变量。

在本节，ADaM变量以表格格式描述。在最右边的列“核心”和“CDISC注释”提供了关于变量的信息，帮助用户准备他们的数据集。这些列不是在define.xml中提交的元数据。“核心”列描述了这个变量是必需的、有条件的或许可的。“CDISC注释”列提供了关于变量的更多信息。另外，“类型”列用于定义该变量是字符型或数值型。更具体的信息将在元数据中提供（例如，文本、整数、浮点）。

#### ADaM中属性“核心”的值

- **Req** =必需的，必须包含在数据集中的变量。
- **Cond** =有条件要求的，某些情况下，必须包含在数据集中的变量
- **Perm** =许可的，可能会包含在数据集中的变量，但并不是必需的。  
除非另加说明，所有的ADaM变量可根据实际情况加以填充，也即允许使用null。

#### 通用变量命名惯例

1. 在一对相关的变量中（例如，TRTP和TRTPN、AVAL和AVALC），最主要或最常用的变量是没有后缀或者扩展符的（例如，N是数字，或者C是字符）。
2. 日期填补标志变量的名称以DTF结尾，时间填补标志变量的名称以TMF结尾。
3. 所有其它字符标志（或指示）变量以FL结尾，相应的数字标志（或指示）变量的名称以FN结尾。如果需要使用标志变量，字符版本(\*FL)是必须的，同时也可以包括数字版本(\*FN)。
4. 任何在ADaM中的变量，如果在SDTM中变量名相同，则必须是SDTM变量的复制，并且它的标签、涵义和值保持不变。ADaM遵循“相同名称、相同涵义、相同值”的一致原则。
5. 为了保证SAS传输文件格式和Oracle中约束的兼容性，所有的ADaM变量名必须是不多于8个字符的长度，以字母开始（不是下划线），并且只能由字母(A-Z)、下划线(\_)、和数字(0-9)组成。所有的ADaM变量标签长度必须不超过40个字符。所有的ADaM字符变量值必须不超过200个字符长度。
6. 在第3节中，一个星号(\*)有时用于一个变量名的前缀或后缀，在实际应用中，出现在变量名中的星号必须由一个合适的字符串代替，这样的变量名才有意义并保证还要符合上述的命名与长度限制。
7. 出现在变量名或标签中的小写字母“xx”、“y”、和“zz”必须根据如下约定，在实际的变量名或者标签中被替换。一个变量名称(例如，TRTxxP, APxxSDT)中的字母“xx”指的是具体的一个阶段，其中“xx”用两位整数代替[01-99]，个位前面加0。变量名中的小写字母“y” (例如，SITEGRy)是指一个分组或者其它分类、一个分析判据、或者一个分析范围、可用一个个位数字[1-9]代替。变量名中的小写字母“zz”（例如，ANLzzFL）是对第zz个记录选择算法的一个索引，其中“zz”由两个数字代替，个位数前加0。
8. 以Gry结尾的变量名是分组变量名，y表示分组方案或算法（不是分组中的类别）。例如，SITEGR3是包含试验中心群（合并试验中心）名字的变量的名称，是指分组是根据第三种试验中心分组算法完成的，SITEGR3不是指第三组试验中心。
9. 一般来说，如果SDTM字符变量在ADaM数据集中若被转化为数值型变量，那么它们的命名应该和SDTM的中一样，并添加一个N作为后缀。例如，数值型版本的DM SEX变量在ADaM数据集中变量名是SEXN，RACE的数值型版本名称是RACEN。如果为了保证8字符变量名长度的限制，为了加上N，最后一个字符可以先删除。也要注意到这个规定不适用于日期/时间变量。
10. 如果SDTM字符类别值在ADaM中进行了合并，衍生的ADaM字符归组变量的名称应该以Gry结尾，它的数值型等价变量应该以GryN结束，其中y是一个整数1-9，代表分组方案。例如，如果从包含5个类别的SDTM RACE变量值来创建一个字符分析变量只包含到高加索人和非高加索人，那么应该命名为RACEGRy，它的数值型等价变量命名为RACEGRyN (例如，RACEGR1, RACEGR1N)。在添加后缀片段Gry或者GRyN时，如果必要的话，原来变量名要截尾处理。

#### 通用时间变量惯例

1. 数值型日期、时间和日期/时间应该有固定格式来保证可读性，同时没有精确度方面的损失。用于所有其他日期的计数的参照日期应该在元数据中被清楚的确定。
2. 名称以DT结尾的是数值型日期变量。
3. 名称以DTM结尾的是数值型日期时间变量。
4. 名称以TM结尾的是数值型时间变量。
5. 如果一个\*DTM和与其相关的\*TM变量存在，那么\*TM变量必须和\*DTM变量的时间部分相同。如果一个\*DTM和与其相关的\*DT变量存在，那么\*DT变量必须和\*DTM的日期部分一样。
6. 名称以DTF结尾的是日期填补标志变量。\*DTF变量代表基于SDTM DTC变量来源的\*DT变量的填补程度。如果整个日期都是填补的，则\*DTF =Y；如果月和日是填补的，则\*DTF =M；如果只有天是填补的，则\*DTF =D。如果\*DT等同于SDTM DTC变量日期部分，则\*DTF = null。只要有一个日期是填补的，\*DTF变量是必须的而且应该有填充值。如果\*DTM需要进行填充，\*DTF和\*TMF可能需要在用来描述需要填补的程度。
7. 名称以TMF结尾的变量是时间填补标志。\*TMF变量代表基于SDTM DTC变量来源的\*TM（和\*DTM）的填补程度。如果全部时间都是填补的则\*TMF = H，如果分钟和秒是填补出来的，则\*TMF = M，如果只有秒是填补的，则\*TMF = S，如果\*TM和SDTM DTC变量的时间部分相等，则\*TMF = null。对于一个给定的SDTM DTC变量，如果只是收集了小时和分钟，\*DTM的秒数被设定为00，那么就没有必要把\*TMF的值设为“S”。但是如果秒数通常是采集的，但是在DTC变量的一个给定的值中缺失，并且设定为00，或者如果一个收集到的秒值在创建\*DTM时被改变了，那么差异会是明显的，此时应该在\*TMF中进行设定。如果时间是通过填补得到，那么\*TMF是必须的并且有值。如果已经对\*DTM值做了填补，为了描述填充程度，\*DTF和\*TMF可能都会被用到。
8. 名称以DY结尾的变量是相对日变量。在ADaM中和在SDTM中一样，没有第0天。如果需要创建一个包括第0天的相对日变量，名称一定不要以DY结尾。ADaM相对日变量不一定以SDTM RFSTDTC为参考，当SDTM.RFSTDTC不是参考日期的时候，那么参考日期必须存在于ADaM数据集中。
9. 除非在第三节其它地方具体指出，开始时间变量的名称用S和紧跟其后的表示时间类型的两个字符结尾（例如SDT、STM）。
10. 除非在第三节其它地方具体指出，结束时间变量的名称用E和紧跟其后的表示时间类型的两个字符结尾（例如，EDT和ETM）。
11. 表3.2.3.1的最后一节介绍了标准的后缀命名惯例，用于用户定义的支持性的变量，包含数值型日期、时间、日期时间和相对的天数，及日期和时间填补标志。这些惯例对于ADSL和BDS数据集都是适用的。

读者应该注意的是，用户定义的支持性ADaM日期/时间变量名称的根或前缀必须小心选取，以防会与其它类似的名称和可能的SDTM变量名称的数值型版本发生意外冲突。特别的，可能导致潜在问题的用户定义的根/前缀（\*）的数值包括：

- 一个字母的前缀  
作为这个问题的一个实例，如果\*是Q，那么一个日期\*DT会是QDT。但是一个开始日期\*SDT就会是QSDT，如果用户打算用QSDT来表示SDTM变量QSDTC的数值型版本之外的其它内容，则可能会导致混淆。
- 两字母前缀，除了有意选择清楚的引用一个特定的SDTM域和它的--DTC, --STDTC, 及/或--ENDTC变量的时候。  
比如一个合适的有意使用两字母前缀的例子，如果\*是LB，那么\*DT是LBDT，即SDTM LBDTC的数值型日期版本。一个可能会导致问题的例子，比如，如果\*是QQ，那么一个日期\*DT会是QQDT，如果用户想用QQDT来代表SDTM变量QQDTC数值型日期版本之外的其它内容，则会导致潜在的混淆。
- 三字母前缀，以S和E结尾  
作为这个问题的一个例子，如果\*是QQS，那么一个日期\*DT会是QQSDT，如果用户想用QQSDT来代表SDTM变量QQSTDTC数值型日期版本之外的其它内容，则可能会导致混淆。

## 通用标识 变量惯例

1. 术语“标识”和“指示”同义，且“标识变量”有时就简单的称为“标识”。
2. 如果数据集是对给定的人群进行分析，人群标志必须被包括在数据集中，用于分析的数据集至少要求一个人群标识。所有适用受试者水平人群标识必须包括在ADSL中。
3. 字符型和数值型的受试者水平人群标识名称分别以FL和FN结尾。相似的，参数水平的人群标识名称以PFL和PFN结尾，记录水平的人群标志名称以RFL和RFN结束。

4. 对于受试者水平的字符型人群标识变量：N=否（不包括在人群中），Y=是（包括），不允许空值。
5. 对于受试者水平的数字型人群标识变量：0=否（不包括在人群中），1=是（包括），不允许空值。
6. 对于参数水平和记录水平的字符型人群标识变量：Y=是（包括），允许空值。
7. 对于参数水平和记录水平的数值型人群标识变量：1=是（包括），允许空值。
8. 除了在第3节定义的人群标识变量，其它人群标识变量可能会根据需要添加到ADaM数据集中，并且名称必须服从以上约定。
9. 对于不是人群标识的字符型标识，可能会指定是/否/空或者是/空配置。如在表3.2.6.1中指出的，一些通用的字符标识使用 是/空 配置。相对应的，1/0/空和1/空 配置应用到不是人群指示变量的数值型标识。
10. 可以添加附加的标识，它们的名称和值需要符合上述约定。

### 关于第3节的补充信息

一般来说，变量必需使用第3节表格中指定的标签，但有两个例外：

- 1) 对变量名中包含“y”或者“zz”的变量，在变量标签的结尾允许添加描述性的文字
- 2) 指定的变量标签中带星号(\*)和省略号(...)，这种情况应由申办者用适当的文字加以代替。

值得注意的是，标准变量标签的使用并不代表不同研究及/或申办者中该变量采用了相同的衍生算法。

一些受控术语是为了特定的ADaM变量值而开发的。最新的CDISC受控术语集可以通过CDISC网站 ([www.cdisc.org](http://www.cdisc.org)) 获得。在第3节的表中，如果适用，在列“编码列/受控术语”中会有相对应的外部编码列名称，并用括弧加以标识。本文档中使用的受控术语只是作为示例，最新、权威的CDISC受控术语集可以从官方网站获得。

应该注意的是CDISC外部受控术语集中，有效术语中是不允许包含空值（null值不存在），除非在以下的一个特定变量定义中具体指出，空值（null）是允许的。

在第3节中没有定义的有关变量可能是进行分析或支持数据的可溯源性所必需的，因此可能被添加到ADaM数据集中，条件是它们应该遵循本文档中所约定的ADaM命名惯例和规则。

## 3.1 ADSL 变量

在ADaM文档中，注意到ADaM的一个要求是：在基于CDISC的临床试验数据提交中，即使不需要提交其它分析数据集时，ADSL和它的相关元数据也是必需的。无论临床试验的类型是什么，ADSL的结构都是每个受试者一条记录。ADSL用于提供描述一个受试者属性的所有变量。本节列出了每个ADSL中必须有的标准变量，同时ADSL中还包括一些其它的，在试验中对于描述一个受试者经历来说重要的受试者水平变量。

另外，虽然技术上是可行的，即把研究中每个单独的数据值都拿出来并包括在ADSL中作为变量，但那不是创建ADSL的目的。对于研究过程中那些随时间变化的变量的数据，正确的做法是存放在BDS数据集中，例如正常情况下不会把关键终点值包括在ADSL中。

498

表3.1.1 ADSL变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
研究标识符					
STUDYID	研究标识符	Char		Req	必须与SDTM变量DM.STUDYID、DM.USUBJID、DM.SUBJID及DM.SITEID完全一致。
USUBJID	受试者唯一标识符	Char		Req	
SUBJID	本研究受试者标识符	Char		Req	
SITEID	研究中心编号	Char		Req	
SITEGRy	研究中心组y	Char		Perm	为了分析目的所做的关于临床研究的一个归组或者合并的字符描述。例如，SITEGR3是包含中心组类（或合并中心）名称的一个变量的名称，在变量元数据描述中，该组类是根据第3种中心归组算法来进行合并的，SITEGR3并不是指的第3组中心。
SITEGRyN	研究中心组y(N)	Num		Perm	SITEGRy的数值型编码，与SITEGRy相对应。
受试者人口特征					
AGE	年龄	Num		Req	受试者年龄是ADSL中必需有的变量，如果该变量及其值并非来自于DM.AGE，则必须新增一个名称不同的类似变量。
AGEU	年龄单位	Char	(AGEU)	Req	受试者年龄的单位是ADSL中必需有的变量，如果该变量及其值并非来自于DM.AGEU，则必须新增一个名称不同的类似变量。
SEX	性别	Char	(SEX)	Req	受试者性别是ADSL中必需有的变量。如果该变量并非来自于DM.SEX，则必须新增定义一个不同名称的类似变量。
RACE	人种	Char	(RACE)	Req	受试者人种是ADSL中必需有的变量。如果该变量及其值并非来自于DM.RACE，则必须新增一个不同名称的类似变量。
RACEGRy	种群y	Char		Perm	为了分析目的对于受试者人种的一个归组或者合并的字符描述。
RACEGRyN	种群y(N)	Num		Perm	RACEGRy的数值型编码，便于分析和报告时，基于受试者种族群按一定顺序排序，与RACEGRy相对应。

499

500

表3.1.1 ADSL变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
人群标示变量					
FASFL	全分析集人群标帜	Char	Y, N	Cond	在统计分析计划中定义的针对每个人群的一个字符型标示变量。每个临床试验要求至少有一个受试者级别的人群标示变量。根据需要可添加其他人群标示变量。受试者级别的人群标示变量不能为空。如果使用其中一个标帜，其相应的数值型变量(*FN)也可以被包括在内。
SAFFL	安全集人群标帜	Char	Y, N	Cond	
ITTFLL	意向性治疗人群标帜	Char	Y, N	Cond	
PPROTFL	符合方案人群标帜	Char	Y, N	Cond	
COMPLFL	完成试验人群标帜	Char	Y, N	Cond	

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
RANDFL	随机人群标帜	Char	Y, N	Cond	
ENRLFL	入组人群标帜	Char	Y, N	Cond	
治疗变量					
ARM	计划分组描述	Char		Req	DM.ARM
TRTxxP	xx阶段计划治疗	Char		Req	一受试者水平的治疗标识符，代表xx阶段的计划治疗。 在一个单阶段的随机化试验，TRT01P可以是受试者随机化分配到的治疗。TRTxxP可能会从SDTM DM变量ARM衍生。至少会要求TRT01P。
TRTxxPN	xx阶段计划治疗(N)	Num		Perm	TRTxxP的数值型编码，与TRTxxP相对应。
TRTxxA	xx阶段实际治疗	Char		Cond	受试者水平的治疗标识符，表示xx阶段受试者的实际治疗。当实际治疗和计划治疗不同时或要求按实际治疗进行分析时，该变量是必需的。
TRTxxAN	xx阶段实际治疗(N)	Num		Perm	TRTxxA的数值型编码，与TRTxxA相对应。
TRTSEQP	计划治疗序列	Char		Cond	当涉及多阶段治疗分析时该变量是必需的，例如在交叉设计试验中，ARM可以为与治疗分析不直接相关的描述，亦或是不完全的治疗信息描述（例如，“组1”，“组2”）等，这种情况下，TRTSEQP可以和ARM不同。不管是否和ARM相同，只要是适用的，TRTSEQP都是必需的。
TRTSEQPN	计划治疗序列(N)	Num		Perm	TRTSEQP的数值型编码，与TRTSEQP相对应。
TRTSEQA	实际治疗序列	Char		Cond	如果在进行一个试验时受试者接受的一系列治疗并不是计划的治疗，则TRTSEQA是必需的。
TRTSEQAN	实际治疗序列(N)	Num		Perm	TRTSEQA的数值型编码，与TRTSEQA相对应。
TRxxPGy	xx阶段计划治疗合并组y	Char		Perm	计划的xx阶段的合并治疗。这个变量用在当在一个特定的xx阶段的计划的治疗(TRTxxP) 根据合并算法y被合并在一起分析时，。例如在阶段2，第一个合并算法规定所有药A (TR02PG1=“药A的所有剂量”)的剂量被合并在一起与所有药B的剂量(TR02PG1=“药B的所有剂量”)比较时。每个TRTxxP值只能合并最多一个TrxxPGy值内。
TRxxPGyN	xx阶段计划治疗合并组y(N)	Num		Perm	TrxxPGy的数值型编码，与TrxxPGy相对应。
TRxxAGy	xx阶段实际治疗合并组y	Char		Cond	实际的xx阶段的合并治疗y，在TrxxPGy和TRTxxA同时存在时是必需的。
TRxxAGyN	xx阶段实际治疗合并	Num		Perm	TrxxAGy的数值型编码，与TrxxAGy相对应。

501  
502

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
	组y(N)				

表3.1.1 ADSL变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
试验日期					
RANDDT	随机日期	Num		Cond	在随机化试验中是必需的。
TRTSDT	治疗开始日期	Num		Cond	在一个研究中的一个受试者第一次暴露于治疗的日期。如果涉及到研究的药品，TRTSDT 及/或 TRTSDTM 是必需。
TRTSTM	治疗开始时间	Num		Perm	在一个研究中的一个受试者第一次暴露于治疗的时间。
TRTSDTM	治疗开始日期/时间	Num		Cond	在一个研究中的一个受试者第一次暴露于治疗的日期和时间。如果涉及到研究的药品，TRTSDT 及/或 TRTSDTM 为必需。
TRTSDTF	治疗开始日期填补标帜	Char	(DATEFL)	Cond	第一次暴露于治疗的日期的填补程度。请参见常规时间变量惯例#6。
TRTSTMF	治疗开始时间填补标帜	Char	(TIMEFL)	Cond	第一次暴露于治疗的时间的填补程度。请参见常规时间变量惯例#7。
TRTEDT	治疗结束日期	Num		Cond	一个研究中一个受试者最后一次暴露于治疗的日期。如果涉及到研究的药品，TRTEDT 及/或 TRTEDTM 为必需。
TRTETM	治疗结束时间	Num		Perm	一个研究中一个受试者最后一次暴露于治疗的时间。
TRTEDTM	治疗结束日期/时间	Num		Cond	一个研究中一个受试者最后一次暴露于治疗的日期时间。如果涉及到研究的药品，TRTEDT 及/或 TRTEDTM 为必需。
TRTEDTF	治疗结束日期填补标帜	Char	(DATEFL)	Cond	最后一次暴露于治疗的日期的填补程度。请参见常规时间变量惯例#6。
TRTETMF	治疗结束时间填补标帜	Char	(TIMEFL)	Cond	最后一次暴露于治疗的时间的填补程度。请参见常规时间变量惯例#7。
TRxxSDT	xx 阶段治疗开始日期	Num		Cond	xx 阶段中第一次暴露于治疗的日期。如果同一受试者接受多个治疗，则 TRxxSDT 及/或 TRxxSDTM 在试验设计中为必需，例如在一个交叉试验中。这个变量在对于不同阶段的的同一治疗的设计也是有用的（即相同研究治疗的多

503  
504

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
					个周期)。

表3.1.1 ADL变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
TRxxSTM	xx 阶段治疗开始时间	Num		Cond	xx 阶段中暴露于治疗的开始时间。如果同一受试者接受多个治疗，则 TRxxSTM 及/或 TRxxSDTM 在试验设计中必需。例如一个交叉试验设计。时间对于分析来说重要的考虑。
TRxxSDTM	xx 阶段治疗开始日期/时间	Num		Cond	在 xx 阶段中第一次暴露于治疗的日期时间。如果同一受试者接受多个治疗，则 TRxxSDT 及/或 TRxxSDTM 在试验设计中必需比如一个交叉设计。
TRxxSDTF	xx 阶段治疗开始日期 填补标帜	Char	(DATEFL)	Cond	在 xx 阶段第一次暴露于治疗的日期的填补程度。请参见常规时间变量惯例#6。
TRxxSTMF	xx 阶段治疗开始时间 填补标帜	Char	(TIMEFL)	Cond	在 xx 阶段第一次暴露于治疗的日期的填补程度。请参见常规时间变量惯例#7。
TRxxEDT	xx 阶段治疗结束日期	Num		Cond	在 xx 阶段最后一次暴露于治疗的日期。如果同一受试者接受多个治疗，则 TRxxEDT 及/或 TRxxEDTM 在试验设计中是必需的，比如一个交叉设计。
TRxxETM	xx 阶段治疗结束时间	Num		Cond	xx 阶段的暴露结束时间。如果同一受试者接受多个治疗，则 TRxxETM 及/或 TRxxEDTM 在试验设计中是必需，，比如一个交叉设计。并且结束时间对于分析很重要。
TRxxEDTM	xx 阶段治疗结束日期/时间	Num		Cond	在 xx 阶段最后一次暴露于治疗的日期时间。如果同一受试者接受多个治疗，则 TRxxEDT 及/或 TRxxEDTM 在试验设计中必需比如交叉试验。
TRxxEDTF	xx 阶段治疗结束日期 填补标帜	Char	(DATEFL)	Cond	在 xx 阶段最后一次暴露于治疗的日期的填补程度。请参见常规时间变量惯例#6。
TRxxETMF	xx 阶段治疗结束时间 填补标帜	Char	(TIMEFL)	Cond	在 xx 阶段最后一次暴露于治疗的日期的填补程度。请参见常规时间变量惯例#7。
APxxSDT	xx 阶段开始日期	Num		Perm	xx 阶段的开始日期。
APxxSTM	xx 阶段开始时间	Num		Perm	xx 阶段的开始时间。
APxxSDTM	xx 阶段开始日期/时间	Num		Perm	xx 阶段的开始日期时间。
APxxSDTF	xx 阶段开始日期填补	Char	(DATEFL)	Cond	xx 阶段开始日期的填补等级，请参见常规时间变量惯例

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
	标帜				#6。

表3.1.1 ADSL变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
APxxSTMF	xx 阶段开始时间填补标帜	Char	(TIMEFL)	Cond	xx 阶段开始时间的填补程度，请参见常规时间变量惯例 #7。
APxxEDT	xx 阶段结束日期	Num		Perm	xx 阶段结束的日期。
APxxETM	xx 阶段结束时间	Num		Perm	xx 阶段结束的时间。
APxxEDTM	xx 阶段结束日期/时间	Num		Perm	xx 阶段结束的日期时间。
APxxEDTF	xx 阶段结束日期填补标帜	Char	(DATEFL)	Cond	xx 阶段结束日期的填补程度，请参见常规时间变量惯例 #6。
APxxETMF	xx 阶段结束时间填补标帜	Char	(TIMEFL)	Cond	xx 阶段结束时间的填补程度，请参见常规时间变量惯例 #7。

## 3.2 ADaM 基本数据结构（BDS）变量

ADaM文档介绍了ADaM基本数据结构，一个BDS数据集中，每个受试者在每个分析参数及每个分析时间点上可包含一条或多条记录。分析时间点信息在特定条件下是必需的，其取决于所作分析。在没有分析时间点信息的情况下，结构为每个受试者在每个分析参数水平可包含一条或多条记录。通常在一个研究中会创建多个BDS数据集，ADaMIG在本节定义了用在BDS数据集的标准变量。有关分析数据集ADSL中的标准变量，请参见本文档3.1章节，它们中的任何一个变量都可以复制到BDS分析数据集中，确保数据的可溯源性或实施相关分析。

### 3.2.1 BDS 数据集的受试者标识变量

表3.2.1.1 BDS数据集的受试者标识变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
STUDYID	研究标识符	Char		Req	SDTM DM.STUDYID
USUBJID	受试者唯一标识符	Char		Req	SDTM DM.USUBJID
SUBJID	本研究受试者标识符	Char		Perm	SDTM DM.SUBJID。在 ADSL 中必需，也可以出现在其它数据集中。
SITEID	研究中心编号	Char		Perm	SDTM DM.SITEID。在 ADSL 中必需，也可以出现在其它数据集中。



516

517 **3.2.2 BDS 数据集的治疗变量**

518 **表3.2.2.1 BDS数据集的治疗变量**

变量名称	变量标签	类型	编码列表/受控术语	核心	注释
TRTP	计划治疗	Char		Req	TRTP 是记录水平的标识符，代表归于一组以分析为目的的记录的计划的治疗，TRTP 指出治疗怎样在一个受试者的所有记录间变化并且启动交叉设计和其它设计的分析。TRTxP（从 ADSL 拷贝）也可能为了分析的目的而被需要，且可能对于可溯源性是有用的并能提供背景。
TRTPN	计划治疗(N)	Num		Perm	TRTP 的数值型编码，与 TRTP 相对应。
TRTA	实际治疗	Char		Cond	TRTA 是一个记录级别的标识符，代表归于一组以分析为目的的记录的实际治疗，TRTA 指出治疗怎样在一个受试者的所有记录间变化并且启动交叉设计和其它多期设计的分析。TRTxTA(从 ADSL 拷贝) 也可能为了分析的目的而被需要，且可能对于可溯源性是有用的并能提供背景。当数据以治疗数据进行分析且至少有一个受试者具有和计划的治疗不同的一个治疗相关联的任何数据时，TRTA 是必需的。
TRTAN	实际治疗(N)	Num		Perm	TRTA 的数值型编码，与 TRTA 相对应。
TRTPGy	计划治疗合并组 y	Char		Perm	计划的合并治疗 y。相对于一个特定的合并方案，“y”代表一个个位整数[1-9]。当计划的治疗合并在一起进行分析时有用，例如当药 A 的所有剂量 (TRTPG1=药 A 的所有剂量)和药 B 的所有剂量(TRTPG1=药 B 的所有剂量)相比较，TRTP 的每个值最多会在 TRTPGy 的一个值内部合并，对于一个受试者因记录不同而有可能不同。
TRTPGyN	计划治疗合并组 y(N)	Num		Perm	TRTPGy 的数值型编码，与 TRTPGy 相对应。
TRTAGy	实际治疗合并组 y	Char		Cond	实际的合并治疗 y。相对于一个指定的合并方案，“y”代表一个整数 [1-9]。在 TRTPGy 和 TRTA 同时存在时必需，可能在同一受试者的不同记录变化。
TRTAGyN	实际治疗合并组 y(N)	Num		Perm	TRTAGy 的数值型编码，与 TRTAGy 相对应。

519

520 **3.2.3 BDS 数据集中的时间变量**

521 如果有助于支持数据的可溯源性及/或显示ADaM时间变量与SDTM数据的对比，SDTM中的时间变量(包括但是不限于EPOCH, --DTC, --DY, VISITNUM, VISIT, 和 VISITDY)  
522 可以被复制引入到分析数据集中。

523  
524 **表 3.2.3.1** 定义 BDS 数据集的分析时间变量。名称以字母 A 开始的时间变量直接和分析数据集中的 AVAL 和 AVALC 变量相关联。

525  
526 不直接标示 AVAL 特征的时间变量(例如\*DT)应该以一个字符串作为前缀，而不是表 3.2.3.1 中呈现的占位符星号，所以它们实际的名称遵循在第三节开始时描述的变量命名规则。在许多情形，这些日期时间变量的前缀和一个 SDTM 的--DTC, --STDTC 或者 --ENDTC 变量名称一样。例如，如果一个数字型日期变量是从--STDTC 创建，它会被命名为--SDT。但是，如果--DTC 或--STDTC 是与 AVAL 和 AVALC 相关联的日期，它的数字型等价变量应该被命名为 ADT。第 3 节开始时常规的时间变量惯例文档在这里也适用。

527  
528  
529  
530 **表3.2.3.1 BDS数据集的时间变量**

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
ADT	分析日期	Num		Perm	数字形式的与 AVAL 及/或 AVALC 相关联的日期。
ATM	分析时间	Num		Perm	数字形式的与 AVAL 及/或 AVALC 相关联的时间。
ADTM	分析日期/时间	Num		Perm	数字形式的与 AVAL 及/或 AVALC 相关联的日期/时间。
ADY	分析相对日	Num		Perm	AVAL 及/或 AVALC 的相对日，是从一个参考日期(不必是 DM.RFSTDTC)到 ADT 的天数。参考日期应该是在 ADY 的变量级别元数据中指出，并且参考日期应该被包括为一个给定的分析数据集中的变量或者 ADSL 中的变量。
ADTF	分析日期填补标识	Char	(DATEFL)	Cond	基于 SDTM DTC 来源变量的 ADT 的填补级别，请参见常规时间变量惯例#6。
ATMF	分析时间填补标识	Char	(TIMEFL)	Cond	基于 SDTM DTC 来源变量的 ATM 的填补级别，请参见常规时间变量惯例#7。
ASTDT	分析开始日期	Num		Perm	与 AVAL 及/或 AVALC 相关联的开始日期。ASTDT 和 AENDT 在 AVAL 总结的是一个时间区间内收集的数据或当 AVAL 收集的是一个持续阶段时可能有用。
ASTTM	分析开始时间	Num		Perm	与 AVAL 及/或 AVALC 相关联的开始时间。ASTTM 和 AENTM 在 AVAL 总结的是一个时间区间内收集的数据或当 AVAL 是一个持续时间段时可能有用。
ASTDTM	分析开始日期/时间	Num		Perm	与 AVAL 及/或 AVALC 相关联的开始日期时间。ASTDTM 和 AENDTM 在 AVAL 总结的是一个时间区间内收集的数

531  
532

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
					据或当 AVAL 是一个持续时间段时可能有用。

表3.2.3.1 BDS数据集的时间变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
ASTDY	分析开始相对日	Num		Perm	从一个参考日期（不必是 DM.RFSTDTC）到 ASTDT 的天数。参考日期变量应该在 ASTDY 的变量级别的元数据中指出，并且参考数据变量应该被包括为分析数据集中一个变量（通常是但不必是 ADSL）。
ASTDTF	分析开始日期填补标帜	Num		Perm	基于 SDTM DTC 来源变量的 ASTDT 的填补级别，请参见常规时间变量惯例#6。
ASTTMF	分析开始时间填补标帜	Char	(DATEFL)	Cond	基于 SDTM DTC 来源变量的 ASTTM 的填补级别，请参见常规时间变量惯例#7。
AENDT	分析结束日期	Char	(TIMEFL)	Cond	与 AVAL 及/或 AVALC 相关联的结束日期，请同时参见 ASTDT。
AENTM	分析结束时间	Num		Perm	与 AVAL 及/或 AVALC 相关联的结束时间，请同时参见 ASTTM。
AENDTM	分析结束日期/时间	Num		Perm	与 AVAL 及/或 AVALC 相关联的结束日期时间，请同时参见 ASTDTM。
AENDY	分析结束相对日	Num		Perm	从一个参考日期（不必是 DM.RFSTDTC）到 AENDT 的天数，请同时参见 ASTDY。
AENDTF	分析结束日期填补标帜	Char	(DATEFL)	Cond	基于 SDTM DTC 起源变量的 AENDT 填补级别，请参见常规时间变量惯例#7。
AENTMF	分析结束时间填补标帜	Char	(TIMEFL)	Cond	基于 SDTM DTC 起源变量的 AENTM 填补级别，请参见常规时间变量惯例#7。

533  
534

表3.2.3.1 BDS数据集的时间变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
AVISIT	分析访视	Char		Cond	AVISIT 可以包含观察到的访视名称（即来自 SDTM VISIT）、衍生的访视名称、时间窗名称、概念描述（比如平均值、终点等等）或者这些的任意组合。 AVISIT 是一个衍生的字段，并且不必从 SDTM 映射到 VISIT。AVISIT 表示记录的分析访视，但并不意味着记

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
					录被分析了。经常一个受试者会有多条记录和具有相同 AVISIT 值的参数。可能需要 ANLzzFL 和其它变量来确定任何给定分析选择的记录，请参见 3.2.6 节中关于标识变量的元数据。对于一个给定分析访视时间窗，AVISIT 应该是唯一的。在一个事件，一个记录不落在任何提前定义的分析时间点时间窗内，AVISIT 可以用任何申办者选择来标示这个事实的方式填充（即，空或“不在时间窗内”）。AVISIT 计算的方法，包括用在它的衍生中的变量，应该在 AVISIT 的元数据变量中指出。这些用于衍生 AVISIT 的值和规则在同一数据集中可能会对不同的参数有所不同。AVISIT 是申办者定义的，且经常直接在临床研究报告展示中可用。
AVISITN	分析访视(N)	Num		Perm	AVISIT 的数字表达，这可能是一个方案访视号、一个星期或者周期号、一个分析访视号、或者任何其它的与 AVISIT 逻辑相关的号、或者分析需要的对于排序有帮助的号。 在一个参数中具有 AVISITN 和 AVISIT 之间的一对一映射，所以 AVISIT 的每个不同值都对应着和它相同的 AVISITN 值。在一个事件，一个记录不落在任何提前定义的分析时间点时间窗内，AVISITN 可以用任何申办者选择来标示这个事实的方式填充（例如，可以为空）。AVISITN 的值是申办者定义的。
ATPT	分析时点	Char		Perm	分析时间点描述，如果分析时间是衍生的则为必需。时间点是相对于 ATPTREF 的，ATPT 可以是在一个分析访视内（例如，在用药后 10 分钟、20 分钟、30 分钟的血压评估，AVISIT=Week 1）或者可以和 AVISIT 没有关系（例如，对于第一次头痛突发，用药后 30 分钟、60 分钟、和 120 分钟的偏头痛症状）。
ATPTN	分析时点(N)	Num		Perm	ATPT 的数值型编码。
ATPTREF	分析时点参照	Char		Perm	由 ATPT/ATPTN 引用的对于固定参考点的描述。
APERIOD	分析阶段(N)	Num		Perm	数字值，特征化数据记录属于的时间段。APERIOD 值必须和 TRTxxP、TRTxxA 和名称以 TRxx 及 Apxx 开始的所有的变量中的 xx 值相一致。

表 3.2.3.1 BDS 数据集的时间变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
APERIODC	分析阶段(C)	Char		<u>Perm</u>	特征化记录属于的时间段的文字，一对一映射到 APERIOD。
APHASE	分析分期	Char		<u>Perm</u>	一般来说，一个更高级别的 APERIOD 分类。不代替 APERIOD，因为 APERIOD 提供 TRxx 和 Apxx 变量的索引。
ARELTM	分析相对时间	Num		<u>Perm</u>	相对于参照的时间。当 ARELTM 存在，参照时间变量和 ARELTMU 也必须包括在数据集中，并且锚时间变量必须在 ARELTM 的元数据中标识。
ARELTMU	分析相对时间单位	Char		Cond	ARELTM 的单位。例如，“小时”或者“时间”。如果有 ARELTM 变量，则 ARELTMU 为必需。
APERSDT	阶段开始日期	Num		<u>Perm</u>	由 APERIOD 定义的阶段的开始日期。
APERSTM	阶段开始时间	Num		<u>Perm</u>	由 APERIOD 定义的阶段的开始时间。
APERSDTM	阶段开始日期/时间	Num		<u>Perm</u>	由 APERIOD 定义的阶段的开始日期/时间。
APERSDTF	阶段开始日期填补标帜	Char	(DATEFL)	Cond	基于 SDTM DTC 来源变量的 APERSDT 的填补级别，请参见常规时间变量惯例#6。
APERSTMF	阶段开始时间填补标帜	Char	(TIMEFL)	Cond	基于 SDTM DTC 来源变量的 APERSTM 的填补级别，请参见常规时间变量惯例#7。
APEREDT	阶段结束日期	Num		<u>Perm</u>	由 APERIOD 定义的时间段的结束日期。
APERETM	阶段结束时间	Num		<u>Perm</u>	由 APERIOD 定义的时间段的结束时间。
APEREDTM	阶段结束日期/时间	Num		<u>Perm</u>	由 APERIOD 定义的时间段的结束日期/时间。
APEREDTF	阶段结束日期填补标帜	Char	(DATEFL)	Cond	基于 SDTM DTC 变量的 APEREDT 的填补级别，请参见常规时间变量惯例#6。
APERETMF	阶段结束时间填补标帜	Char	(TIMEFL)	Cond	基于 SDTM DTC 变量的 APERETM 的填补级别，请参见常规时间变量惯例#7。
如下时间变量不直接描述分析值（AVAL 及/或 AVALC）只是为了用来支持评审而可能被包括在数据集里。这样的变量以*作为前缀。请参见常规时间变量惯例#11 获得关于“*”前缀的重要注释。					
*DT	*日期	Num		<u>Perm</u>	分析日期，没有以数字格式直接特征化 AVAL 及/或 AVALC。
*TM	*时间	Num		<u>Perm</u>	分析时间，没有以数字格式直接特征化 AVAL 及/或 AVALC。
*DTM	*日期/时间	Num		<u>Perm</u>	分析日期/时间，没有以数字格式直接特征化 AVAL 及/或

536  
537

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
					AVALC。
*ADY	*相对日	Num		Perm	分析相对日，没有直接特征化 AVAL 及/或 AVALC。
*DTF	*日期填补标识	Char	(DATEFL)	Cond	基于 SDTM DTC 来源变量的*DT 的填补等级。请参见常规时间变量惯例#6。

表3.2.3.1 BDS数据集的时间变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
*TMF	*时间填补标识	Char	(TIMEFL)	Cond	基于 SDTM DTC 来源变量*TM 的填补级别，请参见常规时间变量惯例#7。
*SDT	*开始日期	Num		Perm	开始分析的日期，不直接以数字格式特征化 AVAL 及/或 AVALC。
*STM	*开始时间	Num		Perm	开始分析的时间，不直接以数字格式特征化 AVAL 及/或 AVALC。
*SDTM	*开始日期/时间	Num		Perm	开始分析的日期/时间，不直接以数字格式特征化 AVAL 及/或 AVALC。
*SDY	*开始相对日	Num		Perm	开始分析的相对日，不直接以数字格式特征化 AVAL 及/或 AVALC。
*SDTF	*开始日期填补标识	Char	(DATEFL)	Cond	基于 SDTM DTC 来源变量*SDT 的填补级别，请参见常规时间变量惯例#6。
*STMF	*开始日期填补标识	Char	(TIMEFL)	Cond	基于 SDTM DTC 来源变量*STM 的填补级别，请参见常规时间变量惯例#6。
*EDT	*结束日期	Num		Perm	结束分析的日期，不直接以数字格式特征化 AVAL 及/或 AVALC。
*ETM	*结束时间	Num		Perm	结束分析的时间，不直接以数字格式特征化 AVAL 及/或 AVALC。
*EDTM	*结束日期/时间	Num		Perm	结束分析的日期/时间，不直接以数字格式特征化 AVAL 及/或 AVALC。
*EDY	*结束相对日	Num		Perm	结束分析的相对日，不直接以数字格式特征化 AVAL 及/或 AVALC。
*EDTF	*结束日期填补标识	Char	(DATEFL)	Cond	基于 SDTM DTC 来源变量*EDT 的填补级别，请参见常规时间变量惯例#6。
*ETMF	*结束时间填补标识	Char	(TIMEFL)	Cond	基于 SDTM DTC 来源变量*ETM 的填补级别，请参见常

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
					规时间变量惯例#7。

## 3.2.4 BDS 数据集的分析参数变量

表 3.2.4.1 BDS 数据集的分析参数变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
PARAM	参数	Char		Req	分析参数的描述。例子包括：“仰卧收缩血压（毫米汞柱）”、“Log10（重量（千克））”、“发生第一次高血压事件的时间（天）”、“估计的肿瘤生长率”等等。PARAM 对于不含糊的描述 AVAL 及/或 AVALC 的内容应该是充分的。PARAM 必须包括检验，单位（如果合适的话）、标本类型、地点、位置、和任何其它适用的需要的合格信息，任何附加的信息比如变换函数、和任何真正需要的文本。PARAM 长度可以多于 40 个字符。PARAM 经常是在临床研究报告展示中直接可用，注意到在 ADaMIG 中，“参数”是“分析参数”的一个同义词。
PARAMCD	参数编码	Char		Req	PARAM 中分析参数的短名称，PARAMCD 应该遵循 SAS 5 变量命名惯例（8 字符或更少，以一个字符开始，只包含字母和数字）。和 PARAM 必须是一一对应的映射。例如：SYSBP, LWEIGHT, HYPEREVT。
PARAMN	参数(N)	Num		Perm	对于排序和程序操作有用，必须和 PARAM 是一一对应的映射，必须是一个整数。
PARAMTYP	参数类型	Char	(PARAMTYP)	Perm	指示变量，是否参数是作为一个或多个其它参数的一个函数衍生。这不应该与 DTYPE 混淆，DTYPE 和衍生的 AVAL 及/或 AVALC 相关。
PARCATy	参数类别 y	Char		Perm	PARAM 的一个分类，例如，PARCAT1 的值可能归组和一个特定的调查问卷、实验室标本类型、或者研究领域有关的参数。
PARCATyN	参数类别 y(N)	Num		Perm	PARCATy 的数字化表示，这可以用于关于 PARCATy 的操作，PARCATy 和 PARCATyN 之间应该是一一对应的关

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
					系。
AVAL	分析值(N)	Num		Req	由 PARAM 描述的数字分析值。
AVALC	分析值	Char		(至少一个)	由 PARAM 描述的字符分析值。AVALC 可以是映射到 AVAL 的字符串，但是如果这样，必须在一个给定的 PARAM 内存在 AVAL 和 AVALC 之间的一个映射。AVALC 不应用于给 AVAL 的值分类。
AVALCATy	分析类别 y	Char		Perm	AVAL 及/或 AVALC 的类别表示，不一定是到 AVAL 及/或 AVALC 的一对一映射。例如，如果 PARAM 是“严重头痛”，且 AVAL 具有值 0、1、2、或 3，AVALCAT1 可以把 AVAL 分类成“没有或轻微”（AVAL 值 1 或 2）和“中度或重度”（AVAL 值 3 或 4）。

表 3.2.4.1 BDS 数据集的分析参数变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
BASE	基线值(N)	Num		Cond	基线分析值。如果数据集支持基线值或基线值的函数的分析或审阅则此变量为必需。一个基线的记录可以是衍生的（例如可以是一个平均值）、在此情况 DTYPE 必须也被填充。如果填充了一个参数的 BASE，且 BASE 对一个受试者的那个参数非空，那么那个受试者和参数必须具有一个由 ABLFL 标帜的记录。
BASEC	基线值	Char		Perm	AVALC 的基线值，当对 AVALC 感兴趣时可能需要。如果 BASE 和 BASEC 两变量值同时被填充，在一个给定的 PARAM，他们之间必须是一一对一的映射。AVALC 的基线记录必须和 AVAL 的基线记录一样。
BASECATy	基线值类别 y	Char		Perm	BASE 的一个分类表示，和 BASE 不必是一一对一的映射。例如，如果 PARAM 是“头疼严重性”且 AVAL 有值 0、1、2 或 3，BASECAT1 可以把 BASE 分类为“没有或轻微”（BASE 值为 1 或 2）及“中度或严重”（BASE 值 3 或 4）。
BASETYPE	基线类型	Char		Cond	申办者定义的描述当前记录中和 BASE 值相关的基线定义的文字，当基线可以多种方式定义的时候是必需的。如果用于一个给定的 PARAM，应该填充那个 PARAM 的所有记录。请参考 4.2.1 节规则 6 获取一个实例。
CHG	相对基线变化值	Num		Perm	相对于基线分析值的变化，和 AVAL-BASE 相等。如果用



543  
544

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
					于一个给定的 PARAM, 应该填充那个 PARAM 的所有基线之后的记录。至于如何填充 CHG 基线前的值和基线值, 留给申办者做决定。
CHGCATy	相对基线变化值类别 y	Char		Perm	CHG 的一个分类表示, 不必和 CHG 有一一对应的映射关系。CHGCATy 的定义可能会因为 PARAM 变化。例如, CHGCAT1 可以用于 CHG 相对于 SYSBP 变化范围的分类, “-10 到-5 毫米汞柱”类和“-5 到 0 毫米汞柱”类。
PCHG	相对基线变化百分比	Num		Perm	相 对 于 基 线 分 析 值 的 变 化 百 分 率 , 等 于 $((AVAL-BASE)/BASE)*100$ 。如果用于一个给定的 PARAM, 应该填充那个 PARAM 的所有记录 (当可计算的时候)。至于如何填充 PCHG 基线前和基线的值, 留给申办者做决定。
PCHGCATy	相对基线变化百分比类别类别 y	Char		Perm	PCHG 的一个分类表示, 不必和 PCHG 是一一对应的映射。PCHGCATy 的定义会根据 PARAM 而变化。例如, PCHGCAT1 可能用于相对于 SYSBP 变化范围的 PCHG 的分类: “>5%”类和“>10%”类。
R2BASE	与基线的比值	Num		Perm	AVAL / BASE

表3.2.4.1 BDS数据集的分析参数变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
R2AyLO	与分析范围 y 下限的比值	Num		Perm	AVAL / AyLO, AyLO 在分析数据集中必须存在。
R2AyHI	与分析范围 y 上限的比值	Num		Perm	AVAL / AyHI, AyHI 在数据集中必须存在。
SHIFTy	移动 y	Char		Perm	值的一个移动, 依赖于群“y”定义的配对。SHIFTy 可以是基于任何如下配对值的变化(BASECATy, AVALCATy), (BNRIND, ANRIND), (BTOXGR, ATOXGR), (BASE, AVAL) or (BASEC, AVALC)。对移动表格有用。例如“正常到高”。至于如何填充 SHIFTy 的基线和基线前的值, 留给申办者做决定。
SHIFTyN	移动 y(N)	Num		Perm	SHIFT 的数字版本, SHIFTN 和 SHIFT 之间有一一对应的映射关系。至于如何填充 SHIFTN 的基线和基线前的值, 留给申办者做决定。

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
CRITy	分析准则 y	Char		Perm	标识一个事先确定的判据的文本串，例如 SYSBP > 90。在有些情况，文本串的存在意味着这个记录满足判据，空值意味着判据没有被满足。在其它一些情况，文本串确定被评估的判据，但是是否满足判据由 CRITyFL 的值记录。请参见 3.2.6 节的 CRITyFL 和 CRITyFN。请参见 4.7 节里关于 CRITy、CRITyFL 和 CRITyFN 的附加讨论。

注意到可以添加更多变量，它们是 AVAL 和在同一行的 BASE 的参数不变函数。请参考 4.2 节中的在衍生被添加为行和衍生被添加为列时的管理规则。

### 3.2.5 BDS 数据集的分析描述符变量

表 3.2.5.1 BDS 数据集的分析描述符变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
DTYPE	衍生类型	Char	(DTYPE)	Cond	分析值衍生方法。DTYPE 用作说明，在 AVAL 或 AVALC（因此整条记录）被填补、衍生、或从其它记录复制得来时需要被填充，甚至在衍生记录 AVAL 和 AVALC 为空是也需要被填充。DTYPE 不用于指出一个分析参数是被衍生的。PARAMTYP 可以用于指出一个完全的参数是被衍生的。对于 DTYPE 的每个值，精确的衍生算法必须在分析变量元数据中定义，甚至对于在受控术语中 DTYPE 的值也是一样。请参见第 4 节获取 DTYPE 的例子。DTYPE 的实例值 LOCF=末次观测值向前结转，WOCF=最差的观测值向前结转，AVERAGE=值的平均。

如果分析时间点通过相对日或小时数时间窗定义，那么表3.2.5.2中的变量可能和ADY或ARELTM一起使用，以澄清这个记录如何代表从可能的候选者中选定的每个分析时间点。选定的记录由被分析的记录标识ANLzzFL指出（请参见表3.2.6.1）。注意表3.2.5.2中的变量可能不是在所有的情形下都适用，是作为选项而展示的。

表3.2.5.2 BDS数据集的分析访视窗变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
AWRANGE	分析时间窗范围	Char		Perm	对于一个给定的分析时间点有效的值的范围（AVISIT 的一个给定值）。例如，“5-9 天”。
AWTARGET	分析时间窗目标	Num		Perm	对于 AVISIT 一个给定值的目标或者最需要分析的相对日（ADY）值或者分析相对时间（ARELTM）值。
AWTDIFF	分析时间窗与目标的	Num		Perm	ADY 或者 ARELTM 和 AWTARGET 之间的绝对差异。在

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
	差异				事件 ADY 和 AWTARGET 符号不同中没有零天，将需要为此事实进行校正。如果差异的符号是重要的，那么选择记录时 AWTDIFF 可能不得和 ADY 或 ARELTM 且可能与 AWTARGET 联合使用。
AWLO	分析时间窗开始时点	Num		Perm	开始时间点的值（包括所述的限度）需要和 AWRANGE 联合使用，例如，如果 AWRANGE 是“5-9 天”，那么 AWLO 是“5”。
AWHI	分析时间窗结束时点	Num		Perm	结束时间点的值（包括所述的限度）需要和 AWRANGE 联合使用。例如，如果 AWRANGE 是“5-9 天”，那么 AWHI 是“9”。
AWU	分析时间窗单位	Char		Perm	用于 AWLO 和 AWHI 的单位，例如：天、小时。

表3.2.5.3 BDS数据集中的达到事件时间变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
STARTDT	受试者到达事件时间起始参照	Num		Perm	为到达事件时间分析记录的处于风险最初日期，这通常是受试者第一次处于感兴趣事件的风险时的时间（如在方案或统计分析计划中定义的），例如，这可能是随机化日期或者第一次研究治疗用药的日期。
CNSR	删失	Num		Cond	定义是否事件为删失（在事件被观测到之前观测阶段被截断）。强烈推荐使用 0 作为事件的指示变量，且用正整数作为删失的指示变量。也推荐使用唯一的正整数删失原因的编码描述。CNSR 对于到达时间的参数是必需的。
EVNTDESC	事件或者删失描述	Char		Perm	感兴趣的事件或者删失原因的描述。

表3.2.5.4 BDS数据集中的实验室相关分析变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
ATOXGR	分析毒性等级	Char		Perm	用于分析的毒性等级，可能会基于 SDTM--TOXGR 或者一个填补的或分配的值。
BTOXGR	基线毒性等级	Char		Perm	由 ABLFL 确定的基线记录的 ATOXGR。
ANRIND	分析参考范围标志	Char		Perm	用于分析的正常范围指示变量，可能是基于 SDTM--NRIND 或者一个填补的或分配的值。
BNRIND	基线参考范围标志	Char		Perm	由 ABLFL 确定的基线记录的 ANRIND。

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
ANRLO	分析正常范围下限	Char		Perm	用于分析的正常范围的下限，可能是基于 SDTM --NRLO 或者一个填补的或分配的值。
ANRHI	分析正常范围上限	Char		Perm	用于分析的正常范围的上限，可能是基于 SDTM --NRHI 或者一个填补或分配的值。
AyLO	分析范围 y 下限	Char		Cond	AyLO 及/或 AyHI 在有多个范围用于分析的时候使用，创建 AyLO 及/或 AyHI 是为了收集不同级别的中止值，这些中止值用于决定是否一个分析是在临床可以接受的值范围或是在那个范围之外。AyLO 及/或 AyHI 通常是但不必是常数、参数特定常数、或者受试者特定常数。如果 R2AyLO 被包括在数据集中则必须包括 AyLO。
AyHI	分析范围 y 上限	Char		Cond	参见 AyLO。例如，如果 ECG QTc 值是基于值为 450、值大于 480、和值大于 500 进行总结，一共需要 3 个“嗨，值”范围变量，为了对照着下面的三个值进行计算：A1HI=450, A2HI=480, A3HI=500。如果 R2AyHI 被包括在数据集中，AyHI 必须被包括。

### 3.2.6 BDS 数据集的标识变量

参见3.2.7节关于ADaM人群和基线标识及SDTMIG 3.1.1 and 3.1.2中标识之间差异的讨论。

表3.2.6.1 BDS中的标识变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
ABLFL	基线记录标识	Char	Y	Cond	字符型标识变量，为每个参数识别基线记录，或者在如果每个参数和基线类型具有多条或者多个基线定义时识别基线记录。参见表 3.2.4.1 中的 BASETYPE，如果 BASE 存在于数据集中，则 ABLFL 是必需的。 基线记录可以是原始基线数据的衍生（比如可以是多个值的平均值），在那种情形下，DTYPE 也必须被填充。如果 BASE 存在，且对于受试者和这某个参数的 BASE 值非空，那么该此受试者及该参数必须有一条记录由 ABLFL 标识。
ABLFN	基线记录标识(N)	Num	1	Perm	ABLFL 的数值型。
ANLzzFL	分析记录标识 zz	Char	Y	Cond	ANLzzFL是一个有条件要求的标识变量，当其它选择变量组合在一起也不足以确定用于一个或多个分析的记录的确切集合时作用为其它选择变量的附加。经常一个 ANLzzFL将用于支持为多于一个的分析支持正确的记录选择。当在定义用于一个特定分析或分析族的记录集合时，ANLzzFL是其它选择变量的补充，并且意图是和其它选择变量一起使用，比如受试者级别、参数级别和记录级别的人群标识，AVISIT，DTYPE，归组变量比如SITEGRY，和其它的。每个记录选择算法“zz”（即填充ANLzzFL的每个算法）必须在变量元数据中定义。当算法“zz”操作的记录集是事先使用其它判据过滤的，比如记录级别的人群标识，那么元数据中的选择算法定义必须如此指定。 注意到 Y 的 ANLzzFL 值指出记录履行了算法的要求，但是不必暗指记录被实际的用于一个或多个分析，因为一个记录是否会被使用也依赖于应用的其它选择变量。 ANLzzFL 标识在许多其它情形是有用的，一个例子是当在一个受试者和参数内具有用于分析时间点的多于一个的记录，就像它能用于确定选择的记录来为分析表示分析时间点。“zz”是一个记录选择算法的索引，比如“AVISIT”的与目标相对日最近的记录，以最新记录打破相同值，对于每个<AVISITS 列表>中的 AVISIT”。
ANLzzFN	分析记录标识 zz(N)	Num	1	Perm	ANLzzFL 的数值型。

563  
564

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
ONTRTFL	治疗记录标帜	Char	Y	Perm	字符型标帜变量，指出观测值是否在受试者处于治疗时发生。

表3.2.6.1 BDS数据集中的标帜变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
ONTRTFN	治疗期记录标帜(N)	Num	1	Perm	ONTRTFL 的数字型。
LVOTFL	治疗期末次记录标帜	Char	Y	Perm	字符型指示变量，标示某个参数在治疗期间最后一非缺失数据值。
LVOTFN	治疗期末次记录测值标帜(N)	Num	1	Perm	LVOTFN 的数值型。
ITTRFL	意向性治疗记录标帜	Char	Y	Cond	字符型指示变量，指出受试者某一特定记录是否纳入意向性治疗分析。
ITTRFN	意向性治疗记录标帜(N)	Num	1	Perm	ITTRFN 的数值型。
ITTPFL	意向性治疗参数标帜	Char	Y	Cond	字符型指示变量，指出受试者某一特定参数是否纳入意向性治疗分析。
ITTPFN	意向性治疗参数标帜(N)	Num	1	Perm	ITTPFN 的数值型。
SAFRFL	安全分析记录标帜	Char	Y	Cond	字符型指示变量，指出受试者某一特定记录是否纳入安全分析。
SAFRFN	安全分析记录标帜(N)	Num	1	Perm	SAFRFL 的数值型。
SAFPFL	安全分析参数标帜	Char	Y	Cond	字符型指示变量，指出受试者某一特定参数是否纳入安全分析。
SAFPFN	安全分析参数标帜(N)	Num	1	Perm	SAFPFN 的数值型。
FASRFL	全分析集记录标帜	Char	Y	Cond	字符型指示变量，指出受试者某一特定记录是否纳入全分析集分析。

565  
566

表3.2.6.1 BDS数据集的标帜变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
------	------	----	-----------	----	---------

567  
568

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
FASRFN	全分析集记录标帜(N)	Num	1	Perm	FASRFN 的数值型。
FASPFL	全分析集参数标帜	Char	Y	Cond	字符型指示变量，指出受试者某一特定记录是否纳入全分析集分析。
FASPFN	全分析集参数标帜(N)	Num	1	Perm	FASPFN 的数值型。
PPROTRFL	符合方案记录标帜	Char	Y	Cond	字符型指示变量，指出受试者某一特定记录是否纳入符合方案集分析。
PPROTRFN	符合方案记录标帜(N)	Num	1	Perm	PPROTRFL 的数值型。
PPROTPFL	符合方案参数标帜	Char	Y	Cond	字符型指示变量，指出受试者某一特定参数是否纳入符合方案集分析。
PPROTPFN	符合方案参数标帜(N)	Num	1	Perm	PPROTPFN 的数值型。
COMPRFL	完成试验记录标帜	Char	Y	Cond	字符型指示变量，指出受试者某一特定记录是否纳入完成试验受试者分析。
COMPRFN	完成试验记录标帜(N)	Num	1	Perm	COMPRFL 的数值型。
COMPPFL	完成试验参数标帜	Char	Y	Cond	字符型指示变量，表示受试者某一特定参数是否纳入完成试验人群集分析。
COMPPFN	完成试验参数标帜(N)	Num	1	Perm	COMPPFL 的数值型。

表3.2.6.1 BDS数据集的标帜变量

变量名称	变量标签	类型	编码列表/受控术语	核心	CDISC注释
CRITyFL	准则 y 结果标帜	Char	是或者 是，否	Cond	字符型指示变量，指出某一特定参数值是否满足 CRITy 中定义的标准。请参阅 3.2.4 节中相关变量 CRITy。如果 CRITy 存在则该变量是必需的，详见 4.7 节中的讨论。
CRITyFN	准则 y 结果标帜(N)	Num	1 或1, 0	Perm	CRITyFL 的数值型。

569  
570

### 3.2.7 SDTM 及 ADaM 人群和基线标识之间的区别

SDTM实施指南包括受试者水平人群标识的一些补充修饰语值的受控术语。从其它术语到ADaM变量的概念性映射在表3.2.7.1中介绍。

表3.2.7.1 相对于SDTM补充修饰语的ADaM受试者级别人群标识

SDTM QNAM	SDTM QLABEL	ADaM Subject-Level Population Flags
COMPLT	完成试验人群标识	COMPFL
FULLSET	全分析集人群标识	FASFL
ITT	意向性治疗人群标识	ITTFL
PPROT	符合方案人群标识	PPROTFL
SAFETY	安全性分析人群标识	SAFFL

ADaM受试者水平人群标识可能和它们在SDTM中的概念性对应部分不匹配，例如，SDTM ITT的补充修饰语可能和一个给定受试者的ADaM ITTFL指示变量不匹配。这些人群标识可能不匹配是因为操作方面的问题，一个公司完全有可能因为各种无法改变的原因而继续使用SDTM的数据。

在这种情况下，那些为了回到和“修改”SDTM人群补充修饰语而创建的分析数据集不是必须的，并且不做的原因也很充分。

得到ADaM小组认可的是，如果SDTM受试者水平人群补充修饰语和ADaM人群指示变量能够协调是最好，但更重要的是认识到可能会有些情况它们不相同。有一些添加的ADaM受试者水平人群标识在SDTM中没有相应的部分，ADaM也支持参数水平和记录水平的人群标识，这在SDTM中是不存在的。

相似的，一个在SDTM中确定的基线记录可能不是一个ADaM数据集中确定的记录，导致这发生的原因很多。例如，可能同时具有基线血糖和另一基线尿糖是必要的。这在ADaM数据集中为两个不同的参数，每个都有自己的基线，但是在SDTM中，可能只有一个葡萄糖基线。另外，ADaM中有一些高度衍生的参数，在一个发现域中并没有简单的相应部分。一个ADaM参数可能从SDTM数据通过跨越多个域类衍生。这样的参数不会在SDTM中存在，所以它的基线只能在ADaM数据集中存在。此外，在研究中可能对于不同的阶段具有不同的基线是必要的，例如，为了支持相对于筛选基线、双盲治疗基线、和不设盲扩展基线发生的变化分析（请参见4.2节，规则6）。当存在记录水平人群标识的时候，可能对于两个不同的分析人群设两个不同的基线是必要的。最后，想要的是对于不同的基线定义进行分析。ADaM的基线标识ABLFL，和BASE及BASETYPE列一起，加上人群指标，可以处理所有这些实际的复杂情况。为了分析的目的，人群标识和用于分析的基线指标等可以在分析数据集中找到。ADaM标识应该在ADaM元数据中加以描述。

### 3.2.8 其它变量

#### 分析使能变量

这一类变量，使得设计这些数据集想要支持的一个或多个分析可以进行。通常，这些使能变量会包括指示变量和上面提到的分析描述符变量，它们经常需要把分析数据集整理到只是经过一个统计过程就能得到分析结果的程度。使能变量也可能包括为了进行分析需要存在的分层和归组变量、模型协变量和任何其它变量。

#### 数据点可溯源性变量

支持数据点的可溯源性变量不管是否使用都应该被包含。SDTM中可作为数据点可溯源性的主要来源包括SDTM DOMAIN变量值、SDTM来源变量的名称、和相关的SDTM域--SEQ值。在事件类中，变量AVAL或AVALC值则来源于SDTM中的补充修饰变量值，ADaM 数据集中--SEQ两字母的域前缀是相关的域缩写（SUP--或SUPPQUAL中RDOMAIN的值），并且--SEQ的值是相关的有关系的域记录的序号。即使--SEQ不是溯源到SDTM源数据的关键链接变量，但--SEQ仍可以作为溯源SDTM源数据链接键的一部分，因为--SEQ会指向至少一个有效的SDTM记录作为ADaM数据的来源。表3.2.8.1定义了特定状况有用的附加变量来促进数据点的可溯源性。4.4节包含一个如何使用这些变量的例子。



616

表3.2.8.1 数据点可溯源性变量

SDTM QNAM	SDTM QLABEL	ADaM受试者级别人群标识	SDTM QNAM	SDTM QLABEL	ADaM受试者级别人群标识
SRCDOM	来源域	Char		Perm	SDTM域的两字母标识符，联系到AVAL或AVALC。
SRCVAR	来源变量	Char		Perm	列的名称（在SDTM域由SRCDOM识别），联系到AVAL或AVALC。
SRCSEQ	来源序号	Num		Perm	行的序号SEQ（在SDTM域由SRCDOM识别），联系到AVAL或者AVALC。

617

## 4 执行过程中的问题与解决方案及示例

ADaM标准变量（列）在第3节进行了详细描述。但是，更多的应该是依从ADaM，而不是遵守第3节。本章节的目的是对于如何正确实施ADaM标准数据集提供额外的指导，辅以例子来说明。

4.1节提供常用的试验设计的治疗变量的例子。

4.2至4.8节涉及BDS，这些节里提供了对于BDS实施问题的标准解决方案，辅以例子说明。与ADSL相比，在一个研究中通常有多个数据集服从BDS。如在ADaM文档讨论的，分析数据集的个数应该对于支持分析和审阅为“最优”。

分析数据集元数据的类属性指明一个数据集应该遵从的结构：ADSL、BDS、或其它。限于篇幅，为节省空间，第4节中的例题必要时省略了许多必需的和允许的ADaM变量，只列出了有助于帮助理解所述内容需要用到变量。

### 4.1 常用试验设计治疗变量的例子

如下几个例子为几种不同的试验设计，包括平行设计、交叉设计、和含开放扩展的平行设计，阐释了ADSL中围绕治疗变量的概念。请注意，只有被选择的变量才被阐述；这些例子并不表示它们是ADSL中仅有的变量。

例一（表4.1.1），阐释一个平行设计研究中三个受试者的治疗变量。  
请注意，第三个受试者被随机化到研究用药物治疗组，但实际接受的是对照组安慰剂治疗。

表4.1.1 随机化的平行设计

行	USUBJID	ARM	TRT01P	TRT01A	TR01SDT	TR01EDT
1	1001	药X5毫克	药X5毫克	药X5毫克	2007-10-23	2007-12-07
2	1002	安慰剂	安慰剂	安慰剂	2006-07-19	2007-09-20
3	1003	药X5毫克	药X5毫克	安慰剂	2007-11-01	2007-11-20

例二（表4.1.2），解释一个两阶段交叉设计临床试验中三个受试者的治疗变量。请注意，变量TRTSDT和TRTEDT没有被列出，但是因为一些受试者可能在接受TRT02P之前就中止试验了，所以TRTSDT=TR01SDT，TRTEDT取TR01EDT和TR02EDT中的最大值。应该注意，对于两个试验阶段，受试者1002和1003（在行2和3）都使用了安慰剂。

表4.1.2 两阶段交叉设计

USUBJID	TRTSEQP	TRT01P	TRT02P	TRTSEQA	TRT01A	TRT02A	TR01SDT	TR01EDT	TR02SDT	TR02EDT
1001	安慰剂-药X	安慰剂	药X	安慰剂-药X	安慰剂	药X	2006-02-15	2006-05-03	2006-05-10	2006-08-15

USUBJID	TRTSEQP	TRT01P	TRT02P	TRTSEQA	TRT01A	TRT02A	TR01SDT	TR01EDT	TR02SDT	TR02EDT
1002	安慰剂-药X	安慰剂	药X	安慰剂-安慰剂	安慰剂	安慰剂	2006-03-01	2006-06-12	2006-06-20	2006-09-23
1003	药X-安慰剂	药X	安慰剂	安慰剂-安慰剂	安慰剂	安慰剂	2006-02-03	2006-04-25	2006-05-01	2006-08-04

例三(表4.1.3), 解释一个三阶段交叉设计临床试验中三个受试者的治疗变量。请注意, 变量TRTSDT和TRTEDT没有被列出, 但是因为一些受试者可能在接受TRT03P之前就中止试验了, 所以TRTSDT=TR01SDT, TRTEDT取TR01EDT、TR02EDT和TR03EDT中的最大值。在这个试验中, 所有的受试者在每个阶段都是接受计划的治疗, 所以不需要TRTxxA变量。

表4.1.3 三阶段交叉设计

行	USUBJID	TRTSEQP	TRT01P	TRT02P	TRT03P	TR01SDT	TR01EDT	TR02SDT	TR02EDT	TR03SDT	TR03EDT
1	1001	安慰剂-药X-药Y	安慰剂	药X	药Y	2006-02-15	2006-05-03	2006-05-10	2006-08-15	2006-08-23	2006-11-14
2	1002	药Y-安慰剂-药X	药Y	安慰剂	药X	2006-03-01	2006-06-12	2006-06-20	2006-09-23	2006-10-01	2006-12-05
3	1003	药X-药Y-安慰剂	药X	药Y	安慰剂	2006-02-03	2006-04-25	2006-05-01	2006-08-04	2006-08-12	2006-10-15

例四(表4.1.4), 阐述的是一个开放性、不设盲的平行设计临床试验中两个受试者的治疗变量, 对于开放性研究, 变量TRT01P用于受试者在双盲随机化试验阶段的治疗, TRT02P用于开放性阶段的治疗。

表4.1.4 一个平行设计的不设盲扩展试验

行	USUBJID	TRTSEQP	TRT01P	TRT02P	TR01SDT	TR01EDT	TR02SDT	TR02EDT
1	1001	药X 5毫克 -药X 5毫克	药X 5毫克	药X 5毫克	2007-08-14	2007-09-20	2007-09-21	2008-03-15
2	1002	安慰剂 -药X 5毫克	安慰剂	药X 5毫克	2007-07-05	2007-08-15	2007-08-17	2008-02-04

## 4.2 衍生列 vs 衍生行的创建

在ADaM BDS中, 受试者、分析参数、和分析时间点决定了数据行, 并用标准的列名来识别。受试者、参数和时间点的组合可能不足以作为自然的关键字(唯一的记录标识符)。在一个给定的组合中, 基于收集或衍生的观测值个数、基线定义等不同, 可能有多行数据。

创建标准列可实现多种目的, 比如为实现数据的可溯源性而设创建SDTM记录标识符、为实现灵活多样化分析而创建的人群标识及其它记录选择标识、分析值及基于分析值的某种标准函数值等。允许的列并不仅仅限于在第3节中讲述的变量名, 可能包括特定研究的分析模型协变量、亚组变量、支持可溯源性的变量、和其它分析需要或者有利于审阅的变量。然而, 对于何时将数据衍生为列有一些限制。4.2节中的主题是详细讲述何时分析值的函数应被衍生为列, 何时应被衍生为行。根据实际操作和特定研究的需要, 创建一个BDS分析数据集所涉及的详细步骤有所不同。基于本次讨论的目的, 先讲讲两个基本步骤。

第一步, 根据SDTM不同域类别, 直接或适当作些改变创建一系列的行和列, 并将这些行和列合理的排列。这一步可以包括创建分析参数(如PARAM等)、分析时间点(如AVISIT等)和分析变量(如AVAL和AVALC)。也可以包括其它的标识符(如STUDYID, SITEID, USUBJID, SUBJID)和其它SDTM域中的可溯源性变量(如VISIT、--SEQ等)。

第二步, 在第一步创建的数据集行和列的基础上进一步衍生一些行和列, 4.2节将着重讲述。

具体来讲, 4.2节中定义的衍生行和衍生列是根据已经包含在分析数据集中的数据来创建的行和列, 而不是基于(1)直接从SDTM域复制或衍生的数据; (2)直接从其它分析数据集或元数据复制或衍生的数据。本节只讲述如何利用这些已在数据集中“内部生成”的数据来创建新的行和列。

本节将讨论与上述有关的ADaM规则, 这些规则决定了何时要用这些“内部生成”的数据来创建新列, 以及何时用它们来创建新行。这些规则是BDS的定义中的必不可少部分。

### 4.2.1 创建行和列的规则

为了遵从BDS, 有必要对何时允许创建新列作一些规定。规则1描述了一种应该创建新数据列的情况。规则2至6则描述了应该创建新行, 这些新产生的行可以是全新的参数, 也可以是已有参数的附加行。

**规则1. 对于处于同一行的关于AVAL和BASE的一个和参数无关的函数, 如果不包含对BASE的变换, 应该被添加为一个新列。**

规则1关于何时将关于AVAL和BASE的函数添加为一个新列有三个条件:

1. 该函数是针对AVAL的, 也可以同时针对同一行的BASE;
2. 该函数与参数无关;
3. 该函数不涉及转换BASE。

下面来详细解释这些条件。

PARAM能详尽描述AVAL或AVALC的内容。通常AVAL本身不是分析所必须。例如, 在相对于基线的改变的分析中, 被分析的是相对于基线的变化值CHG。相对于基线的变化列CHG应该根据规则1创建, 因为它满足三个条件:

1. CHG由处于同一行的AVAL和BASE的差值衍生而来;
2. 相同的计算应用于同一数据集中的其它行的CHG(函数CHG=AVAL-BASE不会随PARAM而变化)。

第2个条件称为“参数无关性”特性; 除了在第3节所说的, 针对AVAL的一个函数(或者也同时针对BASE), 如果它的目的是为了包含一组与参数相关的函数, 不可以作为一个列被衍生。

3. 函数CHG=AVAL-BASE中BASE没有经过变换。

表4.2.1.1演示了CHG列。请注意, 并不要求在所有行计算CHG的值。如果需要, 规则1允许CHG和其它函数列可以仅仅在对于该研究的分析和审阅比较适合或者有潜在用处的行和分析参数里计算。基线标识列ABLFL用来标记衍生BASE列的那些行。

表4.2.1.1 规则1的示例：创建包含针对AVAL和BASE的处于同一行的参数无关函数的新列

行	PARAM	PARAMCD	AVISIT	ABLFL	AVAL	BASE	CHG
1	体重（千克）	WEIGHT	筛选		99	100	.
2	体重（千克）	WEIGHT	入组		101	100	.
3	体重（千克）	WEIGHT	基线	是	100	100	0
4	体重（千克）	WEIGHT	第24周		94	100	-6
5	体重（千克）	WEIGHT	第48周		92	100	-8
6	体重（千克）	WEIGHT	第52周		95	100	-5
7	脉搏频率（次/每分钟）	PULSE	筛选		63	62	.
8	脉搏频率（次/每分钟）	PULSE	入组		67	62	.
9	脉搏频率（次/每分钟）	PULSE	基线	是	62	62	0
10	脉搏频率（次/每分钟）	PULSE	第24周		66	62	4
11	脉搏频率（次/每分钟）	PULSE	第48周		70	62	8
12	脉搏频率（次/每分钟）	PULSE	第52周		64	62	2

现在考虑潜在的函数列 $\text{LOG10} = \text{Log10}(\text{AVAL})$ ，这个函数满足所有规则1的三个条件，因此允许作为一个函数列。

但是，如果希望对 $\text{LOG10}$ 进行相对于基线的变化的分析，那么 $\text{LOG10}$ 列、 $\text{LOG10}$ 的基线和相对于 $\text{LOG10}$ 的基线的改变也可能被分析和审阅，那么这个 $\text{LOG10}$ 变换应该被创建为一个新参数，这样就可以使用通常的列AVAL、BASE和CHG。

这是因为基线 $\text{LOG10}$ 和相对于基线 $\text{LOG10}$ 的变化的列不满足规则1的条件。基线 $\text{LOG10}$ 违反了第一个条件，因为它不是通常所说的针对处于相同行的AVAL的一个函数（一般不随AVAL变化），而只是处于基线行的AVAL的一个函数。“相对于基线 $\text{LOG10}$ 的变化”= $\text{LOG10}(\text{AVAL}) - \text{LOG10}(\text{BASE})$ 违反了第三个条件，因为它包含了对BASE的 $\text{Log10}$ 变换。

其目的是尽量多的使用标准列，尽量保持结构标准，并且避免不当的水平变换，同时仍旧允许函数列的高效使用。

任何满足规则1的三个条件的函数，都允许成为一列，如果该函数在第3节有列出，那么应使用ADaM标准列名称，就如同在表4.2.1.1中使用CHG一样。

**规则2. 对AVAL的变换，如不满足规则1的条件，则应该被添加为一个新参数，而且AVAL应该包含所变换的值。**

如果目的是为了根据AVAL的变换重新定义AVAL、BASE、CHG等等，那么就必须添加一个新的参数，用PARAM描述该变换。创建一个新参数，产生一组新的行。

例如，在规则1的讨论中描述过，在对体重的对数变换的进行相对于基线变化的分析中，AVAL值应该是体重的对数，BASE值应该是体重的对数的基线值，而CHG值应该是二者的差值。PARAM应该描述AVAL值所对应的变换数据，例如，“ $\text{Log10}(\text{体重(千克)})$ ”。以这种方式，ADaM标准可以在标准列中涵盖变换数据的分析，而无须为不同目的创建一个新的特别用途列。

在表4.2.1.2我们看到，申办者将AVISITN的值设置为星期的编号，这样能很好的满足排序和画图的需要。VISITNUM是SDTM中的访视编号。

请注意，当SDTM变量，比如USUBJID, SUBJID, SITEID, VISIT, VISITNUM和--SEQ，以SDTM变量名形式被包含在一个ADaM数据集中时，它们的值必须保留原样，不能以任何方式被改变。

为了清楚说明PARAM  $\text{Log10}(\text{体重(千克)})$ 是衍生的，许可变量PARAMTYP已被填充为DERIVED。PARAMTYP不是必需的。

表4.2.1.2 规则2示例：产生一个新参数来处理一个变换

行	PARAM	PARAMCD	AVISIT	AVISITN	VISITNUM	ABLFL	AVAL	BASE	CHG	PARAMTYP
1	体重（千克）	WEIGHT	筛选	-4	1		99	100	.	

行	PARAM	PARAMCD	AVISIT	AVISITN	VISITNUM	ABLFL	AVAL	BASE	CHG	PARAMTYP
2	体重 (千克)	WEIGHT	入组	-2	2		101	100	.	
3	体重 (千克)	WEIGHT	基线	0	3	是	100	100	0	
4	体重 (千克)	WEIGHT	第24周	24	4		94	100	-6	
5	体重 (千克)	WEIGHT	第48周	48	5		92	100	-8	
6	体重 (千克)	WEIGHT	第52周	52	6		95	100	-5	
7	Log10 (体重 (千克))	L10WT	筛选	-4	1		1.9956	2	.	衍生
8	Log10 (体重 (千克))	L10WT	入组	-2	2		2.0043	2	.	衍生
9	Log10 (体重 (千克))	L10WT	基线	0	3	是	2	2	0	衍生
10	Log10 (体重 (千克))	L10WT	第24周	24	4		1.9731	2	-0.0269	衍生
11	Log10 (体重 (千克))	L10WT	第48周	48	5		1.9638	2	-0.0362	衍生
12	Log10 (体重 (千克))	L10WT	第52周	52	6		1.9777	2	-0.0223	衍生

规则2还可以用在需要用两个不同系统的单位做分析和报告的情况下。在SDTM测量域中，比如LB, QS, EG等等，--STRESN列是仅有的数字结果列，并且也是仅有的标准化数字结果列。--ORRES列包含收集结果的以字符型表示的值，单位放在--ORRESU列中。--ORRES列不是标准化的。所以例如如果数据通常以传统的单位收集，SDTM不能同时以传统的单位和国际系统单位（SI）包含标准化的数据。在SDTM中，对于任意给定的--TEST，申办者只能在标准化数据中使用一种单位，而不能同时使用两种单位。如果既要用传统的单位又要用SI单位来分析标准化的结果，则需要对分析数据集做变换。每当遇到每个这样的情况，为了包括另外一种单位的标准化数据，就必须创建一个新的参数（Parameter）。

在PARAM列的描述中必须包含单位，以及任何其它信息，比如场所和标本类型，以此来保证PARAM唯一的描述了在AVAL值中到底是什么，这样就可以根据中和需要区分不同参数。对于不同的单位，PARAM不能相同。

以表4.2.1.3的数据为例，支持低密度脂蛋白（LDL）胆固醇在传统的单位（毫克/分升）和SI单位（毫摩尔/升）的分析。在这个研究中，SDTM中胆固醇数据以毫克/分升为单位标准化；在分析数据集中，从每个初始SDTM记录产生两个记录，每种系统单位一条记录。

768

表4.2.1.3 规则2示例：产生一个新参数以应付第二种单位系统

行	PARAM	PARAMCD	AVISIT	AVISITN	VISITNUM	LBSEQ	ABLFL	AVAL	BASE	CHG	PCHG
1	低密度脂蛋白胆固醇（毫克/分升）	LDL	筛选	-2	1	2829		206.3	213.4		
2	低密度脂蛋白胆固醇（毫克/分升）	LDL	入组	-1	2	2830		202.1	213.4		
3	低密度脂蛋白胆固醇（毫克/分升）	LDL	第0周	0	3	2831	是	213.4	213.4	0.0	0.00
4	低密度脂蛋白胆固醇（毫克/分升）	LDL	第5周	5	4	2832		107.4	213.4	-106.0	-49.67
5	低密度脂蛋白胆固醇（毫克/分升）	LDL	第11周	11	5	2833		90.2	213.4	-123.2	-57.73
6	低密度脂蛋白胆固醇（毫克/分升）	LDL	第17周	17	6	2834		96.8	213.4	-116.6	-54.64
7	低密度脂蛋白胆固醇（毫克/分升）	LDL	第23周	23	7	2835		104.0	213.4	-109.4	-51.27
8	低密度脂蛋白胆固醇（毫摩尔/升）	LDLT	筛选	-2	1	2829		5.3349	5.5185		
9	低密度脂蛋白胆固醇（毫摩尔/升）	LDLT	入组	-1	2	2830		5.2263	5.5185		
10	低密度脂蛋白胆固醇（毫摩尔/升）	LDLT	第0周	0	3	2831	是	5.5185	5.5185	0.0000	0.00
11	低密度脂蛋白胆固醇（毫摩尔/升）	LDLT	第5周	5	4	2832		2.7773	5.5185	-2.7412	-49.67
12	低密度脂蛋白胆固醇（毫摩尔/升）	LDLT	第11周	11	5	2833		2.3326	5.5185	-3.1859	-57.73
13	低密度脂蛋白胆固醇（毫摩尔/升）	LDLT	第17周	17	6	2834		2.5032	5.5185	-3.0153	-54.64
14	低密度脂蛋白胆固醇（毫摩尔/升）	LDLT	第23周	23	7	2835		2.6894	5.5185	-2.8291	-51.27

769

### 规则3. 为了创建一个分析时间点，相同参数中一个行或多个行的函数应该被添加为此为参数的一个新行

为了分析的目的，经常会需要填补缺失数据，或者需要创建一个衍生概念的时间点，这样的衍生应该是相同参数内新衍生记录的创建。

作为一个一般性规则，当一条记录是由数据集中的一条单独记录衍生时，在衍生的记录中保留来自任何初始记录的没有改变的变量值，和那些在新纪录背景下有意义的变量值（例如，--SEQ, VISIT, VISITNUM, --TPT,协变量等等）。当一条记录是从多条记录衍生时，那么在衍生的记录中保留那些和初始记录一致的、没有改变的、且在新的记录背景有意义的变量值。请注意，确实有些情形，保留来自源记录的值对于衍生的记录没有任何意义，那么就没必要保留那些值。

例如，假定分析的终点值被定义为取最后两个可用的基线后观测值的平均值。平均。在这种情况下，应该添加一条新的行，在AVISIT中做相应的描述，并且设置DTYPE（衍生类型）来描述那一行，比如“平均值”，说明此行是衍生的，同时也指出了衍生方法。和AVISIT=Endpoint相关联的元数据应该充分的描述衍生平均值需要用到的哪几条记录。请注意，虽然体重的对数变换的一组记录被衍生了，但是DTYPE不是不在每一行都填充，DTYPE应该用来说明在一个给定PARAM值的内部衍生的行，而不是不用来说明记录是否在SDTM中存在。许可变量PARAMTYP可以用来说明整个参数是衍生的。

在表4.2.1.4中，VISITNUM没被保留在衍生记录中，因为VISITNUM对于前身记录来说不是一个常数，并且在衍生分析时间点也没有任何意义，分析时间点是一个平均值，在大多数情况将跨越多个访视。同样的，VSSEQ对于多个初始记录不是一个常数，所以VSSEQ没有在衍生的记录被填充。PARAM和BASE应该被保留，因为他们在前身记录是常数，在新的记录背景下有意义。对于新的记录，AVAL和变化被重新计算，且AVISIT, AVISITN, 和DTYPE被适当填充。请注意，元数据需说明将指用于计算的算法（在此例中，行被平均）。

AVISIT和AVISITN由申办者定义，在一个数据集内，AVISIT和AVISITN不必对于单独的参数定义为相同。AVISIT值的定义和衍生，及任何对于参数的依赖，都应该在元数据中描述清楚。在此例中，申办者决定对于衍生的AVISIT=Endpoint的记录，AVISITN的值设为9999。



800 表4.2.1.4 规则3示例：创建一个新的行以处理衍生的分析时间点

行	PARAM	AVISIT	AVISITN	VISITNUM	VSSEQ	ABLFL	AVAL	BASE	CHG	PARAMTY P	DTYPE
1	重量（千克）	筛选	-4	1	1164		99	100	.		
2	重量（千克）	入组	-2	2	1165		101	100	.		
3	重量（千克）	基线	0	3	1166	是	100	100	0		
4	重量（千克）	第24周	24	4	1167		94	100	-6		
5	重量（千克）	第48周	48	5	1168		92	100	-8		
6	重量（千克）	第52周	52	6	1169		95	100	-5		
7	重量（千克）	终点	9999				93.5	100	-6.5		平均值
8	Log10（重量（千克））	筛选	-4	1	1164		1.9956	2	.	衍生	
9	Log10（重量（千克））	入组	-2	2	1165		2.0043	2	.	衍生	
10	Log10（重量（千克））	基线	0	3	1166	是	2	2	0	衍生	
11	Log10（重量（千克））	第24周	24	4	1167		1.9731	2	-0.0269	衍生	
12	Log10（重量（千克））	第48周	48	5	1168		1.9638	2	-0.0362	衍生	
13	Log10（重量（千克））	第52周	52	6	1169		1.9777	2	-0.0223	衍生	
14	Log10（重量（千克））	终点	9999				1.9708	2	-0.0292	衍生	平均值

801 在有记录值级别（记录级别）人群标识的情况，对规则3进行扩展是必要的。例如，假定在统计分析计划中说到，如果受试者在一个访视之前7天没有用药，那个  
802 访视收集的测量值将不被包括在符合方案分析中。那么对于某些受试者，最后两个可用的值可能对于意向性治疗（ITT）和符合方案分析（PP）是不同的，所以计算出来的  
803 终点平均值就会不同。对于这样的受试者，需要衍生两个不同的终点行，对于每种分析则由记录级别的人群标识ITTRFL和PPROTFL标出适当的行。

804  
805 在表4.2.1.5中，被分析的终点值根据人群而变化。例如，对于PARAM=体重（千克），最后两个可用的ITT值是92和95，平均值为93.5。然而在符合方案集  
806 中的最后两个值是94和92，平均值为93。这就是为什么对于这个受试者需要衍生两个终点行。对于其他的受试者，计算出的终点平均值，ITT和PP的数据可能  
807 相同，在那种情况，只需要一个终点记录，ITTRFL和PPROTFL值都标为“Y”“是”。 AVISIT和AVISITN的值都是申办者控制的。和在表4.2.1.4中的例子一样，  
808 申办者决定对于衍生的AVISIT=Endpoint记录，AVISITN的值设为9999。请注意，元数据将说明用于计算的算法（在本例，行被平均）。  
809  
810

811 表4.2.1.5 规则3示例：在具有记录值级别的人群标识时，产生新的行以应对一个衍生的分析时间点

行	PARAM	AVISIT	AVISITN	VISITNUM	VSSEQ	ABLFL	AVAL	BASE	CHG	DTYPE	ITTRFL	PPROTFL
1	体重（千克）	筛选	-4	1	1164		99	100	.		是	是
2	体重（千克）	入组	-2	2	1165		101	100	.		是	是
3	体重（千克）	基线	0	3	1166	是	100	100	0		是	是
4	体重（千克）	第24周	24	4	1167		94	100	-6		是	是
5	体重（千克）	第48周	48	5	1168		92	100	-8		是	是
6	体重（千克）	第52周	52	6	1169		95	100	-5		是	
7	体重（千克）	终点	9999				93.5	100	-6.5	平均值	是	

行	PARAM	AVISIT	AVISITN	VISITNUM	VSSEQ	ABLFL	AVAL	BASE	CHG	DTYPE	ITTRFL	PPROTRFL
8	体重 (千克)	终点	9999	1	1164		93	100	-7	平均值		是
9	Log10 (体重 (千克))	筛选	-4	2	1165		1.9956	2	.		是	是
10	Log10 (体重 (千克))	入组	-2	3	1166	是	2.0043	2	.		是	是
11	Log10 (体重 (千克))	基线	0	4	1167		2	2	0		是	是
12	Log10 (体重 (千克))	第24周	24	5	1168		1.9731	2	-0.0269		是	是
13	Log10 (体重 (千克))	第48周	48	6	1169		1.9638	2	-0.0362		是	是
14	Log10 (体重 (千克))	第52周	52				1.9777	2	-0.0223		是	
15	Log10 (体重 (千克))	终点	9999				1.9708	2	-0.0292	平均值	是	
16	Log10 (体重 (千克))	终点	9999				1.9685	2	-0.0315	平均值		是

在表4.2.1.6的例子中，缺失的基线后值由最后一个观测值向前结转，也用了最差观测值向前结转。在此研究中，在第8周，有一个计划好的访视（访视号6），在那个访视中，应该采集血压。但是对于这个受试者，或者没有访视6，或者有访视6但是没有血压数据被采集。统计分析计划说，缺失的基线后数据应该用两种方法填补（衍生）：LOCF（最后观测值向前结转）和WOCF（最差观测值向前结转）。

对于最后观测值向前结转（LOCF），缺失的第8周的结果（VISITNUM 6）是通过把最近的较早基线后观测值向前结转，也就是VISITNUM 5的值。第8周的值是被填充的，可以由衍生类型(DTYPE)列的LOCF看出来。

对于最差观测值向前结转（WOCF），虽然非计划的VISITNUM 4.1的值没有被用来代表第2周的结果，但是它被用来填充第8周的结果，因为它是最差的基线后的观察值。

用于记录衍生方法中的精确算法（在此情形LOCF和WOCF）必须在DTYPE的元数据中指出。

可溯源性由添加SDTM VISITNUM和--SEQ列来增强。USUBJID和VSSEQ的组合提供了一个到SDTM VS域中的精确输入记录的链接。在衍生的LOCF和WOCF行中，VISITNUM和VSSEQ清楚地描述了这些值是从哪里来的。

在此例中还介绍了几个其它概念。分析相关天数(ADY)在此方案中被定义为是相对于第一次用药日期。在很多方案中（但是不是所有的方案），ADY和SDTM --DY变量（或者对于某些类型的数据是---STDY变量）相等。这里列出的数据说明，这个特定的受试者直到随机化两天之后才开始服药，所以ADY的值在随机化访视，访视3 (VISITNUM 3) 时是-2。对于SDTM研究日的情况是一样的，ADY没有0天。

在此方案，如果在一个分析时间窗有多个数据点，和事先确定的目标计划对相关天数最近的观察值就是是选来表示分析时间点的值。对于这个研究和参数，AWTARGET = 来自SDTM的VISITDY（计划的研究日），并且ADY=VSDY。AWTDIFF是ADY – AWTARGET的绝对值，实际上因为为没有0天进行了校正（所以如果ADY和AWTARGET具有不同的符号，那么AWTDIFF = |ADY - AWTARGET - 1|）。

对于AVISIT=第2周，观察到两个值，在研究的第13天和第17天（行4和5），第13天和目标天数14更近些，所以第13天记录（4行）被选来做分析，并用分析记录标识ANL01FL = Y标记。为了获得用于分析的更精确记录集合，一起使用了ANL01FL和其它选择变量。

AVISIT自己充当一个分析时间窗的描述功能。为识别在给定分析中要使用的记录，AVISIT、DTYPE和ANL01FL都需要用到用于。

在衍生的AVIST=第8周的记录中，AWTARGET被设为第8周的目标，并且相应的计算AWTDIFF。保留来源于原始记录的AWTARGET和AWTDIFF值没有意义。

表4.2.1.6 规则3的示例：产生新的行，以LOCF和WOCF方法填补缺失值

行	PARAM	AVISIT	AVISITN	VISITNUM	VSSEQ	ABLFL	AVAL	BASE	CHG	DTYPE	ADY	AWTARGET	AWTDIFF	ANL01FL
1	收缩压（毫米汞柱）	筛选	-4	1	3821		120	114	.		-30	-28	2	是
2	收缩压（毫米汞柱）	入组	-2	2	3822		116	114	.		-16	-14	2	是
3	收缩压（毫米汞柱）	第0周	0	3	3823	是	114	114	0		-2	1	2	是
4	收缩压（毫米汞柱）	第2周	2	4	3824		118	114	4		13	14	1	是
5	收缩压（毫米汞柱）	第2周	2	4.1	3825		126	114	12		17	14	3	
6	收缩压（毫米汞柱）	第4周	4	5	3826		122	114	8		23	28	5	是
7	收缩压（毫米汞柱）	第8周	8	5	3826		122	114	8	LOCF	23	56	33	是
8	收缩压（毫米汞柱）	第8周	8	4.1	3825		126	114	12	WOCF	17	56	39	是
9	收缩压（毫米汞柱）	第12周	12	7	3827		134	114	20		83	84	1	是

表4.2.1.7的数据是偏头痛的例子，支持相对于基线的改变的分析。在此研究，缺失的基线后数据由基线观察向前结转（BOCF）和最后观测向前结转（LOCF）方法填补。

当偏头痛发生时，受试者自我服用一份单剂量的盲态研究药物。受试者在如下时间点对偏头痛进行评分：用药前，用药后30分钟，用药后1小时，以及用药后2小时。收集的偏头痛数据放置在SDTM临床发现（CF）域中。

ATPT是对分析时间点的描述。ATPTN是分析时间点数。CFTPTNUM是来自SDTM的收集的时间点数。AVALC包含疼痛评价，以及AVAL包含评估的数字编码值。AVAL和AVALC之间是一一对应的映射关系。用药一小时后受试者000276不再继续提供数据。对于这个受试者，用药2小时后的那个计划的观测值必须被填补。

受试者001863具有完整的数据，所以不必填补。

受试者000276在用药后两小时从偏头痛的观察病例分析中排除。

这两个受试者的数据都包括在用药后两小时偏头痛的BOCF和LOCF分析中。

**表4.2.1.7 规则3示例：产生新行，处理缺失数据的基线观测向前结转和最后观测向前结转的填补问题**

行	USUBJID	TRTP	PARAM	ATPT	ATPTN	FATPTNUM	FASEQ	ABLFL	AVAL	AVALC	BASE	CHG	DTYPE
1	000276	安慰剂	偏头痛	用药前	0	1	14	是	3	严重疼痛	3	0	
2	000276	安慰剂	偏头痛	用药后30分钟	0.5	2	22		2	中度疼痛	3	-1	
3	000276	安慰剂	偏头痛	用药后1小时	1	3	27		1	轻微疼痛	3	-2	
4	000276	安慰剂	偏头痛	用药后2小时	2	1	14		3	严重疼痛	3	0	BOCF
5	000276	安慰剂	偏头痛	用药后2小时	2	3	27		1	轻微疼痛	3	-2	LOCF
6	001863	Soma 30毫克	偏头痛	用药前	0	1	638	是	3	严重疼痛	3	0	
7	001863	Soma 30毫克	偏头痛	用药后30分钟	0.5	2	639		1	轻微疼痛	1	-2	
8	001863	Soma 30毫克	偏头痛	用药后1小时	1	3	640		1	轻微疼痛	1	-2	
9	001863	Soma 30毫克	偏头痛	用药后两小时	2	4	641		1	轻微疼痛	1	-2	

表4.2.1.8包含的示例，是一个数据集中的一些列，用来支持一个两阶段交叉研究的分析。

在一个交叉试验设计中，按计划所有的受试者都要接受所有的研究治疗，治疗顺序是随机的。随机化。 如果在一个交叉设计研究中有两个治疗，就需要两个治疗阶段。

在这个例子在此例中，计划的访视是1（筛选并且开始安慰剂入组阶段）、2（第-2周，安慰剂入组阶段中途）、3（第03周，安慰剂入组结束和随机化）、4（第4周，第一个治疗阶段结束）和5（第8周，第二个治疗阶段结束）。在统计分析计划中，基线被定义为第-2周2周（访视2）和第03周（访视3）测量的平均值平均。这个基线在第一个和第二个交叉阶段的分析中都要被用到。USUBJID 0987\_4252没有访视2的测量，所以平均值是平均第0周3周（访视3）的测量。

在任何基线后整周时间窗，最后一个观察值被用来代表那个星期。例如，对于USUBJID 0987\_3984，访视5（第7行7）的值被用来代表AVISIT=第8周，而不是早些的访视4.1的值（第6行），因为这个访视4.1的值也是在第8周的时间视窗里被观测到的。在这个研究中，变量ANL01FL被用来标记对于给定的AVISIT具有多条记录时为进行分析而选择的记录，并且为了标记用于一个给定分析或总结的更精确记录集合，必须和其它的选择变量一起使用。

APERIODC是交叉阶段特征描述。

请注意，一般来说APERIODC和EPOCH不同。比如在没有治疗可分析的基线前这样的时间阶段，不会定义APERIOD/APERIODC。而且，在有些情形下，可能APERIOD的界限和EPOCH的界限不完全一致。举个简单的例子，比如要用于分析的最近治疗阶段与中止后记录相关联。来自ADSL的TRTSEQP是交叉治疗的计划的顺序，TRTP是一个给定阶段的被分析的计划的计划的治疗，只有只在两个阶段都有数据的受试者，才需要衍生两个终点记录。用在AVISITN的惯例是由申办者定义的。在这个例子在此例中，申办者已经决定对AVISITN变量，衍生的基线记录的值设为-8888，衍生的终点记录的值设为9999，其它记录的值设为星期号。

**表4.2.1.8 规则3示例：创建终点行，以便于一个交叉设计的分析**

行	USUBJID	PARAMCD	AVISIT	AVISITN	VISITNUM	DTYPE	ANL01FL
1	0987_3984	ALT	筛选	-4	1		Y
2	0987_3984	ALT	第-2周	-2	2		Y

行	USUBJID	PARAMCD	AVISIT	AVISITN	VISITNUM	DTYPE	ANL01FL
3	0987_3984	ALT	第0周	0	3		Y
4	<b>0987_3984</b>	<b>ALT</b>	<b>基线</b>	<b>-8888</b>		平均值	<b>Y</b>
5	0987_3984	ALT	第4周	4	4		Y
6	0987_3984	ALT	第8周	8	4.1		
7	0987_3984	ALT	第8周	8	5		Y
8	<b>0987_3984</b>	<b>ALT</b>	<b>终点</b>	<b>9999</b>	<b>4</b>	终点	<b>Y</b>
9	<b>0987_3984</b>	<b>ALT</b>	<b>终点</b>	<b>9999</b>	<b>5</b>	终点	<b>Y</b>
10	0987_4252	ALT	筛选	-4	1		Y
11	0987_4252	ALT	第0周	0	3		Y
12	0987_4252	ALT	基线	-8888		AVERAGE	Y
13	0987_4252	ALT	第4周	4	4		Y
14	0987_4252	ALT	第8周	8	5		Y
15	<b>0987_4252</b>	<b>ALT</b>	<b>终点</b>	<b>9999</b>	<b>4</b>	AVERAGE	<b>Y</b>
16	<b>0987_4252</b>	<b>ALT</b>	<b>终点</b>	<b>9999</b>	<b>5</b>	AVERAGE	<b>Y</b>

881

行	TRTP	APERIOD	APERIODC	TRTSEQP	AVAL	ABLFL	BASE	CHG
1 (继续)				药物 B, 药物 A	16		17	.
2 (继续)				药物 B, 药物 A	16		17	.
3 (继续)				药物 B, 药物 A	18		17	.
4 (继续)				药物 B, 药物 A	17	是	17	0
5 (继续)	药物B	1	1期	药物 B, 药物 A	14		17	-3
6 (继续)	药物A	2	2期	药物 B, 药物 A	10		17	-7
7 (继续)	药物A	2	2期	药物 B, 药物 A	12		17	-5
8 (继续)	药物B	1	1期	药物 B, 药物 A	14		17	-3
9 (继续)	药物A	2	2期	药物 B, 药物 A	12		17	-5
10 (继续)				药物 A, 药物 B	12		11	.
11 (继续)				药物 A, 药物 B	11		11	.
12 (继续)				药物 A, 药物 B	11	是	11	0
13 (继续)	药物A	1	1期	药物 A, 药物 B	14		11	3
14 (继续)	药物B	2	2期	药物 A, 药物 B	15		11	4
15 (继续)	药物A	1	1期	药物 A, 药物 B	14		11	3
16 (继续)	药物B	2	2期	药物 A, 药物 B	15		11	4

882

#### 规则4. 同一个参数内多个行的函数应该被添加为一个新参数

规则4是规则2的一个特例。这条规则包含的函数违反了规则1的第2个条件（它们不是AVAL的相同行函数），并且也可能违反第1和第3个条件。

例如，在一个HIV疫苗的临床试验中，每次访视都要抽血样，并且要测量CD4的细胞数。为了评估有效性，给药之后的随访过程中，观察CD4细胞数随时间的累计效应就显得很重要。

设AVAL(t)等于在基线后访视t的CD4细胞数，并且设VISITDY(t)为访视t的计划的研天数。

在任何给定的基线后访视以如下方式计算CD4AUC（随访中累计的每日CD4数）：

- 将基线访视的CD4AUC设为0
- $CD4AUC(t) = CD4AUC(t-1) + [0.5 * AVAL(t-1) + 0.5 * AVAL(t)] * [VISITDY(t) - VISITDY(t-1)]$

CD4AUC不是一个简单的BASE和AVAL的同行函数，它基于来自多个观察值（行）的CD4的数据的计算，所以它应该被添加为一个新的行，而不是新的列。

CD4AUC在基线前没有定义，所以这个参数没有第-1周。CD4AUCMB（随访中CD4数相对于基线的累计平均变化）计算方法为：

- $CD4AUCMB(t) = CD4AUC(t) / [VISITDY(t) - 1] - CD4\text{细胞数的基线值}$

CD4AUCMB是CD4AUC和CD4的基线值的函数，所以它也必须以它自己作参数（参见下面的规则5）。CD4AUCMB在基线前和基线记录没有定义，因此这些记录在PARAM的这个值没有体现。

表4.2.1.9 规则4示例：产生一个新参数来处理一个参数多于一行的一个函数

行	PARAM	PARAMCD	AVISIT	VISITDY	ABLFL	AVAL	BASE
1	CD4（数/立方毫米）	CD4	第-1周	-7		75	76
2	CD4（数/立方毫米）	CD4	第0周	1	是	76	76
3	CD4（数/立方毫米）	CD4	第2周	15		128	76
4	CD4（数/立方毫米）	CD4	第4周	29		125	76
5	CD4（数/立方毫米）	CD4	第8周	57		191	76
6	CD4（数/立方毫米）	CD4	第12周	85		167	76
7	CD4（数/立方毫米）	CD4	第16周	113		136	76
8	CD4累计AUC	CD4AUC	第0周	1	是	0	0
9	CD4累计AUC	CD4AUC	第2周	15		1428	0
10	CD4累计AUC	CD4AUC	第4周	29		3199	0
11	CD4累计AUC	CD4AUC	第8周	57		7623	0
12	CD4累计AUC	CD4AUC	第12周	85		12635	0
13	CD4累计AUC	CD4AUC	第16周	113		16877	0
14	CD4累计AUC	CD4AUCMB	第2周	15		26	.
15	CD4累计AUCMB	CD4AUCMB	第4周	29		38.25	.
16	CD4累计AUCMB	CD4AUCMB	第8周	57		60.125	.
17	CD4累计AUCMB	CD4AUCMB	第12周	85		74.4167	.
18	CD4累计AUCMB	CD4AUCMB	第16周	113		74.6875	.

## 规则5. 多于一个参数的一个函数应该被添加为一个新参数

为了进行分析，经常有需要衍生一个没有收集的参数。这样的参数可能是多个SDTM域和域类的数据的很复杂的函数。规则5处理的情况是一个参数是从数据集中已经存在  
的其它参数衍生。

例如，一个调查问卷所有域总分由多于一个观察到的问题的函数计算。所有域总分应该被添加为一个新函数，和它的相应的衍生行的集合。为了这个衍生的参数，PARAM  
的值会是例如“所有域总分”，并且所有域总分的会被存储在标准的AVAL列，像往常一样，基线值会被存储在标准BASE列，相对于基线的变化会被存储在CHG。

在表4.2.1.10的例子，在每次访视都会抽取血样，并且实验室检验测量总的胆固醇和高密度脂蛋白胆固醇可以在SDTM LB域中找到。方案号召每一种单个实验室分析物的分析，也号召对于总的胆固醇（C）和高密度脂蛋白胆固醇（HDL）胆固醇比率的分析。分析数据集包含两个测量的实验室检验中每一个的参数，也包含一个新的衍生行的集合，  
在PARAM的描述是“总胆固醇：HDL-C比率”，并且AVAL在每个时间点包含计算的比率。

对于三个参数，感兴趣的是相对于基线变化的百分比，因此对所有记录都填充。但是，一般来说，如果变化百分比对于PARAM的一个特殊值没有被分析，那么没有  
必要在  
那些行填充PCHG。

表4.2.1.10 规则5图解：创建新参数来处理多于一个参数的一个函数

行	PARAM	PARAMCD	AVISIT	AVISITN	VISITNUM	LBSEQ	ABLFL	AVAL	BASE	CHG	PCHG
1	总胆固醇（毫克/分升）	CHOL	筛选	-2	1	39394		265	266	.	.
2	总胆固醇（毫克/分升）	CHOL	入组	-1	2	25593		278	266	.	.
3	总胆固醇（毫克/分升）	CHOL	第0周	0	3	23213	是	266	266	0	0.000
4	总胆固醇（毫克/分升）	CHOL	第2周	2	4	32952		259	266	-7	-2.632
5	总胆固醇（毫克/分升）	CHOL	第4周	4	5	12768		235	266	-31	-11.654
6	总胆固醇（毫克/分升）	CHOL	第8周	8	6	18773		242	266	-24	-9.023
7	总胆固醇（毫克/分升）	CHOL	第12周	12	7	28829		217	266	-49	-18.421
8	高密度脂蛋白胆固醇（毫克/分升）	HDL	筛选	-2	1	32437		44	42	.	.
9	高密度脂蛋白胆固醇（毫克/分升）	HDL	入组	-1	2	26884		40	42	.	.
10	高密度脂蛋白胆固醇（毫克/分升）	HDL	第0周	0	3	52657	是	42	42	0	0.000
11	高密度脂蛋白胆固醇（毫克/分升）	HDL	第2周	2	4	38469		43	42	1	2.381
12	高密度脂蛋白胆固醇（毫克/分升）	HDL	第4周	4	5	12650		47	42	5	11.905
13	高密度脂蛋白胆固醇（毫克/分升）	HDL	第8周	8	6	24345		46	42	4	9.524
14	高密度脂蛋白胆固醇（毫克/分升）	HDL	第12周	12	7	23484		47	42	5	11.905
15	总胆固醇:HDL-C比率	CHOLH	筛选	-2	1			6.023	6.333	.	.

行	PARAM	PARAMCD	AVISIT	AVISITN	VISITNUM	LBSEQ	ABLFL	AVAL	BASE	CHG	PCHG
16	总胆固醇:HDL-C比率	CHOLH	入组	-1	2			6.950	6.333	.	.
17	总胆固醇:HDL-C比率	CHOLH	第0周	0	3		是	6.333	6.333	0.000	0.000
18	总胆固醇:HDL-C比率	CHOLH	第2周	2	4			6.023	6.333	-0.310	-4.896
19	总胆固醇:HDL-C比率	CHOLH	第4周	4	5			5.000	6.333	-1.333	-21.053
20	总胆固醇:HDL-C比率	CHOLH	第8周	8	6			5.261	6.333	-1.072	-16.934
21	总胆固醇:HDL-C比率	CHOLH	第12周	12	7			4.617	6.333	-1.716	-27.100

#### 规则6. 当具有多于一个的基线定义时，每个附加的基线定义都要求创建它自己的行集合

在具有多于一个的基线定义的情况，必须为每一个另外附加的基线定义创建新的行。因此会有多组行，每组行相对于基线的一个特定的定义。无论何时具有多于一个的基线定义，BASETYPE列都是必需的。BASETYPE指定和每个行BASE的值对应的基线的定义。只有一个BASE列，并且对于每个具有资格的AVAL和BASE的函数只有一列。

表4.2.1.11中的例子介绍了一个数据集，它支持相对于三个不同基线的移动分析，因此它使用上面描述的BASETYPE变量，也阐释了ANRIND, BNRIND, 和SHIFTy变量。因为空间的缘故，没有显示ANLzzFL变量，虽然在一个分析时间点具有多个观察记录的情况需要确定选择哪个记录，对于这个受试者和参数，AVISIT=WEEK 12 (DB)就是这种情况。

表4.2.1.11 规则6的图解：创建新的行来处理多基线定义

行	BASETYPE	EPOCH	AVISIT	VISIT	AVAL	ANRLO	ANRHI	ANRIND	ABLFL	BASE	BNRIND	SHIFT1
1	入组	入组	基线（入组）	基线	34.5	15.4	48.5	正常	是	34.5	正常	正常到低
2	入组	入组	第8周（入组）	第57天	11.6	15.4	48.5	低		34.5	正常	正常到低
3	入组	入组	终点（入组）	第57天	11.6	15.4	48.5	低		34.5	正常	正常到低
4	入组	稳定性	第14周（稳定性）	第99天	13.1	15.4	48.5	低		34.5	正常	正常到低
5	入组	稳定性	终点（稳定性）	第99天	13.1	15.4	48.5	低		34.5	正常	正常到低
6	入组	双盲	基线（双盲）	第99天	13.1	15.4	48.5	低		34.5	正常	正常到低
7	入组	双盲	第12周（双盲）	第184天	13.7	15.4	48.5	低		34.5	正常	正常到正常
8	入组	双盲	第12周（双盲）	访视98天	19.7	15.4	48.5	正常		34.5	正常	正常到正常
9	入组	双盲	终点（双盲）	访视 98 天	19.7	15.4	48.5	正常		34.5	正常	正常到正常
10	入组	不设盲	基线（不设盲）	访视 98 天	19.7	15.4	48.5	正常		34.5	正常	正常到正常
11	入组	不设盲	第24周（不设盲）	第169天	28.1	15.4	48.5	正常		34.5	正常	正常到正常
12	入组	不设盲	终点（不设盲）	第169天	28.1	15.4	48.5	正常		34.5	正常	正常到正常
13	双盲	双盲	基线（双盲）	第99天	13.1	15.4	48.5	低	是	13.1	低	
14	双盲	双盲	第12周（双盲）	第184天	13.7	15.4	48.5	低		13.1	低	低到低
15	双盲	双盲	第12周（双盲）	访视98天	19.7	15.4	48.5	正常		13.1	低	低到正常
16	双盲	双盲	终点（双盲）	访视98	19.7	15.4	48.5	正常		13.1	低	低到正常



行	BASETYPE	EPOCH	AVISIT	VISIT	AVAL	ANRLO	ANRHI	ANRIND	ABLFL	BASE	BNRIND	SHIFT1
17	双盲	不设盲	基线（不设盲）	访视98	19.7	15.4	48.5	正常		13.1	低	低到正常
18	双盲	不设盲	第24周（不设盲）	第169天	28.1	15.4	48.5	正常		13.1	低	低到正常
19	双盲	不设盲	终点（不设盲）	第169天	28.1	15.4	48.5	正常		13.1	低	低到正常
20	不设盲	不设盲	基线（不设盲）	访视98	19.7	15.4	48.5	正常	是	19.7	正常	
21	不设盲	不设盲	第24周（不设盲）	第169天	28.1	15.4	48.5	正常		19.7	正常	正常到正常
22	不设盲	不设盲	终点（不设盲）	第169天	28.1	15.4	48.5	正常		19.7	正常	正常到正常

## 4.3 分析包含一个参数所有观察的和衍生记录 vs 记录的子集

本节讨论分析数据集是否应该包含一个分析参数的所有行（即记录），或者只是把其的一部分（即子集）用于分析。在一个具体的时间点对于一个具体分析参数，其AVAL或AVALC中的一个值可能被观察到（即从病例报告表或在那个时间点的电子日记中收集），它若缺失则可以填补，或者可以衍生自其它值的组合。

为了解释提出的问题，假定调查问卷A的总分（在访视1、2和3给药）收集如下表所示的SDTM QS域中。任何缺失的总分采用基线后（访视1后）最后一个总分向前结转的方法填补。而访视3的总分将被用于分析。

在如下表所示的SDTM QS域数据中，受试者0001在访视1、2和3均有数据；受试者0002因没有基线后数据，将不被包含在分析中；而受试者0003在访视1和2有数据，但是访视3数据缺失。

表4.3.1 SDTM QS数据集中问题数据示例

行	DOMAIN	USUBJID	VISITNUM	QSSEQ	QSCAT	QSTESTCD	QSSTRESN
1	QS	0001	1	101	QUES-A	TOTSCORE	7
2	QS	0001	2	201	QUES-A	TOTSCORE	12
3	QS	0001	3	555	QUES-A	TOTSCORE	14
4	QS	0002	1	91	QUES-A	TOTSCORE	4
5	QS	0003	1	156	QUES-A	TOTSCORE	2
6	QS	0003	2	300	QUES-A	TOTSCORE	6

问题描述：受试者0002因没有基线后数据，而不被纳入分析，但它的数据是否应被包含在分析数据集中；如果被分析的数据来自访视3，访视1和访视2的数据亦是否亦应包含在分析数据集中。

### 4.3.1 ADaM 方法和实例

ADaM采用的方法是对一个给定的分析参数包括所有的观察行与衍生行。在分析数据集中包含所有行，也包括那些在分析中没有用到的，这就要求为一个特定的分析确认要用到的行方式。这个方法增加了数据集的规模，且引入了一个潜在的风险，即用户可能选择了不合适的行，从而产生不正确的分析结果。包含所有行的好处是能更容易的验证所选择的数据，及某时间点衍生数据的准确性，提供了有用的数据可溯源性。另外，这些数据也可以用来帮助启动其它分析，包括敏感性分析。

如果可能并且是合理的，监管审阅者们倾向于创建及/或选择分析行遵循的路径被清晰的描述，且能一直溯源到SDTM域中的初始行。只是简单的把算法包括在元数据中是不充分的，因为任何复杂的数据处理方式可能不能被清楚的加以表达（例如，如何处理输入数据缺失）。在一个标准的BDS数据集中，保留分析参数的所有观察行和衍生行，能以最灵活的、最清楚的方式提供数据的可溯源性，该分析数据集也为审阅者为检验某一分析的稳健性而提供了最大的灵活性（例如，使用不同的填补方法）。

例1：在上面讨论的例子中（表4.3.1），分析数据集将包含如下行（表4.3.1.1）：

表4.3.1.1 例1：分析数据集

行	PARAMCD	USUBJID	VISITNUM	AVISITN	AVISIT	AVAL	DTYPE
1	TOTSCORE	0001	1	1	访视1	7	
2	TOTSCORE	0001	2	2	访视2	12	
3	TOTSCORE	0001	3	3	访视3	14	
4	TOTSCORE	0002	1	1	访视1	4	
5	TOTSCORE	0003	1	1	访视1	2	
6	TOTSCORE	0003	2	2	访视2	6	
7	TOTSCORE	0003	2	3	访视3	6	LOCF

对于上面讨论的分析，分析数据将通过条件AVISITN = 3 (或者AVISIT=访视3)加以选择。

应该注意到这个方法并不要求包含来自输入数据的所有行。例如，如果输入数据集包含几个不同调查问卷的数据，外部数据（例如，不是重点提到的调查问卷的其它调查问卷）不必一定包含在数据集中。

**例2：**在下例中（表4.3.1.2和表4.3.1.3），Q01评估被安排在访视1、3、5和7进行，并且针对这些访视结果进行分析。受试者1099在访视1、2和7有评估数据。（注意到虽然评估没有安排在访视2进行，但数据显示该受试者在那个时间点进行了评估）。受试者2001不在全分析集中。受试者3023在访视5具有两个评估，但研究分析计划中指出，只有一个访视内的最早观测将被分析。但因该受试者没有访视7数据，访视5最后一行记录将被结转到访视7。SDTM域作为分析数据集的基础，具有如下观测行：

**表4.3.1.2 例2：SDTM QS数据集中的数据**

行	QSTESTCD	USUBJID	QSEQ	VISITNUM	VISIT	QSSTRESN	QSDTC
1	Q01	1099	111	1	基线	25	2005-04-04
2	Q01	1099	121	2	访视2	24	2005-05-02
3	Q01	1099	132	7	访视7	15	2005-08-22
4	Q01	2001	150	1	基线	27	2005-02-05
5	Q01	3023	117	1	基线	31	2005-06-30
6	Q01	3023	123	3	访视3	29	2005-07-25
7	Q01	3023	134	5	访视5	28	2005-08-20
8	Q01	3023	135	5	访视5	25	2005-08-21

分析数据集包含相应于SDTM中能找到的那些行，也包括用LOCF方法为缺失的访视评估创建的衍生行，连同标识及识别所需分析行的其它列：

**表4.3.1.3 例2：分析数据集**

行	USUBJID	VISITNUM	VISIT	AVISITN	AVISIT	AVAL	DTYPE	ANL01FL	FASFL
1	1099	1	基线	1	基线	25		是	是
2	1099	2	访视2			24			是
3	1099	2	访视2	3	访视3	24	LOCF	是	是
4	1099	2	访视2	5	访视5	24	LOCF	是	是
5	1099	7	访视7	7	访视7	15		是	是
6	2001	1	基线	1	基线	27		是	是
7	3023	1	基线	1	基线	31		是	是
8	3023	3	访视3	3	访视3	29		是	是
9	3023	5	访视5	5	访视5	28		是	是
10	3023	5	访视5	5	访视5	25			是
11	3023	7	访视7	7	访视7	25	LOCF	是	是

应用于上述例子的数据选择标准为：

- DTYPE为空：说明数据能在SDTM域中发现。
- DTYPE="LOCF"：具体指出了用于添加衍生行的方法，并且标记哪些行是衍生的。
- FASFL="是"：识别全分析数据集成员的受试者。
- ANL01FL="是"：识别被选择用来代表每个AVISIT的行。对于受试者3023在AVISITN=5有多个观察值，因此在此例中，ANL01FL="是"的行是那些已经被选择代表各自分析时间点的行。
- 对于受试者1099的VISIT="访视2" (行2)ANL01FL=空，因为访视2是这个调查问卷没有被安排的访视，且访视2将不被显示在分析中，AVISITN和AVISIT也为空，因为根据研究分析计划，访视2将不被分析。
- "(ANL01FL="是" 与FASFL="Y" 与AVISITN=5)"识别访视5数据FAS分析中用到的行。

### 考虑过但没采用的方法

考虑过的另外一个方法是只在分析数据集中包括实际用于参数分析的那些行。在上面的例1，只有或者被观察、或者用LOCF方法衍生的访视3行会被包括在分析数据集中。这个方法的主要优点是能够简化某些分析，因为不需要对分析的行加以选择。但主要缺点是评审者失去了对分析稳健性加以验证的灵活性，同时导致数据的不可溯源性。因为这些缺点，没有采纳这种方法。

## 4.4 在分析数据集中包含不用于分析但与衍生数据有关的输入数据

4.3节说明对于一个给定的分析参数，数据集中应该包括那个参数的所有观察和衍生行，而不只是在分析中用到的行。

4.3节举了一个简单的例子，在本节（4.4节）里将讨论更为一般性的问题。

本节将讨论更广泛的问题，即一个分析数据集是否应该包含用于生成分析数据的输入数据以及用于分析的实际数据，包括：

- 输入数据行和列，用于追踪分析所用的行和列，及
- 自身没有被分析，但是被用于衍生一个分析参数的原始参数或者衍生参数的前身。

上面的输入数据的行和列根据生成新的分析数据的不同需要，可能是来自SDTM的一个域或者多个域，如分析参数所描述的一样。

### 4.4.1 ADaM 方法和实例

创建分析数据集是为了便于分析的进行，SDTM是源数据，因此评审者期望SDTM域和分析数据集之间具有一定程度的可溯源性是符合逻辑的。

为了获得期望的可溯源性，ADaM方法是在元数据中描述衍生算法，并且如果实际可行，可适当地增加一些行来支持这种可溯源性。为了将输入数据作为行包含于分析数据集中，应该在适当的地方添加列以指出输入数据的来源 – 域、变量名称、和序号。尽管这个方法使数据集变大，并增加了选择分析所需行的复杂性，但它却提供了输入数据，并可直接对输入数据进行操作。另外，如果用适当的标帜区分中间值，也可以保留这些中间值。

一般来说，强烈推荐包括尽量多的可溯源性所需要的支持性数据，但是这样做可能有时行不通。例如，如果一个被分析的参数是从大量的原始电子日记输入记录中得到的一个总结，则要分析数据集中包含所有的原始电子日记记录可能既没用也不实际。

本节的剩余部分讨论了分析数据集中不仅包含分析数据，也包含了输入数据，用于更清楚地跟踪生成这些分析数据的算法。除了用于分析的实际值，数据集可能包括没用于分析的行、包含输入数据的行、和包含在分析数据衍生过程中计算中间值的行。标帜或其它列被用于区分各种数据类型，也提供了一个从输入数据到用于分析的值的一个跟踪路径。分析结果元数据规定了如何识别适当的行（通过一个特定的选择条件），4.5节和4.6节将讨论如何确定分析中所用的行。

除非在数据行中已经有列变量（如协变量或者支持性变量）来源于输入数据，否则输入数据仍将作为行保留在分析数据集中。所保留的输入数据行上的分析值列（AVAL和/或AVALC）含有所要分析的参数值。分析数据集并不需要包含来自输入数据集的全部列，而只需包含一些变量来指明输入数据的来源- 域、变量名称和序列号。这种方法允许包括来自多个域的输入数据。如果是所要分析的参数行中已经含有列变量（如协变量或者支持性变量）来源于输入数据，就无需再重复包含那些输入数据集中的相应的行。所以，将输入数据作为行保留还是作为列保留取决于输入数据的类型以及输入数据在分析数据集中是否有其它用途。将决定分析参数值的所有数据放在一个数据集里可在标准的ADaM BDS中以最灵活的方式提供最清晰的可溯源性。这个大的数据集也使得审阅者能够灵活地验证分析的稳健性。

如果觉得这个数据集太大太复杂，申办者可以提供两个数据集，一个包含所有的行，另一个是第一个的子集，只包含用于特定分析的行。为确保可溯源性，子分析数据集的元数据中要指明其前身是完全分析数据集。虽然这种方法提供了所需的可溯源性，也提供了一个不需要指定选择条件的分析数据集，但总的文件大小却变大了。而且，数据集的创建者要确保两个数据集之间的一致性，并对两个数据集都要进行验证，如果未提供分析结果的元数据，也可能会分不清是用哪个数据集来分析的。

#### 例1

该分析数据集是为了分析高血压事件的发生时间。分析参数是发生高血压事件的研究天数，定义为从受试者进入研究到发生下列事件中的最早者之间的研究天数：入院、舒张压高于90mmHg、和收缩压高于140mmHg。如果一个受试者没有发生这些事件，分析时该受试者将被删失处理，研究天数为从研究开始到他/她退出研究的那天为止。

表4.4.1.1 例1: SDTM VS数据集中的数据

行	USUBJID	VISITNUM	VSSEQ	VSDTC	VSDY	VSTESTCD	VSSTRESN
1	2010	1	22	2004-08-05	1	SYSBP	115
2	2010	1	23	2004-08-05	1	DIABP	75
3	2010	2	101	2004-08-12	8	SYSBP	120
4	2010	2	102	2004-08-12	8	DIABP	90
5	2010	3	207	2004-08-19	15	SYSBP	135
6	2010	3	208	2004-08-19	15	DIABP	92
7	2010	4	238	2004-08-25	21	SYSBP	138
8	2010	4	239	2004-08-25	21	DIABP	95
9	3082	1	27	2004-09-08	1	SYSBP	120
10	3082	1	28	2004-09-08	1	DIABP	80
11	3082	2	119	2004-09-15	8	SYSBP	125
12	3082	2	120	2004-09-15	8	DIABP	84

表4.4.1.2 例1: SDTM DS数据集中的数据

行	USUBJID	DSSEQ	DSSTDT	DSSTDY	DSDECOD	DSTERM
1	2010	25	2004-08-05	1	RANDOM	受试者被随机化
2	2010	99	2004-08-13	9	HOSPSTRT	受试者入院
3	2010	140	2004-08-15	11	HOSPEND	受试者出院
4	2010	199	2004-08-20	16	HOSPSTRT	受试者入院
5	2010	225	2004-08-22	18	HOSPEND	受试者出院
6	2010	301	2004-08-26	22	COMPLETED	受试者完成试验
7	3082	20	2004-09-08	1	RANDOM	受试者被随机化
8	3082	130	2004-09-17	10	COMPLETED	受试者完成试验

分析数据集包含用于推导分析参数“HYPEREVT”的子事件数据。表4.4.1.3举例说明了这种ADaM方法。使用这个方法，可以包括所有的子事件作为分析参数（即行），并且创建输入域、输入变量、和输入次序列（SRC\*列）来识别这些行的来源。PARAMCD=“HOSPADM”的AVAL是入院的最早相对日。PARAMCD=“DBP”的AVAL是舒张压超过90的最早相对日。PARAMCD=“SBP”的AVAL是收缩压超过140的最早相对日。如果一个受试者没有经历一个特定的子事件，仍然会为那个子事件创建一行说明受试者在退出研究的那天被删失（CNSR=1），而SRC\*开头的几列引用了DS数据集。

PARAMCD=“HYPEREVT”的AVAL被推导为HOSPADM, DBP和 SBP三个事件中的最早者（即三者中CNSR=0时AVAL最小的值将被定为三类事件的最早相对日）；受试者只要满足这三个条件之一并且当PARAMCD=“HYPEREVT”的CNSR=0时即说明受试者经历了一个事件。如果受试者不满足三个条件中的任何一个条件（即所有的三个记录都有CNSR=1），那么这个受试者被删失，也就是PARAMCD=“HYPEREVT”的AVAL被推导为受试者退出研究的相对日并且用CNSR=1来表示受试者被删失。分析将以HYPEREVT为中心，但是数据集里也包括了变量HOSPADM, DBP和SBP以支持可溯源性，如果想对子事件进行进一步分析，也是可以的。这个结构的主要好处是它仅仅通过增加3个标准的支持性列就能处理来自多个域的子事件输入行。该方法之所以比较好是因为它是标准化的、可度量的、且支持子事件的分析。

表4.4.1.3 例1: 分析数据集

行	USUBJID	PARAM	PARAMCD	AVAL	CNSR	EVNTDESC	SRCDOM	SRCVAR	SRCSEQ
1	2010	到达第一次入院的时间	HOSPADM	9	0	第一次入院	DS	DSSTDY	99
2	2010	第一次DBP>90的时间（天数）	DBP	15	0	第一次DBP>90	VS	VSDY	208
3	2010	第一次SBP>140的时间（天数）	SBP	22	1	完成试验	DS	DSSTDY	301
4	2010	第一次高血压事件的时间（天数）	HYPEREVT	9	0	高血压事件	DS	DSSTDY	99
5	3082	第一次入院的时间（天数）	HOSPADM	10	1	完成试验	DS	DSSTDY	130
6	3082	第一次DBP>90的时间（天数）	DBP	10	1	完成试验	DS	DSSTDY	130

行	USUBJID	PARAM	PARAMCD	AVAL	CNSR	EVNTDESC	SRCDOM	SRCVAR	SRCSEQ
		间（天数）							
7	3082	第一次SBP>140的时间（天数）	SBP	10	1	完成试验	DS	DSSTDY	130
8	3082	第一次血压高事件的时间（天数）	HYPER EVT	10	1	完成试验	DS	DSSTDY	130

## 例2

分析参数是肾小球滤过率（GFR），用MDRD（Modification of Diet in Renal Disease Study Group，肾脏疾病膳食改进研究）方程从血清肌酐估计。该方程要用到来自LB域的血浆肌酐、血中尿素氮（BUN）、白蛋白值，也要用到年龄、种族和性别。

表4.4.1.4 例2：SDTM LB数据集中的数据

行	USUBJID	VISITNUM	LBSEQ	LBTEST	LBTESTCD	LBSTRESN	LBSTRESU
1	3000	3	98	肌酐	CREAT	78.2	微摩尔/升
2	3000	36	115	血中尿素氮	BUN	9.1	微摩尔/升
3	3000	3	120	白蛋白	ALB	40	克/升

因为年龄、种族和性别是分析数据集中的协变量，所以数据集中不对这些输入数据再新建行。分析记录通过PARAMCD=MDRD\_GFR来标识，相应的参数编码是PARAM=肾小球滤过率（GFR）（毫升/分/1.73平方米）（请注意此处因为空间有限，表4.4.1.5中没有列出PARAM列）。

表4.4.1.5 例2：分析数据集

行	USUBJID	AGE	SEX	RACE	PARAMCD	VISITNUM	AVAL	SRCDOM	SRCVAR	SRCSEQ
1	3000	52	F	亚裔	CREAT	3	78.2	LB	LBSTRESN	98
2	3000	52	F	亚裔	BUN	3	9.1	LB	LBSTRESN	115
3	3000	52	F	亚裔	ALB	3	40	LB	LBSTRESN	120
4	3000		F	亚裔	MDRD_GFR	3	76.77			

## 考虑过但没采用的方法

第二种被考虑过的方法是在元数据中描述推导算法并且在分析数据集中将输入数据作为列添加进来。还要增加一些指针列以标识输入数据的来源：变量名称和编号。这种方法允许所有的相关输入数据出现在相关的分析行中（即所有的子事件会在同一行作为一个复合事件呈现），这可能会有助于简化对分析参数的计算进行核对，但是这个方法会明显的增加分析数据集中的列数，且要求对变量的命名要简单明了。这个方法也假设需要保留的数据只是原始输入数据。此方法的另一个弊端是如果在将来需要分析子事件，还必要在数据集中添加子事件参数，以得到一个符合ADaM标准的数据结构以支持子事件的分析。鉴于上述原因，最终该方法没被采用。

第三种被考虑的方法是在元数据中描述衍生算法，但分析数据集中不包括输入数据或输入数据的标识。这个方法的好处是简化分析数据集。但是因为分析数据集的结构过于简化，使得研究中收集的数据（即SDTM域）和分析的数据（即分析数据集）之间的可溯源性无法保持。除非元数据中描述的推导算法简单明了，否则对分析数据计算的核对可能是非常困难甚至是不可能的，也不应使用这种方法。

## 4.5 分析所用行的标识

本节处理如何标识用于分析的一个数据集的行，要解决的四个具体问题包括：1）最后观测向前结转（LOCF）分析的行的标识；2）包含基线值的行的标识；3）基线后概念时间点行的标识，比如终点、最小值、最大值、或者平均值；和 4）分析中所用到的某些特定行的标识。

### 4.5.1 用于时间点填补分析的行的标识

本节讨论如何标识与时间点相关的填补分析中所用到的行以及如何在分析数据集中包含缺失时间点的所填补的数据。最后观测向前结转（LOCF）是最常用的一个时间点相关的填补分析之一，因此这里特别加以说明。但是所用的方法仍具有普遍性，并不只是限于LOCF分析。为了强调普遍性，本节也对最差

观测向前结转（WOCF）分析进行了讲解。

#### 4.5.1.1 ADaM方法和实例

当一个分析时间点缺失，ADaM所采取的方法是在分析数据集中去创建一个新行来表示这个缺失的时间点，并且通过填充推导类型变量DTYPE来标识这些新建的填补行。例如当进行LOCF/WOCF分析时，在LOCF/WOCF分析时间点缺失时要创建LOCF/WOCF行，并且通过对推导类型变量DTYPE赋值LOCF或WOFC来标识这些填补的行，所有的原始行的DTYPE值为null。分析时选择所需的行就会很简单，如要对实际观测数据（DAO）进行分析时只需选择DTYPE = null的行，要进行LOCF分析时只需选择DTYPE = null或者LOCF的行，而要进行WOFC分析时只需选择DTYPE = null或者WOFC的行就行了。对这种方法需要有足够的理解和沟通，如果DTYPE标识没有被正确的引用，分析会默认使用所有的行，包括DAO行，加上由LOCF和WOFC方法所填补的行。要进行一个正确的DAO分析，需要明确的选择DTYPE = null。

##### 例1：标识LOCF分析所用的行

在下例中（表4.5.1.1.1），一些受试者具有完全的数据，另一些受试者的一些数据行则是通过一种方法（LOCF）填补的。没有缺失数据的受试者所有的数据行都是观察到的，其DTYPE值都为空。受试者1001具有完全的数据，其所有行的DTYPE都为空，说明这些行都不是填补的。在此例中，AVISIT和VISIT（来自SDTM）相同。但即使在没有缺失数据的情况下，AVISIT和来自SDTM的VISIT也不总是完全一致的。受试者1002缺失第二周的评估，第二周的数据使用LOCF方法进行了填补。AVISIT=第2周但是VISIT=第1周，所以我们可以看到填补值是用原始数据的哪些数据进行填补的。受试者1003缺失了第二周和第三周的数据，DAO分析时可以通过选择那些DTYPE = null的行来实现，而LOCF分析则应选择所有行（DTYPE=null或者DTYPE="LOCF"）。

表4.5.1.1.1 例1： 标识LOCF分析所用行的分析数据集

行	USUBJID	VISIT	AVISIT	ADY	PARAM	AVAL	DTYPE
1	1001	基线	基线	-4	仰卧位收缩压(mmHg)	145	
2	1001	第1周	第1周	3	仰卧位收缩压(mmHg)	130	
3	1001	第2周	第2周	9	仰卧位收缩压(mmHg)	133	
4	1001	第3周	第3周	20	仰卧位收缩压(mmHg)	125	
5	1002	基线	基线	-1	仰卧位收缩压(mmHg)	145	
6	1002	第1周	第1周	7	仰卧位收缩压(mmHg)	130	
7	1002	第2周	第2周	7	仰卧位收缩压(mmHg)	130	LOCF
8	1002	第3周	第3周	22	仰卧位收缩压(mmHg)	135	
9	1003	基线	基线	1	仰卧位收缩压(mmHg)	150	
10	1003	第1周	第1周	8	仰卧位收缩压(mmHg)	140	
11	1003	第1周	第2周	8	仰卧位收缩压(mmHg)	140	LOCF
12	1003	第1周	第3周	8	仰卧位收缩压(mmHg)	140	LOCF

##### 例2：标识LOCF及WOFC分析所用的行

以下例子中的这些行（表4.5.1.1.2）显示了一种情况，即使用的填补方法不止一种。在这个例子中，为每种类型的填补都新增了数据行。DAO分析可以通过只选择那些DTYPE=null的那些行进行。对于LOCF分析，使用DTYPE=null或者DTYPE="LOCF"的那些行，对于WOFC分析，使用DTYPE=null或者DTYPE="WOFC"的那些行。

表4.5.1.1.2 例2： 用于LOCF也用于WOFC分析的行的识别的分析数据集

行	USUBJID	VISIT	AVISIT	ADY	PARAM	AVAL	DTYPE
1	1002	基线	基线	-4	仰卧位收缩压(mmHg)	145	
2	1002	第1周	第1周	3	仰卧位收缩压(mmHg)	130	
3	1002	第2周	第2周	9	仰卧位收缩压(mmHg)	138	
4	1002	第3周	第3周	18	仰卧位收缩压(mmHg)	135	
5	1002	第3周	第4周	18	仰卧位收缩压(mmHg)	135	LOCF
6	1002	第2周	第4周	9	仰卧位收缩压(mmHg)	138	WOFC
7	1002	第5周	第5周	33	仰卧位收缩压(mmHg)	130	
8	1003	基线	基线	-1	仰卧位收缩压(mmHg)	145	



行	USUBJID	VISIT	AVISIT	ADY	PARAM	AVAL	DTYPE
9	1003	第1周	第1周	7	仰卧位收缩压(mmHg)	140	
10	1003	第2周	Week 2	15	仰卧位收缩压(mmHg)	138	
11	1003	第2周	Week 3	15	仰卧位收缩压(mmHg)	138	LOCF
12	1003	第2周	Week 4	15	仰卧位收缩压(mmHg)	138	LOCF
13	1003	第2周	Week 5	15	仰卧位收缩压(mmHg)	138	LOCF
14	1003	第1周	Week 3	7	仰卧位收缩压(mmHg)	130	WOCF
15	1003	第1周	Week 4	7	仰卧位收缩压(mmHg)	130	WOCF
16	1003	第1周	Week 5	7	仰卧位收缩压(mmHg)	130	WOCF

## 考虑过但没有采用的方法

另一种曾被考虑过的方法是为每种类型的分析创建一个相应的行的“全集”(或者一个单独的数据集), 并通过为分析时间点描述AVISIT分配唯一值(例如“第4周”, “第4周(LOCF)”和“第4周(WOCF)”)来指明不同的分析类型。这种方法在进行DAO, LOCF,和WOCF分析时更为简单明了, 因为只需要引用AVISIT一个步骤就行了。但是该方法最大的不利是生成的分析数据集特别大, 因为要产生许多重复的行。另外, 这个方法可能对于分析工具不太友好, 因为可能需要检查AVISIT, 在其中搜索一个关键字子串, 例如“(LOCF)”。所以不应该使用这种方法。

创建一个标识变量(LOCFFL/LOCFFN)来指明某一数据行是用LOCF创建的, 对WOCF也用类似的方法处理。这种方法与ADaM规定的方法相似, 区别是这种方法为每个推导类型分别创建了一个的标识, 而不是在一个列(DTYPE)中指出数据行的推导类型。这种方法所创建的数据集的行数可能比推荐的方法少(例如, 如果WOCF行和LOCF行相同)。在其它方面, 这种方法和推荐的方法有相同的优缺点。不推荐使用这种为每个推导类型分别创建标识的方法。

## 4.5.2 标识基线行

许多统计分析要求标识一个基线值, 本节描述如何标识基线的记录。

### 4.5.2.1 ADaM方法和实例

ADaM方法是创建一个基线标识列来指出用于基线的行(AVAL值用来填充BASE变量的行)。如果基线行不是通过推导而产生的, 这种方法并不需要在数据集中产生重复的行。虽然可以创建基线行标识变量ABLFL来标识基线行, 这也不妨碍再增加一行, 该行的AVISIT具有唯一的值(如“基线”), 来表明用于分析的基线行, 即使增加这一行显得有些多余。对于比较复杂的基线定义(多行的函数), 应尽可能创建一个推导的基线行。这种方法要求为基线行变量提供清楚的元数据, 让基线行变量的值能被精确的再生。

#### 例1: 标识基线行 – 用筛选访视填补基线行

本例(表4.5.2.1.1)举例说明了用一个基线标识变量ABLFL来标识基线行, 同时也说明在数据集中为基线分析时间点增加一行(第6行)。在本例中, AVISIT为分析所用的基线记录定义了一个唯一的值。受试者1001具有完全数据, 在源数据中受试者1002没有合格的基线值记录。在数据集中添加了一个推导的基线记录(AVISIT=“基线”), 其DTYPE=“LVPD”(用药前的最后值结转), 指出该条记录是通过填补产生的, 将被用作基线记录。

表4.5.2.1.1 例1: 当用到填补时具有基线行标识的分析数据集

行	USUBJID	VISIT	AVISIT	ADY	ABLFL	PARAM	AVAL	DTYPE
1	1001	筛选	筛选	-12		仰卧位收缩压(mmHg)	144	
2	1001	基线	基线	1	是	仰卧位收缩压(mmHg)	145	
3	1001	第1周	第1周	6		仰卧位收缩压(mmHg)	130	
4	1001	第2周	第2周	12		仰卧位收缩压(mmHg)	133	
5	1002	筛选	筛选	-14		仰卧位收缩压(mmHg)	144	
6	1002	基线	基线	-14	是	仰卧位收缩压(mmHg)	144	LVPD
7	1002	第1周	第1周	8		仰卧位收缩压(mmHg)	130	
8	1002	第2周	第2周	14		仰卧位收缩压(mmHg)	133	



## 例2：标识基线行 - 用多个访视的平均值推导一个基线行

此例（表4.5.2.1.2）举例说明了当基线值为用药前收集的非缺失数据平均值时，使用一个基线标识变量ABLFL来标识分析中所用的基线记录。第3行是一个使用行1和行2的平均值而推衍生的“基线记录”，DTYPE = “AVERAGE”说明行3是推衍生的。基线标识(ABLFL=“是”)指出第3行的AVAL被用于填充BASE（基线）列。VISIT（来自SDTM）在第3行为空，因为那条记录的AVAL不是从另一条记录复制而来。

表4.5.2.1.2 例2：当基线是一个平均值时具有基线行标识的分析数据集

行	USUBJID	VISIT	AVISIT	ADY	ABLFL	PARAM	AVAL	BASE	DTYPE
1	1001	筛选	筛选	-12		仰卧位收缩压(mmHg)	144	144.5	
2	1001	基线	基线	1		仰卧位收缩压(mmHg)	145	144.5	
3	1001		基线		是	仰卧位收缩压(mmHg)	144.5	144.5	平均值
4	1001	第1周	第1周	12		仰卧位收缩压(mmHg)	130	144.5	
5	1001	第2周	第2周	-14		仰卧位收缩压(mmHg)	133	144.5	

## 例3：标识基线行 - 使用多个访视平均值推导一个基线行

本例（表4.5.2.1.3）和例2基本相同，区别是分析时间点的描述是“筛选/基线组合”，这样有助于区分推衍生的平均基线记录和一个已存在的时间点描述也是“基线”的实际观察记录。这对分析和报告是有帮助的，因为除了要总结平均基线访视之外，还要总结所有的预先安排的访视。AVISIT用不同的描述使得分析很直观明了，AVISIT值的选择由申办者决定。

表4.5.2.1.3 例3：具有基线行标识的分析数据集，分析时间点变量中含有基本行的描述

行	USUBJID	VISIT	AVISIT	ADY	ABLFL	PARAM	AVAL	BASE	DTYPE
1	1001	筛选	筛选	-12		仰卧位收缩压(mmHg)	144	144.5	
2	1001	基线	基线	1		仰卧位收缩压(mmHg)	145	144.5	
3	1001		筛选/基线组合		是	仰卧位收缩压(mmHg)	144.5	144.5	平均值
4	1001	第1周	第1周	12		仰卧位收缩压(mmHg)	130	144.5	
5	1001	第2周	第2周	-14		仰卧位收缩压(mmHg)	133	144.5	

## 4.5.3 基线后概念性时间点行的标识

当分析包含跨时间点的推导值，比如终点、最小值、最大值、和基线后平均值，应该考虑有些问题，比如“是不是总是应该创建使AVISIT具有唯一值的不同行，即使这一行与一个观测到的记录完全相同而导致冗余？还是只需标识这些行就行了？”在本节介绍了两种方法。

### 4.5.3.1 ADaM方法和实例

ADaM方法是当分析基于AVISIT时，创建一个新行使AVISIT具有唯一值，这种方法的好处是简单且易于分析，当然，对一些概念时间点的定义来说，这些新行可能和实际观察到的行相同而造成冗余。创建一个新行，该行的AVISIT具有唯一的值来表明这是用于分析的行（如“终点”、“基线后最小值”、“基线后最大值”）。这样做的好处是一旦理解了AVISIT值的涵义，审阅者及软件能够充分使用这些AVISIT值。这是一般情况下的一种做法，因为通过这种方法新建的行可以是任何跨时间点的推导行，其AVISIT描述是唯一的。这样做的坏处是数据集会包含更多的行，并且对常用的推导方法也必须进行沟通和理解。当分析不是基于AVISIT时，这两种方法都可以。当然，如果分析文档中未定义AVISIT的值时，那么添加一个标识会更适合。在一个“分析访视”值没有被定义的情况下适宜采用哪种方法取决于如何进行分析。如果只是分析数据集中的子集（如仅仅是治疗期间的最小值），那么标记适合分析的值可能比创建一个包含最小值的新行更好。但是，当在有更多的集合数据的环境下分析这个子集时，新建一行来包含最小值会有助于即可分析方法的使用和审阅。

#### 例1：标识终点行

此例（表4.5.3.1.1）显示了创建一个新行，其AVISIT值是唯一的，来表明用于分析的终点记录。受试者1001在第2周中止，在数据集中增加了用第2周的访视推衍生的一个终点记录（AVISIT=“终点”），DTYPE=“LOV”（最后观察值）指出AVISIT=“终点”记录是如何被填充的。受试者1002不具有任何基线后的访视，且因此没有终点记录。

表4.5.3.1.1 例1：具有终点行标识的分析数据集

行	USUBJID	VISIT	AVISIT	ADY	PARAM	AVAL	DTYPE
1	1001	筛选	筛选	-12	仰卧位收缩压(mmHg)	144	
2	1001	基线	基线	1	仰卧位收缩压(mmHg)	145	
3	1001	第1周	第1周	6	仰卧位收缩压(mmHg)	130	
4	1001	第2周	第2周	12	仰卧位收缩压(mmHg)	133	
5	1001	第2周	终点	12	仰卧位收缩压(mmHg)	133	LOV
6	1002	筛选	筛选	-14	仰卧位收缩压(mmHg)	144	
7	1002	基线	基线	-1	仰卧位收缩压(mmHg)	144	

例2：标识终点及基线后最小、最大和平均值行

此例(表4.5.3.1.2)显示了创建一些新行，其AVISIT具有唯一值，来标明终点记录和基线后最小值、最大值和平均值。受试者1001在第1周具有最小的基线后结果，在第2周具有最大的基线后结果，基线后结果的平均值是第1周和第2周的平均。该受试者在第2周中止。用第2周的访视推衍生一条终点记录(AVISIT="终点")。DTYPE="LOV"（末次观测值）指出AVISIT="终点"的这条记录是一条推衍生的记录。受试者1002没有任何的基线后访视，因此对该受试者不能推衍生基线后最小、基线后最大、基线后平均和终点行。

表4.5.3.1.2 例2：标识终点和基线后最小值、最大值和平均值的分析数据集

行	USUBJID	VISIT	AVISIT	ADY	PARAM	AVAL	DTYPE
1	1001	筛选	筛选	-12	仰卧位收缩压(mmHg)	144	
2	1001	基线	基线	1	仰卧位收缩压(mmHg)	145	
3	1001	第1周	第1周	6	仰卧位收缩压(mmHg)	130	
4	1001	第2周	第2周	12	仰卧位收缩压(mmHg)	133	
5	1001	第1周	基线后最小值	6	仰卧位收缩压(mmHg)	130	最小值
6	1001	第2周	基线后最大值	12	仰卧位收缩压(mmHg)	133	最大值
7	1001		基线后平均值		仰卧位收缩压(mmHg)	131.5	平均值
8	1001	第2周	终点	12	仰卧位收缩压(mmHg)	133	LOV
9	1002	筛选	筛选	-14	仰卧位收缩压(mmHg)	144	
10	1002	基线	基线	-1	仰卧位收缩压(mmHg)	144	

例3：标识基线后最小行和最大行

此例(表4.5.3.1.3)显示基线后最小和最大行的标识。受试者1001在第1周具有最小基线后结果（以ANL01FL=是来标识），在第2周具有最大基线后结果(以ANL02FL=是来标识)。受试者2没有任何的基线后访视，因此无法为此受试者标识基线后最小和基线后最大值。

表4.5.3.1.3 例3：标识基线后最小行和最大行的分析数据集

行	USUBJID	VISIT	AVISIT	ADY	PARAM	AVAL	ANL01FL	ANL02FL
1	1001	筛选	筛选	-12	仰卧位收缩压(mmHg)	144		
2	1001	基线	基线	1	仰卧位收缩压(mmHg)	145		
3	1001	第1周	第1周	6	仰卧位收缩压(mmHg)	130	是	
4	1001	第2周	第2周	12	仰卧位收缩压(mmHg)	133		是
9	1002	筛选	筛选	-14	仰卧位收缩压(mmHg)	144		
10	1002	基线	基线	-1	仰卧位收缩压(mmHg)	144		

## 4.5.4 标识用于分析的行 - 一般情况

标识用于分析或从分析中排除的行是很重要的，应该用标识来标识分析中所用的行还是通过分析时间点描述变量AVISIT的唯一值来标识分析所用的行？

### 4.5.4.1 ADaM方法和实例

ADaM方法是使用一个分析记录标识变量（ANLzzFL）来标明符合一种或多种特定分析要求的行。例

如, ANLzzFL=“是”指出符合分析要求的行, 并且在其它行(如分析中没用到的重复行或事先指定的不包括在分析中的研究后时间点)为空(null), 这允许一个参数内的多个行具有相同的AVISIT值。但是, 它 also 要求在数据集中添加一个标识变量用来为分析选择合适的行。为了产生正确的分析结果, 要理解这些标识。除了ANLzzFL, 可能还需要其他标识, 比如基于行的人群标识, 如ITTRFL和PPROTFL。

请注意可能会有多个ANLzzFL变量, 在这种情况下, 必须要有清楚稳健的元数据来指明创建和填充ANLzzFL变量的依据。

#### 例1: 标识用于分析的行 – 一个访视窗内有多个访视

此例(表4.5.4.1.1) 举例说明了当“基线”和“第2周”分析时间窗内有多个访视时用分析标识变量ANLzzFL来指明要选出哪些行进行分析。根据分析窗的定义, 受试者1001具有两条基线观测和两条第2周观测, 用于分析的那条观测以ANL01FL=“是”标出。使用这个方法是因为所有的原始访视(行)都在数据集中, 而那些选出来做分析的必须加以标识。为了便于溯源, 适当的时候建议添加3.2.5节介绍的AW\*列(例如AWTARGET等等), 以更清楚的标明是如何从每一个分析窗的多个行中选出分析所用的行。(参考表4.2.1.6中使用这些变量的例子)

表4.5.4.1.1 例1: 当一个访视窗中有多个访视时, 标识分析所用行的的分析数据集

行	USUBJID	VISIT	AVISIT	ADY	PARAM	AVAL	DTYPE	ANL01FL
1	1001	筛选	基线	-5	仰卧位收缩压(mmHg)	144		
2	1001	基线	基线	1	仰卧位收缩压(mmHg)	145		是
3	1001	第1周	第1周	7	仰卧位收缩压(mmHg)	130		是
4	1001	第2周	第2周	12	仰卧位收缩压(mmHg)	133		是
5	1001	第3周	第2周	17	仰卧位收缩压(mmHg)	125		
6	1001	第4周	第4周	30	仰卧位收缩压(mmHg)	128		是

#### 例2: 标识分析所用的行 – 落在目标时间窗外的访视

在此例(表4.5.4.1.2), 受试者1001第3周的访视落在了第3周分析窗之外, 所以“研究后”被赋值给AVISIT。这个访视以及第一个基线访视不纳入分析。“基线后最差”分析时间点(第6行)由最差的观测填补(DTYPE=WC)。“终点”行用“第2周”访视推导而成, 因为根据统计分析计划它是最后一条合格的观测值。两个推导的行都由ANL01FL=“是”标识, 因为这两行要选出来进行分析。

表4.5.4.1.2 例2: 当访视落在一个目标时间窗之外时, 标识分析所用行的的分析数据集

行	USUBJID	VISIT	AVISIT	ADY	VISITDY	PARAM	AVAL	DTYPE	ANL01F
1	1001	筛选	基线	-5	1	仰卧位收缩压(mmHg)	144		
2	1001	基线	基线	1	1	仰卧位收缩压(mmHg)	145		是
3	1001	第1周	第1周	7	7	仰卧位收缩压(mmHg)	150		是
4	1001	第2周	第2周	12	14	仰卧位收缩压(mmHg)	133		是
5	1001	第3周	研究后	40	21	仰卧位收缩压(mmHg)	140		
6	1001	第1周	基线后最差	7	7	仰卧位收缩压(mmHg)	150	WC	是
7	1001	第2周	终点	12	14	仰卧位收缩压(mmHg)	133	终点	是

#### 例3: 标识分析所用行 – 用未用于分析的行创建LOCF填补行

此例(表4.5.4.1.3) 举例说明了两个访视发生在同一访视窗(第2周)中的情况。第一个记录(在第4行)在被分析时不变(它被选择代表第2周的分析时间点), 第2周的第2个时间点记录(在第5行)则被作为推导分析时间点第3、4和5周(第6、7和8行)LOCF的基础。在LOCF分析中, 第2周是基于在第4行的观察数据, 而第3、4和5周的数据则是用第5行的最后可用的观测值填补而成。

表4.5.4.1.3 例3: 分析数据集, 其中一个值向前结转但没包括在分析中

行	USUBJID	VISIT	AVISIT	ADY	PARAM	AVAL	DTYPE	ANL01F
1	1001	筛选	基线	-5	仰卧位收缩压(mmHg)	144		
2	1001	基线	基线	1	仰卧位收缩压(mmHg)	145		是
3	1001	第1周	第1周	7	仰卧位收缩压(mmHg)	130		是
4	1001	第2周	第2周	12	仰卧位收缩压(mmHg)	133		是

行	USUBJID	VISIT	AVISIT	ADY	PARAM	AVAL	DTYPE	ANL01F
5	1001	第3周	第2周	17	仰卧位收缩压(mmHg)	125		
6	1001	第3周	第3周	17	仰卧位收缩压(mmHg)	125	LOCF	是
7	1001	第3周	第4周	17	仰卧位收缩压(mmHg)	125	LOCF	是
8	1001	第3周	第5周	17	仰卧位收缩压(mmHg)	125	LOCF	是

## 考虑过但是没有被采用的方法

另一个曾考虑过的方法是时间为时间窗描述变量AVISIT创建唯一值，例如如果没有被分析，在AVISIT的结尾处添加一个星号比如“第2周\*”。因为用户不需要知道一个标识，这个方法可能更不容易混淆。缺点是需具有对于AVISIT值的一个约定，并且为了产生正确的结果，需要分析AVISIT值的工具。因为这些缘故，没有选择这个方法。

## 4.6 特定人群分析行的标识

在临床试验的统计分析中常常对多个感兴趣的人群进行重复分析，感兴趣的人群可以在受试者水平定义，也可以在行（测量）水平定义，或者同时在这个两个水平上定义。例如，当定义一个分析人群时，一个受试者可能被包括在一个分析人群中，比如意向性治疗人群，但是可能被排除在另一个分析人群比如符合方案人群外。分析人群也可以用单个测量的特征来定义，例如一个在事先指定的时间窗之外的特定访视的测量可能不包括在访视水平的符合方案人群中。在本节中假定行水平分析人群的定义取决于受试者水平人群的定义，换句话说，如果一个受试者被排除在受试者水平的符合方案人群之外，那么这个受试者的所有行都不会被包含在行水平的符合方案人群的分析中。考虑到人群定义的多样性，一个分析数据集中的同一行可能包含在一个分析中但是在另一个分析中被排除，这取决于受试者作为一个整体的特征以及单个测量的特征。因此，这个问题演变为如何为每一种分析选择最合适的行。

### 4.6.1 ADaM 方法和实例

对于这个分析问题的ADaM方法是创建一个可以进行多种分析的分析数据集，用特定人群的指示变量来标识每种分析所用的行。这种做法的好处是只要一个分析数据集就可以进行多种分析，而标识变量的使用不必为每种不同的分析复制相同的行。这提高了电子提交的操作效率，也使得分析很清晰，并且便于FDA审阅者对不同人群所选定的值进行比较。但是这个方法也要求为指示变量提供清晰的元数据，才能使每个具体的分析都能精确地再现。下面是几个使用特定人群指示变量来标识用于不同分析行的几个例子。

#### 例1: 使用受试者级水平指示变量(ITTFL和PPROTFL)和行(测量)水平指示变量(ANL01FL, ITTRFL, 和PPROTFL)

下面的分析数据集(表4.6.1.1)能对多个感兴趣人群进行重复分析，分析可以在受试者水平进行，也可以在行（测量）水平进行。ITTFL和PPROTFL是受试者水平的分析人群标识。如果一个受试者属于意向性治疗人群，那么ITTFL列值将为“是”（如果不是则为“否”）。在表4.6.1.1中，受试者1001，1002和1003属于意向性治疗人群。与此相似，如果一个受试者属于符合方案人群，则PPROTFL列值将为“是”（如果不是则为“否”）。在表4.6.1.1受试者1001和1003属于符合方案集，而PPROTFL=空的受试者1002被排除在任何符合方案分析之外。这些指示变量用来标识每个受试者属于哪个受试者水平的人群。

与受试者水平的人群标识相对，列ITTRFL和列PPROTFL是行水平的分析标识。如果一行符合意向性治疗分析的条件，变量ITTRFL值被设为“是”，如果此行不合适作这一分析，变量ITTRFL值被设为“空（null）”。在表4.6.1.1，所有行的列ITTRFL的值都被设为“是”。与此相似，如果一行适合作符合方案分析，则变量PPROTFL值被设为“是”，如果不符合此分析的条件，变量PPROTFL值为空（null）。

在表4.6.1.1，受试者1002的所有三行和受试者1003四行中的两行不是行水平符合方案的数据，在我们应用子集条件PPROTFL=“是”时不会选其进入符合方案分析。不是表4.6.1.1的所有行都用于分析。此例中，被分析的行标识ANL01FL有一行为空（USUBJID=1003, VISIT=Week 1, AVISIT=Week 1, AVAL=999），因为它的值被下一行的重新检测结果代替（USUBJID=1003, VISIT=Retest, AVISIT=Week 1, AVAL=49），重新检查记录的分析记录标识为“是”。

**表4.6.1.1 例1: 具有受试者水平和行水平指示变量的分析数据集**



行	USUBJID	ITTFL	PPROTFL	VISIT	AVISIT	PARAMCD	AVAL	ANL01FL	ITTRFL	PPROTRFL
1	1001	是	是	第0周	第0周	TEST1	500	是	是	是
2	1001	是	是	第1周	第1周	TEST1	400	是	是	是
3	1001	是	是	第2周	第2周	TEST1	600	是	是	是
4	1002	是	是	第0周	第0周	TEST1	500	是	是	
5	1002	是	是	第2周	第1周	TEST1	48	是	是	
6	1002	是	是	第2周	第2周	TEST1	46	是	是	
7	1003	是	是	第0周	第0周	TEST1	999	是	是	是
8	1003	是	是	第1周	第1周	TEST1	999	是	是	是
9	1003	是	是	重新 检验	第1周	TEST1	49		是	
10	1003	是	是	第2周	第2周	TEST1	499	是	是	

由于统计分析的目的不同，即使一个受试者被包括在符合方案人群中，在一个特定数据集中那个受试者的一些或全部数据可能并不适合做符合方案分析。考虑HIV研究中的一种情况，其中一个符合方案分析排除了永久中止研究药物或添加其他抗逆转录病毒治疗之后的所有数据。在上例中受试者1003的最后一次用药是在第1周，所以重新检测的和第2周的数据的PPROTRFL列为空值，将会被排除在任何行水平的符合方案数据分析之外。

要标识第1周时参数编码为“TEST1”的意向性治疗分析的那些行，只需要如下的选择条件：

AVISIT=“第1周” & PARAMCD=“TEST1” & ANL01FL=“是” & ITTRFL=“是”；

相似的，为了标识TEST1 <=400的符合方案分析的那些行，选择条件变为：

PPROTFL=“是” & PARAMCD=“TEST1” & AVAL <=400 & ANL01FL=“是” & PPROTRFL=“是”；

因为上述两种情况下只要选择条件一出错，都将会产生不正确的结果，所以对每个指示变量来说，其元数据要很清楚。另外，ADaM分析结果元数据中会指定选择条件，以提供清楚的文档，说明对某分析是如何用指示变量来选择分析所用的行。

## 4.7 标识符合预定分析标准的行

为了分析的目的，经常会定义一些准则，来根据所收集数据值与一个或多个算法条件的关系将结果分组。例如，具有高于正常值上限五倍的结果的受试者或者具有收缩血压值大于160mmHg且比基线增加25mmHg以上的受试者。除了创建受试者亚群，清单、列表或统计建模中（作为协变量或是反应变量）也常依据某个条件是否成立来进行分类。

### 4.7.1 ADaM 方法和实例

ADaM方法要求使用一个分析条件变量CRITy和一个条件评估结果标识变量CRITyFL一起来标识是否满足一个条件。这两个变量分别在3.2.4节和3.2.6节定义。

CRITy中存放用来定义使条件成立的描述文本，CRITy的定义可以使用该行的任何变量，且定义必须对具有相同PARAM值的所有行保持不变。一个条件可以很复杂的，从多个行得出（参见例3：复合条件）并且可以用到AVAL, AVALC, CHG, PCHG等等。

CRITyFL“条件评估结果标识”是字符指示变量，用来指出是否满足CRITy所描述的条件。如果数据集中有CRITy变量，那么变量CRITyFL也必须同时在数据集中存在。如果需要一个数字型结果标识，那么CRITyFN是允许的。

如果有一行满足CRITy条件，那么ADaM方法允许只填充那一行的CRITy（参见例1）。在那种情况下，只要CRITy中设了条件，CRITyFL的值就设为“是”，否则为空。如果不用这种方法，则对某参数CRITy的所有行都要定义条件，而CRITyFL被设为“是”或者“否”或者为空null（参见例2）。

CRITy和CRITyFL方便了亚组分析。ADaM方法还可以为CRITy条件在BDS中创建新行（除了增加

CRITy列和CRITyFL列以外）。但如果CRITy一旦定义则必须在具有相同PARAM值的所有行上保持一致。  
CRITy, CRITyFL和CRITyFN不具有参数不变性。

#### 例1: CRITy只在满足条件时被填充

采用这种方法, 当对一个PARAM定义条件时, 如果在某行上条件不满足, CRITy和CRITyFL被设为空。如果对某条件有一个或多个输入数据缺失而导致条件无法进行判断 (是否无法判断是申办者定义的, 并不一定由输入数据缺失引起), CRITy和CRITyFL也被设为空。

这样做的目的是当对符合条件的受试者亚组感兴趣时, 根据某个参数进行亚组分析将很方便。当需要得到符合条件的简单计数时, 这种方法也是有用的。采用这种方法时下面的条件肯定成立:

1. 变量CRITy和CRITyFL在数据集中存在;
2. 分析变量元数据根据特定的参数来定义CRITy;
3. 当条件不满足或不可评估时, 在该参数所在行上, CRITy和CRITyFL被设为空。

表4.7.1.1举例说明了ADaM方法中的一种即“只在条件满足时填充CRITy”。CRIT1值的存在说明受试者1001满足条件。采用这种方法时, 当对符合该条件的受试者亚组感兴趣时, CRIT1方便了进行分组。因为受试者1002不满足这个条件, CRIT1的值为空。对于受试者3来说, CRIT1的值为空是因为缺失输入数据而使条件无法评估。

表4.7.1.1 例1: 仅当条件满足时CRITy被填充的分析数据集

行	USUBJID	PARAM	AVAL	BASE	CHG	CRIT1	CRIT1FL
1	1001	收缩压 (毫米汞柱)	163	148	15	收缩压>160	是
2	1002	收缩压 (毫米汞柱)	140	148	-8		
3	1003	收缩压 (毫米汞柱)		120			

#### 例2: 对某参数所有行都填充CRITy值

采用该方法, 对某参数CRITy在所有行内均被填充, 而CRITyFL被设为“是”或者否则设为“否”或空。采用该方法的目的是当要列表显示和/或要在统计模型中使用该参数时, 可以便于分析的进行。表4.7.1.2举例说明了ADaM方法中的一种即“对某参数所有行均填充CRITy”。因为在本例中这个条件要用于设定模型或进行分析, 有必要填充那些不满足此条件的行。CRIT1FL表明是否受试者满足这个条件。对于受试者1005 CRIT1FL被设为空, 因为输入数据的缺失导致条件不可评估。

表4.7.1.2 例2: 分析数据集, 对参数CRITy在所有行被填充

行	USUBJID	PARAM	AVAL	BASE	CHG	CRIT1	CRIT1FL
1	1001	收缩压 (毫米汞柱)	163	148	15	收缩压>160且相对于基线的收缩压变化>10	是
2	1002	收缩压 (毫米汞柱)	140	148	-8	收缩压>160且相对于基线的收缩压变化>10	否
3	1005	收缩压 (毫米汞柱)	120			收缩压>160且相对于基线的收缩压变化>10	

#### 例3: 复合条件

如果一个条件的定义使用到位于多个行的数值 (不同的参数或同一参数的多个行), 那么必须添加一个新行, PARAM的值是条件的文字描述, 且PARAMTYP值被设为“推导” (参见4.2.1节规则5)。  
PARAM (和CRITy) 的文本由申办者定义, 长短自定 (但最长不得超过200), 但要意义明确。对于复合的条件行, AVALC必须总是用值是/否/空来填充。如果一个分析也需要一个数字型的指示变量, 可以采用下面两种方法其中之一:

1. CRITy可以被设为和PARAM相同的条件文本, CRITyFL则被设为和AVALC相同的是/否/空, 而CRITyFN则被设为1/0/null。
2. AVAL可以被设为数字型1/0/null这些指示值。

如果一个分析只要求简单的划分是否满足某复合条件, 也可以只在数据集中添加“复合条件满足”

(AVALC=“是”)行。如果采用这种方法，对某参数来说，当复合条件的评估结果是AVALC=“否”或空时，相应的行不用添加。请注意如果已经定义了复合条件，那么参与条件定义的变量不一定再保留在数据集中，除非还要用这些变量划分子集、输出显示或设定模型，或者是为了保持可溯源性。

表4.7.1.3举例说明了一个复合条件（行3），与非复合条件（行1和行2）同时包含在同一数据集中。

表4.7.1.3 例3：包含复合条件和非复合条件的分析数据集

行	USUBJID	PARAM	PARAMTYP	AVAL	AVALC	BASE	CHG
1	1001	收缩压（毫米汞柱）		163		148	15
2	1001	舒张压（毫米汞柱）		96		87	9
3	1001	收缩压>160且舒张压>95	衍生的		是		

行	CRIT1	CRIT1FL	CRIT1FN	CRIT2	CRIT2FL	CRIT2FN
1（继续）	收缩压>160	是	1	收缩压相对于基线的变化>10	是	1
2（继续）	舒张压>95	是	1			
3（继续）						

注意条件“舒张压>95”（行2）能够和“收缩压>160”（行1）同时存在于同一列CRIT1，每个条件都分别对应于相应的PARAM行的子集。

## 4.8 其它要考虑的问题

在前述章节提出的一些问题中包括了在创建分析数据集时经常会碰到的一些分析决策。但是ADaM小组认识到那并不全面，在本节里继续对其它可能出现的问题提供一些看法。

### 4.8.1 添加记录，为每个受试者创建所有分析时间点

某个受试者在某个分析时间点没有数据并不罕见，例如假定要对一个具有4次访视的数据按访视进行分析，对缺失数据不进行填补。对于没有参加所有四个访视的受试者，可以在分析数据集中为那些缺失的评估创建记录，其AVAL和AVALC为缺失（空null），并用合适的变量来指明这些添加的记录。例如DTYPE可能包含一个申办者定义的值比如“PHANTOM”。每个受试者在分析数据集中具有相同数目观测值具有一些好处，例如所有的受试者都具有相同数据维数使得编程更加便利，编程时可以明确地表示缺失数据，而不是通过缺失整条记录来“暗示”。这也允许ADaM数据集支持列表的创建，特别是在SDTM中没有的数据（例如添加的分析参数）的列表创建。对于一些分类数据的分析，分母可以直接通过分析数据集获得，而不是另一个输入数据集如ADSL获得。这个方法的不便之处在于它可能需要另外的元数据来解释如何使用这些推衍生的空白记录，且在一些情况下，要用选择语句来排除那些AVAL缺失的行。ADaM小组既不提倡也不阻止这种做法。

### 4.8.2 创建多个数据集，支持相同类型数据的分析

统计分析计划通常会规定一个分析将要用稍微不同的方法进行。例如，主要疗效分析可能会对缺失值使用两种不同的填补算法进行。申办者必须决定是将两组填补观测值包括在一个分析数据集里还是创建两个分析数据集，每个数据集包含一种填补算法。ADaM提供的变量能用来标识用于不同分析目的的记录。但是，这不意味着申办者不该或不能提交具有相似内容的多个分析数据集，每个数据集是为一个特定的分析而设计。

附录

附录 A 缩略词

以下是本文档中用到的缩写和首字母简略词，但不包括引用的相关SDTM域(例如QS, DM) 及其变量解释或描述。

ADAE	ADaM不良事件分析数据集
ADaM	CDISC分析数据模型
ADaM document	分析数据模型文档
ADaMIG	分析数据模型实施指南
ADSL	ADaM受试者水平分析数据集
BDS	ADaM基本数据结构
BOCF	基线观测结转
CDASH	临床数据采集协调标准
CDISC	临床数据交换标准协会
DAO	观测到的数据
eCTD	电子通用技术文档
FDA	美国食品药品监督管理局
ITT	意向性治疗
LOCF	前次观测结转
LOV	末次观测值
LVPD	用药前末次观测值
SAP	统计分析计划
SDS	递交数据标准
SDTM	研究数据列表模型
SDTMIG	研究数据列表模型实施指南
WC	最差观测值
WOCF	最差观测结转
XML	可扩展标记语言



---

1484 附录 B 免责声明

1485

1486 译者注：未翻译，请参考原文。注：CDISC 知识产权政策可在以下网站找到：

1487 [http://www.cdisc.org/stuff/contentmgr/files/0/4c011269a1cb48cb0c8eab2c3e172694/misc/cdisc\\_2](http://www.cdisc.org/stuff/contentmgr/files/0/4c011269a1cb48cb0c8eab2c3e172694/misc/cdisc_2)

1488 0ip\_20policy\_final.pdf