

患者登记数据库构建技术规范



谭婧^{1#}, 彭晓霞^{2#}, 舒啸尘^{3#}, 王丽⁴, 黎国威⁵, 张玲⁶, 毛琛⁷, 郭新峰⁸, 孙鑫¹,
代表中国真实世界数据与研究联盟 (ChinaREAL)

1. 四川大学华西医院中国循证医学中心 (成都 610041)
2. 国家儿童医学中心, 首都医科大学附属北京儿童医院, 临床流行病学与循证医学中心 (北京 100045)
3. 苏州大学公共卫生学院流行病与卫生统计学系 (江苏苏州 215123)
4. 中国医学科学院基础医学研究所, 北京协和医学院基础学院, 流行病学系 (北京 100005)
5. 广东省第二人民医院临床流行病与方法学中心 (广州 510317)
6. 首都医科大学公共卫生学院 (北京 100069)
7. 南方医科大学公共卫生学院流行病学系 (广州 510515)
8. 广东省中医院, 广东省中医药科学院广州中医药大学第二附属医院 (广州 510120)

【摘要】 患者登记数据库作为一种重要的真实世界数据来源, 在药械评价和疾病管理等领域应用广泛。本文作为系列技术规范的第二篇, 从患者登记的概念、适用范围、患者登记数据库的策划、构建流程、与基于既有健康医疗数据的研究型数据库的比较等维度提出了构建患者登记数据库的基本要求, 并同时提出了患者登记数据库质量评价关键指标, 以期规范和指导同类研究的开展。

【关键词】 患者登记; 真实世界数据; 数据库; 技术规范

Technical guidance for developing patient registry databases

TAN Jing¹, PENG Xiaoxia², SHU Xiaochen³, WANG Li⁴, LI Guowei⁵, ZHANG Ling⁶, MAO Chen⁷,
GUO Xinfeng⁸, SUN Xin¹, On behalf of China REal world data and studies Alliance (ChinaREAL)

1. Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu, 610041, P.R.China
2. National Center for Children's Health, China; Clinical Epidemiology and Evidence-based Medicine Center, Beijing Children's Hospital, Capital Medical University, Beijing, 100045, P.R.China
3. Department of Epidemiology and Biostatistics, School of Public Health, Medical College Soochow University, Suzhou, 215123, P.R.China
4. Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine Peking Union Medical College, Beijing, 100005, P.R.China
5. Center for Clinical Epidemiology and Methodology (CCEM), Guangdong Second Provincial General Hospital, Guangzhou, 510317, P.R.China
6. School of Public Health, Capital Medical University, Beijing, 100069, P.R.China
7. Department of Epidemiology, School of Public Health, Southern Medical University, Guangzhou, 510515, P.R.China
8. Guangdong Provincial Hospital of Chinese Medicine, Guangdong Provincial Academy of Chinese Medical Sciences & 2nd Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, 510120, P.R.China

Corresponding author: TAN Jing, Email: tanjing84@outlook.com; SUN Xin, Email: sunx79@hotmail.com

【Abstract】 A patient registry database is an important source of real-world data, and has been widely used in the assessment of drug and medical devices, as well as disease management. As the second part of the serial technical guidance for real-world data and studies, this paper introduces the concept and scope of potential uses of patient registry databases, proposes recommendations for planning and developing a patient registry database, and compares existing health and medical databases. This paper further develops essential quality indicators for developing a patient registry database, in expect to guide future studies.

【Key words】 Patient registry; Real-world data; Research database; Technical guidance

DOI: 10.7507/1672-2531.201904161

基金项目: 国家“青年千人计划”项目 (编号: D1024002); 国家科技重大专项 (编号: 2017ZX09304023); 四川大学华西医院学科卓越发展 1.3.5 工程项目 (编号: ZYYC08003)

通信作者: 谭婧, Email: tanjing84@outlook.com; 孙鑫, Email: sunx79@hotmail.com

#共同第一作者



患者登记 (patient registries) 作为一种重要的真实世界数据来源, 日益受到国内外临床和卫生管理决策者的重视。通过预先的设计, 在多种质量控制措施下通过主动收集数据过程建立患者登记数据库, 可在药械评价和疾病管理中发挥重要作用, 尤其是在药械上市后监测、支持新药上市、医疗器械监管、慢性病和罕见病管理等研究领域^[1-6]。现有研究表明, 患者登记数据库质量在许多领域仍不甚满意^[7-9]。国内许多研究者对患者登记的概念、应用和数据库质量仍有不少疑问, 包括如何定义患者登记数据库? 其与基于既有健康医疗数据的研究型数据库的区别? 患者登记数据库是否一定代表高质量数据来源? 其主要应用领域是哪些? 因此, 中国真实世界数据与研究联盟 (ChinaREAL) 工作组发布系列技术规范 2, 重点阐释患者登记数据库构建和应用中的技术要点。

1 概念和适用范围

目前对患者登记的权威解释来自美国医疗保健研究与质量局 (Agency for Healthcare Research and Quality, AHRQ) 2007 年首次发布的《评估患者结局的登记指南》(Registries for evaluating patient outcomes: a user's guide) 及其在 2014 年的第 3 次更新版本^[10]。该指南指出“登记”是指收集、储存数据和数据记录产生的过程。通过上述登记形式而形成的数据库, 我们称为“患者登记数据库”^[10]。

登记研究 (registry study) 是基于一种或多种以研究、临床或卫生政策制定为目的, 采用观察性研究的方法收集一致性数据的组织系统, 用于评估具有某种疾病、状态或暴露人群特定结局的过程。由于研究对象一般是患者, 也称为患者登记研究^[10, 11]。

对于患者登记数据库的定义和识别, 我们提出如下技术要点: ① 患者登记数据库是以观察性研究的形式收集数据, 研究者不干预常规临床实践过程; ② 在收集和分析数据前, 需要形成相对清晰的研究目的和研究计划; ③ 至少部分数据需要主动收集, 而不仅仅是基于各类既有健康医疗数据库 [比如: 电子病历数据 (electronic medical records, EMR)、医保数据等] 收集数据; ④ 可基于不同研究目的在患者登记数据库中开展不同流行病学设计的研究, 如队列研究、病例-对照研究、基于患者登记的随机对照试验等; ⑤ 数据来源、定义、编码过程的标准化和一致性是保障患者登记数据库质量的核心。

根据登记对象的不同, 患者登记大致可分为 3

类: ① 特定疾病或者医疗状态的患者登记。通常以患特定疾病或具备特定医疗状态的诊断为此类登记的纳入标准。如急性缺血性脑卒中患者登记^[12]; ② 医疗产品登记。这是指患者在常规临床实践中使用了某种医疗产品, 包括药品、医疗器械、手术等。如接受药物洗脱支架患者的登记研究^[13]; ③ 综合医疗服务登记。通常将接受了某种综合医疗服务的患者作为纳入对象, 用于评估某种疾病管理方案、质量控制措施的效果^[14]。

患者登记数据库的适用范围, 从不同的维度划分有所不同, 总体来说可以划分为药械评价和疾病管理两大类。在药械评价领域, 包括产品有效性、安全性、经济性和依从性等不同维度, 可用于特定产品上市后安全性的常规监测和评价、特定产品有效性和安全性的上市后再评价 (条件上市药品、重点监测药品)、孤儿药的药品审评与评价和扩展适应症、基于患者自报数据的药物经济学研究等。在疾病管理领域, 包括了解疾病自然史, 确立疾病在特定人群中的发生、发展和转归的过程; 了解疾病负担, 调查人群患病率和地区、时间、空间分布; 诊断和筛查研究, 探索不同诊断方法的诊断价值; 评估预防和治疗措施的效果, 评价相对疗效和安全性; 疾病预后研究等。

我们总结以下情景特别适宜于建立患者登记数据库, 但不局限于以下情景: ① 观察疾病自然史; ② 开展疾病预后研究, 尤其是风险预测、预警; ③ 开展患者自报结局 (patient-reported outcome research) 研究; ④ 评估复杂干预的效果; ⑤ 罕见病防治与管理研究; ⑥ 如开展传统临床试验不符合伦理, 如孕妇、儿童等特殊人群; ⑦ 评估在真实诊疗环境下的疗效和治疗依从性, 研究人群异质性; ⑧ 评估亚组人群疗效。

在患者登记数据库种类的选择和应用中, 技术要点如下: ① 基于特定疾病或医疗状态的患者登记是开展疾病管理研究的优先选择; ② 仅纳入目标医疗产品的登记数据库可用于上市后药品安全性监测及依从性评价, 但因缺乏对照医疗产品评价, 难以比较有效性和安全性; ③ 研究对象可来自单中心或多中心, 但单中心来源的患者登记因人群代表性不足而局限性明显; ④ 在特定范围内连续纳入符合纳入标准的研究对象 (符合知情同意原则) 是减少选择性偏倚的关键举措。

2 患者登记数据库的策划

由于患者登记数据库通常需要前瞻性的纳入、

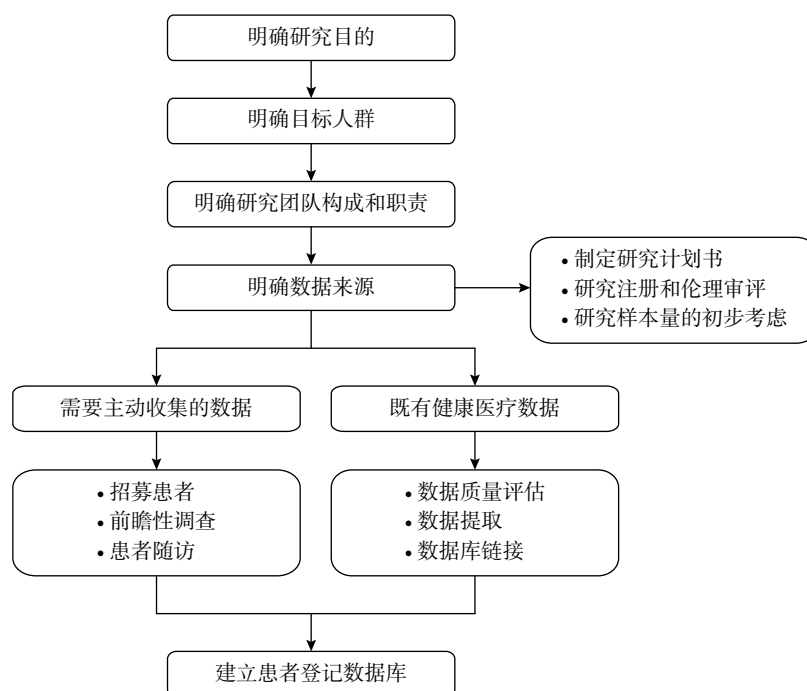


图1 患者登记数据库整体策划流程图

随访患者并收集数据,因此,需要的人力、财力、物力往往显著大于基于既有健康医疗数据的研究型数据库。其整体策划流程如图1所示。

在数据库策划阶段,有如下技术要点:①明确研究目的;②明确目标人群;③明确研究团队构成和相应职责;④明确核心数据的种类、来源和数据质量;⑤制定可实施的研究计划书;⑥伦理审批和研究注册;⑦研究样本量的初步考虑。

研究者应明确患者登记的研究目的,虽然研究目的可能是多样化的,如研究者既想了解疾病自然史,也想评估多种干预措施的疗效和安全性,但应有一定的边界。一个患者登记研究并不能解决与研究疾病相关的所有问题,研究目的增多所需的人力、财力显著增加,而相对聚焦和清晰的研究目的能减少研究成本、提高数据质量。同时应注意,开展患者登记的研究目的与研究问题是有明显差异的。研究目的是多样化的、框架性和方向性的,而研究问题则是单一的、具体的和针对性的。基于明确的研究目的建立患者登记数据库可解决多个研究问题,这些研究问题可能是预设的,也可能是在数据库建立后,研究者基于研究数据而产生的新的假设。

根据研究目的的不同,患者登记的目标人群有显著差异,也与患者登记的种类有关。如研究者希望了解某种心血管支架的安全性,可将使用该支架的患者作为目标人群,则此时建立的患者登记数据库属于医疗产品登记。但应注意,由于仅纳入使用

特定支架的患者,缺乏对照,研究者难以了解使用特定支架的不良事件发生率是否高于对照,故无法进行因果推断。

因此,可考虑将在一定纳入排除标准下的冠心病患者作为目标人群,开展特定疾病或者医疗状态的患者登记,由此建立的登记数据库不仅包含了使用特定支架的患者,还包括了使用其它类型支架的冠心病患者。

患者登记数据库从策划、设计、患者招募、随访、数据管理和分析,至少需要临床工作者、流行病学和统计学的方法学工作者和信息工程师的参与。临床工作者往往是患者管理的主要负责人,在患者随访和调查的过程中,可能还需要专业临床研究团队的协助;流行病学和统计学工作者需要负责研究设计、数据管理和分析等内容,是保障研究设计科学性和规范性的关键;建立电子化的在线(internet-based)患者登记数据库,以及从既有健康医疗数据中提取和链接数据,通常还需要信息工程师的参与和协助。在研究策划阶段明确上述研究团队各个角色的职责,是保证患者登记数据库长期、稳定运行的关键。

在数据库策划阶段应明确核心数据的种类和来源并初步评估数据质量。首先,若既有健康医疗数据(如医院EMR)能提供核心数据,并在数据质量(如变量定义、完整性)上能满足研究需求,研究者应权衡此时开展患者登记研究的必要性。其次,若

部分数据来自既有健康医疗数据,如通过医院 EMR 中的医嘱信息获得药品使用的情况,研究者还需通过主动收集的形式获得其他数据,如将患者自报的生命质量作为研究结局。此时,研究者一方面需要通过主动收集的方式获得数据,包括患者招募、前瞻性调查和患者随访等步骤,同时评估不同的调查方式,如面谈、电话访问等的可及性、准确性和实施难易程度。另一方面需要明确既有健康医疗数据的来源(如来自医院 EMR),并按照本技术规范 1 的内容开展数据质量评估;同时明确数据获取方式(如电子化提取、人工摘录)、是否需要链接其它数据库(如链接居民出生死亡登记库)等。为节约成本和提高研究可操作性,现有大多数患者登记研究的数据来源均由两部分构成,一部分来自前瞻性主动收集,一部分来自既有健康医疗管理数据。

研究计划书由多学科研究团队共同制定,内容至少包括:研究目的、研究内容、研究团队及其职责、如何获得研究对象、研究对象纳入和排除标准、如何开展随访、数据来源、数据收集方法、收集流程、数据变量定义和规则、研究进度安排、质量控制措施等。与临床随机对照试验不同,患者登记目前不需要强制在公共平台注册,但我们仍推荐研究者在研究正式开始前注册患者登记,而且有多个注册平台可供选择^[15,16]。同时将研究方案提交当地伦理审查委员会(institutional review board, IRB)接受审查和批准。若为多中心研究,建议在各个中心均提交方案至当地伦理审查委员会审查;对无伦理委员会的单位,一般可由项目负责单位完成伦理审查。患者登记研究通常需要研究对象签署患者知情同意书,向其说明此次研究目的、研究内容、研究期限、患者获益和风险等内容^[11,17]。应特别注意对患者个人信息保密,采取措施保障患者隐私和数据安全^[17]。待伦理委员会正式批准后,方可开始纳入患者。

基于研究目的,依据研究设计类型、主要研究结局、计划招募患者的时间、患者随访期限、拟分析的数据结构和研究预算等,估算合适的样本量。需注意的是,基于患者登记数据库的目的不同,样本量估算不一定必须。比如开展罕见病患者登记,其目的是不断收集和积累罕见病诊疗资料,同时管理患者,因此,并不需要计算明确的样本量。在资源、能力不足时,盲目扩大样本量并不可取。

3 患者登记数据库的构建过程

患者登记数据库的构建过程涉及患者管理和

数据收集两个部分。患者管理流程包括确定目标人群、设定纳入排除标准、招募患者、随访和维持。数据收集包括制作病例报告表(case report form, CRF)、制定标准化的数据收集手册、制作在线数据收集系统(electronic data-capturing, EDC)、开展预试验、开展调查员培训、数据提取、数据录入、数据审核、数据清理和数据储存。上述详细内容可参考已发表文献^[11]。其中,标准化数据收集手册至少应包括明确数据来源、变量字典、调查员培训内容、EDC 操作流程、研究者权限、主动收集/调查方式、数据提取方式、数据录入方式、数据核查流程、数据储存要求和研究进度安排(图 2)。需要说明的是,上述流程是构建患者登记数据库的主要步骤,但在不同的环境中、基于不同的研究目的和数据来源,数据库构建流程可能存在差异。

3.1 患者管理

可考虑如下技术要点:①充分考虑拟招募患者的人群代表性,跨地域、多中心招募为佳;②在特定范围内,符合纳入条件的患者应在知情同意原则下连续性纳入,防止主观选择患者;③明确由于不同的招募方法可能导致的偏倚类型,如志愿者偏倚;④对所有患者采用相同的随访方式,如经培训的调查员电话随访;⑤采用多种方式维持患者随访,可根据主要研究结局发生率确定可接受的最高失访率。

患者招募根据招募层次不同可有多种实现方式:一是医院招募,将目标医院所有符合纳入标准的患者纳入研究;二是医生招募,通常将在参加研究的医生处就诊的所有符合纳入标准的患者纳入;三是志愿者招募,通过网络等方式,由患者或负责医生将患者数据上报至登记系统。根据研究目的、疾病分布不同可采用不同的招募层次。应充分考虑不同招募层次对目标人群代表性的影响,及由此带来的选择性偏倚对研究结果的影响。

患者的随访和维持是患者登记数据库的难点之一。与传统队列研究类似,研究者将失访率设置在 20% 是主观的,可根据主要关注结局的发生率来确定可接受的最高失访率。过高的失访不仅降低检验效能,同时失访原因不同可能造成研究结论的重大偏误。因此,研究者在关注失访率的同时更应探讨失访的原因,采取多种措施减少患者失访。

3.2 数据库建设

3.2.1 设计阶段 ①应建立中央化的登记数据库;②在明确研究目的后确定 CRF 表的核心变量,包括纳入患者基本信息、联系方式、主要的暴

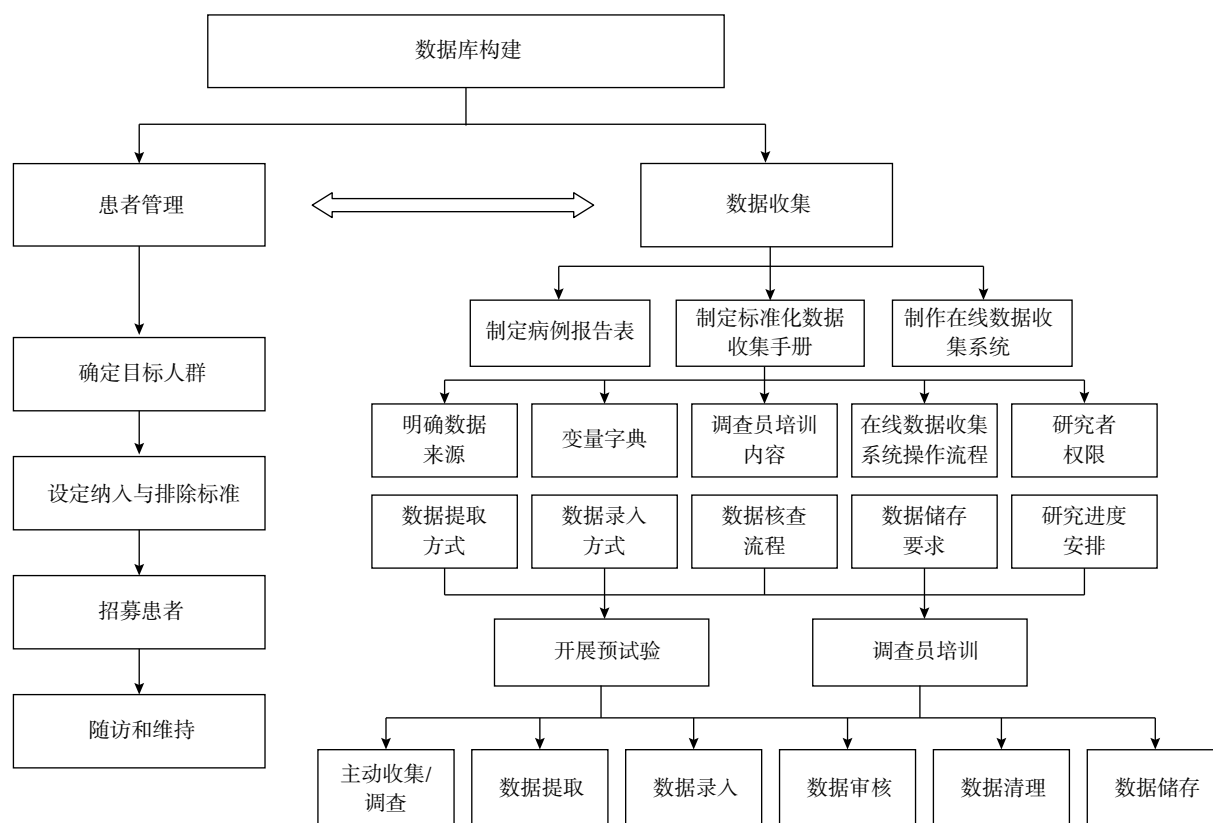


图2 患者登记数据库构建流程图

露因素、主要和次要结局指标、混杂因素等；③ CRF 表的内容设计应平衡目标和资源，切忌过于冗长；④ 制定随访进度表，随访次数过多和随访时间过长会显著增加研究执行难度，导致更大比例的失访；⑤ 标准化数据收集手册是开展患者登记的重要基础材料；⑥ 明确数据来源，明确需要主动收集/调查（如患者自报数据）和来自既有健康医疗数据（如 HIS、LIS 数据库、医保数据库、出生/死亡登记数据库等）；⑦ 明确不同来源数据的收集方式，如采用信息化编程方式从既有健康医疗数据中提取数据，采用人工提取（如 chart review）的方式从纸质病历中提取数据；⑧ 制定变量字典，包括变量定义、测量时间、测量方式、变量来源、编码信息等；⑨ 制定调查员手册，建立清晰和统一的调查流程，明确不同职责研究人员数据库使用权限；⑩ 建立结构化 EDC 和标准化 EDC 操作流程；⑪ 优化 CRF 表和 EDC 系统页面设计，便于填写和数据录入；⑫ 制定不同来源数据更新规则，如外部链接数据更新规则和随访数据更新规则。

3.2.2 数据收集阶段 ① 尽量在所有分中心开展预试验，进一步完善数据收集流程；② 采用合适的数据提取措施，如计算机编程提取、背对背提取，应开展一致性检查；③ 采用合适的数据录入

措施，如经双人录入，并核查一致性；④ 在研究过程中，可能根据实际情况更新和修改数据收集的方式、频率等，需要经核心研究团队讨论后以书面形式明确，如修改研究计划书；⑤ 制定数据核查流程，包括数据核查对象、数据核查规则、数据核查时间、生成质疑报告等；⑥ 需要长期随访的研究对象，需分阶段（如年度、项目中期）制定数据核查计划，及时处理在数据收集中可能出现的各种偏误。

3.2.3 数据清理阶段 ① 需建立标准化数据清理流程；② 无论采用信息化方式从电子系统中提取的数据，或采用人工阅读纸质材料方式提取的数据，均需抽取部分样本，采用人工核查方式验证并报告提取数据的准确率；③ 锁定数据库后，数据不再修改；若有修改，需书面向数据库负责人提出修改申请。

CRF 表的制作是建立患者登记数据的要点之一。除明确 CRF 的核心要素外，通常需要区分不同来源数据的填报对象和填报时间。如需要患者在门诊调查时填报的数据应在当次门诊就诊时完成，不恰当的调查时间和调查方式可能影响数据真实性和数据质量。

变量字典通常应至少包含以下变量：研究核心变量（暴露、结局、重要混杂）、尚无明确诊断标

准的变量、诊断标准可能存疑的变量、存在多次重复测量的变量、可能从多个来源获得的变量。上述变量应明确变量定义、测量时间、测量方式、变量来源、编码信息等,充分考虑临床实践情况是制定变量字典的基础。

数据核查可在不同的阶段完成。当需要使用既有健康医疗数据时,首先应核查既有数据的准确性和完整性;对于主动收集数据,可通过EDC系统设置数据核查规则,在数据录入时,同步对极端值、异常值和逻辑关联开展核查;最后对于各中心上传数据,数据管理单位还应基于预设的数据核查规则开展数据核查。存疑数据应生成数据质疑表,返回至数据调查和录入单位,核查数据并返回质疑结果。

我国各地的医疗条件和诊疗流程差异显著,对需要主动收集的数据建议各参与中心开展预试验,确保研究流程科学、可行;对于需要从既有健康医疗数据中提取的数据,由于各医院电子化信息平台

各异,应分别提取部分样本数据,验证数据的准确性和完整性。

将文本数据转化为结构化数据主要有两种方式:一是通过预设的标准,通过有经验的调查员阅读文本信息,从中摘取信息,实现结构化转化;二是基于多种基于机器学习和人工智能技术实现转化。无论哪种技术,准确性的高低都与原始文本数据的质量密切相关,不同的医疗机构和医务工作者书写习惯存在显著差异,因此,无论采取何种文本数据转化方式,均应报告文本数据转化为结构化数据的方法,并通过小样本数据验证数据准确性,报告验证准确率。

基于已有全民健康系统信息化程度的差异,不同的国家和地区在开展患者登记数据库研究时,可能不同程度地链接外部数据库,甚至将广泛地链接不同来源数据库作为该类患者登记的主要特征。在这类研究中,研究者应明确与外部数据库的链接规则和数据更新规则,尤其是针对需要长时间随访

表1 患者登记数据库与基于既有健康医疗数据的研究型数据库的比较

数据库类型	覆盖人群	覆盖数据变量	优势	局限	适宜的研究领域	备注
患者登记数据库						
特定疾病或者医疗状态的患者登记	具备特定疾病或者医疗状态的患者的患者	据研究目的确定数据范围	能获得研究目的下较完整数据;人群代表性较好,数据准确性较高;能获得患者自报数据;能获得长期随访数据	资源需求高;研究时间长;协调管理难度相对更大	疾病自然史;疾病负担评估;病因探索;防治策略评估;预后管理策略;药械评价	数据质量与研究设计、执行情况密切相关;患者登记数据库数据若部分来自既有健康医疗数据,则同时具备其特征
医疗产品登记	使用特定医疗产品的患者	医疗产品暴露数据、结局和其他数据	能获得患者自报数据;能获得长期随访数据	可能缺少对照医疗产品;难以开展归因研究	新上市药械安全性评价研究;支持医疗器械审评;治疗依从性;探索新适应症	
综合医疗服务登记	使用特定综合医疗服务的患者	综合医疗服务暴露、结局和其他数据	能获得患者自报数据;能获得长期随访数据	可能缺少未接受医疗服务人群数据	评估综合医疗服务效果	
基于既有健康医疗数据的研究型数据库						
单一医疗机构电子病历数据	医院门诊和住院患者	患者特征、病史、体格检查、诊断、门诊和住院医嘱、病程记录、手术、护理、检验、影像等	诊疗过程数据相对完善;医嘱信息完整,能获得院内用药数据	无院外诊疗数据;通常无法获得患者生活行为习惯数据;无院外随访数据;数据缺失可能较多;非结构化数据需要转化和验证	疾病负担评估;防治策略评估;诊断价值评估;预后管理策略;药械评价	数据完整性与医疗机构管理制度和医生填写习惯密切相关
区域性医疗数据	区域内户籍人口或常住人口	区域内医院EMR数据、区域内疾病防控数据、出生/死亡登记、环境数据等	能获得患者在多家医疗机构的就诊、转诊数据;能获得患者既往患病和就诊信息;能链接多个数据库(如预防接种、出生登记等)	医疗机构间数据质量有差异;患者链接可能不完善;药编码可能不完善;数据缺失可能较多	疾病负担(区域内人群);病因探索;预后管理策略;部分疾病自然史;药械评价	理论上是最全面和丰富的数据来源,但数据完整性、结构化程度与区域化数据库的成熟度密切相关
医保数据	参保患者	医保报销疾病诊断、部分诊疗数据、报销疾病费用数据	人群代表性好;费用数据准确性高	疾病诊断可能不准确;缺乏次要诊断数据;缺乏诊疗过程数据;无患者自报数据	药物经济学评价;疾病负担评估;部分病因探索;部分疾病自然史	可能缺少多种混杂变量数据

表2 患者登记数据库质量评价关键指标

内容	关键考核指标	解释	核对
数据库筹备	是否发表患者登记数据库研究计划书	患者登记数据库研究计划书应在公共平台发表	是/否
	是否经伦理委员会批准建立	研究方案经伦理委员会审查并批准	是/否
	是否与患者签署知情同意书	患者接受调查前应签署知情同意书	是/否
	是否有方法学人员(流行病学和统计学)参与数据库构建	从研究策划阶段开始应有方法学人员的参与	是/否
	是否明确患者登记数据库的种类	特定疾病或健康状态患者登记、医疗产品患者登记和综合医疗服务患者登记中的一类	是/否
	是否制作结构化的病例报告表	根据研究目的制定结构化的病例报告表,并经临床团队和方法学团队审核	是/否
	是否制作在线数据采集和管理系统	以病例报告表为模板,制作可在线录入和管理数据的电子化数据采集系统	是/否
	是否制定标准化数据收集手册	制定标准化数据收集手册,以统一、标化的方式收集数据	是/否
	是否制定变量字典	制定变量字典,包括核心变量定义、数据来源、测量方式等	是/否
	是否制定数据核查方案	制定数据核查方案,包括数据范围设定、逻辑核查等	是/否
	是否开展预试验	应在一家或多家研究中心开展预试验	是/否
患者招募	是否开展调查员培训	应对所有参与研究人员开展培训	是/否
	纳入人群能否代表目标研究人群	应明确目标研究人群和纳入人群的特征和范围,跨地域、多中心研究可提高人群代表性	是/否
患者随访和维持	是否采用恰当的患者招募方式	为减少选择偏倚,采用患者连续入组、随机化入组等患者入组方式	是/否
	是否有专人管理患者随访	应有专人或团队管理患者随访,包括随访进度、患者提示、接诊医生提示等	是/否
	是否采用多种措施减少患者失访	包括采取电话随访、微信/短信推送、召开研究者会议、定期报告研究进程等多种措施	是/否
	完整随访人群能否代表目标人群	尽量控制失访、分析失访原因	是/否
主动收集/调查	随访时间是否足够	随访时间是否足够观察预期研究结局	是/否
	是否明确数据来源	包括调查对象、调查时间、调查方式等	是/否
数据提取	是否采取标准和统一的方法收集数据	对需要主动收集或调查的数据,采用标准和统一的方法进行数据收集	是/否
	是否明确数据来源	数据所在数据库、字段、时间(多次测量)	是/否
	是否评估待提取结构化数据的完整性	抽取部分样本,初步评估待提取结构化数据的完整性、异常值和极端值比例	是/否
	是否评估待提取文本数据的准确性	抽取部分样本,初步评估待提取文本数据的准确性,在医疗文书中不同字段文本数据的准确性可能有显著差别	是/否
	是否验证提取数据的准确性	信息化方式从电子系统中提取数据,需抽取部分样本,采用人工核查方式验证提取数据准确性;人工阅读纸质材料方式提取数据,需经双人背对背提取	是/否
	是否采用恰当的数据链接方式	尽量采用患者唯一识别码链接数据;若采用算法链接数据,需报告算法的准确率	是/否
	是否采用恰当的方法进行文本数据结构化	采用各种技术将文本数据转化为结构化数据时,应抽取部分样本,人工验证文本数据结构化准确率	是/否
数据录入	是否核查录入数据准确性	经纸质问卷调查数据,需经双人录入,并核查数据准确性	是/否
数据传输	是否实现在线数据传输	应实现在线数据传输,便于管理数据和防治数据丢失	是/否
数据核查	是否有专人负责数据核查	应有专人或团队在数据录入阶段和数据录入完成后开展数据核查	是/否
	是否对疑问数据生成数据质疑表发送至数据调查和录入单位	对极端值、异常值、逻辑不一致等数据应生成数据质疑表发送至数据调查和录入单位,要求核实数据	是/否
	数据调查和录入单位是否返回核实后的意见	数据调查和录入单位应返回核实后的意见,并报告数据质疑表回收率	是/否
	是否制定阶段化数据核查计划	针对需要长时间随访的研究对象和数据收集过程,应制定阶段化数据核查计划,如年度计划、中期计划等	是/否
数据更新	是否有制定数据更新规则	包括外部链接数据更新规则和随访数据更新规则	是/否
数据清理	是否制定清晰明确的数据清理规则和流程	应制定清晰明确的数据清理规则和流程,保证数据清理透明、可重复	是/否
	是否保存完整的数据清理记录	应记录和保存完整的数据清理记录	是/否
	是否采用明确和恰当的极端值和异常值处理方法	采用明确和恰当的极端值和异常值处理方法,处理应结合临床实际	是/否
	是否采用明确和恰当的矛盾数据处理方法	采用明确和恰当的矛盾数据处理方法,建立矛盾数据处理优先级	是/否
研究进展	是否报告缺失数据比例	应报告数据库全部变量缺失数据比例	是/否
	是否有定期公布数据库研究进展	应预先设定时间节点,定期公布数据库建设进展,包括患者招募和随访情况、数据收集情况等	是/否

的研究对象和数据收集过程。

4 患者登记数据库与基于既有健康医疗数据的研究型数据库的比较

患者登记数据库与基于既有健康医疗数据的研究型数据库的比较见表1。不同设计和来源的数据库在覆盖人群类型、数据变量种类、适宜应用领域方面有显著不同,各有优势和不足。

5 患者登记数据库质量评价关键指标

见表2。

6 小结

综上所述,患者登记作为一种数据收集的方式,优势在于可通过预先设计和前瞻性的数据收集流程,获得研究者所需的全面、完整和在严格质控条件下收集的数据,而研究设计、实施质量直接关系到最终患者登记数据库质量。患者登记数据库与既有健康医疗数据的研究型数据库在质量评估和技术要点方面存在差异,基于既有健康医疗数据的研究型数据库构建的核心在于评估原始数据是否能满足研究需求,并通过多种方式获得数据并验证数据的可靠性、完整性和准确性。由于患者登记数据库的部分数据可能来自各种既有健康医疗数据,因此,在上一篇基于既有健康医疗数据的研究型数据库技术规范中的相应内容在此部分同样适用。最后,不同类型数据库并不代表数据质量的绝对高低,科学的设计、严格的实施和分析才是高质量研究数据的关键。研究者应回归研究本身,基于研究目的选择在现有资源条件下最适合的数据库类型。

参考文献

- 1 Bouvy JC, Blake K, Slattery J, *et al.* Registries in European post-marketing surveillance: a retrospective analysis of centrally approved products, 2005-2013. *Pharmacoepidemiol Drug Saf*, 2017, 26(12): 1442-1450.
- 2 Jonker CJ, van den Berg HM, Kwa MSG, *et al.* Registries supporting new drug applications. *Pharmacoepidemiol Drug Saf*, 2017, 26(12): 1451-1457.
- 3 Zhang S, Gaiser S, Kolominsky-Rabas PL. Cardiac implant registries 2006-2016: a systematic review and summary of global experiences. *BMJ Open*, 2018, 8(4): e019039.
- 4 U.S. Food & Drug Administration. Use of real-world evidence to support regulatory decision-making for medical devices. Available at: https://www.sogou.com/link?url=hedJjaC291MKYTE9c75enrRl1re67CaI2smYvubWPDFTrmuzC6_HxuOpabFoY9DDAtF2Uxnf_K7H6UfOzzMu_peCDarnS0mCuc_NSvjIMbO2TiRDw9Lmp63ypS7ArUC4tv2Zulat5pUc3zy0dg.
- 5 Resnic FS, Majithia A, Marinac-Dabic D, *et al.* Registry-based prospective, active surveillance of medical-device safety. *N Engl J Med*, 2017, 376(6): 526-535.
- 6 Peterson ED, Shah BR, Parsons L, *et al.* Trends in quality of care for patients with acute myocardial infarction in the National Registry of Myocardial Infarction from 1990 to 2006. *Am Heart J*, 2008, 156(6): 1045-1055.
- 7 Goodney PP. Improving the quality of data in surgical registries. *JAMA Surg*, 2015, 150(7): 636.
- 8 Arts D, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *J Am Med Inform Assoc*, 2002, 9(6): 600-611.
- 9 Arts D, de Keizer N, Scheffer GJ, *et al.* Quality of data collected for severity of illness scores in the Dutch National Intensive Care Evaluation (NICE) registry. *Intensive Care Med*, 2002, 28(5): 656-659.
- 10 Agency for Healthcare Research and Quality. Registries for evaluating patient outcomes: a user's guide. US Department of Health and Human Services: Agency for Healthcare Research and Quality (AHRQ), 2014: 1-356.
- 11 谭婧,程亮亮,王雯,等.患者登记研究的策划与患者登记数据库构建:基于观察性设计的真实世界研究. *中国循证医学杂志*, 2017, 17(12): 1365-1371.
- 12 Yang X, Li Z, Zhao X, *et al.* Use of warfarin at discharge among acute ischemic stroke patients with nonvalvular atrial fibrillation in China. *Stroke*, 2016, 47(2): 464-470.
- 13 Han Y, Jing Q, Li Y, *et al.* Sustained clinical safety and efficacy of a biodegradable-polymer coated sirolimus-eluting stent in "real-world" practice: three-year outcomes of the CREATE (Multi-Center Registry of EXCEL Biodegradable Polymer Drug Eluting Stents) study. *Catheter Cardiovasc Interv*, 2012, 79(2): 211-216.
- 14 Wandling MW, Ko CY, Bankey PE, *et al.* Expanding the scope of quality measurement in surgery to include nonoperative care: Results from the American College of Surgeons National Surgical Quality Improvement Program emergency general surgery pilot. *J Trauma Acute Care Surg*, 2017, 83(5): 837-845.
- 15 Langan SM, Schmidt SA, Wing K, *et al.* The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*, 2018, 363: k3532.
- 16 Spitzer E, Cannon CP, Serruys PW. Should real-world evidence be incorporated into regulatory approvals? Expert Opin Drug Saf, 2018, 9: 1-5.
- 17 李洪,魏来,郭晓蕙,等.真实世界研究伦理审查初探. *中国循证医学杂志*, 2018, 18(11): 1198-1202.

收稿日期: 2019-05-30 修回日期: 2019-06-14

本文编辑: 熊鹰