

临床试验数据管理的流程优化策略探究

赵振凯

百济神州(北京)生物科技有限公司 北京 100022

摘要: 在医药研发中, 临床试验至关重要, 其数据管理流程的优劣直接影响试验结果。但传统流程依赖人工和简单技术, 在数据采集、录入、审核、存储与归档环节问题频出, 如采集方式落后、审核标准不统一等, 严重制约医药研发进程。本文针对这些问题, 提出基于信息化技术、自动化手段和标准化体系的优化策略。包括优化 EDC 系统、应用 OCR 技术和智能清洗算法、构建标准化审核体系、改进存储与归档管理等, 并阐述了实施保障措施。有助于提升数据管理的效率、质量与合规性, 为临床试验的高效开展提供有力支持。

关键词: 临床试验; 数据管理; 流程优化; 信息化技术

1. 引言

在医药研发领域, 临床试验是确保新药或新治疗方法安全性与有效性的关键环节。随着医学研究的深入和医药产业的发展, 临床试验的规模不断扩大, 复杂性日益增加。数据管理作为临床试验的核心支撑, 其流程的科学性和高效性直接决定了试验的成败。准确、完整且及时的数据是得出可靠试验结论的基础, 能够为药物审批、临床应用提供坚实依据。然而, 传统的临床试验数据管理流程多依赖人工操作和相对简单的信息技术手段, 在面对现代临床试验的复杂需求时, 逐渐暴露出诸多弊端, 严重影响了数据管理的质量和效率, 制约了医药研发的进程。

本文旨在深入剖析现有临床试验数据管理流程中各个环节存在的具体问题, 从实际操作和行业发展需求出发, 提出具有针对性、可操作性的优化策略。通过这些策略的实施, 实现提升数据管理效率, 降低人为错误, 提高数据质量, 确保数据的合规性存储与便捷检索。

2. 临床试验数据管理流程现状分析

2.1 数据采集阶段

2.1.1 数据采集方式

目前常见的数据采集方式主要有纸质病例报告表(CRF)和电子数据采集(EDC)系统。纸质 CRF 在一些小型临床试验或资源有限的地区仍被使用, 其操作相对简单, 研究人员可以直接手写记录数据。然而, 这种方式存在明显缺陷, 数据容易因书写潦草导致难以辨认, 在后续录入过程中极易出错, 而且人工录入速度慢, 耗费大量人力和时间, 严重影

响数据采集的及时性。EDC 系统则利用电子设备和网络技术实现数据的实时采集和传输, 大大提高了数据采集效率, 能实时监控数据质量^[1]。

2.1.2 数据来源多样性

临床试验的数据来源广泛, 除了医疗机构的患者诊疗信息外, 还包括各种临床试验专用设备产生的数据, 如血糖仪、血压计等监测设备记录的患者生理指标数据。此外, 患者自我报告的数据, 如症状感受、生活质量评价等也十分重要。然而, 这些不同来源的数据在格式、精度、时间戳等方面存在差异, 整合过程复杂。

2.2 数据录入与清洗阶段

2.2.1 人工录入问题

人工录入是将采集到的数据输入到电子系统中的常见方式, 但这一过程容易出现各种错误。录入人员可能因疲劳、疏忽导致数据录入错误, 如数字颠倒、小数点错位等。而且不同录入人员对数据的理解和判断存在差异, 对于一些模糊或不完整的数据, 处理方式不一致, 影响数据的准确性和一致性。

2.2.2 清洗规则与方法

现有的数据清洗规则通常基于简单的逻辑判断, 如检查必填字段是否为空、数值是否在合理范围内等, 清洗方法也多为人工肉眼检查和简单的程序筛选。这种方式对于一些明显的错误能够有效识别和纠正, 但对于复杂的数据错误, 如数据之间的逻辑矛盾、异常数据点的判断等, 往往无能为力。

2.3 数据审核阶段

2.3.1 审核流程与标准

目前的数据审核流程一般是在数据录入完成后, 由专业的数据审核人员进行人工审核。审核标准主要从数据完整性、准确性、一致性三个方面进行考量。

2.3.2 人工审核局限性

人工审核主要依赖审核人员的专业知识和经验, 效率较低^[2]。面对大量的数据, 审核人员难以在短时间内全面、细致地检查所有数据。而且人工审核容易受到主观因素影响, 对于一些复杂的数据关联错误, 如多个变量之间的相互影响关系出现异常, 人工审核很难及时发现。

2.4 数据存储与归档阶段

2.4.1 存储方式与安全性

数据存储主要采用数据库存储和文件存储两种方式, 数据库存储适用于结构化数据, 能够方便地进行数据查询和管理, 但在面对大规模数据和高并发访问时, 可能出现性能瓶颈^[3]。

2.4.2 归档流程与检索便利性

数据归档是将完成临床试验的数据进行长期保存的过程。现有的归档流程通常是按照试验项目或时间顺序进行整理存储, 但在归档数据的检索方面存在不足。缺乏完善的数据索引和元数据管理, 导致在需要查询历史数据时, 检索过程繁琐, 耗费大量时间和精力。

3. 流程优化策略

3.1 基于信息化技术的采集流程优化

3.1.1 EDC 系统优化

为解决 EDC 系统存在的问题, 需进一步完善其功能。增加数据自动验证功能, 在数据录入时, 系统根据预设的规则实时检查数据的准确性和完整性。

3.1.2 数据采集接口整合

建立统一的数据采集接口是实现不同数据源与 EDC 系统无缝对接的关键。制定标准化的数据接口规范, 要求各数据源按照规范进行数据输出^[4]。

3.2 自动化数据录入与智能清洗

3.2.1 光学字符识别 (OCR) 技术应用

引入 OCR 技术可以有效解决纸质数据录入效率低和错误率高的问题。将纸质 CRF 通过扫描设备转化为电子图像, 利用 OCR 技术对图像中的文字进行识别和转换, 自动录入

到 EDC 系统中。在 OCR 识别过程中, 结合机器学习算法对识别结果进行优化, 提高识别准确率。对于一些模糊不清的文字或特殊符号, 通过人工辅助标注的方式进行训练, 使 OCR 模型能够更好地识别这些字符^[5]。

3.2.2 智能清洗算法应用

运用数据挖掘和机器学习算法建立智能清洗模型。首先, 收集大量的历史临床试验数据, 对其中的错误数据进行标注和分类。然后, 使用这些数据训练机器学习模型, 如决策树、神经网络等。模型训练完成后, 能够自动识别和纠正新数据中的错误。智能清洗模型可以发现数据之间的潜在逻辑关系, 对不符合逻辑的数据进行检查和修正, 提高数据质量。

3.3 构建标准化审核体系

3.3.1 制定统一审核标准

联合行业协会、监管机构和专业的临床试验机构, 从国内外相关法规、行业最佳实践出发, 制定涵盖各类临床试验数据的统一审核标准。明确数据完整性、准确性、一致性的具体审核细则, 对于必填字段、数据格式、取值范围等制定详细的规范。

3.3.2 引入自动化审核工具

结合人工审核, 引入自动化审核工具可以大大提高审核效率和准确性。使用数据比对软件, 对不同来源或不同时间点的数据进行比对, 快速发现数据的差异和变化。运用逻辑校验工具, 根据预设的逻辑规则对数据进行自动检查, 如检查患者的用药剂量与体重、年龄之间的关系是否合理。自动化审核工具能够快速处理大量数据, 将发现的问题标记出来, 供人工审核人员进一步核实和处理, 减少人工审核的工作量, 提高审核的准确性和效率。

3.4 优化存储与归档管理

3.4.1 数据存储架构优化

采用分布式存储架构可以提高数据存储的安全性和读写性能。将数据分散存储在多个存储节点上, 避免因单个节点故障导致数据丢失。对于结构化数据, 采用分布式数据库 (如 TiDB、CockroachDB 等) 进行存储, 充分发挥分布式数据库的高可用性和扩展性优势。在数据备份方面, 采用异地多活备份策略, 将数据备份到不同地理位置的存储中心, 确保在发生自然灾害、硬件故障等意外情况时数据的安全性和可用性。

3.4.2 建立数据索引与元数据管理

为提升归档数据的检索效率,建立完善的数据索引和元数据管理体系。根据数据的关键属性,如试验项目名称、患者编号、时间等,建立多维度的数据索引。通过元数据管理,用户可以更准确地理解数据内容,提高检索的准确性。在检索过程中,利用全文检索技术数据进行快速搜索,用户可以通过关键词、时间范围、数据类型等条件进行灵活检索,快速获取所需的数据^[6]。

4. 优化策略的实施与保障措施

4.1 人员培训与组织架构调整

4.1.1 人员培训计划

针对不同岗位人员制定详细的培训计划。对于数据录入人员,开展 OCR 技术应用和数据录入规范培训,使其熟练掌握新的数据录入方式和操作流程,提高录入准确性和效率。培训内容包括 OCR 软件的使用方法、常见识别错误的处理、数据录入的质量控制要点等。

4.1.2 组织架构调整建议

分析现有组织架构在数据管理流程中的不足,如职责分工不明确、沟通协作不畅等问题。建议设立专门的数据管理小组,由数据管理经理负责统筹协调。

4.2 质量管理与风险控制

4.2.1 建立质量控制体系

从数据质量指标设定、质量监控流程制定方面建立全面的质量控制体系。设定明确的数据质量指标,如数据错误率、数据完整性率、数据一致性率等,并定期对这些指标进行统计和分析。

4.2.2 风险识别与应对策略

识别流程优化过程中的技术风险、人员风险、法规风险等。技术风险包括新系统和工具的稳定性、兼容性问题,如 EDC 系统与其他系统对接时出现的数据传输错误。通过加强人员培训、建立人才储备机制等方式进行应对。法规风险主要是指流程优化是否符合国内外相关法规要求,如《药物临床试验质量管理规范》(GCP)等。建立法规跟踪和解读机制,确保流程优化始终符合法规要求。

4.3 法规与合规性保障

4.3.1 法规政策解读与遵循

深入解读国内外临床试验数据管理相关法规政策,如

欧盟的《通用数据保护条例》(GDPR)、美国食品药品监督管理局(FDA)的相关规定以及我国的 GCP 等。

4.3.2 内部合规性审查机制

建立内部合规性审查机制,定期对数据管理流程进行审查和改进。成立合规性审查小组,由法律专家、质量管理专家等组成,对数据管理流程的各个环节进行检查,评估是否符合法规要求和内部制定的标准。

5. 结论与展望

本文通过对临床试验数据管理流程现状的深入剖析,提出了一系列切实可行的优化策略,涵盖数据采集、录入、审核、存储与归档全流程。通过真实案例验证,这些策略有效提升了数据管理的效率与质量,显著缩短数据处理时间,降低错误率,增强数据的安全性与可检索性,同时实现成本控制,为临床试验的顺利开展提供了有力保障。

未来研究可聚焦于拓展优化策略的适用范围,针对不同规模和类型的临床试验机构,制定个性化、可操作性强的实施指南。

参考文献:

- [1] 付海军. 临床试验数据管理与统计分析 [C]// 中国药理学学会药物临床试验专业委员会. 中国药理学学会药物临床试验专业委员会第五次学术年会汇编. 上海: 韧致医学, 2023: 217–234. DOI: 10.26914/c.cnkihy.2023.122917.
- [2] 曹丽亚, 谢林利, 谢江川, 等. 2020 版《药物临床试验质量管理规范》实施后药物临床试验数据现场核查的要点与浅析 [J]. 中国新药杂志, 2023, 32(03): 264–269.
- [3] 张明辉, 李响, 孔德娟, 等. 基于医院大数据平台的临床试验智能管理系统的构建与应用 [J]. 中国食品药品监管, 2022, (06): 26–33.
- [4] 陈采婷, 陈小嫦. 临床试验数据管理研究 [J]. 中国卫生事业管理, 2020, 37(08): 599–602.
- [5] China-Based Contract Research Organization Selects Oracle Cloud to Optimize Clinical Trial Data Management [J]. M2 Presswire, 2016.
- [6] Time and Cost Reduction in Clinical Trials – Outsourcing and IT to Optimize Patient Recruitment, Logistics, Data and Site Management [J]. M2 Presswire, 2012.