

药物临床试验数据管理与统计分析的计划和报告指导原则

(征求意见稿)

国家食品药品监督管理总局药品审评中心

目 录

一、前言.....	3
二、数据管理的计划和报告.....	3
(一) 一般考虑.....	3
(二) 数据管理计划的基本内容.....	4
1. 试验概述.....	4
2. 数据管理流程及数据流程.....	4
3. 采集/管理系统.....	5
4. 数据管理步骤与任务.....	5
5. 质量控制.....	7
(三) 数据管理报告的基本内容.....	7
1. 参与单位/部门及职责.....	8
2. 数据管理的主要时间节点.....	8
3. CRF 及数据库设计.....	8
4. 数据核查和清理.....	8
5. 医学编码.....	9
6. 外源数据管理.....	9
7. 数据管理的质量评估.....	9
8. 重要节点时的数据传输记录.....	9
9. 关键文件的版本变更记录.....	10
10. 报告附件.....	10
三、统计分析的计划和报告.....	10
(一) 一般考虑.....	10
(二) 统计分析计划的基本内容.....	11
1. 试验概述.....	11
2. 评价指标.....	12
3. 分析数据集.....	12
4. 缺失数据和离群值的处理.....	12
5. 统计分析方法.....	12
6. 图表模板.....	14
(三) 统计分析报告的基本内容.....	14
1. 试验概述.....	14
2. 统计分析方法.....	14
3. 统计分析结果.....	15
4. 统计学结论.....	16
5. 报告附件.....	16
四、名词解释.....	17
五、参考文献(略).....	19
六、附录.....	20

一、前言

规范的数据管理计划有助于获得真实完整、准确可靠的高质量数据；而详细的统计分析计划则有助于保证统计分析结论正确和令人信服。为保证临床试验数据的质量和科学评价药物的有效性与安全性，必须事先对数据管理工作和统计学分析原则制定详细的计划书。在试验完成时，对试验中的数据管理和统计分析工作进行全面完整的总结至关重要，通过数据管理报告真实反映临床试验过程中的数据质量和试验样本特征，通过统计分析报告为临床试验总结报告的内容和研究结论提供主要依据。因此，在药物上市注册时，监管部门将数据管理计划和报告与统计分析计划和报告视为评价临床试验结果的重要文件和依据。

虽然我国《药物临床试验质量管理规范》(Good Clinical Practice, GCP)中对药物临床试验数据管理与统计分析进行了原则要求，且国家食品药品监督管理局(China Food and Drug Administration, CFDA)已发布的有关药物临床试验及其统计学的相应技术指南也涉及数据管理和统计分析工作的主要环节，但针对数据管理计划和报告、统计分析计划和报告却没有详细的技术规范和指导性建议。因此，本技术指导原则对此进行了较为详细地介绍和阐述，并提出具体要求，旨在为临床试验的数据管理和统计分析人员提供技术指导，帮助其更好地完成相关工作以达到监管要求。

二、数据管理的计划和报告

(一) 一般考虑

数据管理计划(Data Management Plan, DMP)是由数据管理员依据临床试验方案书写的一份动态文件，它详细、全面地规定并记录某一特定临床试验的数据管理任务，包括人员角色、工作内容、操作规范等。数据管理计划应在试验方案确定之后、第一位受试者筛选之前定稿，经批准后方可执行。通常数据管理计划需要根据实际操作及时更新与修订。

数据管理工作涉及多个单位或业务部门，包括数据管理、临床研究者、统计分析、医学事务、临床监查、临床稽查等单位或部门。数据管理的职责可分为负责、参与、审核、批准、告知等，各单位/部门在数据管理各步骤的职责不尽相同。数据管理计划需明确参与数据管理的相关组织及人员职责，并确保各参与人员具备与其职责相对应的资质与技能。数据管理各步骤需建立并遵循相应的标准操作规程（Standard Operation Procedure, SOP），数据管理计划应列出项目所遵循的 SOP 清单。

数据管理报告是在临床研究结束后，数据管理员撰写的研究项目数据管理全过程的工作总结，是数据管理执行过程、操作规范及管理质量的重要呈现手段。通常以定性和定量的参数来表达，如数据量、疑问数等，并与数据管理计划一起作为药物注册上市的申请材料提交给监管部门用于对临床试验结果的评价。

(二) 数据管理计划的基本内容

数据管理计划应全面且详细地描述数据管理流程、数据管理与采集所使用的系统、数据管理各步骤及任务，以及数据管理的质量保障措施。

1. 试验概述

简要描述试验方案中与数据管理相关的内容，一般包括研究目的和总体设计，如随机化方法及其实施、盲法及设盲措施、受试者数量、评估指标、试验的关键时间节点、重要的数据分析计划及对应的数据要求等。

2. 数据管理流程及数据流程

列出数据管理的工作流程以及试验数据的流程，便于明确各环节的管理，可采用图示方式。

数据管理的工作流程应包含数据采集/管理系统建立、CRF 及数据库的设计、数据接收与录入、数据核查与质疑、医学编码、外源数据管理、盲态审核、数据库锁定、解锁及再锁定、数据导出及传输、数据及数据管理文档的归档等数据管理过程。

数据流程应包含临床试验中所有类型的数据的生成、采集、传输、导入、导出、存档等的位置、负责单位/人、期限等。详细列出每一种类型的试验数据流程，便于明确各种类型和介质的数据的管理，如病例报告表数据、中心实验室检测数据、药代动力学检测数据、电子的患者报告结果（electronic Patient-Reported Outcome, ePRO）数据、影像学数据等。

3. 采集/管理系统

列出采集试验数据的方法，如纸质或电子病例报告表（Case Report Form, CRF）、采用的数据采集/管理系统的名称及版本。描述系统用户的权限控制计划，包含权限定义、分配、监控及防止未经授权操作的措施或方法、权限撤销等。

数据采集/管理系统应具备稽查轨迹、安全管理、权限控制及数据备份的功能，并通过完整的系统验证。

4. 数据管理步骤与任务

(1) CRF 及数据库的设计

CRF 的设计必须保证收集试验方案所规定并满足统计分析需求的所有数据。

不论是何种数据记录方式，均需对相应 CRF 填写指南的建立和管理有所阐述。

数据库的设计通常按既定的注释 CRF 和/或数据库设计说明执行，建立逻辑核查，经用户接受测试（User Acceptance Testing, UAT）合格后方可上线使用。数据管理计划中对此过程应进行简要描述和说明。

(2) 数据的接收与录入

数据管理计划应明确阐述数据采集、接收和录入的方式和过程。

临床试验研究者或临床研究协调员（Clinical Research Coordinator, CRC）应依照 CRF 填写指南，准确、及时、完整、规范地填写 CRF。在数据录入前需制

定数据录入说明，确定数据录入的要求及方式。纸质 CRF 常用双人双份录入，电子 CRF 由临床研究者或由其指定的 CRC 直接录入。纸质 CRF 表还需定义完成 CRF 的发送、转运、接收方式，如传真、邮寄、监查员收集等。同时定义收集频率及记录文件接收的格式等。

(3) 数据核查与质疑

在进行数据核查之前，应制定详细的数据核查计划 (Data Validation Plan, DVP)，明确数据核查内容、方式与核查要求。数据核查通常需要数据管理人员、监查员、医学人员及统计师共同完成。

(4) 医学编码

医学编码是把从 CRF 上收集的不良反应、医学诊断、合并用药、既往用药、既往病史等的描述与标准字典中的术语进行匹配的过程。如采用医学编码，数据管理计划需详细描述编码流程、编码工具、编码字典及版本，以及执行编码的相关标准文件。

(5) 外源数据管理

临床试验外源数据包括实验室数据、电子日志、ePRO、随机化数据、盲态数据等。针对外源数据的管理，数据管理计划中应列出数据传输协议，包括数据类别、数据提供者、数据格式、传输方式、传输频率等，以及对外源数据进行质控的措施，如传输测试、一致性核查等。对于盲态数据，需描述盲态数据的管理流程。

(6) 盲态审核

描述数据管理员如何根据数据盲态审核的要求提供所需要的数据。

(7) 数据库锁定、解锁及再锁定

数据管理计划应详细说明数据库锁定的流程、负责人及执行的标准文件。

数据库锁定后的解锁和再锁定，应事先规定并详细说明其条件和流程。

(8) 数据导出及传输

描述数据的导出和传输的文件格式、导出内容（变量名及变量值编码）、提交程序及符合国家法规和监管部门要求的传输介质。

(9) 数据及数据管理文档的归档要求

试验数据及录入/导入数据库的时间、录入者、数据稽查轨迹及数据管理过程形成的文档都需要完整保存。数据管理过程形成的数据通常包括但不限于：临床试验数据、外源数据、数据库元数据信息、实验室检测参考值范围、逻辑检验及衍生数据变更控制列表、数据质疑表和程序代码等。数据管理过程形成的文件通常包括但不限于：数据管理计划、空白 CRF、CRF 填写指南、完成病例报告表的 PDF 格式文件、注释 CRF、数据库设计说明、数据库录入说明、数据核查计划、数据质控核查报告等。

数据管理计划中应明确需要存档的试验数据、管理文件、介质、归档方式及时限。

5. 质量控制

数据管理计划需确定数据及数据管理操作过程的质控项目、质控频率、样本选取方式及样本量、质量要求及达标标准、对未达到预期质量标准的补救措施等。

(三) 数据管理报告的基本内容

数据管理报告应全面且详细陈述与数据管理执行过程、操作规范及管理质量相关的内容，包括参与单位/部门及职责、主要时间节点、CRF 及数据库设计、数据核查和清理、医学编码、外源数据管理、数据质量保障、重要节点时的数据传输记录、关键文件的版本变更记录，并描述与数据管理计划的偏离。

1. 参与单位/部门及职责

数据管理报告应列出数据管理涉及的所有单位/部门及其在数据管理各步骤的职责。

2. 数据管理的主要时间节点

数据管理各步骤的时间节点可体现数据管理工作的时效性及数据质量，数据录入与数据清理不及时可能有损数据质量。可采用列表方式描述各主要时间节点的起止时间，包括数据录入、数据清理、外源数据管理、数据质控、数据锁库、数据传输、文档归档等主要步骤。

3. CRF 及数据库设计

描述 CRF 及数据库设计各主要步骤的执行情况及具体工作内容/方法，包括 CRF 设计、编制 CRF 填写指南和注释 CRF、形成数据库设计说明以及数据录入说明、数据库建库及数据标准、数据库测试情况等。

4. 数据核查和清理

数据管理报告应描述数据质疑的总体情况，包括质疑总数、每位受试者质疑平均数及每个研究中心质疑平均数。并按照疑问类型进行归类汇总，根据质疑内容不同，通常可分类为人口学资料、入排标准、生命体征、既往病史/用药史、研究用药、合并用药、疗效评估、实验室检查、不良事件等。为体现质疑的及时性，数据管理报告应描述质疑生成的时长（中位天数及其范围）以及研究者答疑的时长（中位天数及其范围）。针对质疑管理中的主要异常问题，数据管理报告应描述出现问题的原因或说明，如质疑数量过高/过低的临床中心/研究者、质疑/答疑时间过长等。

数据管理报告应描述是否有不同于临床数据库的严重不良事件数据库，如有则应描述一致性核查情况，包括试验严重不良事件（Serious Adverse Event, SAE）总数、被核查的 SAE 数量及 SAE 核查频率等，对未核查的 SAE 以及经核查不一致的 SAE 应当详细说明其不一致点和修正情况。

5. 医学编码

对所采用的医学编码，数据管理报告应描述各项内容编码采用的字典名称及其版本号，并列出具体的编码数量，包括手工编码与程序编码的数量。

6. 外源数据管理

描述外源数据的种类，并描述各类外源数据的数据源单位、数据传输协议、数据传输起止日期、传输频率及方式，以及是否执行外部数据的一致性核查和核查结果等。对盲态数据需重点描述维持其盲态的措施。

7. 数据管理的质量评估

在数据库锁定前进行数据质量评估，评估并报告的内容应包含计划与实际发生的临床数据录入时间、数据审核所需时间、质控过程发现并纠正的问题的数量等。

描述数据管理过程中进行数据质控核查的次数，每一次质控核查需描述核查时受试者总例数、抽样例数、抽样比例及依据、抽样方法、总错误率和关键指标与非关键指标的错误率。

数据管理应当严格按照数据管理计划执行，如实际操作中有任何不一致，报告中需详细描述其发生原因，并进一步阐述对数据质量的影响。如数据管理接受稽查或视察，应当描述稽查承担单位、稽查时间、稽查发现的主要问题、采取的纠正和预防措施等。

8. 重要节点时的数据传输记录

试验数据管理过程中可能需要多次数据传输，数据管理报告应描述重要节点的传输记录，包括期中分析的数据传输、数据锁定后向统计分析单位或申办者的传输、以及向药政部门的提交等。描述内容应当包含传输的数据集名称、传输日期、接收单位、传输格式、以及原数据集的储存/备份地点、责任单位/人。

9. 关键文件的版本变更记录

数据管理报告应详细列出与数据管理相关的重要文档的版本变更记录，包括试验方案、CRF、数据库（包括 eCRF 与逻辑检验程序）及数据管理计划的版本变更记录，并描述各版本执行日期、修正内容及修正原因等。

10. 报告附件

以下报告附件作为关键性文件，应视为数据管理报告不可缺少的内容。

- 1) 空白 CRF
- 2) 注释 CRF（可提交电子版）
- 3) 数据库锁定清单及批准文件
- 4) 数据核查计划 DVP（可提交电子版）

三、统计分析的计划和报告

（一）一般考虑

统计分析计划（Statistical Analysis Plan, SAP）是包括试验涉及的全部统计学考虑的一份独立文件，应比试验方案中描述的分析要点具有更多技术细节，且具有实际的可操作性。临床试验的统计分析有其特殊性，统计分析计划应当由具有参与临床试验经验的统计学专业人员起草，要求全面而详细地陈述临床试验数据的分析方法和表达方式，以及对统计分析结果的解释。作为独立文件的统计分析计划初稿应形成于试验方案和病例报告表确定之后，在临床试验进行过程中以及数据盲态审核时，可以进行修改、补充和完善，不同时点的统计分析计划应标注版本及日期，正式文件在数据锁定和揭盲之前完成并予以确认。如果试验过程中试验方案有修订，则统计分析计划也应作相应的调整。如果涉及期中分析，则相应的统计分析计划应在期中分析前确定。

统计分析报告（Statistical Analysis Report, SAR）是根据统计分析计划，对试验数据进行统计分析后形成的报告，是临床试验结果的重要呈现手段，是撰写临床研究报告（Clinical Study Report, CSR）的重要依据，并与统计分析计划一起作为药物注册上市的申请材料提交给监管部门用于对临床试验结果的评价。

（二） 统计分析计划的基本内容

统计分析计划的基本内容涵盖了设计的类型、比较的类型、随机化与盲法、主要指标和次要指标的定义与测量、检验假设、数据集的定义、疗效及安全性评价和统计分析的详细计划。确证性试验要求提供主要指标的分析原则及预期分析方法。探索性试验通常描述概括性的原则和方法。

1. 试验概述

试验概述是试验方案中与统计学相关的部分，常可直接摘录。一般包括以下主要内容：

- 1) 研究目的：临床试验的主要目的和次要目的。
- 2) 设计类型：如平行设计、交叉设计、析因设计、成组序贯设计等。
- 3) 对照的类型：如安慰剂对照、阳性对照、剂量组对照等，需说明试验选择的对照类型及理由。
- 4) 随机化方法及其实施：明确随机化方法，如区组随机、分层随机及其分层因素、中央随机化及其所控制的基线因素等。
- 5) 盲法及设盲措施：说明是单盲还是双盲，设盲措施是双盲单模拟、双盲双模拟等，以及保持盲态下执行统计分析的措施。若采用开放设计，需充分说明无法实施盲法的理由。
- 6) 样本量：计划入组的受试者数量及其计算依据。若采用成组序贯设计应说明不同阶段的样本量。

2. 评价指标

统计分析计划中应清晰描述主要指标和次要指标的定义，包括具体观察和测量的方法、观察时点、指标属性。如果主要指标需要通过计算得到，则需给出相应的计算公式。

3. 分析数据集

根据不同研究目的，在统计分析计划中需明确描述数据集的定义。临床试验的分析数据集一般包括全分析集（Full Analysis Set, FAS）、符合方案集（Per Protocol Set, PPS）、安全性数据集（Safety Set, SS）。在定义分析数据集时，需遵循两个原则：①尽可能地减小偏倚；②控制 I 类错误的增加。

4. 缺失数据和离群值的处理

缺失值和离群值是临床试验中潜在的偏倚来源之一，但在实际的临床试验中往往难以避免。因此，一方面在试验的计划、执行过程中应有必要的措施尽量避免其发生，另一方面在统计分析计划中应预先说明主要疗效指标缺失值的填补方法及理由，离群值的处理方法应当从医学和统计学两方面去考虑，并在统计分析计划中明确描述。

5. 统计分析方法

统计分析应建立在真实、可靠、准确、完整的数据基础上，应根据研究目的、试验方案和观察指标的类型选择国内外公认的统计分析方法。应给出不同类型资料的描述及统计推断方法，明确采用的单双侧检验及其水准，并说明所采用的统计软件及版本号。

（1） 比较类型和检验假设

明确临床试验的比较类型，如优效性检验、非劣效性/等效性检验及其界值等。写出主要指标进行统计学检验的原假设和备择假设及其检验水准等。

要注意多个主要指标、多个比较组、多个时间点的比较、期中分析、亚组分析等情况的多重性问题，说明控制 I 类错误率的措施。

(2) 人口学资料和基线特征分析

说明对于人口学等基线资料根据数据性质进行描述统计分析的具体方式。

(3) 依从性和合并用药分析

对于依从性和合并用药的分析，说明所采用描述性统计分析的具体方式，并说明对依从性差、具有合并用药的受试者具体情况的描述方式。

(4) 主要指标的分析

说明主要指标分析采用的统计分析方法和统计分析模型。分析模型的选择要注意考虑指标的性质及数据分布的特性。处理效应的估计应给出效应大小、置信区间和假设检验结果。有些基线在统计分析中可作为协变量处理，但必须在统计分析计划中事先说明。

在确证性试验中，只有统计分析计划中事先规定的统计分析内容才可以作为确证性试验的证据，其他的分析结果只能是探索性的。

(5) 次要指标的分析

对于次要指标的统计分析，处理效应的估计也需要给出效应大小、置信区间和假设检验方法。

(6) 安全性分析

安全性分析的资料主要来源于受试者的主诉、症状、体征以及实验室检查结果等，所有的安全性指标在分析中都需要高度重视，应考虑对不良事件采用统一的编码词典进行编码。对于安全性数据的分析需说明所采用的统计学分析方法。

对不良事件的分析，应按事件发生的频数、频次和发生率描述，并进行组间发生率的比较。分析计划中需说明各种不良事件/反应的分类和汇总方式和所采

用的具体不良事件编码词典名称及其版本号。

(7) 其他分析

除以上的分析之外，有时还考虑期中分析、亚组分析、敏感性分析等。

期中分析的时点（包括日历年时点或信息时点）、具体实施方式和所采用的 α 消耗函数等应当事先制订计划并在试验方案中阐明。如由于特殊情况进行了计划外的期中分析，则在研究报告中应解释其必要性以及破盲的程度和必要性，并提供可能导致偏倚的严重程度以及对结果解释的影响。

当涉及亚组分析时，需要对亚组给出明确定义。对于非预先规定的缺失数据的填补、离群值、亚组分析、不同数据集的分析、不同协变量的调整等，可进行敏感性分析，考察对试验结果的影响。

6. 图表模板

统计分析结果通常以统计分析表或图的形式呈现，计划中应该以简明的格式、精炼的文字描述所有相关信息。

(三) 统计分析报告的基本内容

统计分析报告是对临床试验的统计设计、分析、结果的总结，是临床试验报告的基础和依据，其基本内容包括：试验概述、统计分析方法、统计分析的结果与结论，一般采用统计表和统计图表示。统计分析报告中的所有结论应使用准确的统计学术语阐述。

1. 试验概述

统计分析报告中的试验概述应与统计分析计划一致。

2. 统计分析方法

统计分析报告中的统计分析方法应与统计分析计划一致。

3. 统计分析结果

(1) 受试者的分布

统计分析报告中应写明所有入组的受试者的分布情况，包括筛选例数、剔除例数及原因、参与随机化的例数、各组脱落或剔除受试者的例数、百分比等，以及方案偏离情况、各分析数据集的分布。除文字、表格描述外，应采用流程图的方式描述受试者的分布情况（流程图参见附录）。

详细描述每一位因脱落/剔除等原因未进入各分析数据集的受试者的情况，如受试者编号、中心、入组时间、脱落或剔除原因及时间等。

(2) 人口学资料和基线特征分析

对于人口学资料、既往病史、家族史、药物过敏史以及疗效指标的基线值等数据常采用统计描述的方式进行可比性分析。计量资料一般用均数、中位数、标准差、四分位数、最大值和最小值等进行描述；计数及等级资料一般用频数和百分比描述。

(3) 依从性和合并用药分析

根据依从性定义，报告各受试者完成试验的情况，包括研究时间、药物暴露时间、药物使用量以及方案偏离发生率的情况，列表描述依从性差的受试者、依从性差的具体原因及进入分析数据集情况。

对于合并用药分析，需列出合并药物的详细情况，如受试者编号、中心、组别、合并药物名称、使用原因、开始时间、结束时间等，进行组间合并用药的比较。

(4) 疗效分析

对于主要和次要疗效指标，需根据事先确定的统计分析方法进行统计描述和统计推断，可能包括指标基线情况、治疗后各访视点的测量值及前后变化情况，以及变化值组间差异的描述统计量、置信区间和组间比较的检验统计量及P值等。

对于主要指标，应报告效应大小、置信区间和假设检验结果，根据事先确定的标准，从统计学角度判断主要指标的优效性/非劣效性/等效性的假设是否成立。

(5) 安全性分析

安全性分析应按统计分析计划给出统计分析结果。需要分类汇总各种不良事件/反应、严重不良事件/反应、重要不良事件/反应的发生率、严重程度及可能进行的组间比较。并列表描述每位受试者每项不良事件/不良反应发生的详细情况，包括不良事件的类型、严重程度、发生和持续时间、结局以及与试验药物及药物剂量的关系等。

对实验室指标的比较和评价，主要关注治疗前正常而治疗后异常的发生情况，以及治疗前异常但在治疗后加重的受试者，需列表描述上述两种情况。生命体征、心电图、体格检查以及其他安全性相关指标的分析与实验室检查指标的分析类似。必要时，进行实验室指标前后变化及组间比较。

4. 统计学结论

根据主要指标的统计分析结果，结合研究的设计类型、样本量、试验实施情况、次要指标及敏感性分析结果等阐述证据的充分性和结果的稳健性，并给出统计学结论：明确针对主要指标的统计假设是否成立，并简要描述安全性的主要统计结果。

5. 报告附件

以下报告附件作为关键性文件，应视为统计分析报告不可缺少的内容。

- 1) 原始数据库、分析数据库及相应的变量说明文件（数据库应为 sas7bdat 格式）
- 2) 受试者分布流程图
- 3) 随机化方案（含随机分配表）

- 4) 盲态审核决议
- 5) 补充正文的统计附图和附表
- 6) SAS 分析代码（必要时）
- 7) 统计方法的发表文献（必要时）

四、名词解释

稽查轨迹(Audit Trail): 是计算机系统（如数据管理系统）的基本功能。意旨系统采用安全的和计算机产生的带有时间烙印的电子记录，以便能够独立追溯系统用户输入、修改或删除每一条电子数据记录的日期、时间，以及修改原因，以便日后数据的重现。任何记录的改变都不会使过去的记录被掩盖或消失。只要受试者的电子记录保存不变，这类稽查轨迹文档记录就应当始终保留，并可供监管视察或稽查员审阅和复制。

系统验证(System Validation): 意旨建立计算机化系统生命周期管理的文档化证据，以确保计算机化系统的开发、实施、操作以及维护等环节自始至终都能够高度满足其预设的各种系统技术标准、使用目的和质量属性，和处于监控的质量管理规程中，并能在其投入应用直至退役过程中都能高度再现和维护系统的标准和功能符合监管要求。

权限控制(Access Control): 是指按照临床试验电子系统的用户身份及其归属的某项定义组的身份来允许、限制或禁止其对系统的登录或使用，或对系统中某项信息资源项的访问、输入、修改、浏览能力的技术控制。

注释 CRF (Annotated CRF): 在空白的临床试验病例报告表中记录所收集数据位置的注释，即标示在病例报告表数据库中数据集及其变量名所含每个数据点的位置。

逻辑核查(Edit Check): 是指临床试验数据输入计算机系统后对数据有效性的检查。这种核查可以通过系统的程序逻辑，子程序和数学方程式等方法实现，主要评价输入的数据域与其预期的数值逻辑、数值范围或数值属性等方面是

否存在错误。

用户接受测试 (User Acceptance Testing, UAT) : 用户接受测试是由临床数据管理系统的用户进行的一种检测方式, 检测记录可用以证明所设计系统经过了相关的验证过程。用户应全面检测所有正确和错误数据组合, 记录检测结果。全面的检测文档应包括验证方案、测试细则记录、测试总结报告和验证总结报告等。

数据核查计划 (Data Validation Plan, DVP): 也称逻辑核查计划, 是由数据管理员为检查数据的逻辑性, 依据临床试验方案以及系统功能而撰写的系统设置文件。

盲态数据: 是指可以从中分辨出受试者试验组别的数据, 如血液样品中的药物浓度或某些关键变量等。

盲态审核 (Blind Review): 是指在试验结束 (最后一位受试者最后一次观察) 到揭盲之前对数据进行的核对和评估, 以便最终确定统计分析计划。

方案偏离 (protocol deviation) : 是指任何有意或无意偏离和不遵循未经 IRB 批准的试验方案规定的治疗规程, 检查或数据收集程序的行为。一般来说, 这种偏离只是逻辑的或管理性的偏离试验方案, 不会对受试者的安全和福祉产生实质性的作用, 也不会影响所收集数据的价值。

方案违背 (protocol violation) : 是指严重违规 GCP 原则、国家监管法规、申办者或研究机构 SOP 标准、和伦理委员会批准的试验方案规定的治疗、检查或数据收集程序。这种违规危害研究的科学性和受试者的权益, 安全性或福祉, 改变风险 / 效益比或伦理原则, 和对试验数据的质量和可信性有严重影响。

期中分析 (Interim Analysis): 是指在正式完成临床试验前, 按事先制订的分析计划, 对处理组间的有效性和安全性进行比较的分析。

缺失数据 (Missing Data) : 是指按照研究方案要求收集但未观测到的数据。

离群值 (Outliers) : 是指严重偏离平均水平的观测数据。离群值可能由于变量

的变异较大所致，也有可能由过失误差引起；若是后者，应说明原因后作为缺失数据处理。

亚组分析 (Subgroup Analysis): 是指对整体中根据某种因素分层的部分数据进行分析。

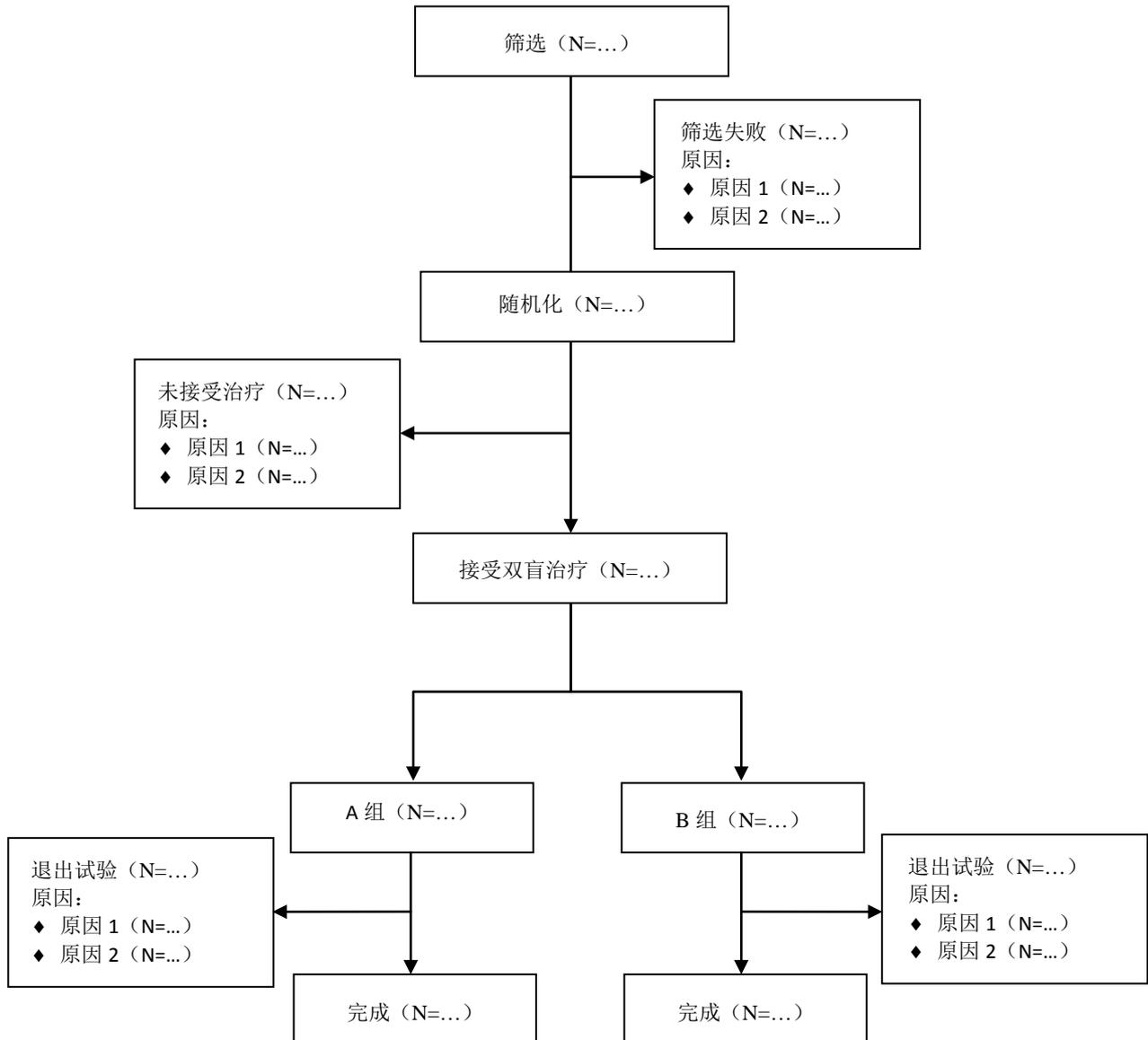
敏感性分析 (Sensitivity Analysis): 是指对非预先规定的试验中可能出现的各种情况进行分析，如缺失数据的填补、亚组分析、不同数据集分析、不同协变量的调整等，并将分析结果作为参考，与事先确定的分析结果进行比较，考察所得结果的一致性和稳定性。敏感性分析可以作为主要分析的附加支持，但不能作为结论的主要依据。

重要不良事件 (Significant Adverse Event): 指的是除严重不良事件外，发生的任何导致采用针对性医疗措施（如停药、降低剂量和对症治疗）的不良事件和血液学或其他实验室检查明显异常。

五、参考文献（略）

六、附录

1. 受试者分布流程样图



2. 统计报告的基线结果模板

分类变量和连续变量基线结果表格参考格式见下：

表 2.1.1 基线情况

指标	试验组	对照组	合计	统计量	P 值
分类变量	类 1				
	类 2				
	合计				
连续变量	N(Nmiss)				
	Mean(SD)				
	Median				
	Min,Max				

3. 统计报告的安全性分析结果模板

- 1) 安全性小结和结论
- 2) 受试者用药与暴露的程度

描述受试者在研究期间的用药持续时间与暴露量,如果有必要,可以分性别、分阶段(例如化疗周期)进行描述。表格参考格式见下：

表 3.2.1 受试者用药与暴露的程度

指标	试验组	对照组	合计	统计量	P 值
分类变量	类 1				
	类 2				
	合计				
连续变量	N(Nmiss)				
	Mean(SD)				
	Median				
	Min,Max				

3) 不良事件

分组描述不良事件/与药物相关不良事件/严重不良事件/重要不良事件/导致脱落的不良事件的发生例数与发生率。

分组描述各系统不良事件/与药物相关不良事件/严重不良事件/重要不良事件/导致脱落的不良事件的发生例数与发生率。

根据不同严重程度和药物暴露量分组描述各系统不良事件/与药物相关不良事件的发生例数与发生率。表格参考格式见下。

表 3.3.1 不良事件总结

	试验组			对照组			P值
	例次	例数	发生率	例次	例数	发生率	
不良事件							
与研究药物相关不良事件							
严重不良事件							
重要不良事件							
导致脱落的不良事件							

表 3.3.2 各系统不良事件发生情况

	试验组			对照组		
	例次	例数	发生率	例次	例数	发生率
合计						
SOC1						
PT1						
PT2						
SOC2						
...						

表 3.3.3 各系统相关不良事件发生情况

	试验组			对照组		
	例次	例数	发生率	例次	例数	发生率
合计						
SOC1						
PT1						
PT2						
SOC2						
...						

表 3.3.4 各系统严重不良事件发生情况

	试验组			对照组		
	例次	例数	发生率	例次	例数	发生率
合计						
SOC1						
PT1						
PT2						
SOC2						
...						

表 3.3.5 各系统重要不良事件发生情况

	试验组			对照组		
	例次	例数	发生率	例次	例数	发生率
合计						
SOC1						
PT1						
PT2						
SOC2						
...						

表 3.3.6 各系统导致脱落的不良事件发生情况

	试验组			对照组		
	例次	例数	发生率	例次	例数	发生率
合计						
SOC1						
PT1						
PT2						
SOC2						
...						

表 3.3.7 各系统不同严重程度不良事件发生情况

	试验组									对照组								
	轻			中			重			轻			中			重		
	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率
合计																		
SOC1																		
PT1																		
PT2																		
SOC2																		
...																		

表 3.3.8 各系统不同严重程度相关不良事件发生情况

	试验组									对照组								
	轻			中			重			轻			中			重		
	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率
合计																		
SOC1																		
PT1																		
PT2																		
SOC2																		
...																		

表 3.3.9 各系统不同暴露量不良事件发生情况

	试验组									对照组								
	暴露量1			暴露量2			...			暴露量1			暴露量2			...		
	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率
合计																		
SOC1																		
PT1																		
PT2																		
SOC2																		
...																		

表 3.3.10 各系统不同暴露量相关不良事件发生情况

	试验组									对照组								
	暴露量1			暴露量2			...			暴露量1			暴露量2			...		
	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率	例次	例数	发生率
合计																		
SOC1																		
PT1																		
PT2																		
SOC2																		
...																		

4) 临床实验室检查

采用前后交叉表的方式描述实验室检查结果。表格参考格式见下。

表 4.1.1 临床实验室检查前后交叉表

组别	治疗前	治疗后			未查	合计
		正常	异常无临床意义	异常有临床意义		
试验组	正常					
	异常无临床意义					
	异常有临床意义					
	未查					
	合计					
对照组	正常					
	异常无临床意义					
	异常有临床意义					
	未查					
	合计					

5) 心电图

心电图描述参考实验室检查。表格参考格式见下。

表 5.1.1 心电图检查前后交叉表

组别	治疗前	治疗后			未查	合计
		正常	异常无临床意义	异常有临床意义		
试验组	正常					
	异常无临床意义					
	异常有临床意义					
	未查					
	合计					
对照组	正常					
	异常无临床意义					
	异常有临床意义					
	未查					
	合计					

6) 其他安全性指标

包括生命体征、体重检查等。

.....

4. 数据管理总结报告范例

数据管理总结报告范例

版本号：[列出方案版本号，如“V1.0”]

版本日期：YYYY-MM-DD

方案名称：[列出方案标题，如“临床试验 CNS1234”]

方案代码：[列出方案编号，如“PHAMA- CNS1234”]

申办单位：[列出申办方名称，如“医药公司 AB”]

数据管理单位：[列出数据管理单位，如“临床合同公司 CD”]

报告签字页

我在此签字，申明已详细阅读了该报告，并确认该报告准确地描述了本临床试验的数据管理过程。

报告编制人

姓名： XXX

日期： YYYY-MM-DD

签名： _____

数据管理负责人

姓名： ZZZ

日期： YYYY-MM-DD

签名： _____

申办方数据管理负责人

姓名： UUU

日期： YYYY-MM-DD

签名： _____

目 录

1. 试验文件记录
2. 数据管理项目参与单位 /部门及职责
3. 数据管理主要时间节点
4. CRF 及数据库设计
5. 数据清理
 - 5.1 疑问的总体情况
 - 5.2 疑问的处理情况
 - 5.3 疑问管理中的主要问题
6. 医学编码
7. SAE 一致性核查
8. 外源数据管理
9. 数据质控和稽查
 - 9.1. 数据质量评估
 - 9.2. 人工数据核查评估
 - 9.3. 数据管理过程稽查
10. 临床试验数据集提交
11. 数据管理实际过程与数据管理计划不一致

临床试验过程中试验方案，病例报告表（CRF），数据库及数据管理计划的制定与修正：

本次数据管理过程中，总共对试验方案进行 n* 次修正。

试验方案原始版本及日期： [版本号]，YYYY-MM-DD

试验方案最终版本及日期： [版本号]，YYYY-MM-DD

本次数据管理过程中，总共对 CRF 进行 n 次修正。

CRF 原始版本及日期： [列出 CRF 初始版本号和版本日期，如“EDC eCRF, Version 1.0 CNS1234 01JUL2013 YYY-MM-DD”]

CRF 最终版本及日期： [列出使用中的 CRF 终板版本号和版本日期，如 “EDC eCRF, Version 2.0 CNS1234 28DEC2013 YYYY-MM-DD”]

本次数据管理过程中，总共对数据库进行 n 次修正。

数据库原始版本及日期： [列出数据库初始版本号和版本日期，如“V1.0，YYYY-MM-DD”]

数据库最终版本及日期： [列出数据库终版版本号和版本日期，如“V2.0，YYYY-MM-DD”]

本次数据管理过程中，总共对数据管理计划进行 n 次修正。

数据管理计划原始版本及日期： [列出数据管理计划初始版本号和版本日期，如“PHAMA-CNS1234V1.0 YYYY-MM-DD “]

数据管理计划最终版本及日期： [列出数据管理计划终稿版本号和版本日期，如“PHAMA-CNS1234V3.0 YYYY-MM-DD”]

注：* n 代表修正的具体次数

版本号： [下注脚列出版本号]

第 N 页

版本日期： YYYY-MM-DD

1. 试验文件记录

试验方案	版本号	主要修正内容及修正理由	执行日期
(列出试验方案名称, 如 PHAMA-CNS1234)	(顺序列出所有版本, 如 V2.0)	(顺序列出各版本修改内容和理由, 如 修正入组条件)	(顺序列出各版本执行日期, 如 YYYY/MM/DD)
...(根据实际方案版本增减行列数)
CRF	版本号	主要修正内容及修正理由	执行日期
(列出 CRF 名称, 如 PHAMA-CNS1234)	(顺序列出所有版本, 如 V2.0)	(顺序列出各版本修改内容和理由, 如 CRF Page 4 增加入组条件选项。根据实验方案修正)	(顺序列出各版本执行日期, 如 YYYY-MM-DD)
...(根据实际 CRF 版本增减行列数)
数据库	版本号	主要修正内容及修正理由	执行日期
(列出数据库名称, 如 CNS1234)	(顺序列出数据库版本号, 如 DB20131228B1.1)	(顺序列出各数据库版本内容与修改理由, 如 外部数据导入测试后数据库版本改变)	(顺序列出各版本执行日期, 如 YYYY-MM-DD)
...(根据实际数据库版本增减行列数)
数据库解锁	版本号	解锁理由及参与人员	执行日期
(列出被解锁的数据库名称, 如 CNS1234-02。如	(顺序列出所有再锁定版本号。	(顺序列出各数据库版本解锁理由与参与人员名	(顺序列出各版本执行日期,

版本号: [下注脚列出版本号]

第 N 页

版本日期: YYYY-MM-DD

没有解锁发生，可以填“未发生”)	如没有则不填)	单。如未发生，可以不填)	如 YYYY-MM-DD)
...(根据实际解锁情况增减行列数)
数据管理计划	版本号	主要修正内容及修正理由	执行日期
(列出计划名称，如 PHAMA1234)	顺序列出所有版本，如 V1.1)	(顺序列出各版本修改内容和理由，如逻辑核查 文件修正)	(顺序列出各版本执行日期， 如 YYYY-MM-DD)
...(根据实际计划版本增减行列)
...(根据实际文件条目增减行列数)
.....

2. 数据管理项目参与单位 /部门及职责

(注：可以在各项目下列出职责部门的角色，如负责，参与，告知，不适用，审核，批准等，下表为试验项目包括申办者、研究者和 CRO 三方的一个参考案例，实际中可根据情况进行调整。)

项目	CRO								研究者	申办者
	数据管理单位/部门	统计分析单位/部门	临床监查部门	项目管理部	医学部门	QA 质保部		
CRF 设计	负责	审核	审核		告知	审核		不适用	告知	审核/批准
CRF 填写说明	负责	审核	审核		不适用	参与		不适用	告知	审核
eCRF 的测试	负责	不适用	参与		参与	参与		不适用	参与	审核

版本号：[下注脚列出版本号]

第 N 页

版本日期：YYYY-MM-DD

数据采集（EDC）系统上线	负责	告知	告知		告知	告知		不适用	告知	审核/批准
数据库建立及测试	负责	审核	不适用		参与	参与		不适用	不适用	审核
数据核查计划	负责	审核	参与		不适用	参与		不适用	不适用	审核/批准
数据管理系统上线	负责	告知	不适用		不适用	不适用		不适用	不适用	审核/批准
数据管理计划	负责	参与	参与		不适用	参与		不适用	不适用	审核/批准
数据录入	负责	不适用	告知		不适用	不适用		不适用	参与	告知
外部数据管理	负责	参与	参与		不适用	不适用		不适用	不适用	审核/批准
数据质疑	负责	参与	参与		不适用	参与		不适用	不适用	审核
数据质疑管理	负责	不适用	参与		不适用	不适用		不适用	参与	审核
医学编码	负责	告知	不适用		不适用	审核		不适用	不适用	审核
期中分析数据	负责	审核	参与		告知	参与		不适用	参与	审核/批准
数据传输	负责	参与	不适用		不适用	不适用		不适用	不适用	审核/批准
数据质控	负责	不适用	不适用		不适用	不适用		不适用	不适用	审核
数据库锁定	负责	审核/批准	审核/批准		告知	审核		不适用	参与	审核/批准
数据文档保存	负责	不适用	参与		告知	参与		不适用	参与	审核
数据管理过程稽查	参与	参与	参与		不适用	参与		负责	参与	参与
.....

版本号：[下注脚列出版本号]

第 N 页

版本日期：YYYY-MM-DD

3. 数据管理主要时间节点*

(*：可以根据各项数据任务条目名称分别列出相关信息，案例见表中。)

任务条目	开始日期	结束日期	备注
数据管理过程	YYYY-MM-DD	YYYY-MM-DD	(如需说明，可以在此注明)
数据录入	YYYY-MM-DD	YYYY-MM-DD	(如需说明，可以在此注明)
数据清理	YYYY-MM-DD	YYYY-MM-DD	(如需说明，可以在此注明)
外部数据管理	YYYY-MM-DD	YYYY-MM-DD	(如需说明，可以在此注明)
数据质控	YYYY-MM-DD	YYYY-MM-DD	(如需说明，可以在此注明)
盲态审核数据提交	YYYY-MM-DD	YYYY-MM-DD	(如需说明，可以在此注明)
数据库锁库	YYYY-MM-DD	YYYY-MM-DD	(如需说明，可以在此注明)
数据文件存档	YYYY-MM-DD	YYYY-MM-DD	(如需说明，可以在此注明)
...(根据实际任务条目增减行列数)

4. CRF 及数据库设计*

(*：根据与 CRF 相关的实际工作，列出相应环节信息和工作性质描述，案例见表中。)

执行工作	具体内容	备注
选择数据采集工具	eCRF	电子病例报告表(EDC)
CRF 填写指南交付	在线帮助系统	
注释 CRF	电子化技术自动标注 CDASH	
数据库设计	依据 CDISC 和注释 CRF	
数据库测试	模拟 CRF 数据 5 份；项目 CRF 数据 0 份	见 UAT 计划，测试报告和结果
...(根据实际任务条目增减行列数)

5. 数据清理

5.1 疑问的总体情况*

(*: 按实际分类顺序列出各类数据集的名称, 各数据集的疑问数量, 以及高疑问率的根源分析。案例见表中。)

疑问类型	疑问数量	最高频率疑问产生原因
人口学资料	2	
入选/排除标准	23	方案的入排条件把控培训不到位
生命体征	12	
病史	45	方案的入组条件名列不清
研究用药	2	
不良事件	12	
血常规	200	实验室测试单位换算错误
心电图	3	
合并用药	55	不良事件的合并用药对比
...(根据实际疑问类别增减行列数)
合计	(列出总的疑问数, 如 354)	
平均 (/受试者)	(列出按受试者计算平均疑问数, 如 4)	
平均 (/研究机构)	(列出按参与的研究机构数量计算平均疑问数 40)	

5.2 疑问的处理情况

疑问产生天数		疑问回复	
平均天数 (天)	范围 (天)	平均天数 (天)	范围
(列出各类疑问发出后等待答复所需的平均天数)	(列出最短天数和最长天数范围)	(顺序列出各类疑问发出后收到答复所需的平均天数)	(列出最短天数、最长天数范围)

5.3 疑问管理中的主要问题*

(*: 根据实际情况列出数据疑问处理中各类主要问题的类别, 所涉研究机构名称, 实际问题发生的天数, 和问题发生的实际原因分析。案例见表中。)

版本号: [下注脚列出版本号]

第 N 页

版本日期: YYYY-MM-DD

疑问重点问题	研究单位名称	时间/数量	原因及说明
疑问回复时间最长研究单位及原因分析	PI001	120 天	研究者之一离职
疑问产生数量最高研究单位及原因分析	PI001	178 (占总疑问50%)	接任研究者培训不及时
...(根据实际案例问题增减行列数)...

6. 医学编码*

(*: 按顺序列出完成编码的数据集名称, 所用的相应编码字典名称和版本, 手工编码和程序编码的条数, 以及各类数据集的编码总条数。案例见表中。)

编码数据 (数据集名称)	编码字典	编码字典版本	编码数量		
			手工编码	程序编码	合计
不良事件 AE	MedDRA	17.0	--	12	12
合并用药 CM	WHODrug	WHODrug 15Jan2014	4	8	12
医学诊断 SG=Surgery or FA=Findings about Event	MedDRA	17.0	--	6	6
既往病史 HX	MedDRA	17.0	2	20	22
...(根据实际编码数据集条目 增减行列数)...

7. SAE 一致性核查

本试验过程中共观察到严重不良事件 (SAE) N 例, 其中与试验药物相关的 SAE N 例; 按 N 次/月的频率, 共进行了 N 次一致性核查, 其中有 N 例经核查不一致的 SAE 详见附表 (SAE 发生率列表附后)。

(注: N 代表实际发生的次数)

8. 外源数据管理*

(*: 按实际外源数据来源列出各类外源数据的名称，相应外源数据提供的单位名称，相应外源数据传输协议名称和版本号，传输频率和相应传输方式等。如果要求盲态，请提供盲态管理规程；如果进行一致性核查，请提供一致性核查结果报告，如果未进行一致性核查，请给出原因。案例见表中。)

外源数据类型	数据源单位	数据传输协议	传输频率	传输日期 (盲态要求		一致性核查		传输方式
				首次	末次	是	否	是	否	
ECG1	国际心电图中心	2013.0630V1.0	每月二次	YYYY-MM-DD	YYYY-MM-DD		√	√		在线传输
化验室 Labs-1	友好医院	2013.0630V1.0	每月一次	YYYY-MM-DD	YYYY-MM-DD	√		√		在线传输
化验室 Labs-2	有爱医院	20130715V1.0	每月一次	YYYY-MM-DD	YYYY-MM-DD	√		√		在线传输
化验室 Labs-3	爱家医院	20130801V1.0	每月一次	YYYY-MM-DD	YYYY-MM-DD	√		√		在线传输
药代 PK	全球药动实验室	20130810V1.1	试验结束	YYYY-MM-DD)	YYYY-MM-DD	√			未导入临床试验数据库	CD
...(根据实际外源数据类型增减行列数)

版本号：[下注脚列出版本号]

第 N 页

版本日期：YYYY-MM-DD

9. 数据质控和稽查

9.1. 数据质量评估*

(*：根据实际质控步骤，列出数据流程中完成的数据项目评价名称，计划完成的时间或数量，实际完成的时间或数量。如果实际与计划有偏差，提供原因或说明；没有的话可以不填。如果质控评估过的条目没有预期计划，可以留空。案例见表中。)

核查项目	计划	实际发生 (平均)	原因及说明
数据录入时间	5-10 天	12 天	PI 研究员离职
数据管理员数据审核时间	5-7 天	5 天	
方案不依从总数 (件)		20	入组条件
严重方案不依从数 (件)		2	心电图数据导入不及时
... (根据实际评估项目增加行列数)			

9.2. 人工数据核查评估*

(*：根据实际发生的人工数据核查情况，顺序列出各次 QA 检查数，总的的数据点待核查数，实际抽样的数据点核查数，抽样数与总列出的比例，抽样比例决策的依据，总错误率，关键指标和非关键指标的数据错误率。案例见表中。)

核查次数	总例数	抽样例数**	抽样比例	抽样比例依据	错误率		
					总率	关键数据	非关键数据
第一次	80	9	$\sqrt{80}$	总病例数小于 100 例，则抽取例数为总病例数的平方根	0.5%	0.2%	0.3%
第二次	200	20	10%	总病例数大于 100，将随机抽取 10%的病例	0.1%	0.05%	0.05%
...(根据实际核查次数增加行列数)							

**：每次核查不能重复抽样。

9.3. 数据管理过程稽查*

(*: 按照稽查发生的时间顺序, 列出负责稽查的单位名称和时间, 相应各次稽查发现的问题及其相应各次发现问题的纠偏和防偏措施。案例见表中。)

稽查负责单位	稽查时间	发现问题**	纠正和预防措施
医药公司AB	2013-11-07	试验方案修正了入组条件但是 CRF 和数据库没有及时更改	立即进行了CRF 及数据库修改; 相关人员进行了SOP 再培训。
...(根据实际稽查事件增减行列数)

**注: 后附稽查报告

10. 临床试验数据集提交*

(*：根据实际数据集的提交情况，列出数据集名称，相应提交日期，提交的单位名称，提交的数据集格式，和提交的对象及其提交原因；对于提交后的原数据集和备份数据集，列出存储地点，相应数据集负责人及其对数据集的权限。案例见表中。)

数据集	提交日期	提交单位	提交格式	提交对象及原因	原数据集			
					储存地点	备份地点	责任人**	责任人权限
PHAMA-CNS1234.dataset.xpt	2014-04-06	合同公司 CD	SAS-SDTM	申办者，锁库后的统计分析	https://datarepository.xx.xxx.com/	https://datarepository.yy.xxx.com/	吴柒	临床信息技术经理
...(根据实际提交的数据集情况增减行列数)								

**注：除责任人外，其他人对数据集的所有权限均已被解除

11. 数据管理实际过程与数据管理计划不一致*

请描述数据管理过程是否存在与数据管理计划的不一致，有任何不一致，需详细描述实际过程，发生原因。同时需要阐述此过程对数据质量的影响。

(*：根据实际情况予以描述。如果没有出现不一致的情况，则可以写成“不适用”等。)

版本号：[下注脚列出版本号]

第 N 页

版本日期：YYYY-MM-DD