

Overview of China Study Data Standards Proposal from C3C/WG

中国研究数据电子递交标准的技术 路线概述—C3C/中国标准化工作 组的视角

John Wang, JNJ

2015-11-13

研究数据递交指导原则草案

- 背景介绍（美国FDA、日本PMDA）
- 和美国、日本的递交指导原则的区别
- 技术路线、可行性分析
- 待翻译列表

研究数据递交的内容

研究列表数据（SDTM/SEND）

- XPT数据集（或DATASET-XML格式）
- 标准化方案（Standardization Plan）
- 标注CRF（又称aCRF）
- 研究数据审阅指南/说明（SDRG）
- 元数据定义表（Define.xml）

分析数据（ADaM）

- XPT数据集（或DATASET-XML格式）
- 分析数据审阅指南/说明（ADRG）
- 分析程序（Analysis Programs）
- 元数据定义表（Define.xml）

美国和日本已发布指导原则和通知

美国FDA – 指导原则Guidance

- 2014-12-17发布3份文档
- 24个月后强制，即2016-12-17之后
- *Section 745A(a) of the Federal Food, Drug, and Cosmetic Act*
- 联邦食品药品和化妆品FD&C法案第745A(a)条

日本PMDA – 通知Notification

- 2015-04-27发布通知，3份文档
- 2016-10-01生效
- 2020-03-31之后强制

和美国、日本的递交指导原则的区别

内容	美国FDA	日本PMDA	中国CFDA	备注
电子格式（标准化研究数据）递交指导原则	X	X	X	将对标准化研究数据的要求整合；额外添加中医药研究的要求
电子格式递交指导原则-标准化研究数据	X			
电子格式标准化研究数据递交指导原则Q&A		X	X	额外添加中医药研究的要求
研究数据技术遵从性指南	X	X	X	额外添加中医药研究的要求

研究数据递交指导原则草案

- 1. 电子格式标准化研究数据递交指导原则
- 2. 电子格式标准化研究数据递交指导原则
Q&A
- 3. 研究数据技术遵从性指南

FDA 研究数据技术遵从性指南

- 汪晨洁 Chenjie Wang, GSK

3. 研究数据技术遵从性指南



Microsoft Word
Document

1. 电子格式标准化研究数据递交指导原则



Microsoft Word
Document

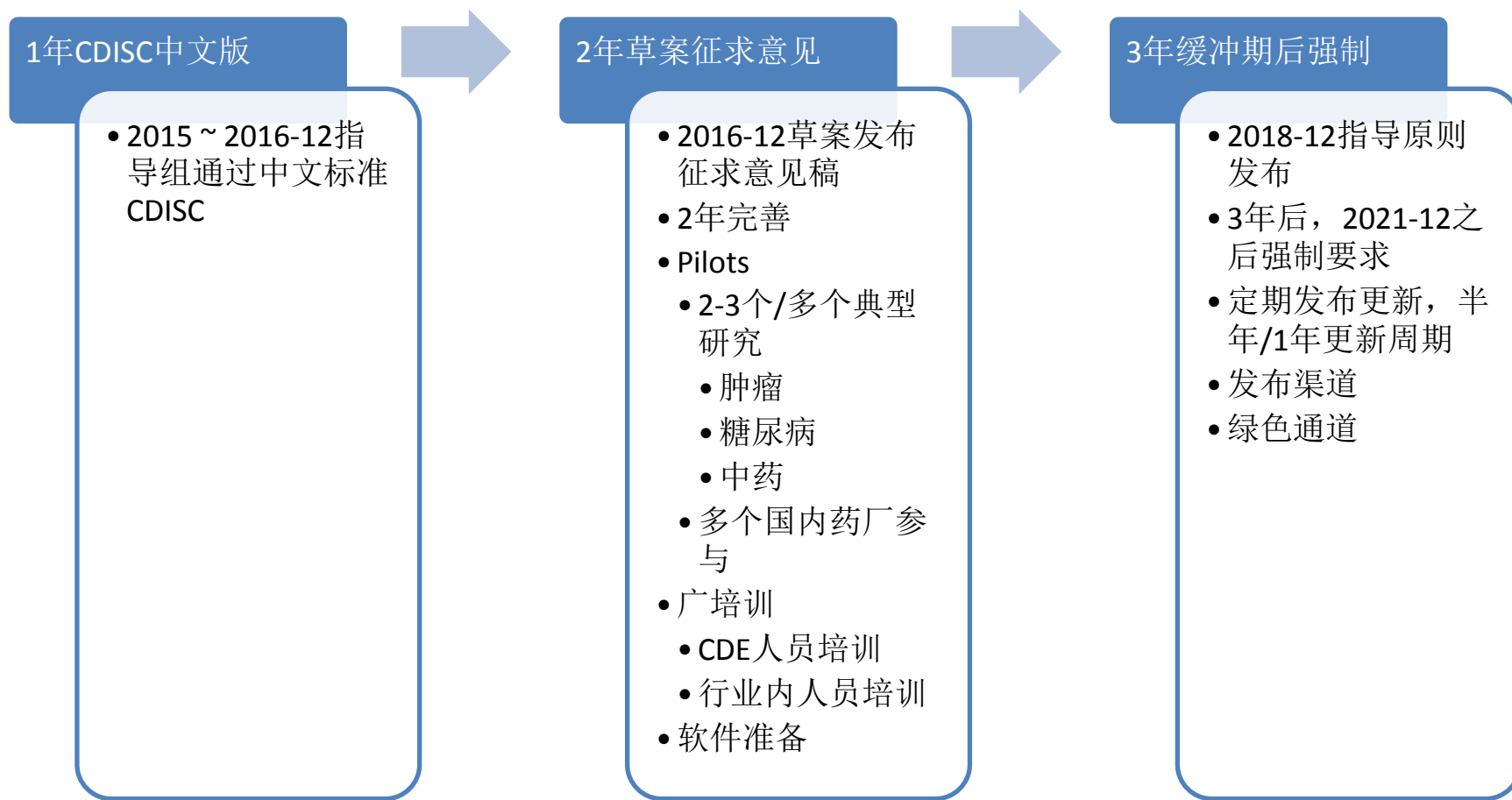
2. 电子格式标准化研究数据递交指导原则Q&A



技术路线、可行性分析

- 实施时间表的要求（正式版发布3年后正式要求）
- C3C官网：持续发布中文CDISC标准、验证规则、审阅指南（SDRG、ADRG）等更新。
- CDE设专门信息渠道发布更新
- CDE设专门网站设置对数据标准的要求
- 绿色通道：以CDISC标准递交的申办方，走绿色通道，优先审评。

技术路线：1+2+3



可行性分析

- CDISC中文版现状
- 2年草案征求意见
- 3年缓冲期

待翻译列表

- 数据标准目录Data Standards catalogs
- 验证规则Validation rules
- 审阅指南Reviewer Guide
 - http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide
 - http://www.phusewiki.org/wiki/index.php?title=Analysis_Data_Reviewer%27s_Guide



MS Office Excel
OpenXML



Microsoft Excel
7-2003 Worksheet



Microsoft Excel
7-2003 Worksheet

致谢

- US FDA guidance
 - 陈丽花(Pamela Chen), Macrostat
 - 辛卫权(Billy Xin), JNJ
 - 郑博, 汪晨洁, GSK
- JP PMDA notification
 - 何昇峰(Shengfeng He), Rundo
 - Yankun Zhang, WuixiPRA
 - 张丽芳(Lifang Zhang), PPD
 - 耿国柱, GSK

加入我们： 致信到 [jwang19 AT its.jnj.com](mailto:jwang19@its.jnj.com)

研究数据技术遵从性指南

行业内电子格式递交申报指导原则- 研究数据标准化

美国药品食品监督管理局, 药品评价研究中心/
生物制品评估和研究中心,
2015年6月, 2.2版

指南摘要

- 第一章: 简介
- 第二章: 标准化研究数据的准备及提供
- 第三章: 数据交换格式- 电子申报
- 第四章: 研究数据申报格式: 临床数据与非临床数据
- 第五章: 治疗领域标准
- 第六章: 受控术语
- 第七章: 电子申报格式
- 第八章: 数据有效性和可追溯性

<http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm>

第一章: 简介

- 适用于研究新药（INDs）, 新药申请（NDAs）, 简略新药申报（ANDAs）以及生物制品许可申请（BLAs）中，动物及人类研究数据的申报
- 早在IND预备会议阶段，申办者就应当讨论：关键数据，数据集中的数据要素，以及数据组内结构组成等问题。
- 所有用来评估项目安全性和有效性的数据应当与FDA支持推广的数据标准相一致

第二章: 标准化的研究数据的准备和提供

- 研究数据标准化设计 (被包含在INDs的总体研究方案中)
 - 计划的试验方案清单
 - 试验类型(例如, I, II or III期)
 - 试验设计(平行,交叉, 开放性延长期研究)
 - 计划执行的数据标准, 格式, 控制术语以及他们的版本, 或者对于那些无法遵照现行标准的试验项目提出的解释
- 研究数据审阅指南 (eCTD 模块4,5)
 - 试验方案名称, 编号, 版本
 - 试验设计
 - 标准, 格式, 受控术语及版本
 - 对研究数据集的说明
 - 数据标准验证规则, 版本, 存在的问题。
 - 申办方关于数据标准的执行中所做过的决定
- 分析数据审阅指南 (eCTD 模块5)
 - 是审阅分析数据集的起点。
(是对define.xml的补充, 以及对ADaM发现的问题的总结, 等等)

第三章：数据交换格式- 电子递交

- 可扩展标记语言(XML)
 - 有助于结构化的数据在不同的信息系统之间实现共享。
例如，define.xml（元数据定义）
- PDF格式
 - 标注CRF
- 文件传输格式
 - SAS 数据传输格式(XPORT)，可用于所有的电子数据集。例如，AE.xpt
 - 数据集大小的要求：>1GB的数据集应该分成数个小数据集以方便传输。
 - 对数据集的栏位长度，变量及数据集描述的长度，变量和数据集名称及标签均有要求。

第四章: 研究数据的递交格式- 临床和非临床研究

- 研究数据表格模型(SDTM): 通过SDTM执行指南为临床试验表格化数据集规定了标准格式 (SDTMIG) 。
- 分析数据模型(ADaM): ADaM执行指南规定了药物临床试验分析数据集的规范, 用以产生和支持用于临床试验报告的结果, 安全性总结报告 (ISS)和有效性总结报告 (ISE).
- 非临床试验数据的交换标准(SEND) 为非临床试验 (例如动物毒性试验) 在药政递交中的表格模型提供数据结构, 格式等规定 (SENDIG) 。
- 标注CRF: 必须被提交。

第五章：治疗领域的标准

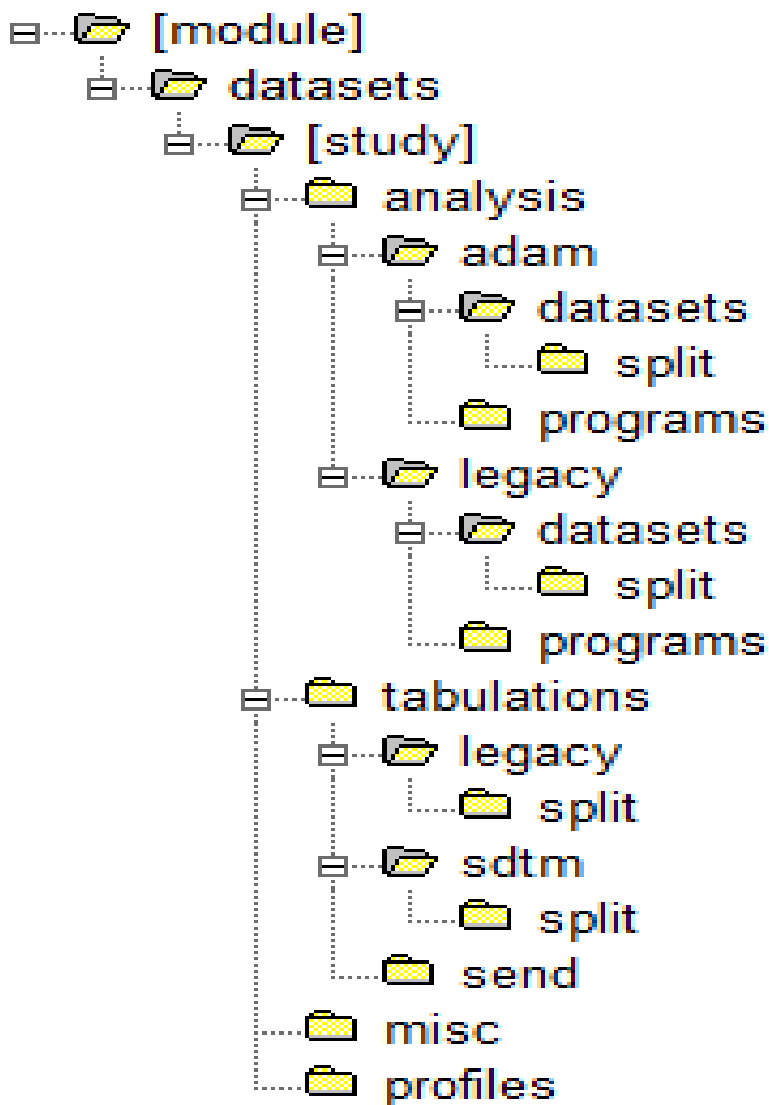
- 以下罗列的这些治疗领域的SDTM域还未完成测试和接受，在目前还不能使用。

SDTM IG Domain		Therapeutic Area User Guide
1.	Healthcare Encounters	Cardiovascular Studies, 1.0; Polycystic Kidney Disease, 1.0; Asthma, 1.0
2.	Microscopic Findings	Tuberculosis, 1.0; Parkinson's, 1.0
3.	Morphology	Cardiovascular Studies, 1.0; Parkinson's, 1.0; Polycystic Kidney Disease, 1.0; Alzheimer's, 1.0; Multiple Sclerosis, 1.0
4.	Procedures	Cardiovascular Studies, 1.0; Polycystic Kidney Disease, 1.0; Alzheimer's, 1.0
5.	Reproductive System	Polycystic Kidney Disease, 1.0
6.	Disease Response	Tuberculosis, 1.0
7.	Skin Response	Asthma, 1.0

第六章: 术语

- 不良事件，合并用药，非药物诊疗，适应证，研究药物和病史应当使用常规字典编码。避免使用申办方自定义的字典。
- CDISC控制术语：除非受控术语不充分， 否则申办方应当使用CDISC受控术语。
- 不良事件：
 - 不良事件的拼写及大小写需符合MedDRA要求
- 药物：
 - 活性成分需使用FDA 唯一成分标识码（UNII）进行识别：SDTM Trial Summary Domain中每种活性成分需有一个对应的UNII
 - WHO Drug字典
- 药理学分类：
 - 国家药物文档-参考术语（NDF-RT）：临床或非临床研究中的所有活性研究成分需用NDF-RT确定药理学分类；该项信息在SDTM Trial Summary domain提供。
- 适应证：
 - 医学系统命名法 — 临床术语（SNOMED CT）：研究药物期望治疗的医学状况需用SNOMED CT编码；该项信息在SDTM Trial Summary domain提供。
- 实验室数据：
 - 逻辑观察标识符名称和代码：LOINC代码是用于实验室和其他临床观察的通用标识。
 - LOINC数据库的实验室部分包含化学，血液学，血清学，微生物学（包括寄生虫学和病毒学），毒理学，和更多的类别。
 - SDTM已经支持使用LBLOINC变量来进行LOINC code的交换。

第七章： 电子化数据提交格式



数据集的文件夹结构：

当使用eCTD格式提交数据时，试验数据集和支持性文件需按特定的文件夹结构进行整理。

第八章：数据验证和可追溯性 – 数据验证

- 数据验证：确保所提交的数据是符合要求并且有效的
- 数据验证规则
 - 一致性验证：数据必须遵照数据标准，比如SDTM
 - 质量检验：比如年龄范围的检测
- 数据验证规则的相关支持
 - 你能在数据标准的网页上找到验证规则：
<http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm>
 - 申办者必须在提交前使用公开的验证规则检验他们的试验数据并且对检验出的错误进行修正，或者在审阅指南（SDRG或ADRG）上阐明不能改正的理由。

第八章：数据验证和可追溯性 – 可追溯性

- 申办方的试验结果可追溯到病例报告(CRF)数据
- 要将原始的试验数据转化为标准数据，可追溯性的建立常常是问题最多的
- **FDA建议：**
 - 准备并且提交既有数据转换计划和报告
 - 提供一份带有两种标注的病历报告：一份基于原始既有数据，一份注释转换后的数据（SDTM）
 - 纪录有意义的数据问题，同时在SDRG（Study Data Reviewers Guide 试验数据审阅指南）内说明
 - 提交既有数据（带注释的病例报告，数据列表和分析数据）

附录： 数据标准和互通性的数据交换

- 互通性：在不同的信息技术系统，软件和网络条件下，实现精确，有效，安全和一致的数据交流与交换.....
 - 技术互通：最低水平的互通，是指能够在两种不同的系统或者机构间简单地交换数据，比如SAS传输文件
 - 语义互通：理解系统分享的信息，那么接受的系统可以处理非数值型的数据，比如使用控制术语和定义元数据的一致性都是支持语义互通的方法
 - 流程互通：一个新兴的概念，已经被定义为一个系统在实际工作中能成功应用的必要条件。简而言之，它涉及到系统是否有能力进行有意义的数据交换，能为接受的系统在业务流程中，在正确的时间自动提供正确的数据。比如，根据一个严重并且意外的不良反应的发生，产生一个加速的不良反应报告。