

促进我国临床试验数据管理规范化



吴泰相¹, 卞兆祥², 李幼平¹, 商洪才³

1. 四川大学华西医院中国循证医学中心(成都 610041)
2. 香港浸会大学中医药学院(中国香港 999077)
3. 北京中医药大学东直门医院(北京 100029)

【摘要】 数据管理制度是影响临床试验质量的主要因素之一, 数据管理制度设计包括指导委员会和数据管理委员会、数据采集、数据库、数据安全和监察、数据库锁定五个基本部分。本文介绍各部分的设计要点和方法, 旨在帮助广大研究者了解临床试验数据管理制度, 吁请广大研究者重视数据管理和了解管理制度的设计, 帮助提高临床试验数据的管理水平。

【关键词】 临床试验数据; 数据管理; 规范化

Promoting standardization of clinical trial data management in China

WU Taixiang¹, BIAN Zhaoxiang², LI Youping¹, SHANG Hongcai³

1. Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu, 610041, P.R.China
2. Hong Kong Chinese Medicine Clinical Study Centre, School of Chinese Medicine, Hong Kong Baptist University, Hongkong, 999077, P.R.China
3. Key Laboratory of Chinese Internal Medicine of Ministry of Education, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, 100700, P.R.China

Corresponding author: WU Taixiang, Email: chict@vip.qq.com; LI Youping, Email: yzmylab@hotmail.com

【Abstract】 Data management system is one of the major elements of quality of clinical trial. Development of data management system include a Steering Group and Data Safety and Monitoring Board, data collection, database, performance of the data safety and monitoring, as well as locking of database. This article provides key issues of five parts aims to help researchers to understand clinical trial data management system.

【Key words】 Clinical trial data; Data management; Standardization

临床试验的所有结果必须并只能通过数据反映出来, 数据既是临床试验结果的证据, 也是临床试验曾经存在过的唯一证据。数据质量既能反映临床试验的设计水平, 也能反映临床试验的管理水平。例如, 2015 年国家食品药品监督管理局发现 1 622 个新药临床试验的申请注册材料中约 80% 涉嫌数据造假^[1-4]。新药临床试验都是在国家食品药品监督管理局认证的临床试验基地实施, 其管理均极为严格。我们以前的研究也证实了他们在撰写研究报告、制订研究计划书等方面都反映出在临床试验方法学和管理方面的训练有素, 他们实施的临床试验质量平均水平高于其他研究者的自拟临床

试验^[5]。虽然我们无从得知最后核实多少个试验是确有实据的数据造假, 但从对这些申请注册材料的核查方式我们可以推测: 至少有相当部分的试验是由于数据管理制度的缺陷导致对数据真实性的置疑。

我们对注册临床试验研究计划书和注册信息的调查发现: 我国大多数研究者在制订临床试验研究计划书时, 对制订数据管理制度重视不够, 这可能是影响我国临床试验质量的主要原因之一。本文旨在帮助广大研究者了解临床试验数据管理制度发展概况, 吁请广大研究者重视数据管理和了解管理制度的设计, 提高临床试验数据的管理水平。

数据管理制度设计包括五个基本部分内容: 指导委员会和数据安全与监察委员会、数据采集、数据库、数据安全和监察、数据库锁定。

DOI: 10.7507/1672-2531.201804096

通信作者: 吴泰相, Email: chict@vip.qq.com; 李幼平, Email: yzmylab@hotmail.com

1 指导委员会 (steering group) 和数据安全与监察委员会 (data safety and monitoring board, DSMB)

这是两个分工不同的委员会, 可根据研究规模或其它因素需要二者分开工作或联合工作。通常, 大型多中心研究是将二者分开各行其责, 较小的研究是两个委员会共司其责。指导委员会是试验的决策机构, 负责临床试验的设计和成果总结, 包括制定和批准治疗方案和研究方案、批准各种标准操作规程, 决定和批准试验结果的传播和撰写研究报告等。DSMB 主要负责设计 CRF 和数据库、评估征集参试者过程、评估效果、评估结局安全性、评估数据采集质量及录入数据等。

以我国两个高质量多中心临床试验为例: ①黑龙江中医药大学附属第一医院妇产科吴效科教授团队的“针刺治疗多囊卵巢综合征不孕症多中心随机对照试验”^[6]; ②山东大学生殖卫生中心陈子江教授团队的“新鲜胚胎与冷冻胚胎治疗多囊卵巢综合征不孕症”^[7]。两个试验均成立了由临床专家、统计学家、方法学家组成的指导委员会, 负责制订研究计划书。仅制订研究方案就经过长达一

年多的讨论和修改完善, 广为征集同行和方法学家的意见。DSMB 由专职数据录入人员、监察人员、统计学家组成, 负责评估研究计划的执行情况、监察数据的采集和质量、数据的安全性、储存和数据库维护管理及录入数据。

因此, 一个临床试验首要的工作就是组建指导委员会和 DSMB, 也可根据试验规模将二者合为一个委员会。不同的试验对两个委员会的职责要求可能各不相同, 我们推荐指导委员会和 DSMB 的组成和职责见表 1。

特别指出: 我们常常见到一些随机对照试验作者只有寥寥数人, 甚至仅一人。可以肯定这样的团队结构难以完成一个高质量的随机对照试验, 因为从方案设计到试验实施, 绝非一、两人之力能够胜任。有些试验虽然规模不大, 所需研究人员不多, 但仍需具备相应的组织功能, 在充分遵循规则的基础上, 如设盲、隐蔽分组和数据库管理员要求必须由专人实施外, 可采取身兼两职或数职的办法, 如指导委员会专家可兼监察员、统计专家等。

2 数据采集

数据采集根据临床试验数据来源和获得途径

表 1 指导委员会和 DSMB 的结构和职责*

委员会名称	人员结构	职责	说明
指导委员会	试验负责人、临床学家、方法学家、统计学家; 应至少有一位副高以上职称人员	制订标准治疗方案	确定治疗措施的强度(剂量)、用法、频度等
		设计试验方案	主要集中于方法学设计
		组织制订和批准各种标准操作规程	各种标准操作规程可由研究团队中相关专业人员制订
		组织制订和批准研究计划	研究计划可由 DSMB 制订, 如数据采集计划
		批准和决定是否继续进行试验	DSMB 可推荐是否继续试验
DSMB	研究人员包括临床学家(临床医生)、统计学家, 不能参与纳入参试者	批准试验结果的传播和撰写结果报告	传播与结果报告须经 DSMB 核准
		设计 CRF	按照研究计划书制定的测量指标和流程图的各时间点, 设计出各时间点要进行的各测量项目
		评估募集参试者过程是否充分	包括知情同意过程和募集方式、参试者是否符合纳入标准及同质性
		评估数据质量	包括数据收集方式、数据完整性和及时性
		结果安全性评估	基于不良事件记录
		结果及效果评估	可作风险-获益分析, 及考虑在肯定效果基础上提前中止试验
		推荐是否继续试验	向指导委员会提出推荐
		评审(review)结果传播和报告稿	基于数据监察和管理情况向指导委员会写出评审意见
		核准试验结果传播和报告稿	指导委员会的批准须基于 DSMB 的核准
		数据库管理员	管理数据库
监察员	监察数据	监察数据的完整性、及时性、正确性	
数据录入员	录入数据	每个分中心可设置至少一名数据录入人员	

*: 本表内容部分参考 David 的文章^[8]; DSMB: 数据安全性与监察委员会。

包括检测技术、方式方法分为两类：①数据通过采集设备如影像、检测仪和探测器如体温计等的测量值获得，采集设备输出的记录称为元数据 (meta data)；②数据来自研究者对从参试者处采集到的临床证据进行分析判断后记录于病例报告表 (CRF)，如查体记录、调查问卷，或试验的管理记录；而元数据必须转记于病例报告表。临床试验数据包括两大部分：一为流程管理信息，二为结果测量指标和观察项目的检测结果。

① 流程管理信息：流程管理指在临床试验进行过程中，试验流程的安排、参试者募集/筛选/纳入和管理、试验物资如药品器材的分配和消耗、生物样本的采集和管理等。准确采集和记录流程管理信息需要仔细规划，制订相应的管理办法和制度，才能保证试验准确实施的可行性和试验结果数据的可靠性和真实性，清楚地反映出试验实施的过程，否则临床试验结果测量指标数据就是无源之水、无本之木。

② 结果测量指标和观察项目的检测数据：结果测量指标指反映治疗效果的指标，其与治疗效果密切相关。这些指标或为治疗的靶向指标、或其变化间接反映病情的变化；通过检测和比较治疗前后这些指标的变化或比较平行不同治疗方案之间这些指标的差异可评估治疗效果。结果测量指标的数据采集一般按照治疗进程和一定的访视时间节点进行。

观察项目指试验进行过程中监测参试者身体状况的一类项目，它们的变化并不直接与治疗有关或并不特异地反映治疗导致的疾病变化。观察项目的数据采集往往根据病情发展变化进行。

结果测量指标和观察项目在有的试验中可能互不相关，分别采集；但在有的试验中可能既是结果测量指标又是观察项目。

制订研究计划书时需要根据结果测量指标、观察项目、流程管理内容针对不同的数据类型选择合适的采集方法和采集时间、记录载体。

我们推荐的数据库设计和数据采集清单见表 2，供设计数据库和数据管理制度时参考。

3 数据库 (data base)

数据库的基本功能是记录和储存数据，数据库的数据分为外部数据和内部数据：外部数据包括元数据和纸质病例报告表等记录的数据；将元数据和纸质病例报告表中的数据输入专门的储存系统，以便于统计分析和集中管理；将外部数据记录

和保存在数据库里的数据为内部数据；数据管理包括对外部数据和内部数据的管理。随着医学研究理念的更新和计算机、互联网等技术的发展，对临床试验数据库的功能要求越来越多，数据管理的标准也随着数据库的发展而更新和发展。

数据库的发展至今大体上可以分为四个阶段：

① 纸质媒介阶段：早期临床试验数据的标准管理是外部数据 (元数据+纸质病例报告表)+内部数据 (纸质表格数据)，纸质表格这种账簿式的数据记录和保管方式成为最原始的数据库，其记录和统计分析的效率均十分低下。

② 计算机化电子数据采集管理系统 (electronic data capture, EDC) 阶段：1980 年代随着计算机技术的普及，出现许多计算机化数据库称为 EDC，如基于 DOS 操作平台的 dBASE 数据库管理系统，基于 ORACLE、SYBASE、IBM、Microsoft 等公司的各种数据库系统开发的各种医学数据库，世界卫生组织专门开发用于流行病学研究的免费数据库 EpiData 等。

2003 年以来，美国 FDA 发布了临床试验的企业电子数据库系列指导原则和条例^[9-11]，临床试验数据管理学会的《临床试验数据管理规范》^[12, 13]，CDISC (clinical data interchange standards consortium) 的《研究数据列表模式》、《分析数据模式》、《XML Schema》、《受控术语集》、《非临床数据交换标准》等系列技术标准，涉及临床试验电子数据的标准、标准术语、交换、向监管机构递交数据的内容规范，及对修改数据的追溯等质控功能和操作标准。临床试验数据管理的计算机化使数据库大大突破了早期表格管理的空间局限性和功能局限性。

③ 互联网阶段：在临床试验数据管理计算机化基础上，互联网技术使中央随机化分组更方便、更隐蔽可靠；使数据监察管理能做到在线质量控制 (online quality control) 和实时质量控制 (real time quality control)，实现临床试验数据管理和流程管理及时、准确和高效。随着临床试验透明化理念的发展，使正在形成的原始数据共享新理念有可能做到医学专业同行和社会公众对临床试验的及时共享和成果的高效共享，将极大地促进临床试验的社会参与度和社会公共伦理道德的发展。

④ 智能化阶段：临床试验数据库的智能化已经起步，现有大量医学检测设备已经实现数字化，解决 ID 辨识系统、检测设备与数据库之间的数据传输均已不存在技术困难。当智能化数据库走向

表 2 临床试验数据库设计和数据采集清单

数据分类	项目	说明
流程管理 数据	流程图	安排每个时间点要进行的项目, 在试验进行中核对每个项目的完成情况
	征募参试者, 筛查 纳入参试者	应记录每位未纳入者的筛查时间、排除原因或未能纳入的原因, 记录排除标准 每位参试者的纳入时间, 记录纳入标准; 纳入医师及签名
	参试者管理	一般信息, 包括姓名、性别、年龄、身高、体重、种族、教育程度、婚姻状况 住院参试者的管理制度 院外参试者的管理制度
	试验完成后未愈参试者的安排	按后续治疗计划治疗的参试者记录
	药品管理	药品种类和总数量, 发放记录, 消耗记录, 核查情况
	器材管理	药品种类和总数量, 发放记录, 消耗记录, 核查情况
	生物样本采集	采集的样本数量: 份数、毫升数或克数; 保存方法; 转运途径或/和方法; 储存时间; 销毁 方法和时间; 根据单中心或多中心、是否集中保存和检测制订相应的管理办法
	医学编码	对试验中各变量名称、类型定义并制定与之相应的编码和赋值规则, 并形成管理文件
	数据接收与录入	需根据单中心和多中心、是否采用基于网络的电子数据库设计不同的数据接收与录入方法 和途径
	相关文件	① 试验开始前所有研究人员的诚信声明; ② 试验完成后监察员对数据完整性和质量的核 查声明; ③ 试验完成后试验负责人对试验全程质量保证的声明; ④ 其他相关声明, 如财 物管理声明(如需要)
测量数据*	观察项目: 通常不是治疗的目标 项目或直接标志物, 用于观察参 试者身体基本状况	既往史: 生活史、生育史、疾病史、过敏史、治疗史(包括用药史和手术史), 及专科或疾病 特异性的既往史 查体; 生命体征; 血常规检查; 大、小便常规检查; 血生化检查, 包括肝功、肾功、血糖、 电解质等常规项目; 凝血功能检查; 及专科或疾病特异性的检查项目。其中, 有些为基础 项目于纳入时一次性检测, 有些项目为研究进行过程中监测身体状况
	结果测量指标: 用于评估治疗效 果和安全性	主要测量指标: ① 终点指标; ② 特异性指标; ③ 客观指标; ④ 治疗方法作用的靶向指 标, 或研究的目标指标 次要测量指标: ① 中间指标; ② 非特异性指标; ③ 主观指标; ④ 可能间接反映疾病变化 与治疗之间的关系的指标 安全性指标: 根据药物代谢途径或毒性的系统特异性作用选择合适的安全性指标, 如肝脏 代谢药物需监测肝功能, 肾脏代谢药物需监测肾功能, 有造血系统毒性药物需监测造血系 统指标、循环系统毒性药物监测循环系统指标等 不良事件管理制度: 不良事件登记表、严重不良事件登记表、上报和鉴定途径

*: 有些试验中, 有些指标既是观察项目, 又是结果测量指标或安全性指标。

成熟时, 相应的管理规则和标准必然会逐渐形成。

我国临床试验应用数据库的使用水平普遍偏低, 有的甚至非常落后: 仅 28.23% 描述了正确的数据管理制度, 包括使用纸质病例记录表和 EDC; 其中知道使用基于互联网 EDC 系统的试验仅占 18.6%; 有些研究者采用报表式数据库如 Microsoft Office 办公系统中的 Access 等, 多数研究者采用表格格式数据库如 Excel、或专业统计软件如 SPSS、SAS、Stata 等中的电子表格。表格或报表式数据库远远不能满足临床试验数据管理的要求。

中国临床试验注册中心致力于推动临床试验数据管理的标准化和帮助广大研究者规范化地进行临床试验数据管理的公益事业, 于 2017 年 12 月 1 日将经过十年研发并成功应用的临床试验公共管理平台(ResMan)向广大研究者免费开放。① ResMan 是基于互联网的临床试验数据采集管理系统, 采用模块化和初步的智能化管理, 将各种测量工具模块化, 如各种量表、各种常规检测项目、不

良事件登记表等近千个模块, 研究者只需按照 CRF 内容进行组合, 就可以自行设计成完整的数据库; ② ResMan 也可以根据临床试验流程图用于设计 CRF; ③ ResMan 还具有中央随机化功能, 帮助研究实现高质量的随机化、隐蔽分组; ④ ResMan 系统已成为我国最大的临床试验公共管理平台, 截止 2018 年 1 月, 已有近 3 000 个在 ChiCTR 注册的中国临床试验使用 ResMan 进行数据管理, 其中已有过试验报告在美国医学会杂志(JAMA)^[6]、新英格兰杂志(NEMJ)^[7]等国际顶级医学期刊发表。

4 数据安全和监察

数据安全指包括外部数据和内部数据在内的所有数据的完整性和准确性及储存的可靠性。数据监察是为保障数据安全的质量管理保障体系。

数据监察按试验进行流程分为开始前、进行中、结束三个部分:

① 临床试验启动前的数据监察一般在现场核

表 3 数据安全和监察推荐清单

数据分类	监察项目	内容	完成情况	情况说明	监察日期
外部数据	生物样本接收使用记录	样本和备份的储存安全性、转运的可靠性等	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	实验室检测记录	如血生化、常规、影像学报告等的完整性和归档管理	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	临床病历报告记录	完整性和归档、隐私信息保护的可靠性	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	药品进出和使用记录	实物和记录一致	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	器械、设备使用记录	耗材实物与使用记录一致	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
内部数据	随机化核查 (仅随机对照试验)	随机方法、入组时参试者的随机化	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	违背方案核查	核查药物使用情况、依从性	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	流程核查	参试者入组、数据采集是否按日程表进行	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	逻辑核查	根据事件关联性识别可能存在的数据错误	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	阈值范围核查	查找极端值是否超出生理范围	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	一致性核查	外部数据与内部数据是否一致	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	盲态核查 (仅对实施盲法的试验)	核查有无破盲和泄盲事件、盲底保管可靠性	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		
	数据更改核查	核查数据更改的轨迹和更改原因	<input type="checkbox"/> 完成 <input type="checkbox"/> 未完成		

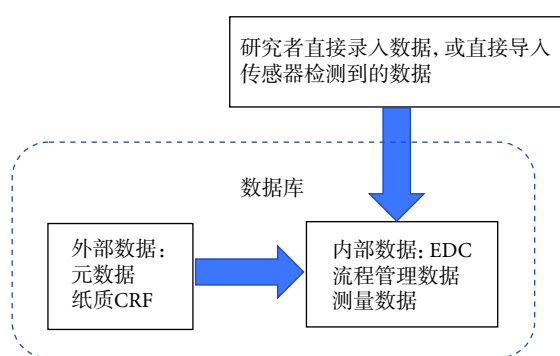


图 1 数据库结构框图

察, 所有数据采集准备工作完成后书面报告课题负责人、DSMB。

② 试验进行中的数据监察根据数据库不同采用不同的方法, 如可采用基于互联网的 EDC 系统; 实时在线监察 (real time monitoring); 在线质控 (online quality control), 即监察工作可随临床试验的进行而每天进行, 以保证纠错的及时性; 如采用计算机本机数据库或单位局域网数据库, 数据监察就只能定期在现场进行; 或通过电子邮件等远距离通讯工具传输数据进行数据监察工作。

要进行中期分析的试验, 一般是纳入参试者一半时进行; 中期分析前要先截止数据录入, 再进行数据监察并编制监察结果清单。

③ 试验数据收集完成后要进行最后的数据监察, 往往在现场对外部数据材料等进行审核验收。

数据监察中发现的问题数据应及时汇总分类, 分别向不同数据来源的责任人发出质询函, 经核对无误后由数据录入人员或数据库管理员进行更正。电子数据库须能追溯任何更改。

我们提出一个数据安全和监察建议清单 (表 3),

供研究者在研究计划书中制订相关计划、在数据库设计、试验进行中和中期分析、试验结束时实施数据监察的参考。

5 数据库锁定

临床试验中期分析和试验结束的数据监察完成后, 为了防止意外或未授权的更改, 均需要将数据库锁定, 即取消所有人员的编辑权限。锁定数据库前必须向 DSMB、课题负责人报告, 并附数据监察清单, 清单内容按照表 3 编制, 在获得书面签字批准后锁定数据库。

数据库锁定后, 如需要重新开锁继续录入数据或对错误数据进行修正, 均需获得 DSMB 的批准并详细记录。研究计划书要规定重新开锁的条件和操作流程, 例如: 中期分析完成后继续开始后半程试验; 数据库锁定后新发现遗漏数据或错误等。由 DSMB 批准授权后, 数据库管理员和监察员实施重新开锁, 修改完成后除数据库自动记录修改操作, 还将纸质授权批件和执行情况记录归档保存。

高质量的临床试验数据必须是可归因 (attributable)、易读 (legible)、及时 (contemporaneous)、原貌 (original)、准确 (accurate)、完整 (complete)、一致 (consistent)、持久 (enduring)、可及 (available when needed), 只有高度重视和严格作好临床试验数据管理, 才能获得高质量的临床试验证据。

参考文献

- 1 临床数据造假, 政策逼出的潜规则. Available at: <http://news.163.com/15/0821/07/B1HCKC0000014JHT.html>.
- 2 新药临床试验数据为何造假成风. Available at: <http://view.news.qq.com/original/intouchtoday/n3273.html>.

- 3 八成新药数据涉假曝医药生态恶劣. Available at: http://www.jjckb.cn/2016-09/09/c_135673951.htm.
- 4 80% of data in Chinese clinical trials have been fabricated. Available at: <http://www.sciencealert.com/80-of-the-data-in-chinese-clinical-trial-is-fabricated>.
- 5 Wu T, Li Y, Bian Z, *et al.* Randomized trials published in some Chinese journals: how many are randomized? *Trials*, 2009, 10: 46.
- 6 Wu XK, Stener-Victorin E, Kuang HY, *et al.* Effect of Acupuncture and Clomiphene in Chinese Women With Polycystic Ovary Syndrome: A Randomized Clinical Trial. *JAMA*, 2017, 317(24): 2502-2514.
- 7 Chen ZJ, Shi Y, Sun Y, *et al.* Fresh versus frozen embryos for infertility in the polycystic ovary syndrome. *N Engl J Med*, 2016, 375(6): 523-533.
- 8 DeMets DL, Ellenberg SS. Data monitoring committees - expect the unexpected. *N Engl J Med*, 2016, 375(14): 1365-1371.
- 9 U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), *et al.* Guidance for Industry: Part 11, Electronic Records; Electronic Signatures - Scope and Application. Pharmaceutical CGMPs, 2003. Available at: <https://www.fda.gov/RegulatoryInformation/Guidances/ucm125067.htm>
- 10 U.S. Department of Health and Human Services, Food and Drug Administration (FDA), Office of the Commissioner (OC). Guidance for Industry: Computerized Systems Used in Clinical Investigations, 2007. Available at: <https://www.fda.gov/ICECI/EnforcementActions/BioresearchMonitoring/ucm135196.htm>
- 11 U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), *et al.* Guidance for Industry: Electronic Source Data in Clinical Investigations. September 2013. Available at: <https://www.fda.gov/training/guidancewebinars/ucm382198.htm>
- 12 Society for Clinical Data Management (SCDM). Good Clinical Data Management Practices (GCDMP), 2007. Available at: <http://www.scdm.org/publications/gcdmp/>
- 13 CDISC. Introducing the CDISC Standards: New Efficiencies for Medical Research, 2009. Available at: <https://www.cdisc.org/>.

收稿日期: 2018-04-18 修回日期: 2018-04-28
本文编辑: 熊鹰