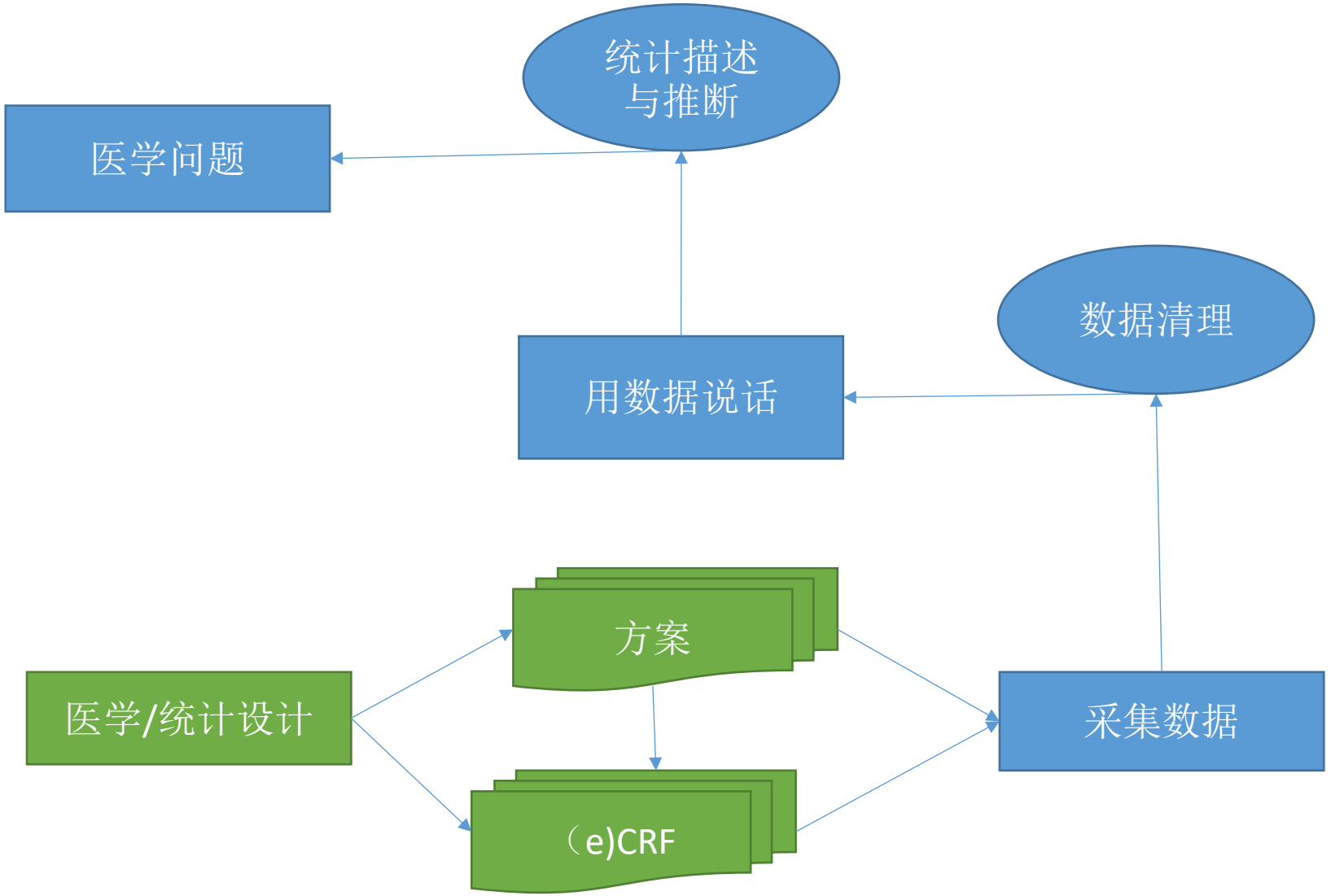
Clinical Data Science

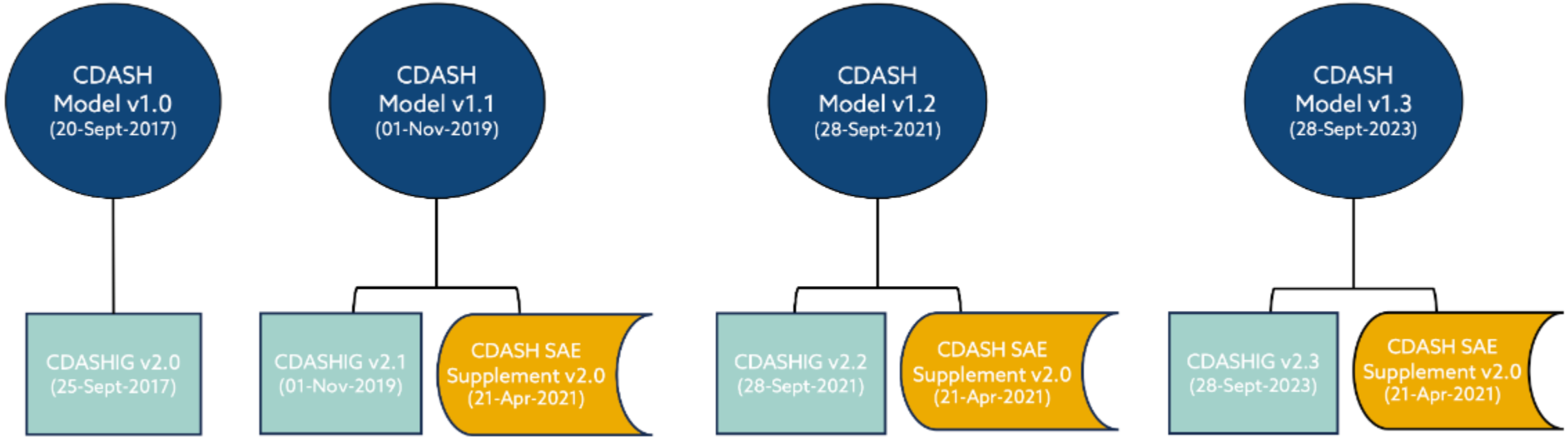
基于CDASH 标准的CRF设计与建库

吴崇胜 (Victor WU)
北京迪时咨询有限公司
Beijing Data Science Express Consulting Co., Ltd
Nov2024

临床研究是在干什么？



CDASH Documents



CRF Design

Clinical Data Science

4. 更容易进行SDTM映射
   (easier sdtm mapping)

CRF Library

3. 更快更好创建新项目
   (set up new studies faster and with higher quality)

重用

2. 减少数据问题
   (avoiding data issues)

数据清理

1. 采集需要的数据
   (necessary data only)

方案

CRF设计要点

Clinical Data Science

- Focus:
  - ✓ 采集需要数据
    - ❖ 方案（解决问题所需要）:科学性
    - ❖ 将来的统计分析与报告（将来怎么用）
  - ✓ 便于清理数据
    - ❖ 确保数据质量
- 易用性
  - ✓ 用起来方便
  - ✓ 符合用户的期望
    - ❖ 方便建库
    - ❖ 易于填写
    - ❖ 方便核查
    - ❖ 方便监查
    - ❖ 便于编程、分析

From routine CRFs to standardized CRF Library

迪时 DATA SCIENCE EXPRESS
Clinical Data Science

- Study CDASH 学习
  - ✓ 原则
  - ✓ 范例
  - ✓ 思考
- Re-construct your CRFs重构
  - ✓ 应用CDASH原则：如该信息是否需要采集?
  - ✓ 遵循CDASH字段与值要求
  - ✓ 思考
- Generalize泛化

Steps for creating a Form

Clinical Data Science

- Which class and which domain
- Data structure

- Leading Questions or not?
- What information to collect (Field)
- Question Text & Field Name

- Code list or free text

Domain 域

Clinical Data Science

Domains are logical groupings of data 域是按逻辑组织在一起的数据

干预类

- 既往与合并用药(CM)
- 暴露(EX)
- 暴露采集(EC)
- 嗜好品使用 (SU)

事件类

- 不良事件(AE)
- 受试者分布(DS)
- 病史(MH)
- 方案偏离(DV)
- 临床事件(CE)

发现类

- 心电图(EG)
- 不满足的入/排标准(IE)
- 实验室检查(LB)
- 体格检查 (PE)
- 问卷与量表 (QS)
- 生命体征(VS)
- 药物发放/回收记录(DA)
- 药代动力学浓度(PC)
- 药代动力学参数 (PP)
- 相关发现 (FA)

特殊用途

- 人口学 (DM)
- 受试者元素(SE)
- 受试者访视 (SV)
- 备注 (CO)

试验设计

- 试验元素(TE)
- 试验分组(TA)
- 试验访视(TV)
- 试验入/排标准(TI)
- 试验总结(TS)

关系描述

- 关联记录(RELREC)

……

Leading Questions or not?

2. The database should contain an indication that a planned exam/assessment was not performed.

3.Data cleaning prompts should be used to confirm that blank CRFs are intentionally blank.

- Usually this will be a "Yes/No" question (e.g., AEYN)

CDASH – Core Designations 核心类别

Clinical Data Science

- **Highly Recommended (HR): 强烈推荐**
  - ✓ A data collection field that must be on the CRF (e.g., a regulatory requirement)
- **Recommended/Conditional (R/C): 推荐/有条件**
  - ✓ A data collection field that should be on a CRF based on certain conditions
  - ✓ "condition" is described in the Implementation Notes
- **Optional (O):可选**
  - ✓ A data collection field that is available for use

Necessary Data Only:

- 4. The same data (i.e., the same information at the same time) should not be collected more than once.
- 11. Manually calculated fields to be recorded? 计算的字段
- 14. The anatomical location of a measurement, position of subject, or method of measurement should be collected only if the protocol specifies the allowable options, or if the parameter is relevant to the consistency or meaning of the resulting data 根据方案，需要时才采集解剖部位、体位或测量方法

Questions text - Clarity

Clinical Data Science

- 7 . CRF questions and completion instructions should be unambiguous, 无歧义 and should not “lead” the site to answer the question in a particular way? 无诱导性
- 8 . CRF questions should be as self-explanatory as possible, thereby reducing the need for separate instructions.自明性

- 日期/时间格式
- 9. Collection of dates should use an unambiguous format, such as DD-MON-YYYY
- 10. To eliminate ambiguity, times should be collected with the use of a 24-hour clock, using the HH:MM:SS format for recording times.

Value: answer list - CT

Clinical Data Science

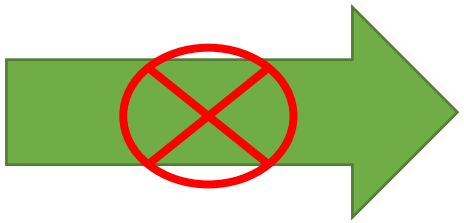
- 12. Questions with free-text responses should be limited to cases of specific safety or therapeutic need in reporting or analysis, such as adverse events, concomitant medications, or medical history—generally in cases where the data will be subsequently coded.? 自由文本

- 6. CRFs should use a consistent order of responses (e.g., “Yes/No”) from question to question, for questions with response boxes or other standardized lists of values. 一致的顺序

避免后续处理困难

Clinical Data Science

Dose Modified
剂量改变？



对研究治疗采取的措施	ACN
	剂量增加
	剂量减少
	频率降低
	无变化
	暂停治疗
	终止治疗
	不适用
	不详

Conformance to the CDASH Standard符合性

Clinical Data Science

- Conformance Rules符合性规则
  - ✓ **字段：Core designations must be followed**
    - ✓ **问题问法：The wording of the CRF questions should be standardized: CDASH Question Text or Prompt must be used to ask the question.**
    - ✓ **字段名称：Variable Names**
    - ✓ **SDTM变量：When CDASH Variables name same as SDTM variable: ideally should require no additional processing**
  - ✓ **受控术语：CDISC Controlled Terminology must be used**
  - ✓ **最佳实践：Best practices must be followed**
  - ✓ **Validated Questionnaires, rating or scales:** restructuring the questions should not be done

Keys to CDASH

- Which Domain?
- What Questions to ask? (Consistent)
- Variable names are harmonized
- Answer list (Consistent)

- Dataset level 数据集水平
  - Name
  - Description
  - Structure
  - Allowed Variables
- Variable level 变量水平
  - Name
  - Label
  - Type
  - Format
- Value level 值水平
  - Controlled Terminology 受控术语

实施CDASH

- Use CDASH standards in (e)CRF design as much as possible尽可能用
- If no CDASH forms/variables are available, take reference to SDTM domains/variables and TA standards 参考SDTM
- When referencing SDTM/TA standards, evaluate the variable first, instead of 100% copying 不是完全照搬
- The more closer your domains/variables are to CDASH/SDTM, the less work your later work 标准化越早越好

终极用法：

-依据标准，建立CRF library

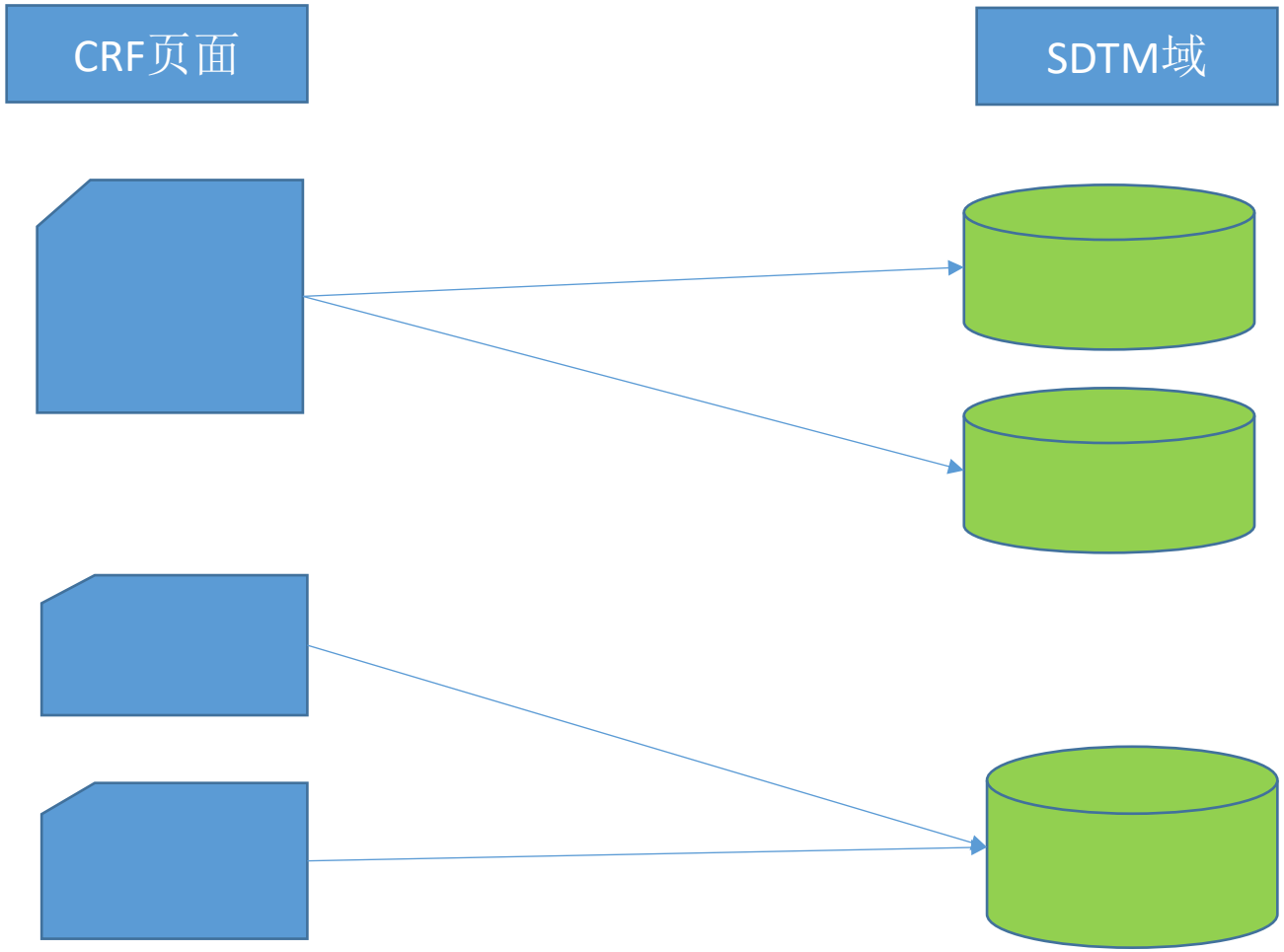
Clinical Data Science

- ✓ 设计质量的保障
- ✓ 设计与审阅的时间减少
- ✓ 更合理的时间与精力分配

- ✓ 培训时间减少：内部、研究中心
- ✓ 便于监查、减少疑问表
- ✓ 录入更快
- ✓ 研究中心填写更省心、省时、省力

- ✓ 再利用、数据整合：CRF、程序、数据的再利用

CDASH vs SDTM



Best Practice Recommendations最佳实践推荐

Clinical Data Science

- 4.1 Best Practices for Creating Data Collection Instruments
  16个常见问题问答
- 4.2 CRF Design Best Practices
  页面/数据库构建 （这里CRF design是狭义指在系统中构建页面/数据库）
- 4.3 Organizational Best Practices to Support Data Collection
  组织层面原原则
- 4.4 General Recommendations on Screen Failures筛选失败

CDASH IG V2.3

4.1 Best Practices for Creating Data Collection Instruments

Necessary Data Only:

- 4. The same data (i.e., the same information at the same time) should not be collected more than once.
- 11. Manually calculated fields to be recorded? 计算的字段
- 14. The anatomical location of a measurement, position of subject, or method of measurement should be collected only if the protocol specifies the allowable options, or if the parameter is relevant to the consistency or meaning of the resulting data 根据方案，需要时才采集解剖部位、体位或测量方法

4.1 Best Practices for Creating Data Collection Instruments

Clinical Data Science

For data clean purpose

1. When a binary response is expected, "Yes/No" responses are preferred over "Check all that apply", because a missing response could lead to a misinterpretation of critical data.

2. The database should contain an indication that a planned exam/assessment was not performed.

3.Data cleaning prompts should be used to confirm that blank CRFs are intentionally blank.

- Usually this will be a "Yes/No" question (e.g., AEYN)

5. A "Check if ongoing" question is recommended to confirm ongoing against an end date. 仍持续

13. Subject-specific data to be pre-populated in the CRF/eCRF? 预填

Clarity

- 7 . CRF questions and completion instructions should be unambiguous, 无歧义 and should not “lead” the site to answer the question in a particular way? 无诱导性
- 8 . CRF questions should be as self-explanatory as possible, thereby reducing the need for separate instructions.自明性

- 日期/时间格式
- 9. Collection of dates should use an unambiguous format, such as DD-MON-YYYY
- 10. To eliminate ambiguity, times should be collected with the use of a 24-hour clock, using the HH:MM:SS format for recording times.

- [unambiguous value: CT/answer list]

Value: answer list - CT

Clinical Data Science

- 12. Questions with free-text responses should be limited to cases of specific safety or therapeutic need in reporting or analysis, such as adverse events, concomitant medications, or medical history—generally in cases where the data will be subsequently coded.? 自由文本

- 6. CRFs should use a consistent order of responses (e.g., “Yes/No”) from question to question, for questions with response boxes or other standardized lists of values. 一致的顺序

- 15. Sites should record verbatim terms for nonsolicited adverse event (AE), concomitant medications(CM), or medical history(MH)-reported terms. Sites should not be asked to select a preferred term from a coding dictionary as a mechanism for recording data.

- 16. An SDTMIG variable name should only be used as a data collection/operational variable name if the collected value will directly populate the SDTMIG variable with no transformation (other than changing case)

4.2 CRF Design Best Practices

Clinical Data Science

1. Place fields that routinely appear on multiple forms at the top of the form 表头
2. Fields in the order … during Clinical Assessment 临床流程
3. Group related fields for a single clinical encounter together 访视
4. Group related fields together (e.g. units)相关字段入一起

5. Data fields that are dependent on other data fields should be placed in the CRF in such a way that this dependence is obvious (e.g. Other, specify). 依赖

6. Lists of values that have a logical order should be provided on the CRF in that logical order. For example, the values of “Low”, “Medium”, and “High” are logically placed in this order.逻辑顺序

4.3 CDASH - Organizational Best Practices 组织层面

Clinical Data Science

1. *Collect necessary data only. 只采集需要的数据*

Controlled Process 流程控制

2. CRF development should be a controlled, documented process that incorporates (as applicable)
3. The CRF design process should include adequate review and approval steps, and each reviewer should be informed on the scope of the review they are expected to provide.
4. Translations of CRFs into other languages should be done under a controlled process

5. *Data that are collected on CRFs should usually be databased.*
6. Establish and use standardized case report forms. 建立与使用标准化CRF

4.4 General Recommendations on Screen Failures

Clinical Data Science

- ICH E3： “It may also be relevant to provide the number of patients screened for inclusion and a breakdown of the reasons for excluding patients during screening, if this could help clarify the appropriate patient population for eventual drug use.

- Although screen-failure data may not be relevant for all studies, it is recommended it be collected based on the needs of the protocol and drug development programs.
- Using CDASH, the minimum data to be collected should include a subject identifier and reason(s) for screen failure. Typically, there is a reason on the End of Study form indicating “Screen Failure”. This information allows overall summarization of all subjects screened/enrolled and, when captured, provides easy subject accountability for the clinical study report. Other data may be considered for collection (e.g., date of informed consent, sex, race, date of birth or age) to further describe the reason for ineligibility (e.g., lab value out of range).

Example for screening failure data

Clinical Data Science

《eCRF填写指南》

筛选失败受试者，请至少完成以下页：

1. 知情同意
2. 人口学
3. 筛选期总结

《筛选期总结》

- 筛选流程完成/退出日期：_____
- 筛选结果
  筛选合格
  筛选失败
- 筛选失败原因：
  不满足入选/排除标准：_____
  受试者主动退出
  其他：_____

CDASH - 最佳实践推荐（Victor 重组）

Clinical Data Science

- **What - Necessary Data ONLY只采集需要的数据**
  - ✓ Support protocol objectives and endpoints.
  - ✓ Calculated?

- **How – for Data Clean Purpose错误/质疑更少**
  - ✓ Question text &completion instructions
  - ✓ Data cleaning prompts/What to database
  - ✓ Free text vs list of values, date/time format
  - ✓ （logical order/Clinical Flow）
  - ✓ Not be pre-populated

- **How – 易用性**
  - ✓ **Layout 版式 :**
    - Header
    - Group related
    - logical order/Clinical Flow

Company level

- **Standardized case Report forms 标准CRF库**
- 流程控制 **Control**
  - ADEQUATE REVIEW
  - TRANSLATIONS
  - When – Stable/Near Final Protocol方案接近定稿

迪时 DATA SCIENCE EXPRESS

欢迎关注迪时咨询

交流数据科学话题

专业的CDISC Solution
基于标准化的数据管理与统计分析服务

victor.wu@datascie.com