

FDA 以电子健康档案作为真实世界数据的监管考虑

李培^a, 陈玉文^{a,b*}[沈阳药科大学, a.工商管理学院, b.药品监管科学研究院(国家药品监督管理局药品监管科学研究基地), 沈阳 110016]

摘要: 电子健康档案作为真实世界数据的重要组成部分, 目前广泛用于安全性研究, 并逐渐用于有效性研究。本文系统介绍了美国食品药品监督管理局对研究者在临床研究中使用电子健康档案数据作为真实世界数据支持药物监管决策时的相关考虑, 以期为中国电子健康档案数据作为真实世界数据在药物研发与注册方面的应用提供参考和借鉴。

关键词: 电子健康档案; 美国食品药品监督管理局; 临床研究; 真实世界数据; 药物监管决策

中图分类号: R95

文献标志码: B

文章编号: 1007-7693(2023)10-1411-05

DOI: 10.13748/j.cnki.issn1007-7693.20222436

引用本文: 李培, 陈玉文. FDA 以电子健康档案作为真实世界数据的监管考虑[J]. 中国现代应用药学, 2023, 40(10): 1411-1415.

FDA's Regulatory Considerations for Electronic Health Records as Real-world Data

LI Pei^a, CHEN Yuwen^{a,b*}[Shenyang Pharmaceutical University, a.School of Business Administration, b.Research Institute of Drug Regulatory Science(Drug Regulatory Research Base of NMPA), Shenyang 110016, China]

ABSTRACT: As an important part of real-world data, electronic health record is widely used in safety research and gradually used in effectiveness research. This article systematically introduces the FDA's considerations when researchers use electronic health record as real-world data to support drug regulatory decision-making in clinical research, with a view to providing reference for the application of electronic health record as real-world data in drug development and registration in China.

KEYWORDS: electronic health record(EHR); Food and Drug Administration(FDA); clinical research; real-world data; drug regulatory decision-making

电子健康档案(electronic health record, EHR)被视为各种临床研究设计中的数据源, 在临床上有广泛的潜在用途, 包括非干预性(观察性)研究和干预性研究。如肿瘤药物开发的单臂临床试验中, 源自 EHR 数据库的外部控制可能会提供额外的背景信息^[1]。随着真实世界数据在药物研发和监管决策中的广泛应用, EHR 数据等日益发挥更大的作用, 如 2019 年辉瑞的 Ibrance(Palbociclib)通过 EHR 和上市后报告数据获批扩大适应证^[2]。

FDA 认识到 EHR 数据等真实世界数据在药物研发和监管决策中的重要价值, 先后发布了《临床研究中使用电子健康档案数据》^[3]、《真实世界数据: 评估电子健康档案和医疗索赔数据以支持药品和生物制品监管决策》草案^[4]、《包含真实数据的药物和生物制品提交的数据标准》草案^[5]、《使用真实世界数据和真实世界证据支持药品和生物制品监管决策的注意事项》草案^[6]等系列指南, 对研究者在临床研究中使用 EHR 数据作为真实世界数据支持药物监管决策的相关考虑进行了系统阐述。

1 EHR 数据转化为真实世界证据的挑战

EHR 在医疗保健中的广泛应用产生了大量真

实世界的的数据, 但由于 EHR 的高度异质性、复杂性、主要用于临床研究以外等特点, EHR 数据在药物研发与监管决策方面的应用仍充满挑战。在 EHR 系统的现状下, 信息主要隐藏在非结构化临床数据中(>80%)^[7], 因此从 EHR 的非结构化文本中提取信息至关重要, 许多信息学方法如自然语言处理(natural language processing, NLP)和机器学习(machine learning, ML)等已被探索利用, 但仍处于起步阶段且缺乏相应的方法标准^[8]。另外, 应关注 EHR 数据质量和局限性, 数据的可靠性和相关性可能难以保证。如 EHR 可能无法捕获或反映患者的完整健康状况, 患者信息可能分散在不同的医院或诊所, 复杂多样的临床数据标准可能妨碍不同的电子系统之间的信息交换^[9]。

2 EHR 作为真实世界数据的监管考虑

2.1 EHR 的定义

FDA 于 2013 年发布的《临床研究中电子源数据》^[10]指导原则中定义 EHR 为“根据公认的互操作性标准, 由医务人员为患者健康创建、输入、储存、并使用的电子临床信息”, 并注明 EHR 具有能够从多个来源获得、可共享、可互操作、可

作者简介: 李培, 女, 博士生, 主管药师 E-mail: lip@cde.org.cn

*通信作者: 陈玉文, 男, 博士, 教授, 博导 E-mail: cywwyc@163.com

供授权方访问等显著特征。指南[3]和[4]中应用了更为广泛的定义,定义 EHR 为“包含患者个人电子健康记录的电子平台”。并说明一份典型的电子健康记录包括患者医疗历史、诊断、治疗计划、免疫接种日期、过敏症、放射图像、药房记录以及实验室和化验结果。

EHR 数据结构类型一般包括结构化数据(包括人口统计、生命体征、实验室数据等)和非结构化数据(自由文本临床笔记、放射学或生物标志物报告等)。

2.2 EHR 和电子数据采集(electronic data capture, EDC)系统的互操作性

FDA 鼓励在临床研究中使用 EHR,并推动 EHR 和 EDC 系统之间交换数据,使用可互操作(将 EHR 数据以电子方式传输到 EDC 系统)或完全集成(将临床研究数据直接输入 EHR)的 EHR 和 EDC 系统,以降低手动转录程序可能会带来数据输入错误的风险,提高数据准确性以及临床研究中收集数据的质量和效率。

一般来说,EHR 系统属于医疗保健提供者、医疗保健组织和医疗保健机构,不受申办者、临床研究人员的控制,FDA 鼓励申办者和临床研究人员与控制 EHR 系统的组织合作。

指南[3]明确指出可以将数据从 EHR 直接传输到电子病例报告表,但也强调了其与医疗设备直接传输数据不同,EHR 可以使用干预过程。因此值得关注的是应保证在临床试验中,作为电子源数据收集和使用的 EHR 数据,具有可归属性、清晰可辨性、即时性、原始性和准确性;符合 FDA 检查、记录保存和记录保留的要求。应符合《临床研究中电子源数据》^[10]规定的临床试验中电子数据的捕获、审查和保留的一般要求。

EHR 和 EDC 系统的互操作性可能存在实际挑战,比如复杂多样的临床数据标准可能妨碍不同的电子系统之间的信息交换。EHR 和 EDC 系统之间的数据交换应利用现有的开放数据标准。FDA 鼓励使用获得国家卫生信息技术协调办公室卫生信息技术认证 EHR 技术以解决数据标准问题,但也允许使用未经国家卫生信息技术协调办公室认证的 EHR 系统,前提是控制措施到位,以确保数据的机密性、完整性和安全性。

此外,FDA 也进行了相关研究,如 2015 年 10 月—2019 年 10 月,FDA 的研究人员和加州大

学旧金山分校(UCSF)-斯坦福大学监管科学与创新卓越中心(CERSI)合作的 OneSource 项目“从电子健康档案(EHR)捕获源数据”^[11],目标是开发方法和工具,以自动将结构化 EHR 数据流向外部系统,从而降低运营成本,节省时间并提高临床试验的数据质量。

2.3 适用性评价

由于 EHR 系统不是为了支持向 FDA 提交监管文件而开发的,其中记录的数据主要取决于每个医疗保健系统的患者护理实践及其提供者的临床实践。基于 EHR 数据的研究需要保证关键数据(暴露、结局、协变量等)的可用性和足够数量具有代表性的患者数据;保证数据准确性、完整性、可追溯性等。

FDA 指南[4]指出解决数据可靠性和相关性须关注的重要问题:①研究问题数据源的适当性和潜在局限性以及支持关键研究要素;②确定研究设计要素的时间段;③研究设计元素的概念定义、操作定义和验证研究的结果;④数据积累、管理和转换为最终研究特定数据集的质量保证和质量控制程序。具体关注点见表 1。

由于 EHR 数据等是在常规护理期间收集,无预先研究方案,FDA 认为选择数据源的关键步骤是确定它是否捕获感兴趣的临床结果,应仔细考虑有关结果数据的可用性、准确性和完整性以及外部数据链接的需要。

关于缺失数据:FDA 建议解决与研究问题相关数据源的全面性和完整性。基于 EHR 的缺失数据主要为收集的 EHR 结构化信息在源数据中缺少信息及信息不在 EHR 数据中收集 2 种情形。FDA 要求应区分不同情形,并指出数据链接是解决缺失数据的一种方法,根据具体情形采取数据链接方法。可根据缺失机制的合理假设采用恰当的填补方法。方案和统计分析计划应详细描述缺失数据的特征和分析缺失数据的影响。

关于操作定义的准确性:应根据概念定义制定操作定义,从数据源中提取最完整、最准确的数据。FDA 建议使用合理的抽样方法评估大样本中操作定义的表现。

FDA 建议在研究制定或选择操作定义时考虑结局错误分类对研究有效性的潜在影响。验证结局变量以尽量减少结局错误分类,若完整验证不可行,应使用合理的抽样策略在研究人群中评估操作结局定义的性能。

表 1 FDA 考虑数据适用性的要素与内容

Tab. 1 Elements and content of data applicability considered by FDA

序号	考虑要素	评估内容
1	研究问题的数据源的适当性和潜在局限性以及支持关键研究要素	①应评估每个数据源,以确定可用信息是否适合解决特定的研究假设;②确定数据来源是否涵盖与研究相关的所有人群;③每个源数据是否都包含研究问题和设计所需完整详细的信息;④应考虑覆盖度及其连续性,应在方案中制定并记录入组或持续覆盖的定义;⑤是否有数据联动和整合;⑥分布式数据网络,应考虑数据转换及通用数据模型收集的数据元素;⑦标准化的可计算表型;⑧数据类型,对非结构化数据的处理;⑨解决缺失数据;⑩如何进行验证
2	确定研究设计要素的时间段	①明确定义与研究设计相关的各个时间段,如确定研究人群、定义纳入和排除标准、评估暴露、评估结果、评估协变量、跟进患者的时间段;②应证明建议的时间段和对研究有效性的潜在影响是合理的
3	研究设计元素的概念定义、操作定义和验证研究的结果	①目标人群的选择,纳入和排除标准;②暴露确认和验证(非干预性研究):持续时间、剂量、转换和其他暴露特征,考虑特殊人群的给药;③FDA 希望验证结局变量以尽量减少结果错误分类,建议使用标准化的病历审查流程,同时在方案中包括定量偏倚分析作为敏感性分析,以证明结局错误分类是否以及如何影响研究结果;④协变量确认和验证,混杂因素和影响修正因素
4	数据积累、管理和转换为最终研究特定数据集的质量保证和质量控制程序	制定整个数据生命周期中使用的管理和转换程序,考虑点包括:①对数据的完整性、一致性和数据值的合理性进行表征;②记录质量保证和质量控制计划包括转换过程;③定义一组确保数据完整性的程序质量控制程序

FDA 建议在方案中包括定量偏倚分析作为敏感性分析,并预先指定将用于定量偏倚分析的指标(敏感性、特异性、阳性预测值、阴性预测值),并描述如何在结果验证中测量所选指标。对于所有关键协变量,包括混杂因素和影响修饰,FDA 建议在方案和研究报告中提供并证明操作定义的有效性。

关于数据真实性:FDA 要求明确数据收集方法和流程,以最大限度地减少缺失和不可信数据的提取,记录数据纠错时间、原因等信息;关键分析数据能够追溯至源数据,任何修改均应记录。

关于数据质量管理:EHR 数据的格式和来源可能不同。研究人员应制定数据积累、管理和转换的流程和相应程序以保证数据完整性和准确性。FDA 明确了标准化数据质量报告要求。同时关注构建分析数据的质量保证和质量控制计划、质量控制方法,以及对数据元素不同数据质量水平的考虑等。并要求从数据收集到最终研究报告的撰写,所有手动和自动数据检索和转换过程都应进行彻底评估,以确保数据完整性。研究人员应确保管理和转换过程不会改变数据的含义或导致重要上下文信息的丢失。

2.4 标准化研究数据

应用当前支持的数据标准来研究源自真实世界数据源的数据。FDA《以电子格式提交申请—标准化研究数据》^[12]指出新药临床试验申请、新药生产上市注册申请、仿制药注册申请、生物制品许可申请等注册申请应采用标准化研究数据电子提交并明确了相关要求,目前 FDA 认可的标准包括临

床数据交换标准联盟的研究数据制表模型等。

FDA 认识到可以使用一系列方法将当前支持的数据标准应用于 EHR 等真实世界数据来源。申办者应参考 FD&C 法案第 745A(a)节要求使用研究数据指南中提供的规范、建议和一般注意事项,通过对所使用的一致性方法及其基本原理的充分记录,来自真实世界数据的研究数据可以转换为研究数据制表模型数据集,并在适用的药物注册申请中提交给 FDA。申办者应在方案、数据管理计划和(或)最终研究报告中详细描述数据转换的方法。

2.5 新技术的应用

在《哨兵系统——2019—2023 年五年战略》中^[13],FDA 将把重点放在新数据科学学科的创新上,例如 NLP 和 ML,并寻求扩大其对 EHR 的访问和使用。由于大量关键临床数据是 EHR 中的非结构化数据(如医师笔记、放射学报告),人工智能领域如 NLP、ML 越来越多地用于非结构化数据提取和转换。对此 FDA 认为所有这些方法目前需要大量的人工辅助管理和决策。如果使用人工智能或其他推导方法,则应明确所使用计算机算法的假设和参数、用于构建算法信息的数据源、算法是否受到监督,以及与方法验证相关指标对数据质量的相关影响。

2.6 与 FDA 沟通交流

《真实世界数据:评估电子健康档案和医疗索赔数据以支持药品和生物制品监管决策》草案^[4]强调了与 FDA 沟通交流的重要性。对于将提交给 FDA 以支持监管决策所使用 EHR 数据的研究,申办者应在进行研究之前提交协议和统计分析计划,

与 FDA 进行沟通交流。应预先定义研究设计、分析、实施和报告的所有基本要素,并描述如何从选定的真实世界数据来源确定该要素及验证研究。

2.7 知情同意与伦理

对于可互操作或完全集成的系统,申办者和临床研究人员应详细了解数据流和数据可见性,以便在双方知情同意允许的情况下访问患者数据。FDA 要求申办者必须确保符合 FDA 法规中保护人类受试者和机构审查委员会的适用要求。

3 思考与建议

3.1 加强顶层设计,推动基础建设

良好的 EHR 建设是 EHR 数据支持真实世界研究的基础。美国 EHR 建设起步较早,早在 1968 年就提出了医疗电子记录方式。后续启动的“我的电子健康档案”计划、Sentinel 计划、EHR 激励计划等对 EHR 的建设和应用起到了促进作用,截至 2020 年,美国约 95% 的医院都能从 EHR 中获取健康数据^[14-15]。中国电子健康档案建设与国外相比起步较晚,但发展迅速,《“十三五”全国人口健康信息化发展规划的通知》(国卫规划发〔2017〕6 号)提出,到 2020 年基本实现城乡居民拥有规范化的电子健康档案。国家积极开展医疗数据中心建设试点工作,如重庆市实现了全市跨区域、跨医疗机构的健康档案和电子病历数据共享和互联互通调阅^[16]。目前,EHR 建设依然存在诸多挑战,例如李静^[17]对北京 EHR 库建设情况进行调查发现存在 EHR 建档率有待提高,EHR 的效用远未发挥,各类健康档案数据需全面整合和共享等。另外,还存在隐私性、标准化和普及率等方面问题^[18]。EHR 建设是系统工程,涉及多部门、多行业、多方面的努力,应加强顶层设计,继续推进居民电子健康档案的普及与应用,不断完善居民电子健康档案、电子病历等数据库建设,制定标准化采集与共享,破除数据壁垒,推动跨部门、跨行业、跨层级的健康医疗大数据开放共享。同时,保证信息安全和个人隐私。

3.2 推动 EHR 在药物研发中的应用

EHR 作为真实世界数据的重要来源,受到国内外广泛关注。随着新一代信息技术与医药领域深度融合,更为真实世界研究提供了有利条件。FDA 大力推动 EHR 数据在药物研发与监管决策中的应用,发布一系列指南引导研究者在临床研究中正确使用 EHR 作为真实世界数据支持药物监管

决策,鼓励临床研究使用 EHR,促进可互操作或完全集成,指导解决数据适用性,采用标准化研究数据等。目前,国内相关研究较少,且尚没有相应指南发布,建议根据国内实际情况,制定完善技术要求,以期更好地推动真实世界研究,促进药物研发上市。

3.3 需要多方合作,加强沟通交流

来源于 EHR 健康医疗大数据具有海量性、复杂性、异构性等特点,且主要用于临床研究以外。如何将 EHR 等应用于真实世界研究,并产生高质量的证据支持药品研发与评价,需要汇聚多方力量,加强医疗机构、研究机构、企业、监管机构等交流合作,共同努力。特别注意的是,鼓励申请人与监管机构及时沟通交流,在真实世界研究正式开始前、研究中、研究完成后和递交资料前等不同阶段均可就相关问题进行沟通,以保证相关研究符合监管要求。

REFERENCES

- [1] CARRIGAN G, WHIPPLE S, CAPRA W B, et al. Using electronic health records to derive control arms for early phase single-arm lung cancer trials: Proof-of-concept in randomized controlled trials[J]. Clin Pharmacol Ther, 2020, 107(2): 369-377.
- [2] WEDAM S, FASHOYIN-AJE L, BLOOMQUIST E, et al. FDA approval summary: Palbociclib for male patients with metastatic breast cancer[J]. Clin Cancer Res, 2020, 26(6): 1208-1212.
- [3] FDA. Use of Electronic Health Record Data in Clinical Investigations[EB/OL]. (2018-07-19)[2022-04-02]. <https://www.fda.gov/media/97567/download>.
- [4] GIRMAN C J, RITCHEY M E, LORE V 3rd. Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products[J]. Pharmacoepidemiol Drug Saf, 2022, 31(7): 717-720.
- [5] FDA. Data Standards for Drug and Biological Product Submissions Containing Real-World Data[EB/OL]. (2021-10-21)[2022-04-02]. <https://www.fda.gov/media/153341/download>.
- [6] FDA. Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products[EB/OL]. (2021-12-07)[2022-04-02]. <https://www.fda.gov/media/154714/download>.
- [7] JUHN Y, LIU H F. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research[J]. J Allergy Clin Immunol, 2020, 145(2): 463-469.
- [8] FU S Y, CARLSON L A, PETERSON K J, et al. Natural language processing for the evaluation of methodological standards and best practices of EHR-based clinical research[J]. AMIA Jt Summits Transl Sci Proc, 2020(2020): 171-180.

- [9] BINKHEDER S, ASIRI M A, ALTOWAYAN K W, et al. Real-world evidence of COVID-19 patients' data quality in the electronic health records[J]. Healthcare(Basel), 2021, 9(12): 1648.
- [10] FDA. Electronic Source Data in Clinical Investigations [EB/OL]. (2013-09-18)[2022-04-02]. <https://www.fda.gov/media/85183/download>.
- [11] FDA. Source Data Capture from Electronic Health Records (EHRs): Using Standardized Clinical Research Data[EB/OL]. (2019-10-30)[2022-04-02]. <https://www.fda.gov>.
- [12] FDA. Providing Regulatory Submissions In Electronic Format-Standardized Study Data[EB/OL]. (2021-06)[2022-04-02]. <https://www.fda.gov/media/82716/download>.
- [13] FDA. Sentinel System Five-Year Strategy: 2019–2023[EB/OL]. (2022-04-02) [2022-04-02]. <https://www.fda.gov/safety/fdas-sentinel-initiative/fdas-sentinel-initiative-background>.
- [14] ZHAO T T. Analysis and experience of foreign electronic health record management model[J]. Digit & Microgr Imaging (数字与缩微影像), 2022(1): 11-13.
- [15] CMS. Promoting Interoperability Programs[EB/OL]. (2022-04-02)[2022-04-02]. <https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms?redirect=/ehrincentiveprograms/>.
- [16] 国家卫生健康委员会. 对十三届全国人大四次会议第10294 号建议的答复[EB/OL]. (2022-02-08) [2022-04-02]. <http://www.nhc.gov.cn/wjw/jianyi/202202/043135aa8fd8405c9c1b6a008a5bf5cd.shtml>.
- [17] LI J. Current situation and prospect of electronic health record construction and management of community residents in Beijing area[J]. Heilongjiang Sci(黑龙江科学), 2022, 13(10): 159-161.
- [18] FAN Y D, WANG Z T, CHEN M X. The construction of electronic health records in developed countries and regions and its enlightenment to China[J]. Mod Hosp(现代医院), 2021, 21(3): 470-473.

收稿日期: 2022-07-08
(本文责编: 沈倩)