

新发重大传染病流行期间临床研究数据治理：挑战与应对策略

弓孟春

刘 莉 彭 郁 管 音

（南方医科大学卫生与健康管理研究院
广州 510000）

（南方医科大学南方医院
广州 510000）

（神州数码医疗科技股份有限公司
北京 100080）

邓光璞

杨 越 杨之洵 母建康 路正鹏 史文钊

（南方医科大学南方医院 广州 510000）

（神州数码医疗科技股份有限公司 北京 100080）

朱 宏

（1 南方医科大学卫生与健康管理研究院 广州 510000 2 南方医科大学南方医院 广州 510000）

〔摘要〕 从隐私、伦理、标准化和统计 4 方面探讨传染病暴发流行期间临床研究数据治理所面临的挑战和应对措施，以期类似传染病流行再次出现时及时、有效开展临床研究数据治理提供参考。

〔关键词〕 COVID-19；临床研究数据；数据治理

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2020.08.001

Clinical Research Data Governance during the Epidemic Period of Emerging Major Infectious Diseases: Challenges and Countermeasures GONG Mengchun, Health Management Institute, Nanfang Medical University, Guangzhou 510000, China; LIU Li, PENG Yu, Nanfang Hospital, Southern Medical University, Guangzhou 510000, China; GUAN Yin, Digital China Health Technologies Co. Ltd., Beijing 100080, China; DENG Guangpu, Nanfang Hospital, Southern Medical University, Guangzhou 510000, China; YANG Yue, YANG Zhixun, MU Jiankang, LU Zhengpeng, SHI Wenzhao, Digital China Health Technologies Co. Ltd., Beijing 100080, China; ZHU Hong, 1Health Management Institute, Nanfang Medical University, Guangzhou 510000, 2 Nanfang Hospital, Southern Medical University, Guangzhou 510000, China

〔Abstract〕 The paper discusses the challenges and countermeasures of data governance in clinical research during outbreaks of infectious diseases from the aspects of privacy, ethics, standardization and statistics. The conclusion provides references for timely and effective data governance of clinical researches when similar epidemics occur again.

〔Keywords〕 COVID-19; clinical research data; data governance

〔修回日期〕 2020-07-16

〔作者简介〕 弓孟春，博士，高级工程师，发表论文 8 篇；通讯作者：朱宏，博士，研究员，硕士生导师。

1 引言

全球范围内人类传染病暴发事件正逐步增加^[1]。近年来全球共同面对埃博拉病毒、H1N1 流感和寨卡病毒等新发疫情时,能否迅速开展设计缜密的临床研究是制定有效医疗对策、尽快控制疫情扩散和预防未来暴发的关键因素^[2]。传染病暴发期间,疫情防控涉及跨机构、跨地区协作,因此如何平衡疫情信息披露、数据共享和患者隐私保护是首先应注意的问题。开展临床研究涉及的患者知情同意应充分考虑疫情状况,结合实际降低暴露风险和时间成本。可信的临床研究结果有赖于高质量研究数据,传染病数据更需要考虑数据采集完备性和可操作性,以产生更具意义的研究结果。统计方案设计应结合疫情实际情况,由于短期内大量新发确诊病例的出现会造成医务人员数量短缺,导致医疗数据质量降低,出现数据完整性、溯源性不足等问题,可利用统计方法学进行弥补。本文依据临床研究开展流程,探讨传染病暴发期间进行临床研究需关注的 4 部分内容:隐私、伦理、数据治理和统计,结合国际及国内经验,寻找解决策略并制定预案,以应对未来可能出现的重大传染病暴发流行。

2 跨机构数据共享及患者隐私保护

2.1 相关法规

随着我国信息技术与医疗行业不断融合,区域医疗大数据共享平台构建、维护与发展逐渐成为跨机构建设重点。数据共享是医疗数据发展内在要求,由于医疗数据特殊性,个人隐私保护在数据共享时面临极大挑战^[3]。《信息安全技术个人信息安全规范》对个人数据收集、存储、传输及共享等各个环节提出具体要求。该规范几乎将所有与医疗行为相关而产生的数据都定义为个人敏感数据,强调收集个人敏感信息时的知情同意,同时考虑到公众安全规定几项医疗行业征得授权同意的例外情况。这些例外情况必须是与公共安全、公共卫生、重大公共利益直接相关的研究,如果对外公布研究结果必须对个人隐私数据进行去标识化处理^[4]。数据隐

私安全是数据共享流程中的根本问题。国内《传染病防治法》、《艾滋病防治条例》、《执业医师法》等均有患者信息保护相关条款,是数据共享时传染病患者隐私保护的有力支撑^[5]。国际上,美国《健康保险流通与责任法案》(Health Insurance Portability and Accountability Act, HIPAA)以及欧盟的《通用数据保护条例》(General Data Protection Regulation, GDPR)都体现出发达国家对个人隐私保护的重视^[6-7]。

2.2 安全策略

数据隐私安全问题离不开信息技术支持,可通过建立医疗领域数字身份验证、数据限制访问信息系统、升级加密手段等加强对隐私数据保护。当前广泛应用的技术有:标准加密方法、数据失真技术、基于密码学可逆置换算法。另外数据共享使用涉及风险再识别问题,在量化隐私再识别风险方面可使用 g-distinct 分析法,该方法的有效性已经在有关文献^[8]中被验证。当前新型冠状病毒肺炎传播面广、潜伏期长,是涉及公共安全的严重事件,根据《信息安全技术个人信息安全规范》,在保护患者的同时有必要对社会共享其部分信息^[6-7]。患者发病前后接触史、活动轨迹、就医情况等是工作人员寻找传染源、传播途径的关键,有助于保护更大范围人群。因此疫情期间保护患者隐私同时公布其部分信息是不可避免的。但是在此过程中要明确划分个人权责,兼顾社会利益,同时实现个人隐私保护与公共卫生安全的最优权衡。

3 伦理审查与知情同意

3.1 公共卫生紧急情况下的伦理审查共识

在应对疫情过程中,通过临床研究寻找并验证特定医学防控措施(Medical Counter Measures, MCM,包括诊断技术、治疗药物及疫苗)的有效性及其安全性极为重要。在新型冠状病毒肺炎疫情发生后短时间内多国科学家发表针对新型冠状病毒感染引起的肺炎病原学、病例报道及流行病学研究^[9-11]。各研究所采取的伦理审查办法各有不同,包括豁免审查和机构审查(Institutional Review

Board, IRB) 等。已有机构开始申报某些药物的随机临床对照试验, 如 Remdesivir^[12]。这些研究的伦理审查是保证受试者权益、确保研究顺利进行、最大化其疫情防控作用的关键举措。世界卫生组织 (World Health Organization, WHO) 在既往甲型流感病毒、埃博拉病毒流行期间曾对公共卫生紧急情况下开展临床研究所采用的伦理审查方法组织了多国专家研讨, 给出相应指导意见^[13]。公共卫生紧急情况下临床研究应遵循的伦理准则与一般生物医学研究大致相同, 不同之处在于: 对于风险、获益及信任的评估存在差异, 必须被纳入伦理审查过程; 研究执行机构可靠性和透明性需被高度重视; 需要考虑紧急情况下完成标准伦理审查时间不足。尽管存在一些差异, 在各类公共卫生紧急情况下, 国际及国内伦理学指南所规定的准则和价值观必须得到体现和遵守。如果需要完成完整的伦理委员会审查流程, 应考虑建立审查快速通道, 但不应缩小审查范围或降低标准。快速通道办法包括: 调整伦理委员会成员面对面会议与线上交流时间比例; 使用疾病暴发前研究方案库中的方案或其中某些部分, 提交后进行预审; 成立紧急状态下特别伦理审查委员会, 可以是国家级或地区级; 无法实施其他措施情况下, 更多依赖事后审查; 同时设置相应措施以核查违规及不达标的伦理实践。

3.2 疫情期间患者知情同意获取

在传染病紧急暴发期间, 如何平衡临床研究开展效率和公众权益保障是面临的主要知情挑战。临床试验需要制定和提交详细研究方案, 获取入组患者知情同意, 响应伦理审查委员会意见, 在疫情期间将极大阻碍临床试验的迅速开展。疫情期间相关监管机构反应能力会降低, 阻碍临床试验开展的灵活性。因此在许多公共卫生实践中, 为应对疫情所进行的患者信息和标本收集工作可以在没有知情同意时开展, 但不允许用作其他疾病和将来的研究。美国食品药品监督管理局 (The US Food and Drug Administration, FDA) 有相应条例允许在紧急情况下进行研究无需知情同意^[14], 但该规则适用于研究参与者无法授予知情同意的临床状况, 如重大创伤、心脏骤停等, 而不是针对传染病紧急响应所进

行的临床研究。对于因为疫情被隔离或者需要危重护理的患者而言, 知情同意的获取需要考虑以下几点: 需尽量减少研究人员与受试者接触时间, 知情同意内容简单易懂; 现场执行知情同意签署的人员培训困难, 要确保核心内容表达清楚; 传递纸质材料可能导致病原传播, 可考虑电子分发和签署; 患者可能危重或被隔离, 考虑家属代为签署。在最小暴露风险和患者负担情况下, 让临床研究入组患者明确风险和义务, 以保障研究顺利开展。

4 数据采集方法和基础设施

4.1 概述

暴发性传染性疾病的流行往往涉及多种类型人群, 包括城市、乡镇、流动人口、特殊聚集人群 (如游轮、定点隔离人群、监狱等)。通过移动设备采集一般人群健康情况信息、利用信息化技术采集各医疗机构临床数据、借助专业人员 (临床研究执行人员、基层行政人员等) 采集特殊人群信息等, 都是可行的数据采集方式。中心化云平台架构是疫情紧急情况下合理高效的选择, 适于在多个医院进行数据共享、整合多种平台的开发环境, 保障多个地区病例上报和协作分析。云架构可以随着疫情进展对硬件资源进行弹性配置, 以随时响应不同数据体量、流量、用户数量及功能需求。例如在一些疫情防控信息综合平台中^[15], 云平台架构能够支撑微信平台健康人群监测数据采集及分析、院内数据整合及分析、抗体实验室病例上报、危重症患者预警及临床研究等工作。

4.2 资料完整性

4.2.1 方法 在疫情流行期间资料收集面临诸多挑战, 体现为不规范、不及时或不完整。其中重要资料或数据缺失对于临床研究影响最大, 提高数据完整性的方法如下: 一是全面收集可获取的信息。从健康人群使用的系统、院内临床信息系统、公共卫生体系信息系统、第3方检测平台系统、媒体舆情监测系统等多个渠道获取资料, 通过个人主索引进行互相关联, 建立尽可能全面的个人信息系统。二是对于院内临床信息系统中未记录的重要临床信

息,寻找可能记录的电子格式,包括医生及护士临时记录电子文档、各设备系统内部记录数据及软件后台数据库可读取数据。三是借助人工对重点人群定向采集信息。例如对疫情地返乡人员、隔离点医学观察人员、发热门诊就诊人员等,可通过疫情期间基层防控体系收集信息,提高个人信息上报系统使用率,确保及时、有效地采集新发症状、体温、不良情绪等信息。

4.2.2 疾病预防控制中心相应措施 全球各国都有以各级疾病预防控制中心(简称疾控中心)为数据节点构建的疫情监测系统,在各地卫生服务机构信息化水平不一致情况下,保障病例上报数据完整性,对疫情披露和防控决策十分重要。在上报过程中应采取一些质控手段保障数据完备性和可溯源,包括:明确数据整合日期(与发病日期、上报日期等区分),在疫情快速发展时间窗内尤为重要;明确报告时间,疾控中心应向各地卫生机构明确上报时间点,防止重复上报;明确数据来源和上报形式,上报数据是手工誊录还是经由医院信息系统、检验信息系统自动生成,需向各地卫生机构明确,统一上报格式对于后续数据整理和分析尤为重要;明晰监测病例定义,需要监测的病例是疑似、确诊或者产生结局(包括转院、治愈、死亡);定义上报管辖权,对于一些交界地带的卫生服务机构需要明确指定上报的疾控中心,以免产生重复数据。数据完备性还可能受到其他因素影响,如可用诊断工具、有效控制措施、公众对特定传染病的认识以及地方公共卫生官员可用资源等。

4.3 数据标准化

4.3.1 概述 数据完备性保障措施保证疫情数据传输过程中的真实度,而标准化则是对数据进行处理保障可操作性。临床症候群总结可能较快完成,但命名及术语标准制定在时效性上往往滞后,使疾病诊断、疾病分型、病原微生物检测、临床药物研发等数据支撑的相关临床研究不够敏捷,因此需要完善数据标准化技术以提高信息可操作性。具有统一临床术语体系后,应确定对关键临床字段标准化处理策略,这需要结合地方一线卫生从业人员习惯,将同一概念不同叫法进行标准化。对新确定的

病原微生物及由其引发的临床症候群进行定名及编码等标准化工作,对于公共卫生政策制定及执行、信息共享及临床研究具有极其重要的意义,要兼顾因地制宜与国际协同,也要考虑语言使用习惯及文化差异等,以科学的方法和扎实的数据为基础。

4.3.2 临床症候群命名 中国国家卫生健康委员会于2020年2月7日发布的暂定命名中,中文为新型冠状病毒肺炎,简称“新冠肺炎”,英文为 Novel Coronavirus Pneumonia (NCP)。这准确反映了目前中文语境下对于这一症候群的最常用描述,有利于我国社会各行各业、各年龄段及不同教育水平人群使用,对于疫情防控指挥、政策宣导贯彻、全民积极应对等各方面具有重大意义。世界卫生组织公布其对于此临床症候群的命名为"Coronavirus Disease 2019",缩写为"COVID-19",中文译法尚无统一标准(学界推荐为“2019 冠状病毒病”或“2019 冠状病毒疾病”)。该命名依据各国际组织间的既有共识,遵循世界卫生组织协调下全球公共卫生工作中对于疾病相关命名的常用做法,避免与地理位置、某种动物、个人或者群体相关联,防止名称使用不准确或是对相关事物“污名化”^[16]。该命名规避对于肺炎的特定指向,这与目前已发表的临床研究中对于该临床症候群的描述是一致的。该疾病存在多个系统受累的临床表现,肺外临床症状及严重程度在多项研究中均得到证实。

4.3.3 临床症候群编码 全球范围内最早对新型冠状病毒引起的临床症候群进行命名和编码的医学术语本体系统是 SNOMED CT,这也是全球使用最广泛的临床术语本体系统。在疫情暴发期间,SNOMED CT 迅速引入由 2019-nCoV(病毒临时命名)感染所导致疾病的临时命名和编码,即 840539006 | Disease caused by 2019 novel coronavirus (disorder) |,在2020年2月13日将病毒临时命名 2019-nCoV 更新为国际病毒分类学委员会命名“SARS-CoV-2”,同时“Disease caused by 2019 novel coronavirus”更新为“COVID-19”,与世界卫生组织统一命名匹配^[17]。SNOMED CT 在全球近 50 个国家用于临床诊疗及公共卫生数据分析处理,其命名逻辑体现了对该疾病临床表现复杂性的认识和各个国家在整理分析数据方面的技术要求,同时

术语体系快速迭代也反映出卫生机构在抗击疫情时的全球协作。国际疾病分类编码 (International Classification of Diseases, ICD) 是由世界卫生组织负责定期维护发布的全球通用疾病诊断编码系统, 为全球各国广泛采用。面对全球范围内暴发性的致病因素未明疾病, ICD 提供 U00 - U49 作为可选的代码字段。在疫情等公共卫生紧急状态下, 快速使用特定编码可保证在电子系统可用情况下, 对疾病及疾病亚型进行正确编码。2003 年 WHO 紧急公布 U04.9 作为 SARS 相关疾病的分类编码, 2016 年发布 U06 作为寨卡病毒感染相关疾病的编码, 作为新型冠状病毒感染所致的急性呼吸系统疾病, WHO 统一发布其 ICD 编码: U07.1。这为准确、有效整理患者临床资料、高效分析临床诊疗情况和寻找疾病诊治规律、及时准确整理及研判疾病流行趋势等提供重要技术手段。

5 统计学考量

5.1 数据质量

当缺失数据较少时 (如 <5%), 可考虑直接删除有缺失数据病例后进行分析。当缺失数据较多时, 可利用多重插补法 (Multiple Imputation) 来减少数据不完整对最终结果的影响。根据国外学者传染病研究经验, 当某变量的数据缺失比例 <25% 时可用多重插补法对数据进行补充, 若缺失比例 >25% 则不应在后续研究和模型拟合时考虑该变量^[18-19]。多重插补方法基本原理是基于回归、决策树、贝叶斯估计等模型, 通过已观测到的值对缺失值进行预测。主流医学统计软件 (如 SAS、R 和 STATA) 都有成熟模块和函数来实现多重插补法, 具体步骤可分为 3 步: 通过现有数据建立插补函数, 估计待插补的值, 加上不同偏差, 生成多组候选的插补值, 与现有数据共同组成多个完整数据集; 利用每个数据集进行研究计划的统计分析; 评价各个数据集结果, 根据模型评分选择最佳补插值。

5.2 随机对照试验

5.2.1 选择偏倚和误差控制 相比于简单随机, 序列改变区组随机化 (Permuted Block Randomiza-

tion) 可以更好地对入组患者进行随机分组。在区组时可分为大小两组, 并对大组和小组设置不同的药物分配比例, 避免负责发放药物的临床医生猜出药物种类, 进而有针对性地将某种药物交给病情更严重的患者^[20], 减小选择偏倚。在暴发性传染病流行期间, 高死亡率、间歇性暴发, 最快速度找到有效治疗方式的迫切需求等因素都支持略宽松的统计设计。因此在针对埃博拉的研究中, 美国学者以双侧 Boschloo 确切检验 ($\alpha = 0.05$) 为标准来分析主要结局变量^[21]。相比于传统的将双侧一类错误概率严格设定为 0.05 的做法, 利用 Boschloo 确切检验需要的样本量更小, 更容易实现。

5.2.2 阶段性监测 在疾病暴发前期与疗效有关的数据几乎为零, 因此必须对研究进行阶段性监测来不断评价和预估实验结果, 以便在发现实验组有确凿的危害或疗效时迅速停止试验。根据美国国立卫生研究院在埃博拉暴发期间的研究经验^[22-23], 两组比较的阶段性监测中, 如果 A 组比 B 组疗效更好的贝叶斯后验概率达到 99.9% 及以上, 即可认为 A 组更优而终止试验。在多种药物比较时也可在阶段性监测中排除明确无效的藥物组。若对照组的死亡人数超过临界值, 则中止该药使用, 不再将新患者分配至该组, 将优于其的实验组视为新对照组。通过这种方式可以在实验正式结束之前停止药效差的药物使用, 既保证 RCT 的正常进行, 又最大化入组患者利益。

6 结语

截至 2020 年 3 月 22 日在中国临床试验注册中心登记的新型冠状病毒相关临床试验多达 509 项, 而其中以治疗为目的的临床试验共有 222 项^[24]。本文分 4 部分探讨传染病暴发流行期间临床研究数据治理所面临的挑战和应对措施。隐私保护方面, 跨机构数据共享时要保证患者人口学信息安全。伦理方面, 特殊时期的“简化协议”或快速审查小组来加快批准传染病相关临床研究。标准化方面, 在临床数据收集、处理和分析过程中要保障数据完备性, 统一症状命名, 对关键字段标准化保证数据有效分析。统计方面, 方法学的考量需要尽早纳入, 以保

证结果真实、可信。这些隐私、伦理、标准化和统计方面的方法学总结可在类似传染病流行再次出现时为及时、有效开展临床研究数据治理提供参考。

参考文献

- Smith K F, Goldberg M, Rosenthal S, et al. Global Rise in Human Infectious Disease Outbreaks [J]. J R Soc Interface, 2014, 11 (101): 20140950.
- Gobat N, Amuasi J, Yazdanpanah Y, et al. Advancing Preparedness for Clinical Research during Infectious Disease Epidemics [J]. ERJ Open Res, 2019, 5 (2): 318–326.
- 关延风, 马骋宇. 基于电子病历的医疗信息隐私保护研究 [J]. 医学信息学杂志, 2011, 32 (8): 36–39.
- 蒋言斌, 李响. 我国医疗大数据患者隐私权保护及其模式选择 [J]. 医学与法学, 2018, 10 (1): 1–7.
- 黄芬, 叶冬青, 郝加虎. SARS 之后《传染病防治法》的教学初探 [J]. 疾病控制杂志, 2003, 7 (5): 384–385.
- 赵蓉, 何萍. 医疗大数据应用中的个人隐私保护体系研究 [J]. 中国卫生信息管理杂志, 2016, 13 (2): 191–196.
- 连凤梅, 沈瑞英. 流行病学调查也应重视伦理问题 [J]. 中国医学伦理学, 2004, 17 (3): 29–30.
- Gong M, Wang S, Wang L, et al. Evaluation of Privacy Risks of Patients' Data in China: case study [J]. JMIR Med Inform, 2020, 8 (2): e13046.
- Holshue M L, DeBolt C, Lindquist S, et al. First Case of 2019 Novel Coronavirus in the United States [J]. N Engl J Med, 2020, 382 (10): 929–936.
- Chen N, Zhou M, Dong X, et al. Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: a descriptive study [J]. Lancet (London, England), 2020, 395 (10223): 507–513.
- Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019 [J]. The New England Journal of Medicine, 2020, 382 (8): 727–733.
- Parsey M. Gilead Sciences Statement on the Company's Ongoing Response to the 2019 Novel Coronavirus (2019-nCoV) [EB/OL]. [2020-01-31]. <https://www.gilead.com/news-and-press/company-statements/>.
- WHO. Research Ethics in International Epidemic Response [EB/OL]. [2019-06-11]. http://www.who.int/ethics/publications/who_hse_gip_itp_10_1/en/.
- U. S. FDA. Protection of Human Subjects: informed consent and waiver of informed consent requirements in certain emergency research (final rules) [J]. Federal Register, 1996, 192 (61): 51498–51533.
- 朱宏, 史文钊, 刘莉, 等. 广东驰援湖北洪湖市医疗队的抗疫实践及思考 [J]. 中华医学杂志, 2020, 100 (16): 1209–1211.
- WHO. World Health Organization Best Practices for the Naming of New Human Infectious Diseases [EB/OL]. [2015-05-29]. <https://apps.who.int/iris/handle/10665/163636>.
- SNOMED International. March 2020 SNOMED CT International Edition Interim Release: up to date COVID-19 content available [EB/OL]. [2020-03-11]. <http://www.snomed.org/news-and-events/articles/march-2020-interim-snomed-release-COVID-19>.
- Hennessey T W, Bruden D, Castrodale L, et al. A Case-control Study of Risk Factors for Death from 2009 Pandemic Influenza A (H1N1): is American Indian racial status an independent risk factor? [J]. Epidemiol Infect, 2016, 144 (2): 315–324.
- Pitman J P, Wilkinson R, Liu Y, et al. Blood Component Use in a Sub-Saharan African Country: results of a 4-year evaluation of diagnoses associated with transfusion orders in Namibia [J]. Transfus Med Rev, 2015, 29 (1): 45–51.
- Follmann D, Proschan M. The Effect of Estimation and Biasing Strategies on Selection Bias in Clinical Trials with Permuted Blocks [J]. Journal of Statistical Planning and Inference, 1994, 39 (1): 1–17.
- Boschloo R D. Raised Conditional Level of Significance for the 2×2 -table When Testing the Equality of Two Probabilities [J]. Statistica Neerlandica, 2008, 24 (1): 1–9.
- Mulangu S, Dodd L E, Davey R T J, et al. A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics [J]. The New England Journal of Medicine, 2019, 381 (24): 2293–2303.
- Proschan M A, Dodd L E, Price D. Statistical Considerations for A Trial of Ebola Virus Disease Therapeutics [J]. Clin Trials, 2016, 13 (1): 39–48.
- 中国临床试验注册中心. 新型冠状病毒肺炎临床研究索引 [EB/OL]. [2020-03-22]. <http://www.chictr.org.cn/>.