

药物流行病学研究中的一次数据源介绍

卓琳^{1#} 贾敏^{2#} 沈智文² 张灵健² 王俊彦^{2,3} 孙凤⁴ 王巍巍⁵ 宫建² 詹思延^{1,4}

(1. 北京大学第三医院临床流行病学研究中心 北京 100191; 2. 沈阳药科大学药物流行病学与临床药物评价课题组;
3. 中国医科大学北部战区总医院和平院区学院药学部; 4. 北京大学公共卫生学院流行病学与卫生统计学系;
5. 首都医科大学附属北京安定医院, 国家精神心理疾病临床医学研究中心)

摘要 健康医疗数据是开展药物流行病学研究的基石,为药物使用、疾病防控、健康管理规划提供支撑。本研究以《中国药物流行病学研究方法学指南》为基础,从数据源和数据收集方法概述、国内外数据源现状、存在的问题、相应解决方案、案例分析等方面进行系统综述,介绍药物流行病学研究中一次数据及数据收集方法的特点,为开展药物流行病学研究时恰当地选择数据源及研究方法提供参考依据。

关键词 药物流行病学;数据源;一次数据;数据采集

中图分类号:R181.3⁺5 **文献标识码**:A **文章编号**:1005-0698(2021)03-0151-04

Primary Data Sources in Pharmacoepidemiology Studies

Zhuo Lin¹, Jia Min², Shen Zhiwen², Zhang Lingjian², Wang Junyan^{2,3}, Sun Feng⁴, Wang Weiwei⁵,
Gong Jian², Zhan Siyan^{1,4}

1. Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing 100191, China; 2. Research Group of Jian Gong on Pharmacoepidemiology and Clinical Drug Evaluation, Shenyang Pharmaceutical University; 3. Department of Pharmacy, School of General Hospital of Northern Theater Command, China Medical University; 4. Department of Epidemiology and Bio-statistics, School of Public Health, Peking University; 5. National Clinical Research Center for Mental Disorders & Beijing Key Laboratory of Mental Disorders, Beijing An Ding Hospital, Capital Medical University

ABSTRACT Data is the cornerstone of scientific research. The data sources in pharmacoepidemiology research can provide rich resources and professional background. The analysis results can support drug use, disease control and prevention, and health management planning. Based on the *Guide on Methodological Standards in Pharmacoepidemiology*, this overview systematically summarized the data source and data collection method for pharmacoepidemiology researches. Aims to provide references for the proper selection of primary data sources and research methods for further investigation, we also exemplified the domestic and foreign data sources accessible for current studies and corresponding pros and cons of the data sources.

KEY WORDS Pharmacoepidemiology; Data source; Primary data; Data collection

健康医疗数据是评价药物安全性、有效性、经济性的重要资源,也是开展药物流行病学研究的基础。国内拥有丰富的医药数据资源,但其综合利用程度尤其是在药物安全性评价领域仍不理想^[1]。充分利用现有的医药数据源,挖掘其医疗价值,可提高临床数据的利用度,更好的为临床医生、临床药师及患者服务。药物流行病学数据源根据来源不同可分为一次数据和二次数据^[2],

本文结合《中国药物流行病学研究方法学指

南》,着重介绍一次数据及数据收集方法的国内外研究进展,通过案例分析,介绍目前存在的问题和解决方案,以期提高我国药物流行病学研究的质量,为促进合理用药、提高临床实践水平、完善公共卫生服务体系提供参考依据。

1 一次数据介绍

医学研究中,在深入研究某特定健康相关问题而没有数据可用时,须组织专题研究收集资料,这种

#同等贡献

基金项目:北京大学医学部教育教学研究课题项目(编号:2020YB03)

通信作者:宫建 Tel:15840064816 E-mail:fanxing1230@163.com

詹思延 Tel:(010)82805162 E-mail:siyan-zhan@bjmu.edu.cn

为特定研究目的所收集的数据称为一次数据。在收集一次数据时,应使用经过验证的工具和测量方法,描述其验证方法并总结相关的完整性和有效性信息。如果需要在预试验中检验数据收集工具和测量方法,应描述预试验计划,如调查问卷的信效度测试。大数据时代之前,药物流行病学的数据库主要是通过观察法、问卷调查法及实验法等形式收集一次数据。随着大数据、云计算的广泛运用,一次数据的来源更加广泛,形成了注册登记和随访系统及健康调查数据库、监测数据库等电子化数据资源。下面将对这些数据库特性进行介绍和分析。

2 一次数据收集方法

2.1 观察法

观察法是指调查者不干扰所观察到的事件,用自己的感官和辅助工具去直接观察被研究对象,以获得信息的方法。根据是否参与被观察对象的活动,观察法可分为参与观察法和局外观察法。观察法直接获得资料,数据比较真实,具有及时性的优点,能捕捉到正在发生的现象。

早在19世纪,皮埃尔·路易斯率先将单纯的观察法用于肺结核、伤寒和肺炎等疾病的研究^[3]。即使到现在,观察法仍然广泛使用,如有研究者采用现场观察法调查菜农的不良工作姿势,进而判断其姿势危险等级^[4]。

2.2 问卷调查法

询问调查是调查人员以询问为手段,从受访者的答复中获取信息的方法。询问调查根据发送问卷的方式可以分为面对面调查、电话调查、邮件调查和邮寄调查。询问调查法区别于直接观察法,直接观察只能获得正在发生的事情的资料,而问卷调查不受时间、空间的限制,可以获得过去的资料。

2017年一项评价止吐药沙利度胺在癌症患者中预防性使用的Ⅲ期临床试验,采用了电话调查观察法来收集患者化疗后的呕吐反应^[5]。也有研究者为了分析临床试验过程中存在的问题,采用自填式问卷获取相关信息^[6]。

2.3 实验法

实验法是根据研究目的,将研究人群分为实验组和对照组,对实验组人为增加或减少某种因素,追踪该因素作用的结果。实验法根据研究对象和实施场所不同分为临床试验、现场试验、社区试验。观察法是实验法的基础,实验是在人为控制研究对象的条件下进行的观察。

目前实验法在分子流行病学中得到广泛应用。例如新型冠状病毒肺炎疫情期间,研究人员通过下一代测序确定病毒进化史,推断其可能的起源^[7]。

3 一次数据源举例

3.1 注册登记和随访系统

注册登记可以为药物流行病学研究提供丰富的数据。注册登记通过观察性研究的方法收集统一数据,以评估某一特定疾病、状况或暴露人群的特定结局,并用于探索预定的临床、科学或政策问题^[8]。注册登记也可以用于追踪疾病自然史、调查影响预后和生活质量的因素、确定治疗干预措施和卫生服务的临床效益或成本效益、测量或监测安全性事件以及评价和改善卫生服务质量,特别适用于罕见病患者的追踪调查。美国联邦疗效比较研究协调委员会(Federal Coordinating Council for Comparative Effectiveness Research)认为注册登记是疗效比较研究数据基础的核心组成部分^[9]。

注册登记根据其目的不同分为:①疾病注册登记(disease registries),关注具有相同诊断的患者;②产品注册登记(product registries),关注特定医疗产品,如药物或器材;③健康服务注册登记(health services registries),关注特定干预、术式或相同健康服务;④特殊病例注册登记,关注具有特定遗传生物标志物的人或双胞胎。注册登记研究可以仅关注某一种疾病/产品,或某一类人群/产品,如孕产妇注册登记、欧洲骨髓移植注册登记(European Bone Marrow Transplant registry, EBMT)。注册登记可以整合多种来源的个体水平数据。这些数据可以与机构或组织数据、医疗保险报销数据、健康统计数据和其他医疗数据相结合。不同注册登记的规模不同,小到在一个诊所内的疾病注册登记,大到可以进行多国家合作的注册登记。个人、卫生体系、私人基金会、政府机构或商业公司均可资助开展注册登记。

由美国政府资助的监测、流行病学及预后(Surveillance, Epidemiology, and End Results, SEER)项目是广泛应用于流行病学研究的注册登记之一^[10]。该项目收集肿瘤患者的人口统计学特征、临床治疗信息和死因信息。而SEER-Medicare数据库的建立真正使其在疗效比较研究中具有了额外的价值。该数据库由SEER与Medicare链接而来,提供Medicare中肿瘤患者从最初的诊断、治疗到长期随访信息,也包括健康相关生活质量(health-related quality of life, HRQOL)的详细信息。除SEER外,建设较为全面的

注册登记库还包括亚特兰大城市出生缺陷项目(Metropolitan Atlanta Congenital Defects Program, MACDP)、中国国家卒中登记数据库(CNSR)等。

国家内部可以通过关联人口信息构建全人群登记注册库,如北欧国家(丹麦、芬兰、冰岛、挪威和瑞典)的注册登记系统,常用于药品上市后的安全性和有效性评价。目前,欧洲国家正在逐步开展一些行动计划,希望进一步整合罕见病注册登记、生物样本信息以及临床信息等多源数据,其中,Orphanet收集了关于罕见病的注册登记。开发和利用多国之间合作的注册登记对罕见病患者信息的收集尤为重要。如戈谢病国际协作组的戈谢病注册登记(International Collaborative Gaucher Group Gaucher Registry, ICGGR)收集了来自56个国家四千余名戈谢病患者的数据,是目前全球最大的戈谢病注册登记^[11]。

目前,注册登记已经广泛应用于药物流行病学研究以及疗效比较研究。2005~2013年,通过欧洲药品管理局(EMA)集中程序批准的医疗器械中,有约9%需要进行注册登记研究^[12]。尤其在罕见病用药的研发中,往往由于临床试验样本量的限制,审批时对安全性的评价并不全面,而产品注册通常被认为是跟踪长期使用特定产品的患者的最有效方法。注册登记也可能包含未在常规诊疗数据中记录的信息,但这部分数据的质量和完整性必须具有一定的稳健性,能否实现预期目的取决于是否具备稳定的数据质量。换言之,研究需要事先掌握登记方案,了解研究过程,清楚数据变量的特性。

3.2 健康相关专题调查

通过调查个人医疗保健资源使用和支出情况的数据也可用于药物流行病学研究。例如,美国医疗支出固定样本调查(Medical Expenditure Panel Survey, MEPS)收集了美国居民的医疗服务使用频率、花费、支付方式等特定信息,以及美国工人医疗保险计划的花费等内容^[13]。MEPS有两个主要组成部分,即家庭调查(Household Component, HC)和保险调查(Insurance Component, IC)。HC由从全国代表性样本家庭(家庭和个人)收集的数据组成,包含人口统计特征、健康水平、健康状况、医疗服务利用情况、医疗费用支付来源、医疗服务可获得性和满意程度、保险范围、收入以及就业信息。IC则是对私营和公共机构雇主样本的单独调查,提供有关雇主健康保险的数据。

除MEPS以外,医疗保健支出和使用计划

(Healthcare Cost and Utilization Project, HCUP)通过美国联邦-州-企业合作的形式,汇集各州数据组织、医院协会、私人数据组织和联邦政府数据,形成了患者个体水平的国家数据资源,是美国最大的纵向医院数据集。HCUP数据集包括美国全国住院患者样本(Nationwide Inpatient Sample)、住院患儿数据库(Kids' Inpatient Database)、全国急诊科样本(National Emergency Department Sample)、州立住院患者数据库(State Inpatient Databases)、州立门诊手术数据库(State Ambulatory Surgery Databases)和州立急诊科数据库(State Emergency Department Databases)等,可用于卫生政策问题、医疗实践模式、医疗保健计划可获得性以及疗效比较等多种研究目的。

美国国家卫生统计中心(National Center for Health Statistics, NCHS)基于人群调查、供应商调查以及电话调查,获取所有州的出生、死亡和预期寿命等生命统计记录,可以为药物流行病学研究提供个体层面数据。其中美国全国健康访谈调查(National Health Interview Survey, NHIS)每年对全国代表性样本进行调查,收集有关健康状态和活动受限、健康行为、功能(functioning)和残疾、免疫接种和特定疾病状况。全国健康和营养调查(National Health and Nutrition Examination Survey, NHANES)在访谈基础上收集了体格检查、实验室检查、营养评估和DNA资料库的数据。

美国行为危险因素监测系统(Behavioral Risk Factor Surveillance System, BRFSS)每年对美国健康状况和危险行为进行跟踪调查,是目前世界上最大的健康电话调查系统。BRFSS每个月在50个州调查包括健康风险行为、预防性健康实践以及慢性病和伤害有关的医疗保健服务。这些数据与医疗保险数据库中商业和公共保险公司数据、美国全国大型数据库和政府数据库共同为药物流行病学研究提供非常宝贵的资源。

我国也有多个健康调查数据库,其中中国健康与养老追踪调查(China Health and Retirement Longitudinal Study, CHARLS)于2007年启动,旨在收集一套代表中国45岁及以上中老年人家庭和个人的高质量微观数据,用以分析我国人口老龄化问题,推动老龄化问题的跨学科研究。CHARLS全国基线调查于2011年开展,每两到三年追踪一次,覆盖150个县级单位,450个村级单位,约1万户家庭中的1.7万人,2018年已完成全国基线样本第三次常规追踪调查。此外,针对65岁及以上老年人及其35~64岁成年子

女的中国老年健康影响因素跟踪调查 (CLHLS) 也免费向学者开放,可自由申请数据进行相关研究。

3.3 自发性药品不良反应报告系统

自发性药品不良反应报告系统 (Spontaneous Reporting Systems, SRS), 简称自发呈报系统, 是应用于药物流行病学研究的重要监测数据资源。它是由参与主体在药品研发、生产经营和使用过程中发生的某种药物所致不良反应 (adverse drug reactions, ADR) 呈报数据组成, 一般由制药企业以及临床医生、药师、护士或者患者自愿呈报, 是发现 ADR 信号, 尤其是在上市前临床试验中未发现的罕见不良反应的重要途径。通过监测上市后药品不良反应的发生, 及时地发现安全信号和新的不良反应, 通过科学评价和有效控制达到加强药品风险管理的作用。

目前多个自发呈报系统数据库可供学者使用, 包括美国食品药品监督管理局 (FDA) 不良事件报告系统 (FAERS)、加拿大药品不良反应监测系统 (CADRMP)、世界卫生组织 (WHO) 乌普萨拉监测中心 (Uppsala Monitoring Centre, UMC) 以及我国国家药品监督管理局药品不良反应监测系统等。其中, 欧洲药品评价局不良反应资料库为限制性开放, 其余数据库均可从官网或经销商处获得^[14]。由于自发呈报系统为自愿而非强制报告, 不可避免会出现漏报情况。同时, 由于被动监测只能收集报告数量, 缺乏现实中整体用药人群基数, 即只有分子, 没有分母, 无法计算药品不良反应的发生率, 因此很难衡量不良反应发生的风险, 也无法分析相应的危险因素。自发呈报系统的另一个局限性是, 安全信息获取存在一定滞后性, 不利于风险早发现早控制。这些都在某种程度上限制了自愿报告系统在药品风险管理中的应用^[15]。

4 总结

在药物流行病学研究中广泛使用一次数据是非常有效的数据获取方式, 具体采用哪种数据收集方式、采用哪种类型的数据源, 主要由研究目的所决定, 根据不同问题发挥各自特长, 选择适合的数据源。一次数据多用于解决药物安全性、疗效、适应证甚至制定临床规范等问题。在面对某一个具体研究问题时, 我们应该根据研究目的、试验方案等合理选择适当的数据来源及数据收集方法。明确在哪些情况下可以联合使用一次数据和二次数据, 充分发挥两者的优势, 取长补短, 利用多源数据库以回答真实世界研究问题。本文概述了一次数据的类型和收集

方法以及存在的局限性, 以期为广大学者的具体研究提供相关参考信息。

参 考 文 献

- 1 宫建, 吴明智, 李翠英, 等. 基于 Web of Science 的药物流行病学领域药物安全性文献计量分析[J]. 药物流行病学杂志, 2018, 27(5): 338-341
- 2 中国药学会药物流行病学专业委员会. 中国药物流行病学研究方法学指南[J]. 药物流行病学杂志, 2019, 28(10): 695-700
- 3 钱柳玉, 吴俊, 叶冬青. 临床流行病学之父: 皮埃尔·路易斯[J]. 中华疾病控制杂志, 2018, 22(1): 94-96
- 4 郑文静, 郭孟杰, 么鸿雁, 等. 大棚蔬菜种植人员肌肉骨骼损伤工作姿势分析和危险等级评价[J]. 中国职业医学, 2019, 46(5): 591-594
- 5 Zhang L, Qu X, Teng Y, et al. Efficacy of thalidomide in preventing delayed nausea and vomiting induced by highly emetogenic chemotherapy: a randomized, multicenter, double-blind, placebo-controlled phase III trial (CLOG1302 study)[J]. J Clin Oncol, 2017, 35(31): 3558-3565
- 6 刘峰, 邓贵新, 李雪芹, 等. 实施药物临床试验存在问题的调查分析与对策研究[J]. 中国新药杂志, 2017, 26(17): 79-83
- 7 Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding [J]. Lancet, 2020, 395(10224): 565-574
- 8 Gliklich RE, Dreyer NA, Leavy MB, et al. Registries for evaluating patient outcomes: a user's guide[M]. Rockville (MD): Agency for Healthcare Research and Quality (US), 2014
- 9 Conway PH, Clancy C. Comparative-effectiveness research--implications of the Federal Coordinating Council's report [J]. N Engl J Med, 2009, 361(4): 328-330
- 10 Hankey BF, Ries LA, Edwards BK. The surveillance, epidemiology, and end results program: a national resource [J]. Cancer Epidemiol Biomarkers Prev, 1999, 8(12): 1117-1121
- 11 Charrow J, Andersson H C, Kaplan P, et al. The Gaucher registry: demographics and disease characteristics of 1698 patients with Gaucher disease[J]. Arch Intern Med, 2000, 160(18): 2835-2843
- 12 Bouvy JC, Blake K, Slattery J, et al. Registries in European post-marketing surveillance: a retrospective analysis of centrally approved products, 2005-2013 [J]. Pharmacoepidemiol Drug Saf, 2017, 26(12): 1442-1450
- 13 Cohen JW, Monheit AC, Beauregard KM, et al. The Medical Expenditure Panel Survey: a national health information resource [J]. Inquiry, 1996, 33(4): 373-389
- 14 Purrington A. 自发报告安全性数据库: FDA、WHO-UMC、Eudravigilance 数据库[J]. 中国药物警戒, 2009, 6(6): 382-384
- 15 杨羽, 詹思延. 上市后大数据药品安全主动监测模式研究的必要性和可行性[J]. 药物流行病学杂志, 2016, 25(7): 401-404

(2021-01-14 收稿 2021-01-27 修回)