

《药物流行病学研究方法学指南（第2版）》 系列解读（6）：研究数据源分类与选择



孟若谷¹, 杨 羽¹, 孙 凤^{2, 3, 4, 5, 6, 7}, 詹思延^{2, 3, 4, 8}

1. 北京大学健康医疗大数据国家研究院（北京 100191）
2. 北京大学公共卫生学院流行病与卫生统计学系（北京 100191）
3. 重大疾病流行病学教育部重点实验室（北京大学）（北京 100191）
4. 北京大学医学部药品上市后安全性研究中心（北京 100191）
5. 北京大学第三医院眼科（北京 100191）
6. 新疆医科大学中医学院（乌鲁木齐 830017）
7. 新疆石河子大学医学院（新疆石河子 832000）
8. 北京大学第三医院临床流行病学研究中心（北京 100191）

【摘要】 适宜的数据源是药物流行病学研究的基石。本文基于《中国药物流行病学研究方法学指南（第2版）》（以下简称“指南第2版”），深入解读药物流行病学研究中数据源的分类与选择。本文首先简要梳理药物流行病学研究中数据源的发展，详细解析一次数据源和二次数据源的定义、适用范围、使用条件、优势和局限性，进一步分析常见数据源的特征并介绍国内外经典数据源。随后，结合指南第2版中的研究问题、研究设计、研究人群、样本量、暴露或干预、结局和协变量等内容，深入探讨药物流行病学研究中数据源的选择标准和策略。最终形成以明晰分类为基础、研究问题为导向、研究方法定适配、数据质量作支柱的数据源选择方法论，旨在促进我国高质量药物流行病学研究的开展。

【关键词】 药物流行病学；方法学；数据源；数据

【中图分类号】 R 181.3+5 **【文献标识码】** A

*Guide on Methodological Standards in Pharmacoepidemiology in China (2nd edition)
and their series interpretation (6): data sources classification and selection*

MENG Ruogu¹, YANG Yu¹, SUN Feng^{2, 3, 4, 5, 6, 7}, ZHAN Siyan^{2, 3, 4, 8}

1. National Institute of Health Data Science, Peking University, Beijing 100191, China
2. Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China
3. Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China
4. Center for Post-Marketing Safety Evaluation of Drugs, Peking University Health Science Center, Beijing 100191, China
5. Department of Ophthalmology, Peking University Third Hospital, Beijing 100191, China
6. College of Traditional Chinese Medicine, Xinjiang Medical University, Urumqi 830017, China

DOI: 10.12173/j.issn.1005-0698.202506019

基金项目：国家自然科学基金面上项目（82473711）

通信作者：孙凤，博士，研究员，博士研究生导师，Email: sunfeng@bjmu.edu.cn

詹思延，博士，教授，博士研究生导师，Email: siyan-zhan@bjmu.edu.cn

7. School of Medicine, Shihezi University, Shihezi 832000, Xinjiang Uygur Autonomous Region, China

8. Clinical Epidemiology Research Center, Peking University Third Hospital, Beijing 100191, China

Corresponding authors: SUN Feng, Email: sunfeng@bjmu.edu.cn; ZHAN Siyan, Email: siyan-zhan@bjmu.edu.cn

【Abstract】 Appropriate data sources are the cornerstone of pharmacoepidemiological research. Based on the *Guide on Methodological Standards in Pharmacoepidemiology in China (2nd edition)*, hereinafter referred to as "the Guideline (2nd edition)", this article provides an in-depth interpretation of the classification and selection of data sources in pharmacoepidemiological research. It begins with a brief overview of the evolution of data sources in pharmacoepidemiological studies, followed by a detailed analysis of the definitions, applicable scopes, conditions of use, and strengths and limitations of primary and secondary data sources. It further introduces the characteristics of common data sources and relevant classic domestic and international databases. Subsequently, integrating aspects of the Guideline (2nd edition), including research questions, study design, study population, sample size, exposure or intervention, outcomes, and covariates, the article discusses the criteria and strategies for choosing data sources in pharmacoepidemiological research. Ultimately, a data source selection methodology is established based on clear classification, guided by research questions, tailored to research methods, and supported by data quality, aiming to promote the development of high-quality pharmacoepidemiological research in China.

【Keywords】 Pharmacoepidemiology; Methodology; Data sources; Data

药物流行病学作为一门交叉学科,将流行病学方法应用于临床药学领域,研究药物在人群中的使用、获益和风险,为制订临床决策和公共卫生政策提供证据^[1-2]。随机对照试验(randomized controlled trial, RCT)作为最经典的药物流行病学研究设计之一,被认为是证明药物有效性和初步评估不良反应风险的金标准。然而受制于RCT研究对象纳入标准严格、样本量小、干预措施明确单一、研究结局固定有限、随访时间短的特点,导致RCT所产生的证据多局限于“理想条件”^[3]。随着大数据时代的到来,电子健康档案(electronic health records, EHR)、医疗保险记录、疾病/医疗产品注册登记、可穿戴设备数据等常规卫生保健相关数据的产生和汇集,使得真实世界数据(real world data, RWD)成为了药物流行病学研究的重要数据来源。2016年美国的《21世纪治愈法案(21st Century Cures Act)》批准美国食品药品监督管理局(Food and Drug Administration, FDA)可利用真实世界证据(real world evidence, RWE)支持药品和医疗器械监管决策的制订,该里程碑事件也正式将药物流行病学研究数据源的重点从以研究目的而收集的数据扩展至因非特定药物流行病学研究目的而收集的数据^[4-5]。

美国FDA^[6]、我国国家药品监督管理局(National

Medical Products Administration, NMPA)^[7],以及国际药物流行病学学会(International Society of Pharmacoepidemiology, ISPE)^[8]、欧盟药物流行病学和药物警戒网络中心(European Network of Centres for Pharmacoepidemiology and Pharmacovigilance, ENCePP)^[9]等国内外机构颁布了多项文件,均对药物流行病学研究中数据源的分类、定义、适用范围、使用条件等内容进行了规定。2019年中国药学会发布的《中国药物流行病学研究方法学指南(第1版)》^[10](以下简称“指南第1版”)也对药物流行病学研究中的数据源提出了指导意见,即将发布的《中国药物流行病学研究方法学指南(第2版)》^[11](以下简称“指南第2版”)在指南第1版的基础上,进一步细化了数据来源、数据质量等内容。本文将围绕指南第2版详细解读药物流行病学研究中数据源的分类与选择,为促进我国药物流行病学研究的科学化和规范化提供参考。

1 药物流行病学研究数据源的分类及定义

从药物流行病学发展历程出发,早年制药企业、医疗机构和监管部门出于监管目的自主收集临床试验数据、医疗产品登记数据、患者处方数

据、不良反应自发呈报数据等用于开展药品有效性和安全性研究。随着卫生信息化不断完善,大数据和人工智能的发展,越来越多基于常规卫生保健服务产生的数据也应用于药物流行病学研究中,如电子病历、医疗保险记录等,研究范畴也扩展至药物利用研究、药物经济学评价等方面。

指南第 2 版中明确了药物流行病学研究中的数据源是确定暴露、结局和与研究目标相关的所有其他变量(如潜在的混杂变量和效应修饰因子)的策略和数据来源。常用的数据源包括:医疗信息系统(如电子病历、实验室信息管理系统等)、患者用药数据(如存储于医院药品管理信息系统、医药电子商务平台、制药企业产品追溯和药品安全性信息数据库以及药品使用监测平台等的数据库)、患者登记(如疾病登记、妊娠登记、出生缺陷登记、手术登记、药品登记、医疗器械登记、人口登记等)、药品安全性监测数据(如药品不良反应自发呈报、生产企业定期安全性更新报告等)、医保支付数据、患者报告结局、社交媒体、临床试验数据库、组学数据、来自移动设备(如智能手机、可穿戴设备等)的个体健康监测数据、公共卫生监测数据、自然人群队列数据、死亡登记、行政管理数据、患者随访数据、血药浓度监测数据等。根据数据源的收集方法,数据源可分类为一次数据源和二次数据源。

1.1 一次数据源

一次数据源是指为特定研究目的主动从患者、医护人员等研究对象中收集的数据。研究问题的性质以及研究设计的类型决定了研究是否需要使用一次数据源,具体包括:①涉及罕见人群;②涉及随机化设计;③需要评估研究终点;④需要开展额外的临床评估;⑤需要患者报告结局(patient reported outcome, PRO);⑥疫苗安全性研究;⑦医疗器械的安全性和有效性研究;⑧有特殊要求、受控分销的医疗产品等。

一次数据源可通过试验、调查、登记注册等途径收集患者报告测量、PRO、临床医生报告结局(clinician reported outcome, ClinRO)和医疗机构报告结局等内容,具体收集方式包括纸质问卷、电子病例报告表等。随着信息技术的发展和移动设备的普及,移动应用程序(applications, APP)也广泛用于药物流行病学研究中一次数据

的收集^[12]。2021 年发表的概括性综述研究^[13]对纵向研究中使用移动 APP 收集数据的情况进行了系统性分析,发现尽管移动 APP 能提高数据收集的效率,但也存在选择偏倚和信息偏倚的问题。正如指南第 2 版中的建议,无论是采用传统的线下数据收集模式还是新兴的线上模式,在收集一次数据时,均应使用经过验证的数据收集工具和测量方法,如已在研究目标人群中进行信效度检验的调查问卷,或遵从行业指南或共识的诊断标准等,并简要描述及引用相关信息。对于未经验证的数据收集工具和测量方法,应首先开展可行性分析或预试验,明确这些工具和方法的科学性和有效性,描述验证方法和结果。对于无法进行验证的数据收集工具和测量方法,应在研究中详细陈述其为数据源带来的潜在偏倚并讨论对研究结果造成的可能影响。

一般而言,一次数据源完整性好、准确性高、规范性强,能够获取 PRO 等常规健康医疗业务数据不涉及的信息,但同时也存在数据量、数据内容和数据时间跨度有限的问题。

1.1.1 试验数据

在药物流行病学研究中,试验数据主要来自于 RCT,以患者个体为单位进行随机分组并施加干预措施,旨在检验药物或治疗方法的效果。此外,数据还可以基于现场试验(field trial)产生,多在未患病个体中进行分组和干预,主要用于预防措施(如疫苗)的评价。试验数据内部效度高,可用于因果推断,但同时数据代表性差,其产生的证据外推性存疑。

1.1.2 调查数据

调查数据主要通过临床或社区中收集研究目标人群相关信息,包括单次横断面调查、多次重复调查和长期随访调查。药物警戒中,调查数据常用于评估风险最小化措施(risk minimisation measures, RMM)和风险评估与减轻策略(risk evaluation and mitigation strategy, REMS)的有效性^[14]。通过调查可获取研究对象用药依从性等主观信息,但也难以规避回忆偏倚和报告偏倚。

1.1.3 注册登记数据

一次数据源中的注册登记数据多指在常规诊疗环境下为评估药物有效性、分析不良反应危险因素等原因而专门系统性收集的观察性的纵向数据。注册登记多聚焦特定的药物或医疗产品保证

了数据的颗粒度,多采用前瞻性设计,其能够减少回忆偏倚,但多基于临床实践场景可能引入选择偏倚,且建立和维护成本高。

1.2 二次数据源

二次数据源是指现有因其他目的(如临床诊疗、药品管理等)而收集且可被再次利用于新研究目的的数据,包括电子健康档案、医疗保险记录、疾病/医疗产品注册登记、自发呈报等。FDA 进一步将从多种来源常规收集的与患者健康状况、卫生保健相关的数据定义为 RWD,主要用于监测和评估药物上市后安全性,此外,近年来也用于支持说明书或标签变更、适应证扩展及新药审批等目的的研究。

在使用二次数据源时,需注意以下几点:

①数据收集目的,决定数据的相关性和可靠性;②数据收集范围,是否覆盖研究目标人群?是否满足研究预期时限?是否覆盖研究所需信息?③数据收集方法,包括编码系统的使用等;④数据的完整性,关键变量、观察值的缺失;⑤药物暴露的准确性及研究结局的有效性,导致的偏倚;⑥已知和未知混杂的识别。在使用二次数据源开展药物流行病学研究前,应先对数据进行可行性评估,随后报告每个研究变量的数据来源和详细的评价(测量)方法^[15-16]。指南第2版中强调涉及多个数据源的研究中数据一致性的问题。若研究还存在数据链接,应报告数据来源、链接变量、链接方法、链接结果及链接评估^[17]。

相较于一次数据源,二次数据源的核心优势在于覆盖人群广、样本量大,使用成本低,能够反映真实环境下的卫生保健实践情况,但数据质量多不可控。

1.2.1 EHR

EHR 为记录患者疾病和/或健康相关信息的数字化记录,包括疾病诊断、过敏史、实验室检查、放射学影像、药物处方和健康结局等信息。EHR 包含在日常卫生保健场景下产生的丰富临床诊疗信息,但由于多为业务数据难以保证数据质量,且数据多成碎片化。临床实践研究数据链(clinical practice research datalink, CPRD)由英国全科医生初级诊疗数据组成,作为全球最知名的 EHR 数据库之一,已开展大量药物流行病学研究用于支持药物和医疗产品监管^[18]。

1.2.2 医疗保险记录

医疗保险记录涵盖参保人群的用药记录和费用信息等,具有大样本、长随访的优势,但缺乏更详细的临床诊疗信息(如实验室检查等)。因此,医疗保险记录更适合开展宏观研究,可用于识别用药模式和不良事件,还常用于药物经济学研究和政策评估。

1.2.3 疾病/医疗产品注册登记数据

疾病注册登记是指追踪特定疾病患者并前瞻、系统地收集的数据,数据一般包括详细的临床特征、治疗方案和预后情况。医疗产品注册登记是指在使用特定医疗产品人群中开展主动监测(如罕见病用药、生物制剂等)。疾病/医疗产品注册登记数据还经常与其他数据源(如死亡监测数据等)链接,在药物流行病学领域常用于研究特定药物在目标人群中的疗效和安全性。

1.2.4 药品不良反应自发呈报

自发呈报多由监管部门、医疗机构和企业用于被动监测药品上市后不良反应,数据包含患者基本信息、用药信息、不良反应信息、报告者信息等,是药物警戒的基石,用于发现安全性信号、识别用药错误和产品质量问题。作为常规监管数据,自发呈报覆盖人群广、持续时间长,能够发现罕见、长期、尤其是严重的不良反应。但数据密度低,缺乏用药全人群数据,无法计算不良反应发生率。此外,自发呈报中迟报、漏报、错报等情况无法避免。FDA 不良事件报告系统(FDA Adverse Event Reporting System, FAERS)包含了 FDA 收集的所有不良事件和用药错误信息,是药品上市后安全性研究的重要数据支撑^[19]。

1.2.5 队列研究数据

队列研究数据是从特定人群中收集所需研究对象某时点的暴露因素并进行随访观察预期结局的发生情况。随着包括英国生物银行(UK Biobank)在内的全球多个大型人群队列的建成,大样本、多组学、长随访的队列数据为开展药物流行病学研究提供了新机遇,尤其是在药物重定位领域^[20]。

此外,可穿戴设备和社交媒体产生的数据等也逐渐应用于药物流行病学研究中,特别是为患者生成的健康数据(patient-generated health data, PGHD)的采集提供了新途径^[21-22]。这些数据能够实时、连续捕捉细颗粒度信息,但同时也对数据质

量和分析能力提出挑战。此外，“数字鸿沟（digital divide）”也是不容忽视的伦理问题^[23]。

2 药物流行病学研究数据源的选择

适宜的数据源是药物流行病学研究的重要保障。尽管研究者可以根据对特定数据源的熟悉程度及使用经验进行选择，但从研究需求出发选择适宜的数据源仍是保证研究科学性和可行性的前提。此外，随着数据互联互通、共享开放程度的提升，单一数据源、多个独立数据源、组合数据源等多种形式也使得研究者面临艰难、复杂的选择。

前文基于指南第 2 版对药物流行病学研究中主要数据源的应用场景、优势和局限性进行了简要介绍，在此基础上，研究者应进一步结合研究问题和研究方法选择适宜的数据源。

2.1 研究问题

指南第 2 版将研究问题简要分类为描述性和分析性。对于描述性研究问题，理论上任何数据源均可以采用，但考虑到药物流行病学研究中描述性研究问题多聚焦于药物利用和安全性信号挖掘等，小样本数据难以满足研究代表性的要求，故建议使用 EHR、医疗保险记录等二次数据源。对于分析性研究问题，可进一步细化为提出假设、检验假设和验证假设^[24]。提出假设旨在揭示药物潜在的非预期效应，包括有益效应和有害效应。

理论上，所有包含暴露 / 干预和结局的数据源均可用于提出假设，但实践中，采用二次数据源研究可行性更强，尤其是医疗保险记录和自发呈报数据。此外，社交媒体数据也成为提出假设的新渠道。检验假设旨在为既有假设提供支持性证据（非确证性证据），通过检验结果决定是否启动假设验证，因此仍建议在二次数据源中开展，以提升研究速度并降低研究成本。验证假设需在提出假设上进一步考虑已知混杂因素，EHR、疾病 / 医疗产品注册登记数据、队列研究数据由于包含变量多，更适宜进行假设检验。检验假设旨在判断药物与已知效应间的因果关联。最有力的研究设计为 RCT，但随着倾向评分、工具变量、目标试验仿真研究（target trial emulation，TTE）^[25-26]等手段的出现，EHR、疾病 / 医疗产品注册登记数据、队列研究数据等数据也应用于验证假设。

2.2 研究方法

不同研究设计、研究人群、样本量、暴露或干预、结局、协变量等因素均会影响数据源的选择。

2.2.1 研究设计

指南第 2 版中将药物流行病学研究设计分为干预性和观察性，干预性研究主要为 RCT 并产生一次数据源，因此，研究设计主要影响二次数据源的选择。表 1 中总结了主要二次数据源适用的研究设计类型^[1, 24]。

表 1 二次数据源适用研究设计
Table 1. Research design on the applicability of secondary data sources

研究设计	数据源
队列研究	EHR、医疗保险记录、疾病/医疗产品注册登记、自发呈报、队列研究数据、可穿戴设备数据、社交媒体数据等
病例对照研究	EHR、医疗保险记录、疾病/医疗产品注册登记、自发呈报、队列研究数据、社交媒体数据等
单纯病例设计	EHR、医疗保险记录、疾病/医疗产品注册登记、队列研究数据、可穿戴设备数据等
横断面研究	EHR、医疗保险记录、疾病/医疗产品注册登记、自发呈报、社交媒体数据等
生态学研究	医疗保险记录、疾病/医疗产品注册登记、自发呈报等

2.2.2 研究人群

对于一次数据源，研究人群纳入排除标准明确。其中，试验数据因受研究成本所限，其代表性相对较差；而调查数据则非常依赖于抽样的设计和实施。

相比之下，二次数据源覆盖人群更广，尤其是 EHR、医疗保险记录和自发呈报多以区域性呈现、代表性好，疾病 / 医疗产品注册登记的人群规模和代表性多取决于原始研究纳排标准，而中国慢性病前瞻性研究（China Kadoorie

Biobank）^[27]、UK Biobank^[28] 和美国全民健康研究项目（All of Us）^[29] 等超大型人群队列人群覆盖广但存在幸存者偏倚。进行选择时研究者需要判断二次数据源是否覆盖研究目标人群，以及为基于人群的（population-based）或基于地点的（site-based），即指南第 2 版中提到的二次数据源使用注意事项之一“数据采集的完整性”——是否准确地采集了患者所有的健康记录，在数据库的覆盖性、信息完整性、时间长度等方面是否有明显欠缺。

2.2.3 样本量

相较于一次数据源,一般而言,二次数据源样本量更大。在使用二次数据源进行研究时,研究人群纳排标准也会直接影响样本量。此外,尽管样本量越大统计功效越高,但在选择数据源时也应考虑数据收集和使用的成本。

2.2.4 暴露/干预

对于暴露或干预为新药和罕见药的药物流行病学研究,建议使用一次数据源或二次数据源中的医疗产品注册登记,以期快速累积足够样本量。

由于一次数据源为特定研究目的收集的数据,因此多能够精准记录或测量暴露或干预情况,尤以 RCT 为甚。不同类型二次数据源暴露或干预数据的有效性存在差异。其中,以 EHR 特别是住院医嘱中的用药数据最为准确、及时,其次为医疗产品注册登记,自发呈报中药物依从性无法保证,医疗保险记录中的药物持有可能受到报销范围的影响,而常见的大型人群队列研究数据仅对常见药物进行调查^[24]。在选择二次数据源时,应尽量选择采用世界卫生组织解剖学治疗学及化学分类系统(anatomical therapeutic chemical, ATC)等术语记录的药物暴露或干预数据,以保证数据的互操作性。指南第2版中还建议,药物流行病学研究中应注意通过二次数据源评价药物暴露时带来的偏倚。

2.2.5 结局

对于发生率较低的研究结局,优选 EHR、医疗保险记录等大规模二次数据源,满足最小样本量需求;对于延迟药物效应产生的结局,因需要长期观察,同样建议选用 EHR、医疗保险记录等流动性低的二次数据源。

一次数据源是否能够准确、及时追踪研究结局取决于研究设计和实施情况,此外,失访问题难以避免。对于二次数据源,指南第2版重点指出了结局的有效性。整体来看,二次数据源中结局数据的有效性以疾病/医疗产品注册登记和自发呈报中不良反应记录为最优^[24]。EHR 和医疗保险记录中的研究结局,尤其是疾病结局的有效性取决于诊断和编码水平,而通过综合检查化验、处方、诊断等信息构建的疾病算法能够提高数据的准确性^[30]。同样,建议选择使用国际疾病分类(International Classification of Diseases, ICD)和《国际医学用语词典》(*Medical Dictionary for Regular Activities*,

MedDRA)等术语编码结局的二次数据源。

2.2.6 协变量

协变量数据的有效性与混杂控制息息相关。一次数据源,特别是 RCT,无论是否记录或测量协变量,均控制了混杂因素。二次数据源中,EHR、疾病/医疗产品注册登记和队列研究数据基于数据丰度的优势,常覆盖部分混杂因素;而医疗保险记录和自发呈报仅收集用药情况和疾病诊断等有限数据,难以控制研究中的混杂^[23]。

除了前文中提到的暴露或干预有效性、结局有效性、协变量完整性等内容外,数据质量的其他维度也会影响二次数据源的选择。FDA 和 NMPA 颁布的相关文件均提出需要基于 RWD 的相关性和可靠性评估数据的适用性^[7],从而判断 RWD 能否用于开展真实世界研究(real world study, RWS)、产生 RWE。此外,数据的合规、安全隐私、伦理问题也是选择数据源时需要考虑的因素。

药物流行病学研究的数据源生态系统日益丰富,从高度控制的一次数据源到反映复杂现实的二次数据源,每种数据源都有其独特的应用价值和固有限制。本文以指南第2版为基石,全面解读药物流行病学研究中数据源的分类。在选择数据源时,应以研究问题为导向、以研究方法定适配、以数据质量作支柱。

利益冲突声明: 作者声明本研究不存在任何经济或非经济利益冲突。

参考文献

- 1 Sabaté M, Montané E. Pharmacoepidemiology: an overview[J]. J Clin Med, 2023, 12(22): 7033. DOI: [10.3390/jcm12227033](https://doi.org/10.3390/jcm12227033).
- 2 Montastruc JL, Benevent J, Montastruc F, et al. What is pharmacoepidemiology? Definition, methods, interest and clinical applications[J]. Therapie, 2019, 74(2): 169-174. DOI: [10.1016/j.therap.2018.08.001](https://doi.org/10.1016/j.therap.2018.08.001).
- 3 孟若谷, 卓琳, 乔瑞, 等. 真实世界证据在上市后药品安全性监测与评价中的应用价值[J]. 中国药物警戒, 2021, 18(7): 624-627. [Meng RG, Zhuo L, Qiao R, et al. Application values of real-world evidence in post-marketing surveillances and evaluation for drug safety[J]. Chinese Journal of Pharmacovigilance, 2021, 18(7): 624-627.] DOI: [10.19803/j.1672-8629.2021.07.06](https://doi.org/10.19803/j.1672-8629.2021.07.06).
- 4 United States of America Congress. 21st Century Cures Act[S/OL]. (2016-12-13) [2025-06-06]. <https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf>.
- 5 Food and Drug Administration. Framework for FDA's real-world

- evidence program[S/OL]. (2018-12-01) [2025-06-06]. <https://www.fda.gov/media/120060/download>.
- 6 Food and Drug Administration. Use of electronic health records in clinical investigations[S/OL]. (2018-07-01) [2025-06-06]. <https://www.fda.gov/media/97567/download>.
- 7 国家药品监督管理局. 真实世界证据支持药物研发与审评的指导原则(试行)[EB/OL]. (2020-01-03) [2025-06-06]. <https://www.nmpa.gov.cn/yaopin/ypggtg/ypqgtg/20200107151901190.html>.
- 8 Public Policy Committee, International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practice (GPP)[J]. *Pharmacoepidemiol Drug Saf*, 2016, 25(1): 2-10. DOI: [10.1002/pds.3891](https://doi.org/10.1002/pds.3891).
- 9 The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (revision 11) [R/OL]. (2023-07-13) [2025-06-06]. https://encepp.europa.eu/encepp-toolkit/methodological-guide_en.
- 10 胥洋, 丁呈怡, 詹思延. 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(1): 5-9. [Xu Y, Ding CY, Zhan SY. Guide on Methodological Standards in Pharmacoepidemiology[J]. *Chinese Journal of Pharmacoepidemiology*, 2019, 28(1): 5-9.] DOI: [10.19960/j.cnki.issn1005-0698.2019.01.002](https://doi.org/10.19960/j.cnki.issn1005-0698.2019.01.002).
- 11 吴昉效, 颜济南, 聂晓璐, 等. 《中国药物流行病学研究方法学指南(第2版)》及其系列解读(1): 概述[J]. *药物流行病学杂志*, 2025, 34(1): 2-11. [Wu YX, Yan JN, Nie XL, et al. Guide on Methodological Standards in Pharmacoepidemiology in China(2nd edition) and their series interpretation(1): an overview[J]. *Chinese Journal of Pharmacoepidemiology*, 2025, 34(1): 2-11.] DOI: [10.12173/j.issn.1004-5511.202412131](https://doi.org/10.12173/j.issn.1004-5511.202412131).
- 12 Menni C, May A, Polidori L, et al. COVID-19 vaccine waning and effectiveness and side-effects of boosters: a prospective community study from the ZOE COVID Study[J]. *Lancet Infect Dis*, 2022, 22(7): 1002-1010. DOI: [10.1016/S1473-3099\(22\)00146-3](https://doi.org/10.1016/S1473-3099(22)00146-3).
- 13 Fischer F, Kleen S. Possibilities, problems, and perspectives of data collection by mobile apps in longitudinal epidemiological studies: scoping review[J]. *J Med Internet Res*, 2021, 23(1): e17691. DOI: [10.2196/17691](https://doi.org/10.2196/17691).
- 14 Smith MY, Russell A, Bahri P, et al. The RIMES statement: a checklist to assess the quality of studies evaluating risk minimization programs for medicinal products[J]. *Drug Saf*, 2018, 41(4): 389-401. DOI: [10.1007/s40264-017-0619-x](https://doi.org/10.1007/s40264-017-0619-x).
- 15 Rivera DR, Gokhale MN, Reynolds MW, et al. Linking electronic health data in pharmacoepidemiology: appropriateness and feasibility[J]. *Pharmacoepidemiol Drug Saf*, 2020, 29(1): 18-29. DOI: [10.1002/pds.4918](https://doi.org/10.1002/pds.4918).
- 16 Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE)[J]. *BMJ*, 2018, 363: k3532. DOI: [10.1136/bmj.k3532](https://doi.org/10.1136/bmj.k3532).
- 17 Pratt NL, Mack CD, Meyer AM, et al. Data linkage in pharmacoepidemiology: a call for rigorous evaluation and reporting[J]. *Pharmacoepidemiol Drug Saf*, 2020, 29(1): 9-17. DOI: [10.1002/pds.4924](https://doi.org/10.1002/pds.4924).
- 18 Yuen ASC, Chen B, Chan AYL, et al. Use of gabapentinoid treatment and the risk of self-harm: population based self-controlled case series study[J]. *BMJ*, 2025, 389: e081627. DOI: [10.1136/bmj-2024-081627](https://doi.org/10.1136/bmj-2024-081627).
- 19 Moore TJ, Glenmullen J, Mattison DR. Reports of pathological gambling, hypersexuality, and compulsive shopping associated with dopamine receptor agonist drugs[J]. *JAMA Intern Med*, 2014, 174(12): 1930-1933. DOI: [10.1001/jamainternmed.2014.5262](https://doi.org/10.1001/jamainternmed.2014.5262).
- 20 Si S, Liu H, Xu L, et al. Identification of novel therapeutic targets for chronic kidney disease and kidney function by integrating multi-omics proteome with transcriptome[J]. *Genome Med*, 2024, 16(1): 84. DOI: [10.1186/s13073-024-01356-x](https://doi.org/10.1186/s13073-024-01356-x).
- 21 Bourke A, Dixon WG, Roddam A, et al. Incorporating patient generated health data into pharmacoepidemiological research[J]. *Pharmacoepidemiol Drug Saf*, 2020, 29(12): 1540-1549. DOI: [10.1002/pds.5169](https://doi.org/10.1002/pds.5169).
- 22 van Stekelenborg J, Ellenius J, Maskell S, et al. Recommendations for the use of social media in pharmacovigilance: lessons from IMI WEB-RADR[J]. *Drug Saf*, 2019, 42(12): 1393-1407. DOI: [10.1007/s40264-019-00858-7](https://doi.org/10.1007/s40264-019-00858-7).
- 23 Lythreitis S, Singh SK, El-Kassar A. The digital divide: a review and future research agenda[J]. *Technol Forecast Soc*, 2022, 175: 121359. <https://doi.org/10.1016/j.techfore.2021.121359>.
- 24 Strom BL, Kimmel SE, Hennessy S. *Pharmacoepidemiology* (6th eds)[M]. Chichester: John Wiley & Sons Ltd, 2019: 165-371. DOI: [10.1002/9781119413431](https://doi.org/10.1002/9781119413431).
- 25 Zhao H, Zhang B, Zhuo L, et al. Association between use of sodium-glucose cotransporter 2 inhibitors and epilepsy: a population-based study using target trial emulation[J]. *Diabetes Care*, 2025, 48(5): 827-836. DOI: [10.2337/dc24-2532](https://doi.org/10.2337/dc24-2532).
- 26 Paterno E, Pawar A, Franklin JM, et al. Empagliflozin and the risk of heart failure hospitalization in routine clinical care[J]. *Circulation*, 2019, 139(25): 2822-2830. DOI: [10.1161/CIRCULATIONAHA.118.039177](https://doi.org/10.1161/CIRCULATIONAHA.118.039177).
- 27 Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up[J]. *Int J Epidemiol*, 2011, 40(6): 1652-1666. DOI: [10.1093/ije/dyr120](https://doi.org/10.1093/ije/dyr120).
- 28 Palmer LJ. UK Biobank: bank on it[J]. *Lancet*, 2007, 369(9578): 1980-1982. DOI: [10.1016/S0140-6736\(07\)60924-6](https://doi.org/10.1016/S0140-6736(07)60924-6).
- 29 All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The "All of Us" research program[J]. *N Engl J Med*, 2019, 381(7): 668-676. DOI: [10.1056/NEJMs1809937](https://doi.org/10.1056/NEJMs1809937).
- 30 Meng R, Ma R, Wang J, et al. Post-marketing surveillance for the safety of the 9-valent human papillomavirus vaccine: a retrospective real-world study in China[J]. *Expert Rev Vaccines*, 2023, 22(1): 696-703. DOI: [10.1080/14760584.2023.2239911](https://doi.org/10.1080/14760584.2023.2239911).

收稿日期: 2025 年 06 月 01 日 修回日期: 2025 年 06 月 12 日
 本文编辑: 冼静怡 杨 燕