Clinical trial database and related data in eCTD
Declaration requirements

(draft for comment)

National Drug Administration Drug Evaluation Center
September 2019
table of Contents

Declaration requirements for clinical trial databases and related materials in eCTD

The clinical trial database and related materials should be included in the eCTD submission, and should be located in the skeleton file after the corresponding clinical summary report and marked with the appropriate label.

First, the clinical trial database

Applicants should submit the original database and analysis database for clinical trials. The original database should normally contain all raw data collected directly from the Case Report Form (CRF) and external files. The analysis database is a newly created database for statistical analysis. It should contain all the variables required for analysis, including derived variables, and all derived variables should be generated by the original database submitted. The submitted database should be SAS XPORT (Version 5 or above) transport format (XPT format).

If the applicant submits data in accordance with the Clinical Data Exchange Standards Association (CDISC) standard, the Research Data Table Model (SDTM) database can

be considered the original database, and the Analytical Data Model (ADaM) database is considered the analysis database.
If the size of a single dataset file in the database exceeds 4G, it needs to be split, and the split rules and results are explained separately in the data description file and data review instructions.
At eCTD filing, the original database is tagged with the "data-tabulation-dataset-legacy" or "data-tabulation-dataset-sdtm" tag, and the analysis database uses the "analysis-dataset-adam" or "analysis-dataset-legacy" tag. Mark it. The use of the label is given in Annex 1.
Second, the data description file
Applicants should submit data description documents for the original database and the analysis database. The data description file should describe the basic information of the data set contained in the database, including the name of the data set, the content description, structure, and so on. The data description file should also indicate the variable name, label, data format, value range, data source, etc. of the variables included in each data set. For derived variables, specific derivative methods need to be specified, including a description of the derived rules and the corresponding programming code.
The data description file is generally XML (including the corresponding XSL file) or PDF format. In addition, it is also recommended to submit data review instructions in PDF format. The data review instructions are further described by another independent document to further assist the reviewer in understanding and using the submitted data more accurately and efficiently.
At the time of eCTD filing, the original database data description file is marked with the "data-tabulation-data-definition" tag, and the analysis database data specification file is marked with the "analysis-data-definition" tag.
Third, other
Applicants should submit a CRF (aCRF) related to the clinical trial report in PDF format. Note CRF is an annotation of the blank CRF, recording the location of each data item of the CRF and its variable name and encoding in the corresponding database.
Applicants should submit key programming code (such as the analysis of the database derivative process and the main analysis results generation process, etc.), the submitted program code should be readable, does not contain complex external program calls, using TXT plain text format.
At the time of eCTD filing, the annotation CRF is tagged using the "annotated-crf" tag and the programming code is tagged using the "analysis-program" tag.

Annex 1 STF file label
label
Description
Data-tabulation-dataset-legacy

Original database (non-CDISC standard)
Data-tabulation-dataset-sdtm
Original database (CDISC standard)
Data-tabulation-data-definition
Original database data description file, data review instructions (if any)
Analysis-dataset-adam
Analysis database (CDISC standard)
Analysis-dataset-legacy
Analytical database (non-CDISC standard)
Analysis-data-definition
Analyze database data description files, data review instructions (if any)
Annotated-crf
Comment CRF
Analysis-program
Programming code

Attachment 2 folder structure

eCTD 中临床试验数据库及相关资料的申报要求

（征求意见稿）

国家药品监督管理局药品审评中心

2019 年 9 月

目 录

eCTD 中临床试验数据库及相关资料的申报要求

在 eCTD 申报资料中应包含临床试验数据库及相关资料，在骨架文件中应位于相应的临床总结报告之后，并使用适当的标签进行标识。

一、临床试验数据库

申请人应递交临床试验的原始数据库及分析数据库。原始数据库通常应包含从病例报告表（CRF）和外部文件中直接收集的所有原始数据。分析数据库为统计分析中衍生新建的数据库，应包含所有分析时所需的变量，包括衍生变量，且所有衍生变量均应能通过递交的原始数据库生成。递交的数据库应为 SAS XPORT（Version 5 或以上版本）传输格式（XPT 格式）。

如果申请人以临床数据交换标准协会（CDISC）标准递交数据，则可将研究数据表格模型（SDTM）数据库视为原始数据库，分析数据模型（ADaM）数据库视为分析数据库。

如果数据库中单个数据集文件大小超过 4G，则需将其进行拆分，并就拆分规则与结果在数据说明文件与数据审阅说明中分别阐述说明。

在 eCTD 申报时，原始数据库使用“data-tabulation-dataset-legacy”或“data-tabulation-dataset-sdtm”标签进行标记，分析数据库使用“analysis-dataset-adam”或“analysis-dataset-legacy”标签进行标记。标签的使用见附件 1。

二、数据说明文件

申请人应递交原始数据库和分析数据库的数据说明文件。数据说明文件中应说明数据库中所包含数据集的基本信息，包括数据集的名称、内容简述、结构等。数据说明文件中还应说明各数据集中所包含变量的变量名、标签、数据格式、取值范围、数据来源等。对于衍生变量，需要明确具体的衍生方法，包括对衍生规则的描述及相应编程程序代码。

数据说明文件一般为 XML（包括对应的 XSL 文件）或 PDF 格式。除此之外，还建议递交 PDF 格式的数据审阅说明。数据审阅说明是通过另一独立文档描述的方式来进一步帮助审评人员更准确、高效的理解与使用所递交数据。

在 eCTD 申报时，原始数据库数据说明文件使用“data-tabulation-data-definition”标签进行标记，分析数据库数据说明文件使用“analysis-data-definition”标签进行标记。

三、其他

申请人应递交临床试验报告相关的注释 CRF（aCRF），格式为 PDF。注释 CRF 是对空白 CRF 的标注，记录 CRF 各数据项的位置及其在相对应的数据库中的变量名和编码。

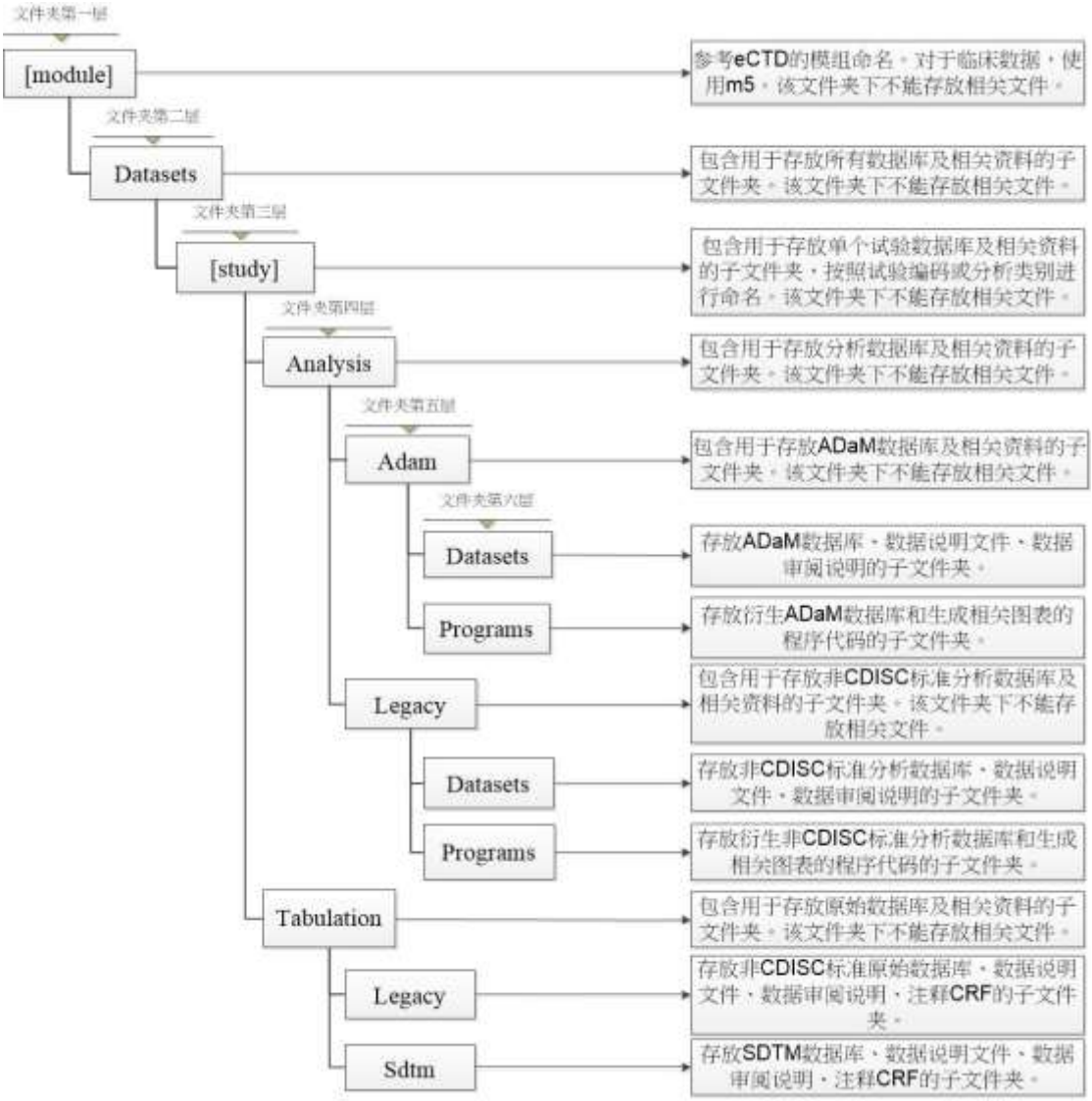
申请人应递交关键编程程序代码（如分析数据库的衍生过程和主要分析结果的生成过程等），所递交的程序代码应可读性强，不包含复杂的外部程序调用，采用 TXT 纯文本格式。

在 eCTD 申报时，注释 CRF 使用“annotated-crf”标签进行标记，编程程序代码使用“analysis-program”标签进行标记。

附件 1 STF 文件标签

标签	说明
data-tabulation-dataset-legacy	原始数据库（非 CDISC 标准）
data-tabulation-dataset-sdtm	原始数据库（CDISC 标准）
data-tabulation-data-definition	原始数据库数据说明文件、数据审阅说明（若有）
analysis-dataset-adam	分析数据库（CDISC 标准）
analysis-dataset-legacy	分析数据库（非 CDISC 标准）
analysis-data-definition	分析数据库数据说明文件、数据审阅说明（若有）
annotated-crf	注释 CRF
analysis-program	编程程序代码

附件 2 文件夹结构


文件夹第一层
[module]
参考eCTD的模组命名，对于临床数据，使用m5，该文件夹下不能存放相关文件。
文件夹第二层
Datasets
包含用于存放所有数据库及相关资料的子文件夹，该文件夹下不能存放相关文件。
文件夹第三层
[study]
包含用于存放单个试验数据库及相关资料的子文件夹，按照试验编码或分析类别进行命名，该文件夹下不能存放相关文件。
文件夹第四层
Analysis
包含用于存放分析数据库及相关资料的子文件夹，该文件夹下不能存放相关文件。
文件夹第五层
Adam
包含用于存放ADaM数据库及相关资料的子文件夹，该文件夹下不能存放相关文件。
文件夹第六层
Datasets
存放ADaM数据库、数据说明文件、数据审阅说明的子文件夹。
Programs
存放衍生ADaM数据库和生成相关图表的程序代码的子文件夹。
Legacy
包含用于存放非CDISC标准分析数据库及相关资料的子文件夹，该文件夹下不能存放相关文件。
Datasets
存放非CDISC标准分析数据库、数据说明文件、数据审阅说明的子文件夹。
Programs
存放衍生非CDISC标准分析数据库和生成相关图表的程序代码的子文件夹。
Tabulation
包含用于存放原始数据库及相关资料的子文件夹，该文件夹下不能存放相关文件。
Legacy
存放非CDISC标准原始数据库、数据说明文件、数据审阅说明、注释CRF的子文件夹。
Sdtm
存放SDTM数据库、数据说明文件、数据审阅说明、注释CRF的子文件夹。