Machine Translated by Google

药物临床试验数据递交指导原则

（试行）

July 2020

Machine Translated by Google

目录

Machine Translated by Google

Machine Translated by Google

药物临床试验数据递交指导原则（试行）

1. Background and Purpose

Drug clinical trial data is one of the important information submitted by sponsors to regulatory agencies. One, it is a valuable resource for both regulators and sponsors. standardized collection, Organizing, analyzing, and presenting clinical trial data is essential for improving the efficiency of drug clinical development and quality, shorten the review time, and benefit the whole life of the drug Life cycle management to facilitate the exchange or sharing of R&D or regulatory information.

If the clinical trial data submitted by the sponsor does not follow certain specifications, be familiar with and Understanding the data structure and content will take up a lot of review resources. In some cases, applying Sponsors or regulators may need to aggregate clinical trial data from multiple sources Analysis, if the data is not normalized, integrated utilization is almost impossible to achieve.

Application materials related to clinical trial data usually include databases and their corresponding Data Documentation, Data Review Instructions, Program Codes, and Annotated Case Report Forms (annotated Case Report Form, aCRF). This guideline mainly addresses the Specific requirements for the content and format of bed trial data submission are intended to guide sponsors Standardize the submission of clinical trial data and related materials, and at the same time contribute to data management, The relevant practitioners such as statistical analysis can better carry out related work in clinical trials.

This guideline is mainly applicable to the key Sexual clinical trials can also be used for reference in clinical trials for non-registration purposes. Book

Machine Translated by Google

The guiding principles are based on the data submission requirements of international regulators and the current situation in China.

Sponsors should prepare relevant materials based on the requirements of this guideline. Sponsors are encouraged to refer to

Clinical Data Interchange Standards

Consortium, CDISC) standard submission of clinical trial data and related application materials

material. With the development of clinical trial data standards and the improvement of their understanding and practice,

These guidelines will be revised and improved as appropriate.

2. Information about clinical trial data and its description

(1) Original database

Original databases often contain direct collections from case report forms and external documents

, and may also contain very small amounts of derived data, such as serial numbers. original number

Missing data in the database should not be imputed. In order to meet the data submission requirements, direct

The data of the receiving set may need to be normalized or encoded as necessary, such as adjusting the number of

The dataset name/label/structure in the database, the variable name/label in the dataset, or in the

Standardized coding of variable values where applicable, such as the Medical Dictionary of Regulatory Activities

ÿMedical Dictionary for Regulatory ActivitiesÿMedDRAÿÿÿ

If the sponsor submits data against the CDISC criteria, the original data criteria may be

Model (Study Data Tabulation Model, SDTM) database treated as original

database.

The original database usually contains multiple original datasets, and the original datasets should be sorted by master

Questions are organized and named, and datasets are usually named with a two-letter code.

Machine Translated by Google

Names such as demographics (dm), adverse events (ae), laboratory tests (lb), etc.

data set. The nomenclature of raw datasets commonly used in clinical trials is detailed in Appendix 1.

All submitted raw datasets must contain the Study Identifier (STUDYID)

variables; datasets that reflect the observations of individual subjects (eg, dm,

ae, lb, etc. datasets) must also contain the Subject Unique Identifier (USUBJID)

variable; in addition, the subject identifier (SUBJID) variable must be included in the dm

data set. Examples of commonly used identifiers are as follows:

Study identifier: variable named STUDYID, character type, unique to the study

Identifier, i.e. study number.

Subject unique identifier: variable named USUBJID, character type, each

The entire trial application (including multiple clinical studies) process of subjects in the same product

should be given the same unique identifier. in all datasets (including raw data

data set and analysis data set), the same subject should have exactly the same unique

identifier. When subjects participated in multiple studies, USUBJID between studies

should be consistent. Follow this rule for pooling the same subject in different studies

It is especially important to have data in (e.g. randomized controlled trials or extension studies).

Subject identifier: variable named SUBJID, character type, SUBJID is

Identifier of the subject participating in the trial.

Visit Name (VISIT, char) and Visit Number (VISITNUM,

Temporal variables such as numeric) should be included in the applicable dataset. planned visit

Machine Translated by Google

VISITNUM should be assigned from small to large according to the time order, and one by one with VISIT correspond.

(2) Analysis database

Analysis database is a new database derived for statistical analysis, used to generate and Support statistical analysis results in documents such as clinical summary reports. Analytical database one It generally includes original data and data derived from original data according to certain rules, such as Data after imputation of missing data, etc. If the sponsor refers to the CDISC criteria to submit The analysis data standard model (Analysis Data Model, ADaM) databases are considered analytical databases.

Analytical databases typically contain multiple analytical datasets. When building an analysis dataset, Collected and derived data (from various raw data sets or other analytical data may be data sets) are merged into one data set, and the following principles should be followed when constructing: ÿ used for Analytical datasets supporting statistical analysis must be clear in their content and sources. ÿAnalysis Data sets must be traceable, and specific rules for data derivation should be as detailed in the documentation. ÿ The structure and content of the analysis data set should meet the requirements Statistical analysis requires very little programming.

The analytical database should contain all variables required for the analysis, including derived variables, And all derived variables should be able to pass the original database and other supporting data files generate. Analysis datasets are usually named "adxxxxxx", the naming of analysis datasets It should try to keep the correspondence with the original data set, such as: adcm, adae, adlb, etc.

Machine Translated by Google

The subject-level analysis dataset (named adsl) is an essential one

Analyze the dataset. In this dataset, there should be only one record per subject, within

Content should include, but is not limited to, demographics, important baseline characteristics/stratification factors, treatment

Information on groups, prognostic factors, important dates, and analysis of populations.

For some endpoints (such as some scale scores), from the original data set to the available

The analysis data set for the final statistical analysis needs to go through a series of derivative processes.

Intermediate variables/datasets derived from the creation of the final analysis dataset should also be used when necessary.

are included in the analysis database together.

(3) Data Description Documents

The submitted original database and analytical database must have corresponding data descriptions

pieces. A data description document is a document used to describe the submitted data and should include at least

Including the name, label, basic structure description and each data set in the submitted database

The name, label, type, source, or derivation of each variable in the dataset.

The data description document is the most accurate understanding of the content of the submitted data when the regulator reviews it.

one of the important documents. Sponsors should ensure that the coded list and source for each variable are

Clearly defined and easy to find. If you use an external dictionary, you need to

The dictionary and version used are specified in the documentation. need to pass the data description file

Establish good traceability between data (e.g. original data set and CRF,

between the analytical dataset and the original dataset) for review by regulators. bid

Parties need to provide relevant details in the data description file, especially in relation to derived variables

If necessary, you can use the key program code to assist the description.

Machine Translated by Google

The data description file is generally extensible mark-up language (Extensible Mark-up Language ÿ XML) or Portable Document Format, PDF) files. If you submit a data description file in XML format, the corresponding Extensible Stylesheet Language (XSL) document should also be submitted together.

(IV) Data Review Instructions

To help reviewers better understand and use the submitted data, applicants are encouraged to The organizer submits the data review instructions. The data review instructions are the further development of the data description file One-step supplement, which includes but is not limited to research data usage instructions, clinical summary Relationship between reports and data, research documentation (e.g. experimental protocols, statistical analysis plan, clinical summary report, etc.), some key information, the application of the submitted program code Description, encoding used in the dataset (such as utf-8, euc-cn, etc.) and other special Situation description, etc. The Data Review Instructions are not intended to replace the Database Data Instructions documents, but help reviewers to be more accurate and efficient through document descriptions Understand and use the submitted databases, related terms, program codes and data descriptions file information, etc. Data review instructions should be in PDF files.

(5) Annotated case report form

Annotated case report forms are based on blank CRFs for collected subjects User data (electronic or paper) information unit (ie field information) and delivery A specific description of the mapping relationship between the corresponding variables or variable values in the original data set. aCRF files should be PDF files.

Machine Translated by Google

In practice, the CRF may collect some data that are not in the submission database. According to the content, such data should be clearly marked on the aCRF as "not submitted" ("NOT SUBMITTED") and state in the data review instructions not to submit these data reason.

(6) Program code

Program codes that sponsors need to submit include, but are not limited to: analysis data sets The derivative process of derived variables, the generation process of the analysis results of efficacy indicators, etc. declare The program code submitted in the documentation should be easy to understand and highly readable, and it is recommended to provide sufficient Comment, avoid external (macro) program calls. The program code generally adopts the TXT file.

3. Format of Clinical Trial Data Related Information

(1) Portable document format

Portable Document Format (PDF) is an open document format that is independent of application using software, hardware and operating systems. Follow the International Human Drugs in the submission materials International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Useÿ ICH) Electronic Common Technical Document, eCTD) format requirements for other documents can be used as PDF files. establish It is recommended to use PDF 1.4 or later for document submission. All PDF files Both should have .pdf as the file extension.

Machine Translated by Google

(2) Extensible Markup Language Format

Extensible Markup Language (XML) was developed by the World Wide Web Consortium (World Wide Web Consortium).

A data interchange language defined by the Wide Web Consortium, W3C). it

Can be opened, edited and created by any text editor for transferring and storing data

according to. XML format files can easily exchange information between different systems.

All XML format files must have .xml as the file extension.

(3) Plain text format

The plain text format document (TXT) has the advantages of simple format, small size and simple storage.

It is also a general-purpose file supported by computers and many mobile terminals.

file format. All TXT files should have a .txt file extension.

(4) Format of research data transmission

The data set in the application materials usually adopts the SAS Data Transfer Format (SAS

Transport Format, XPT for short). An XPT file corresponds to a dataset,

The dataset name needs to be consistent with the XPT file name, and its file suffix is the same

to .xpt. For example, the adverse event data set ae.xpt, the previous and combined medication data set

cm.xpt etc. XPT version 5 (referred to as XPT V5) or above is recommended

This is the data submission format. Sponsors should state the encoding used (e.g. utf-8, euc

cn, etc.) to avoid garbled characters in the submitted dataset.

(5) Data set splitting

When a single data set in the database does not meet the relevant requirements of the application materials due to the storage size

When splitting is required, only the split data set can be submitted. Instructions on Data Review

Machine Translated by Google

, the splitting rules of the dataset and the detailed steps for merging should be explained in detail to ensure that Reviewers were able to generate the same dataset as before the sponsor split.

(6) Data set name, variable name and variable length

The requirements for dataset names and variable names are as follows:

Dataset names can only contain lowercase English letters and numbers, and must start with lowercase letters Start with a letter. And the maximum length of the dataset name is 8 bytes.

Variable names can only contain uppercase English letters, underscores, and numbers, and must Must start with a letter. And the maximum length of the variable name is 8 bytes.

The length of each character variable, which should be set to study all datasets here The maximum actual variable value length of the variable in this variable can effectively control the size of the file.

(7) Dataset labels and variable labels

For ease of review, dataset labels and variable labels should be in Chinese. It is recommended that The length does not exceed 40 bytes, and can contain English characters, underscores or numbers if necessary word, but cannot begin with a number, and cannot contain the following:

ÿ Unpaired half-width or full-width single or double quotation marks

ÿ Unpaired half-width or full-width brackets

ÿ Special characters (eg '>', '<')

Machine Translated by Google

4. Other related matters

(1) Traceability of test data

An important part of the review is an accurate understanding of the source of the data, i.e. the data traceability. Traceability allows reviewers to understand statistical analysis results (clinical total report), analytical data, and relationships to raw data provide

Technology License.

Data traceability ensures that reviewers can accurately:

- Understand the construction of analytical datasets
- Determining observation records and corresponding algorithms for derived variables
- Understand how relevant statistical results are calculated
- Establish links from raw data to corresponding reports

Sponsors should ensure that raw data are available to regulators when submitting databases

The database is derived from the analysis database consistent with the sponsor, and the analysis database can be used to directly Statistical analysis results consistent with those of the sponsor were consistently reproduced. Traceability can also be achieved through A detailed flow chart of the data collection stage to the submission stage is provided to aid in the interpretation.

(2) Data files under electronic general technical documents

All documentation, test data and related support when filing with the eCTD The files need to be organized according to the specified folder structure. All submitted documents are should be in the correct folder with the appropriate study tag file (Study

Machine Translated by Google

Tagging File, STF) for identification. STF and folder structure see Appendix 2 and

Appendix 3.

(3) Foreign language database

The application materials related to clinical trial data should be mainly in Chinese.

The Chinese expressions between the same documents should be consistent, such as the analysis of bad data in the data set

The name of the event and the name of the adverse event in the clinical summary report form should correspond to each other.

In order to improve the review efficiency, the application materials related to clinical trial data are translated from foreign languages

The minimum requirements for Chinese are as follows:

At least the following content in the submitted database should be in Chinese: dataset labels and variables

Labels; names of adverse events appearing in clinical summary reports and other documents, combined use

Drug name, medical history name.

At least the following content in the data description file should be in Chinese: each data in the database

Description/label and description of the set; description/label and derived data of each variable in the dataset

process; a list of values or codes involving efficacy indicators.

At least the following should be in Chinese in the annotated case report form: For data collection

The description of the designed question; the value or coding of the question involving the efficacy index.

Data review instructions should be in Chinese.

(4) Communication with regulatory agencies

According to the characteristics and complexity of the specific clinical trial data, the sponsor may, if necessary,

According to the relevant management measures for drug development and technical review communication, and review

Machine Translated by Google

Institutions communicate with each other on the submission of clinical trial databases and related materials to facilitate review

Reviewers can quickly and accurately understand clinical trial data submitted by sponsors.

5. References

1. CFDA. Technical Guidelines for Clinical Trial Data Management. July 2016.

2. FDA. Study Data Technical Conformance Guide. Mar 2020.

3. PMDA. Revision of Technical Conformance Guide on

Electronic Study Data Submissions. Jan 2019.

4. CDISC. Study Data Tabulation Model Implementation

Guide. November 2018.

5. CDISC. Analysis Data Model Implementation Guide. Oct

2019.

Machine Translated by Google

Appendix 1: Common Raw Data Sets

Table 1 Commonly used original datasets and their names

data set	name	Submit request
Demographic	dm	must submit
History	mh	if applicable
Adverse Events	but	if applicable
Past and Concomitant Drug	cm	if applicable
Exposure Subject Distribution	ex	if applicable
Questionnaire and Scale	ds	if applicable
Protocol Deviations Laboratory	qs	if applicable
Tests Electrocardiogram Vital	dv	if applicable
Signs Clinical Events Physical	lb	if applicable
Examination	eg	if applicable
	vs	if applicable
	this	if applicable
	on	if applicable

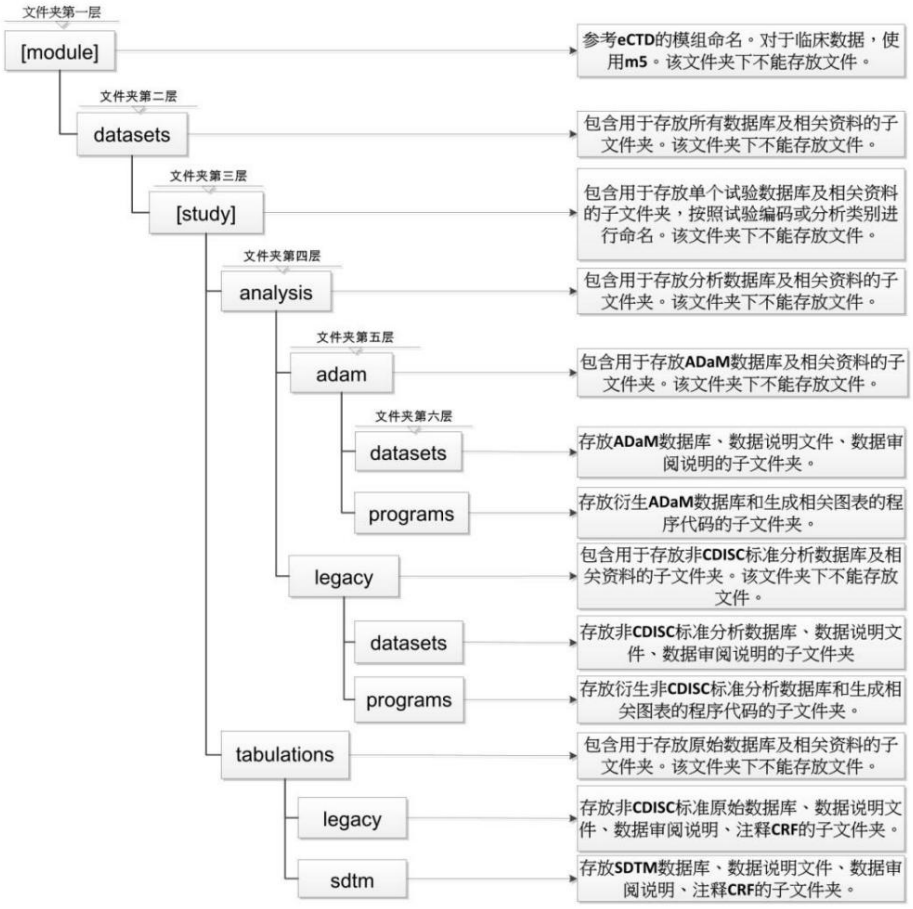
Machine Translated by Google

Appendix 2: Study Label File

title element name attribute value	illustrate
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 description
analysis-dataset-legacy	Analytical database (non-CDISC standard)
analysis-dataset-adam	analysis database (CDISC standard)
analysis-data-definition	Analysis database data description files, data review instructions
annotated-crf	Annotation CRF
analysis-programming	programming code

Machine Translated by Google

Appendix 3: Folder Structure



Machine Translated by Google

Appendix 4: Glossary

Code List: refers to the possible values of the variable, including in the test

The data involves the corresponding standard code, industry general code or the sponsor's self-defined code.

Meaningful encoding.

Case **Report Form (CRF):** refers to the

A paper or electronic document that records subject-related information reported to the sponsor

document.

Electronic Common Technical Document (Electronic Common Technical Document,

eCTD): Electronic registration documents for drug registration application and review. through extensible

The exhibition markup language will electronically upload the drug application materials that comply with the CTD specifications.

Line organization, transmission and presentation.

Data **Definition** File: used to describe the submitted data

The file should at least contain the names, labels, basic results of each dataset in the submitted database.

structure description and the name, label, type and source or derivative of each variable in each dataset

birth process.

Data **Reviewer's** Guide: is a description of the data

A further supplement to help reviewers be more accurate and

Effectively understand and use the submitted database, related terms, program code and data

According to the document information and so on.

Machine Translated by Google

Annotated Case Report Form (aCRF): is based on the blank case report form,

Information units (i.e., text) for collected subject data (electronic or paper)

segment information) and the corresponding variable or variable value in the submitted original data set.

A detailed description of the system.

Machine Translated by Google

Appendix 5: Chinese and English vocabulary comparison

Chinese	English
Portable Document Format	PDF (Portable Document Format)
case report form	CRF (Case Report Form)
Electronic Common Technical Document	eCTD (electronic Common Technical Document)
ADaM (Analysis Data Model)	
International Human Drug Registration Technology Request for Coordination	ICH (International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use)
International World Wide Web Association	W3C (World Wide Web Consortium)
MedDRA (Medical Dictionary for Regulatory Activities)	
Extensible Markup Language	XML (Extensible Mark-up Language)
Extensible Stylesheet Language	XSL (Extensible Stylesheet Language)
CDISC (Clinical Data Interchange Standards Consortium)	
clinical summary report	CSR (Clinical Study Report)
Subject Level Analysis Dataset	ADSL (Subject Level Analysis Dataset)
new drug application	NDA (New Drug Application)
Research Label File	STF (Study Tagging File)
Raw Data Standard Model	SDTM (Study Data Tabulation Model)
Annotated Case Report Form	aCRF (annotated Case Report Form)