附件1

医疗器械临床试验数据递交

要求注册审查指导原则

（征求意见稿）

## 一、前言

医疗器械临床试验数据是评价医疗器械安全有效性的重要支持性资料之一。规范地收集、整理、分析和递交医疗器械临床试验数据有助于提高临床试验实施和管理质量，同时有利于监管机构快速、高效的掌握临床试验的开展情况，提高审评效率。

为促进临床试验信息和数据互通，指导注册申请人规范递交临床试验数据及相关资料，帮助更好地开展临床评价资料审评相关工作，制定本指导原则。

本指导原则是在现行法规和标准体系以及当前认知水平下制定的，随着法规和标准的不断完善，以及科学技术的不断发展，本指导原则的相关内容也将进行相应的调整。

## 二、适用范围

本指导原则适用于以产品注册为目的开展的医疗器械临床试验，不适用于按医疗器械管理的体外诊断试剂临床试验，亦不适用于真实世界研究。本指导原则仅涉及临床试验数据递交相关内容，不涉及临床试验过程中数据管理相关要求。

## 三、基本原则

临床试验数据递交时应遵循“真实”“可追溯”和“易读”三项基本原则。

“真实”是指递交的临床试验数据应与实际病例报告表（Case Report Form，CRF）中记录的数据一致。数据录入过程中，不应擅自更改数据。如确实发现CRF表或原始记录存在异常值或笔误，应通过有记录的形式（如Query质疑程序）通知研究者，在原始记录经过研究者签字确认修改后，方可更新数据。

“可追溯”是指按照注册申请人提交的数据和代码，可从原始数据库重现形成分析数据库、临床试验报告中的统计结果，重现形成的数据和统计结果与注册申请人提交的内容一致。

“易读”是指审阅者熟悉和掌握所递交数据变量及其结构关系、实现统计分析复验的容易程度。按照本指导原则相关规范要求递交临床试验数据有助于提高易读性。

## 四、临床试验数据相关资料及其说明

医疗器械临床试验数据相关资料通常包括原始数据库、分析数据库、说明性文件和程序代码，以下对各申报资料具体格式和内容提出要求。注册申请人亦可参照临床数据交换标准协会（Clinical Data Interchange Standards Consortium, CDISC）标准递交数据。

外文资料提供中文翻译件时，需注意对于原始和分析数据集，至少应翻译变量标签、观测值中的描述性文本（如不良事件描述等）。

### （一）原始数据库

原始数据库中的数据通常来源于CRF表。原始数据库中的数据应与CRF表中的数据保持一致，缺失的数据不应进行填补。

原始数据库通常由多个不同的原始数据集组成，单个原始数据集是相同主题下多个变量的集合，这些变量的观测值共同形成该原始数据集，例如，人口学资料数据集可包括年龄、性别、身体质量指数（Body Mass Index， BMI）等。不同临床试验涉及的原始数据库不完全相同。单个原始数据集应收集相同主题下的变量，不建议将不同主题的变量集合形成一个原始数据集，例如，膝关节Lysholm评分和IKDC2000评分相关变量建议分别形成两个原始数据集。

各数据集需包括受试者唯一标识变量，以实现同一受试者不同数据集观测值的关联。如涉及不同访视时间点观测的数据，应使用访视时间变量进行标识。例如，术后3个月和6个月心脏超声相关观测值，访视时间标识变量可命名为Visit\_3、Visit\_6等进行区分。若涉及两个及两个以上临床试验，数据集中需包括临床研究标识变量。

数据集和变量命名应遵循“易读”的原则，建议在对其命名时参考数据集或变量的英文或拼音，通过名称即可联想到该命名的实际含义。例如，可参考Medical History将病史数据集命名为MH，参考Concomitant Medication将合并用药数据集命名为CM，将变量“性别”命名为Gender，变量“受试者姓名缩写”命名为sub\_abbr等。名称最大长度为8字节。若临床试验采用了随机分组，原始数据库中应包含随机号等变量。

### （二）分析数据库

分析数据库是为便于统计分析形成的数据库，用于产生临床试验报告中的统计结果（包括基线、疗效和安全性指标统计分析等）。分析数据库主要包括原始数据库中的变量数据和按照一定规则（如缺失值填补、量表子项评分加和等）使用原始数据库变量数据衍生的数据。

分析数据库通常由多个不同的数据集组成，其中的数据集一般与临床试验报告中的统计结果项相对应。例如，临床试验报告中NIHSS评分统计结果可对应专门的分析数据集，该分析数据集是为了生成NIHSS评分统计结果而专门创建的，包含生成该统计结果的全部变量数据，其他不相关变量数据不纳入到该数据集中。为了便于统计分析复验，分析数据集中的变量应具有可追溯性，变量结构应清晰，不需繁琐的数据前处理即可开展统计分析。

分析数据集可基于其产生的相应统计结果进行命名，例如，生成不良事件比较结果的数据集可称为ad\_ae（不良事件分析数据集）。建议在分析数据集名称中加入ad（analysis data）前缀，以标识该数据集为分析数据集。

变量命名和变量长度同原始数据库要求。注意明确不同分析集（如全分析集、符合方案集和安全集等）标识变量，以及形成数据库过程中产生的系统变量（如有），如序号、时间等。

### （三）说明性文件

#### 1.数据说明文件

数据说明文件用于描述原始数据库和分析数据库的内容和结构，是帮助监管机构快速掌握各数据库中各数据集、变量及其相互间的结构关系，实现准确理解递交数据内容和统计分析复验的重要工具。

建议采用Excel文件，以表格的形式列明原始数据库和分析数据库中所含的数据集、变量、标签、赋值及其对应关系，具体可参考附录1《变量说明性文件示例》。为了便于审阅，数据集和变量应具有相应的中文标签，标签长度不宜过长。若使用了外部词典(如MedDRA)，应明确使用的外部字典名称和版本号。

分析数据库说明文件需列明衍生变量的生成规则，明确涉及到的变量和计算方法。例如，对缺失值的填补，应明确填补方法，填补的程序代码也应提供。

注册申请人应说明原始和分析数据集所用编码（如UTF-8、EUC-CN等），以避免所递交的数据集出现乱码的情形。

#### 2.程序代码使用说明文件

程序代码使用说明文件用来解释说明程序代码文件使用方法，包括使用代码文件时是否需修改以及如何修改程序代码，同时，以表格形式列明各统计结果所使用到的代码文件和数据集文件名称。

#### 3.注释病例报告表

相对于空白CRF，注释CRF增加了注释内容，反映了数据库中变量与CRF表信息收集的对应关系。例如，在性别空白处注释变量名称Gender。利用注释CRF，审阅者可直观地查阅各变量在CRF表中的位置，可核对是否存在遗漏而未被纳入原始数据库中的重要变量数据。

#### 4.其他说明性文件

除以上说明性文件外，鼓励注册申请人提交其他有利于审评人员快速了解临床试验数据库内容和结构的说明性文件（如概览性文件、其他特殊情况说明文件等）。

### （四）程序代码

需递交的代码主要包括：生成随机化编码的代码、用于原始数据库生成分析数据库的代码、分析数据库生成统计结果的代码等，用于调整格式或生成表格的相关代码不需递交。递交的代码应符合通常的编程格式和编程规范，结构清晰明了，易于阅读。代码中以中文注释的形式描述不同过程步的目的以及其他需解释的内容。代码若涉及宏程序，宏程序代码也需提供。应避免提交复杂嵌套的宏程序，如果提交了复杂嵌套的宏程序，应充分验证以确保在不同的计算机系统环境中均可运行该程序。

## 五、递交方式

所有的文件（原始数据库、分析数据库、说明性文件和代码）按照指定的文件夹结构存放在一个总文件夹中（见附录2）。

原始数据库和分析数据库建议采用XPT数据传输格式进行递交，如果一个XPT文件包含一个数据集，XPT文件名需与数据集名称保持一致。如果一个XPT文件包含多个数据集，需注意避免将来自原始数据库和分析数据库的数据集放在同一个XPT文件中。建议采用XPT第5版本（简称XPT V5）或以上版本作为数据递交格式。

数据说明文件可采用PDF、Word、Excel等文件格式，其中变量词典建议采用Excel文件，注释病例报告表建议采用PDF文件。

程序代码建议采用TXT文件格式。

## 附录1

变量说明性文件示例

注：有下划线文字带有超链接，点击可直接跳转到链接位置。例如，点击视图1中“人口学”，可直接跳转至视图2中“人口学”数据集所含变量视图，点击视图2中“Gender”，可直接跳转至视图3中查看性别具体赋值情况。与变量无关的内容不放入变量词典中。

## 附录2

文件夹结构

