24.3 数据列表标准模型SDTM、SDTM aCRF标注病例调查表和数据映射Data Mapping

**数据列表标准模型（Study Data Tabulation Model，SDTM）**

数据列表标准模型（Study Data Tabulation Model，SDTM），是以标准格式展现研究数据。SDTM最基本的概念是观测（Observation）。一条观测对应数据集中的一行。由特定主题的一组观测组成域（Domain）。比如，一条观测“受试者101在研究的第6天开始出现轻度恶心症状”，转换成AE域（不良事件，Adverse Event）的一条观测：

|  |  |  |  |
| --- | --- | --- | --- |
| USUBJID | AETERM | AESEV | AESTDY |
| 101 | 恶心 | 轻度 | 6 |

上表中的这条观测，有4个变量（Variable），这四个变量担任不同的角色：标识符（Identifier）变量USUBJID，受试者唯一标识符；主题（Topic）变量AETERM，不良事件报告词，指明该观测的主要目的；修饰语（Qualifier）变量AESEV，严重程度，用来进一步描述结果的说明性文字或数值，或观测的更多特征；时间（Timing）变量AESTDY，不良事件开始的研究日，描述观测发生的时间。还有一类规则（Rule）变量，只存在于试验设计模型里。

同一主题并且逻辑上相关的观测结果的集合组成域。受试者的观测结果会在一系列不同的域中采集。对于这么多的域，SDTM 又按照其用途，归类为试验设计（Trial Design）、特殊用途（Special Purpose）、干预类（Interventions）、事件类（Events）和发现类（Findings）。

**干预类**：获取受试者接受的研究性治疗或其他（伴有实际或者期望生理效果），可以是由研究方案规定的治疗（例如暴露）、研究期间的伴随治疗（例如用药），或受试者自我服用的其它物质（如酒精、烟草或咖啡因）。常用域有CM（伴随用药）、EX（暴露）。

**事件类**：获取研究方案计划的重要事件如随机化、研究完成情况、及在试验期间或之前发生的独立于临床研究计划评估状况事件等（前者如不良事件，后者如疾病史）。常用域有AE（不良事件）、DS（研究完成情况）、MH（病史）。

**发现类**：获取临床研究计划评估的观测结果如特定实验室检查、ECG检查或问卷中问题的回答结果等。常用域有EG（心电图）、LB（实验室检查）、QS（问卷调查）、VS（生命体征）等。

**特殊用途**：定义受试者水平的数据，和上面三类的域有所区别。有DM（人口统计学）、CO（注释）、SE（受试者元素）、SV（受试者访视）域等。

**试验设计**：定义了一个标准结构来表示计划的事件序列和治疗方案。它提供了一个标准方法，以定义受试者将要经历的治疗分组、计划访视和评估。有TS（试验总结）、TA（试验分组）、TE（试验元素）、TV（试验访视）和TI（试验入排标准）域等。

在SDTM的设计中，参考了关系型数据库的设计原则，但是又根据临床试验的特点，留有很多冗余，使得每个数据集都有一个主题，可以自我解释，数据集之间的关系又可以通过特殊用途和关联数据集（Relationship）进行关联。

在SDTM模型中，每一类都有规定其可以使用的变量列表。在SDTM IG(实施指南)中，预设了一些常用的域，每个域都有规定其推荐的变量列表，在列表中规定了每个变量的标签、变量类型、如何使用等情况。哪些变量必须使用，不能为空；哪些变量是期望使用，但是如果没有收集的话，是可以空着；哪些变量是可选变量，如果没有收集的话可以不用的。并且还有很多例子供参考。

如果在IG中的域还不够用的话，申办方可以按照SDTM的规则自定义新的域来满足研究的需求。

因为SDTM这些域都是有了标准格式，所以按照这种标准格式来存放研究数据的话，会带来极大的方便：

* + 临床试验过程中的各类人员，如果熟悉了这种格式的话，会减轻熟悉各种格式带来的学习压力，节省时间。
  + 可以开发标准化的工具，来进行SDTM相关的生成、验证、评审的需求，FDA就有可视化的评审工具方便评审员。
  + 便于生成分析数据集。
  + 数据交换时有助于各方快速了解数据的内容。

在SDTM模型递交时，最少需要DM，EX，DS。

**SDTM aCRF标注病例调查表（annotated Case Report Form）**

SDTM aCRF标注病例调查表（annotated Case Report Form），在PDF格式的CRF上，以SDTM的变量名进行标注，方便评审员快速查找对应的数据在CRF上的位置，方便其理解临床试验的数据。这个PDF文件名称必须命名为“acrf.pdf”[[1]](#footnote-1)。

aCRF需要参见CDISC发布的Metadata Submission Guideline[[2]](#footnote-2)。其中规定，标注的内容需要在PDF文件中可搜索，手写后扫描的版本因为无法搜索而不可用。在PDF文件还应该对DOMAIN建立书签索引，按照VISIT和按照DOMAIN两种方式进行书签。CRF上有收集，但是最终没有在SDTM中递交的变量，需标注“NOT SUBMITTED”(“不递交”)。CRF中唯一的页面需要详细标注，详细到每个CRF上收集的条目。重复的页面，可以标注“SEE ANNOTATION ON PAGE XX”（ “见前第xx页标注” ）。域名标注和变量标注可以通过不同字体大小加以区别；同一个页面上的域，可以以不同颜色加以区别。

**数据映射（Data Mapping）**

数据映射（Data Mapping），是将原始数据映射至SDTM数据集，或者将SDTM数据集映射至ADaM分析数据集。

在原始数据映射至SDTM数据集时：一般原始数据因为数据库的限制原因，有一些变量是横向（Horizontal）的，在映射时需要进行转置到纵向（Vertical）标准结构，这个过程我们叫Normalize。

做数据映射时，需要设计一个映射说明文件（Mapping Specs），通常是Excel文件格式。SAS编程人员根据该映射说明文件，通过SAS编程，实现数据的最终映射，转换到SDTM或者ADaM。

举例：

RAW.VS （为展示目的，省略了一些变量）

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| STUDYID | DOMAIN | USUBJID | SYSBP | DIABP | HEIGHT | WEIGHT | PULSE | VISITNUM | VISIT | VSDTC | VSDY |
| 1001 | VS | 10011001 | 122 | 74 | 168 | 70 | 74 | 1 | SCREEN | 2003/4/15 | -14 |

Mapping Specs:

| Variable | Label | Type | Codelist | Core | Source/Derivation/Comments |
| --- | --- | --- | --- | --- | --- |
| STUDYID | Study Identifier | Char |  | Req | RAW.VS.STUDYID |
| DOMAIN | Domain Abbreviation | Char |  | Req | RAW.VS.DOMAIN |
| USUBJID | Unique Subject Identifier | Char |  | Req | RAW.VS.USUBJID |
| VSSEQ | Sequence Number | Num |  | Req | Derived, sort by USUBJID VISITNUM VSTESTCD, first record of a subject assign 1, accumulate by 1 for following records. |
| VSTESTCD | Vital Signs Test Short Name | Char | VSTESTCD | Req | Assigned: RAW.VS, SYSBP DIABP HEIGHT WEIGHT PULSE |
| VSTEST | Vital Signs Test Name | Char | VSTEST | Req | Assigned. |
| VSORRES | Result or Finding in Original Units | Char |  | Exp | Derived: transpose from RAW.VS, SYSBP DIABP HEIGHT WEIGHT PULSE |
| VSORRESU | Original Units | Char |  | Exp | Assigned: SYSBP mmHg DIABP mmHg HEIGHT cm WEIGHT kg PULSE BEATS/MIN |
| VSSTRESC | Character Result/Finding in Std Format | Char |  | Exp | Derived: VSORRES |
| VSSTRESN | Numeric Result/Finding in Standard Units | Num |  | Exp | Derived: input(VSORRES,?best.) |
| VSSTRESU | Standard Units | Char |  | Exp | Derived: VSORRESU |
| VISITNUM | Visit Number | Num |  | Exp | RAW.VS.VISITNUM |
| VISIT | Visit Name | Char |  | Perm | RAW.VS.VISIT |
| VSDTC | Date/Time of Measurements | Char |  | Exp | RAW.VS.VSDTC |
| VSDY | Study Day of Vital Signs | Num |  | Perm | RAW.VS.VSDY |

SDTM.VS （为展示目的，省略了一些变量）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STUDYID | DOMAIN | USUBJID | VSSEQ | VSTESTCD | VSTEST | VSORRES | VSORRESU |
| 1001 | VS | 10011001 | 1 | SYSBP | Systolic Blood Pressure | 122 | mmHg |
| 1001 | VS | 10011001 | 2 | DIABP | Diastolic Blood Pressure | 74 | mmHg |
| 1001 | VS | 10011001 | 3 | HEIGHT | Height | 168 | cm |
| 1001 | VS | 10011001 | 4 | WEIGHT | Weight | 70 | kg |
| 1001 | VS | 10011001 | 5 | PULSE | Pulse Rate | 74 | BEATS/MIN |

（续上表）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| VSSEQ | VSSTRESC | VSSTRESN | VSSTRESU | VISITNUM | VISIT | VSDTC | VSDY |
| 1 | 122 | 122 | mmHg | 1 | SCREEN | 2003/4/15 | -14 |
| 2 | 74 | 74 | mmHg | 1 | SCREEN | 2003/4/15 | -14 |
| 3 | 168 | 168 | cm | 1 | SCREEN | 2003/4/15 | -14 |
| 4 | 70 | 70 | kg | 1 | SCREEN | 2003/4/15 | -14 |
| 5 | 74 | 74 | BEATS/MIN | 1 | SCREEN | 2003/4/15 | -14 |

将SDTM数据集映射至ADaM分析数据集

举例：

Mapping Specs: 和之前的SDTM.VS的映射表类似，只是添加了为可追溯性目的的三个变量SRCDOM, SRCVAR和SRCSEQ。

ADaM.ADVS （为展示目的，省略了一些变量）

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| STUDYID | USUBJID | PARAMCD | PARAM | AVAL | AVISITN | AVISIT | SRCDOM | SRCVAR | SRCSEQ |
| 1001 | 10011001 | SYSBP | Systolic Blood Pressure (mmHg) | 122 | 1 | SCREEN | VS | VSSEQ | 1 |
| 1001 | 10011001 | DIABP | Diastolic Blood Pressure (mmHg) | 74 | 1 | SCREEN | VS | VSSEQ | 2 |
| 1001 | 10011001 | HEIGHT | Height (cm) | 168 | 1 | SCREEN | VS | VSSEQ | 3 |
| 1001 | 10011001 | WEIGHT | Weight (kg) | 70 | 1 | SCREEN | VS | VSSEQ | 4 |
| 1001 | 10011001 | PULSE | Pulse Rate (BEATS/MIN) | 74 | 1 | SCREEN | VS | VSSEQ | 5 |

还有些商业工具可以提供在软件界面中直接进行MAPPING，如SAS ETL Studio，SAS Drug Studio，SAS Clinical Data Integration等产品, XML4PHARMA SDTM-ETL, X-Clinical的Tabulator。

1. FDA conformance guide: <http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm> [↑](#footnote-ref-1)
2. CDISC Website: <http://www.cdisc.org/metadata-submission-guideline-%28msg%29-package-preface> [↑](#footnote-ref-2)