

临床研究的数据管理与质量控制

潘岳松

首都医科大学附属北京天坛医院 国家神经系统疾病临床医学研究中心, 北京 100050

电话: 010-67096699, E-mail: panys2000@163.com

【摘要】 高质量临床研究项目的高效开展离不开规范的数据管理与质量控制。临床研究的数据管理与质量控制应始于研究设计, 并贯穿研究的整个过程。数据管理的过程包括数据管理计划、临床病例报告表设计、数据字典与编码说明书、数据库的构建、数据收集与监查、数据录入、数据核查与清理、生成关键新变量、医学编码、去隐私化处理、数据审核与锁定等过程。通过严格的数据管理与质量控制, 可提高临床研究数据的完整性和准确度, 从而提高整体临床研究质量。

【关键词】 临床研究; 数据管理; 数据库; 质量控制

【中图分类号】 R-1 **【文献标志码】** A **【文章编号】** 1674-9081(2018)05-0458-05

DOI: 10.3969/j.issn.1674-9081.2018.05.016

Data Management and Quality Control for Clinical Research

PAN Yue-song

Beijing Tiantan Hospital, Capital Medical University, National Clinical Research Center for Neurological Diseases,
Beijing 100050, China

Tel: 010-67096699, E-mail: panys2000@163.com

【Abstract】 A clinical research project of high quality cannot be carried out without standard data management and quality control. Data management and quality control of clinical research should be started from the design of the study and performed throughout the whole process of study. The process of data management includes planning, the design of case reporting form, data dictionary and coding specification, construction of the database, data collection and monitoring, data input, data auditing and cleaning, generation of new key variables, medical coding, deidentification, data review and data lock. Strict data management and quality control can improve the integrity and accuracy of clinical research, and further improve the quality of the overall clinical research project.

【Key words】 clinical research; data management; database; quality control

Med J PUMCH, 2018,9(5):458-462

近年来, 随着科研人员对临床研究重视程度的增加, 国内临床研究项目迅速增多, 而临床研究过程中会产生大量医学数据, 同时也伴随出现了研究数据质量低下与孤岛化、研究低效重复等问题^[1]。科研人员逐渐认识到, 规范、完整、准确地收集数据是高质量临床研究项目高效开展必不可少的条件, 而此项工作离不开规范的数据管理及质量控制。

临床研究的数据管理, 是指在临床研究过程中按照临床研究方案规定的流程规范, 完整、准确地采集并录入数据, 按标准化数据库管理数据, 并对数据进行逻辑核查、疑问质询、盲态审核、数据库锁定、数据传输与归档的过程^[2-3]。其目的是为临床研究项目提供可靠的临床研究数据, 方便后期数据分析、共享与利用。临床研究的数据管理整体上分为3个阶

段,包括研究设计阶段、研究实施阶段以及研究数据整理阶段。在此过程中分别采用标准化的数据管理技术并进行严格的质量控制,是临床研究数据管理的关键(图1)。临床试验项目,作为一种特殊类型的临床研究,对数据管理与质量控制的要求比普通临床研究项目更为严格。针对临床试验项目,2016年国家食品药品监督管理总局颁布了《临床试验数据管理工作技术指南(2016年第112号)》^[3]和《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(2016年第113号)》^[4],加强了对临床试验数据的管理与质量控制。

1 研究设计阶段的数据管理与质量控制

1.1 数据管理设计与实施计划

一项规范的临床研究,应在设计研究方案的同时,设计并制定好数据管理制度、流程与实施计划^[5]。数据管理计划是临床研究数据管理工作的纲领性文件,是具体数据管理工作的依据,对数据管理流程各阶段和各环节的工作内容、操作方法及时间计划进行总体设计与规划^[6-7]。然后针对各重要环节,进一步制定相应详细的标准操作规程(standard operating procedures, SOP),作为具体实施的依据^[8]。

1.2 临床病例报告表的设计

1.2.1 基本设计原则

临床病例报告表(case report form, CRF)是临床

研究数据采集最主要的工具,CRF设计的好坏直接影响研究数据采集质量的高低。好的临床研究数据管理始于研究设计阶段,即在设计研究方案和CRF时即体现数据的采集与管理过程。如果CRF的条目与选项设置不合理,则无论后面的数据管理过程多么严格,均得不到正确的数据。CRF的合理设计,与数据采集、录入、核查以及后期的统计分析均息息相关^[9]。从数据管理角度审核CRF的设计,使其更趋于合理,将大大降低后期数据管理与统计分析的难度,提高研究效率。

CRF的条目与版面设计需符合临床工作习惯,以使数据的采集更加清晰和顺畅,不容易出现纰漏和混乱。临床研究最好的数据采集流程应根据现有的临床工作或患者就诊时间顺序及习惯来确定,在患者就诊不同时段分别收集相应的信息。比如,在门急诊时收集患者基本信息、既往史、急救与转运相关信息;在收入院后收集入院当天的住院基本信息;在住院期间或出院前收集辅助检查、用药情况、并发症信息、最终诊断、出院时状态及出院带药等信息;在随访时收集临床结局、预后及随访期间用药等信息。

1.2.2 注意事项

(1) CRF问题的设置应易于理解、不存在歧义,使临床研究的不同参与者(研究者、研究协调员、监查员、数据管理员及统计人员)对同一个问题的理解一致,才能保证数据收集准确可靠。为方便填写过程中正确理解,应编写完整的填写说明,必要时可配以

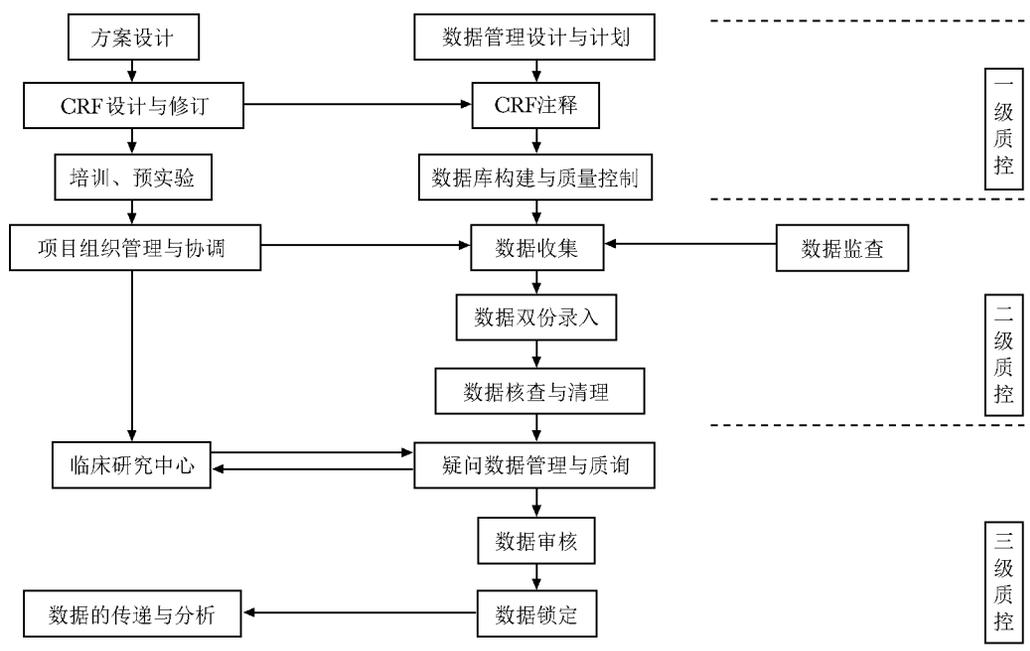


图1 临床研究数据管理流程
CRF: 临床病例报告表

影像等图片说明,放在问题之后,或印于CRF的背面。

(2) 问题条目的设置做到全面且简洁,即应包括后期统计分析时可能用到的全部变量,但多余的变量应全部删除。

(3) 一次只问一个问题,不使用容易引起思维混乱的复合问句,如“患者是否吸烟饮酒?”,而拆分为“患者是否吸烟?”和“患者是否饮酒?”两个问题。

(4) 数据的格式主要有数值型、日期/时间型、文本型与自定义型。在设计CRF时尽可能采用数值型,将问题转化成选择题或直接填入数据(如化验指标值)的形式,而尽量少用文本型格式。这样将方便后期进行数据管理与统计分析。

(5) 选择题的答案选项应具详尽性(exhaustive)和互斥性(exclusive),即包含所有可能的答案,如果还有其他可能的答案,应设置“其他”、“不详”及“不适用”等选项,以适合所有患者勾选;同时,选项应相互之间不重叠,对于每个患者,仅有一个选项适合其勾选。

1.3 数据字典与编码说明书

变量及其属性作为数据管理的最基本单位,其集合即为数据字典。在临床研究开始前,应参考国际标准,编写标准化的通用数据单元,建立数据字典。多个研究项目应按照数据字典建立编码说明书,统一CRF的变量规则,对每一变量的命名规则、数据类型、变量标签和编码规则进行统一、明确的规定(表1)。基于统一的标准化通用数据单元建立数据字典,可极大方便数据库的统一管理并实现多个项目间的数据对接,使多个临床研究项目的数据得以分享,就像来自不同文化、不同地区的人们可以使用普通话进行交流一样,这是多中心数据共享、打破“信息孤岛”的基础。

结合使用习惯,变量名可采用“模块名+变量英文缩写”的方式进行命名。变量的命名尽量做到自明性,即通过变量名可方便地理解此变量所代表的意义。如,变量“H_DM”代表患者既往是否患有糖尿

病。在完成编码说明书后,应根据编码说明书进行CRF的注释,即在空白CRF中标注各数据项的位置及其在对应数据库中的变量名和编码。注释CRF作为数据库与CRF之间的联系纽带,可帮助数据管理员、统计人员及其他数据使用方了解数据库。清晰地注释CRF,可极大方便后期数据管理与统计分析。

1.4 数据库的构建与质量控制

数据库的合理构建是数据管理最重要的环节之一。结构式数据库的基本形式是一行代表一例患者,一列代表一个变量。对于同一例患者多次测量的指标,可在变量名后加上一些后缀进行区分,如“SBP1”、“SBP2”和“SBP3”分别表示3次收缩压值。

目前,大部分临床研究通常采用电子数据采集与管理系统建立研究数据库,通过数据库实现对数据进行规范管理与质量控制^[10-11]。通过电子数据库,可实现和优化数据录入时的逻辑控制,达到数据管理与质量控制的要求。比如为了避免缺失值,可将一些重要的变量设置为必填项,如果不填写就无法提交数据;对于年龄、血压及生化指标等,可设置取值范围限制,如将年龄变量设置成“18~120岁之间”,不在此范围内者无法录入,并对年龄介于100~120岁者进行软错误提示,但经研究者判断后仍可录入;对于有关联的变量,如是否有高血压与是否使用降压药两个变量,其间设置逻辑关系验证。

数据库管理系统应具备可溯源性,对数据进行的任何更改或更正均应签字、注明日期并解释原因,并记录数据修改留痕。此外,电子数据库还应能实现双份录入、实时纠错、在线数据传送、设置多角色功能(实现不同角色具有不同的工作内容与任务)等。对于药物临床试验,国家食品药品监督管理局颁布了《临床试验的电子数据采集技术指导原则(2016年第114号)》^[12],对电子数据采集技术的基本要求,如电子CRF构建、系统设置、数据的安全性、权限控制、用户测试、数据录入、数据核查、数据质疑管理、数据锁库、数据导出等作了详细的要求与规定。

表1 编码说明书举例

变量名	数据类型	变量标签	编码规则
PTID	字符变量	患者编号	四位数据编码
DM_SEX	二分类变量	人口学特征:性别	1=男;2=女
DM_DOB	日期变量	人口学特征:出生日期	<yyyy/mm/dd>年月日
DM_AGE	数值变量	人口学特征:年龄	实际年龄值,单位:岁
DM_EDUC	多分类变量	人口学特征:文化程度	1=大学本/专科或以上;2=高中(包括中专);3=初中;4=小学;5=文盲;98=不详
H_DM	二分类变量	既往史:糖尿病史	0=否;1=是

2 研究实施阶段的数据管理与质量控制

2.1 数据收集与监查

数据采集过程,在有知情同意的情况下,通常由经统一培训的研究护士或研究助理按照事先制定的研究者手册及 SOP 进行采集。在此期间,独立第三方监查员对数据的实时监查,为保证数据采集的准确性和完整性起到了关键作用。

第三方监查员的任务是保证研究遵循研究方案进行,研究记录及报告数据及时填写、并准确完整。其主要审核的内容包括但不限于^[13-14]:(1) 纳入排除标准:实际入组的研究病例与方案规定的纳入排除标准符合情况;(2) 重要时间窗:研究对象入组、用药、随访是否在方案规定的时间窗内完成;(3) CRF 填写:确认所有的数据记录正确完整、字迹清晰整洁、无项目遗漏,并与原始资料一致,所有数据填写的错误或遗漏(如数据缺失、异常值、逻辑关系错误)均已改正或注明,经研究者签名,并注明日期;(4) 用药:剂量改变、治疗变更、合并用药、间发疾病等均应确认并记录;(5) 不良事件:确认研究期间发生的所有不良事件均已记录在案,严重不良事件在规定时间内作出报告,并记录在案;(6) 退出/失访:核实入选受试者的退出与失访,确认并记录;(7) 疑问质询:确认研究期间所有数据疑问质询的填写与回复是否及时。

为便于后期数据查对,所有数据修改应留痕。如为纸版 CRF,更正记录或填写错误时不应改变原始数据,而应采用附加叙述,在原始错误之处划一条线并在旁边写上正确数据,同时写上修改理由,记录修改人员及修改时间。如采用电子数据库,则数据库系统应具有留痕功能,记录修改前的数据、修改理由、修改人员及日期等信息。

2.2 数据录入

为避免数据录入过程中发生数据错误,一般要求由双人背靠背录入双份,如果两人录入不一致,应核对原始数据后再录入正确的数据。参与录入的人员,需进行必要录入培训,包括熟悉项目背景、CRF 的结构与编码、数据库录入操作及录入的 SOP。

2.3 数据复核与实时核查

在研究过程中,研究者也可安排相关数据管理的质量控制。比如,审核数据填写是否有遗漏或存在主要错误、指定访视的所有表格是否均已填写、关键变

量的值在否在允许范围内且无逻辑错误等^[15]。必要时,可由数据管理员定期描述关键变量的分布情况以发现是否存在异常值,或生成相应的数据汇总表格以发现可能潜在的数据错误^[16]。当然,如果是随机对照盲法的临床试验研究,数据的汇总与初步分析需要在盲态下进行。

3 研究数据整理阶段的数据管理与质量控制

3.1 数据核查与清理

在数据录入完成后,导出的数据库并不可以直接用于数据统计分析。在此期间需要进行数据核查与清理,这是保证数据质量的最后一道关卡。数据核查与清理是指由项目数据管理员对 CRF 中的各指标的数值和相互关系根据研究方案要求进行核查,对于存在缺失、逻辑矛盾、错误或不能确定的数据,以疑问表的形式由研究监查员传递给研究者,让研究者对疑问作出回答,经核实后再对数据库数据进行修订的过程。数据核查与清理的过程一般由数据管理员按数据核查计划内容编写数据核查程序,导出可能潜在的数据错误或疑问。数据核查与清理发现的任何问题,均应及时通知监查员,要求研究者作出回答。各种疑问及解答应用疑问表,包括要求对数据的补充和复核等。所有疑问表和错误数据内容及修改结果应有详细记录并妥善保存。

数据核查与清理的内容包括但不限于以下方面^[17]:(1) 数据取值范围:如性别变量“1=男,2=女”,如果数据库出现取值为3,则存在数据错误;一项普通人群的研究,空腹血糖值可设为3~12 mmol/L,超出此范围的数据需进一步进行数据疑问质询与核实。(2) 异常值核实:可通过对关键变量进行简单的描述性统计,分析变量的频数分布表、最大值、最小值、百分位数、茎叶图等以发现异常值;或以指标值超出均数 ± 2 个标准差作为上、下警戒值;亦可由临床医生根据临床经验确定警戒值,如血压超出40~230 mm Hg设置为警戒值,需进一步进行数据疑问质询与核实。(3) 变量间的逻辑关系矛盾核实:如某患者年龄为17岁,同时婚姻状况为已婚,一般情况下需进一步核实。(4) 缺失数据:年龄、性别及其他关键变量缺失需通过疑问质询进行补充。(5) 关键日期与时间的核实:是否在方案规定的日期内完成,随访日期是否在入组和出院时间之后等。

3.2 生成关键新变量、医学编码与去隐私化处理

在数据核查后,还有几项重要工作需要完成,即生成关键新变量、医学编码与去隐私化处理。

3.2.1 生成关键新变量

对于数据库后期统计分析经常会用到的关键变量,需进行整理或生成关键新变量。如由出生日期计算出年龄;由身高、体重计算出体质量指数,并按正常、超重、肥胖将变量进行分组;多次随访结局数据,需将多次随访信息进行汇总生成这段时间内总体是否发生结局事件以及事件发生时间的变量,以满足后期生存分析的需要。

3.2.2 医学编码

临床试验中收集的病史、不良事件、伴随药物治疗应在锁库前完成医学编码。通常建议使用标准的字典进行编码,如使用国际医学用语词典、世界卫生组织(World Health Organization, WHO)不良反应术语集、WHO药物词典等常用术语集。

3.2.3 去隐私化处理

根据健康保险隐私及责任法案的要求,临床研究对象的个人隐私应得到充分保护,任何包含个人身份信息的数据字段应在数据共享前从数据库中删除。个人识别信息可以保存在单独的文件中,由专人保管。共享的数据库中仅有唯一的研究对象识别代码,其他人不能从中识别出具体患者的身份信息。

3.3 数据审核与锁定

在数据清理完成后,由主要研究者、生物统计学家、数据管理员和申办者共同对数据库内数据进行核对和评价,对脱落病例、主要疗效、安全性等数据进行确认和盲态审核。在盲态审核认为所建立的数据库正确,所有数据疑问质询均已进行澄清和解答后,对数据库进行锁定。锁定后的数据文件不允许再作变动,以防止产生误操作及未经授权的修改。将锁定后的数据库进行数据传递或交统计分析人员进行统计分析。

4 小结

临床研究的数据管理与质量控制应从研究设计开始,并贯穿于研究的整个过程。总体上,临床研究的数据管理包括三级质量控制,应紧抓CRF设计、数据库的构建与逻辑审核、实施过程中的数据监查、录入完成后的数据核查与清理等关键环节,提高临床研究数据的完整性和准确性,从而提高整体临床研究质量。

参 考 文 献

- [1] 张崑,应峻. 临床研究数据管理策略[J]. 复旦学报(医学版), 2017, 44: 122-126.
- [2] 谢高强,姚晨. 数据管理在临床研究中的地位和作用[J]. 北京大学学报(医学版), 2010, 42: 641-643.
- [3] 国家食品药品监督管理总局. 临床试验数据管理工作技术指南(2016年第112号)[EB/OS]. <http://www.cfdi.org.cn/resource/news/8011.html>.
- [4] 国家食品药品监督管理总局. 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(2016年第113号)[EB/OS]. <http://www.cfdi.org.cn/resource/news/8012.html>.
- [5] 吴泰相,卞兆祥,李幼平,等. 促进我国临床试验数据管理规范化[J]. 中国循证医学杂志, 2018, 18: 532-537.
- [6] 吴东宁,何丽云,文天才,等. 失眠辨证论治随机对照临床研究中数据管理的设计方法研究[J]. 世界科学技术: 中医药现代化, 2016, 18: 1614-1620.
- [7] 李庆娜,黄秀玲,高蕊,等. 基于电子数据获取的临床研究数据管理计划的制定及其关键因素探讨[J]. 中西医结合学报, 2012, 10: 841-846.
- [8] 陈平雁,胥煜,代因. 临床试验数据管理中的常见问题探讨[J]. 药学学报, 2015, 11: 1485-1487.
- [9] 卜擎燕,熊宁宁,邹建东,等. 从临床研究数据管理角度设计病例报告表[J]. 中国新药杂志, 2007, 16: 339-343.
- [10] 邓亚中,于嘉,刘川,等. 电子化时代临床数据管理现状和未来趋势[J]. 中国新药杂志, 2014, 8: 879-884.
- [11] 王丽,程凯亮. 电子病历数据库在临床研究中的应用及偏向控制[J]. 协和医学杂志, 2018, 9: 77-82.
- [12] 国家食品药品监督管理总局. 临床试验的电子数据采集技术指导原则(2016年第114号)[EB/OS]. <http://samr.cfda.gov.cn/WS01/CL0087/160963.html>.
- [13] 丁丽曼,葛晶晶,赵明月. 初探临床监查员(CRA)职责[J]. 中国民族民间医药, 2012, 21: 88-90.
- [14] 田利华,赵离钟,顾佳,等. 临床监查员在临床试验中的作用[J]. 中国病案, 2014, 11: 17-19.
- [15] 宋歌,李学仲,赵晓峰,等. 中医临床研究中的数据管理和检查[J]. 中国中医药信息杂志, 2012, 19: 7-16.
- [16] Stephen BH, Steven RC, Warren SB, et al. 临床研究设计(第4版)[M]. 彭晓霞,唐讯,译. 北京:北京大学医学出版社有限公司, 2017: 234-259.
- [17] 于永沛,朱赛楠,阎小妍,等. 新药临床试验数据管理常见疑问类型分析[J]. 中国新药杂志, 2011, 11: 1004-1011.

(收稿日期: 2018-06-25)