

R 语言在药物警戒信号检测中的应用

傅雅丽^{1, 2, 3}, 巴磊^{1, 2, 3}, 周健^{1, 2, 3}, 施雯慧^{1, 2, 3}



1. 江苏省卫生健康发展研究中心 (南京 210036)
2. 国家卫生健康委避孕药具警戒和生育力监测重点实验室 (南京 210036)
3. 江苏省生育力保护与卫生技术评估重点实验室 (南京 210036)

【摘要】 药物警戒是保障用药安全的关键，其核心任务是通过数据监测及时识别潜在的药品安全信号。随着自发报告系统、电子健康记录、医保数据及其他真实世界数据的不断积累，信号检测工具快速发展。R 语言凭借其开源生态与强大统计能力等优势，成为药物警戒信号检测的核心工具。本文系统梳理信号检测的进展与主要方法，比较相关 R 软件包功能、应用场景，探讨 R 语言在信号检测中的优势和挑战。结果显示，R 语言生态涵盖经典频数法、似然比检验、自适应 Lasso 和数据驱动模式发现等多类方法，在数据质量要求、可解释性、自动化程度等方面均存在差异。未来应进一步加强多源真实世界数据整合、联合用药信号识别、可解释人工智能和面向监管实践的信号检测工作全流程建设。

【关键词】 药物警戒；信号检测；药品不良反应；不良事件；R 语言

【中图分类号】 R994.1；R195.1 **【文献标志码】** A

Application of R language in pharmacovigilance signal detection

FU Yali^{1,2,3}, BA Lei^{1,2,3}, ZHOU Jian^{1,2,3}, SHI Wenhui^{1,2,3}

1. Jiangsu Health Development Research Center, Nanjing 210036, China
 2. NHC Key Laboratory of Contraceptives Vigilance and Fertility Surveillance, Nanjing 210036, China
 3. Jiangsu Provincial Medical Key Laboratory of Fertility Protection and Health Technology Assessment, Nanjing 210036, China
- Corresponding authors: ZHOU Jian, Email: jsjss277@aliyun.com; SHI Wenhui, Email: codeblue@163.com

【Abstract】 Pharmacovigilance plays a critical role in ensuring medication safety, with its core mission being the timely detection of potential drug-event associations through adverse reaction monitoring. Signal detection tools have advanced rapidly, driven by the accumulation of data from spontaneous reporting systems, electronic health records, health insurance databases, and other real-world sources. With its advantages in open-source ecosystem and powerful statistical capabilities, R language has become a core tool for signal detection in pharmacovigilance. This study systematically reviews the progress and main methods in signal detection, compares the functions and application scenarios of relevant R packages, and discusses the advantages and challenges of using R in signal detection. The results show that the R ecosystem encompasses

DOI: 10.12173/j.issn.1005-0698.202603009

基金项目：国家药品监督管理局药物警戒研究与评价重点实验室 2024 年开放课题 (CDR2024X01001)；江苏省科教能力提升工程 (ZDXYS202210)

通信作者：周健，研究员，Email: jsjss277@aliyun.com

施雯慧，副研究员，Email: codeblue@163.com

<https://ywlbxb.whuznhmedj.com/>

methods such as classical frequency analysis, likelihood ratio tests, adaptive Lasso and data-driven pattern discovery, with each method differing in terms of data quality, interpretability, and automation. In the future, further efforts should focus on integrating multi-source real-world data, detecting drug interaction signals, applying explainable artificial intelligence, and building a full workflow of signal detection tailored to regulatory practices.

【Keywords】Pharmacovigilance; Signal detection; Adverse drug reaction; Adverse events; R language

药物警戒最早由法国学者于1974年提出^[1],其初始定义为对药物潜在风险保持持续警惕,在较长发展阶段中,其含义与药品不良反应监测概念基本一致。WHO于2002年发布的《药物警戒的重要性:药品安全监测》中明确药物警戒是与发现、评价、认识和预防各种不良反应或其他任何与可能的药物问题有关的科学和活动^[2]。2019年新修订的《中华人民共和国药品管理法》将药物警戒活动定义为:对药品不良反应及其他与用药有关的有害反应进行监测、识别、评估和控制的活动。总体上,我国定义和WHO定义的核心一致,药物警戒的研究边界已超传统不良反应监测范畴,覆盖药品研发、审批、上市后监管全生命周期的用药风险与药物损害;但我国定义是对WHO药物警戒理念的法定化承接,更强调监管闭环中的控制环节与主体责任落实。

药品在获批上市前,其安全性信息主要来源于样本量相对有限、研究时间较短的临床试验,罕见、严重的不良事件不易被发现,因此上市后持续的安全性监测成为药物警戒的核心环节^[1]。在药物警戒领域,信号被定义为来源于一个或多个信息源,提示药品与不良事件之间存在潜在新的因果关联或已知关联的新方面的信息^[3]。信号检测的目的,是利用定量或定性的方法,从海量数据库中高效地识别出值得进一步评估的“药品-不良事件”组合(drug-event combination, DEC),为后续进一步验证、评估与分级乃至制定风险控制措施提供依据,保障公众用药安全。

药物警戒中信号检测的实现主要依赖两类工具:一是Oracle Argus、ARISg等专业平台^[4],二是R、SAS、Python等编程语言。R语言是一款专为统计计算和绘图设计的编程语言^[5],经过近30年的发展,凭借其开源免费、软件包丰富等特性成为探索性研究和算法开发的理想工具。既往研究已广泛讨论药物警戒信号检测算法、自动化信号管理和真实世界数据应用^[6-9],部分研究涉及

特定软件包或单一方法的实现^[10-13],但以R语言生态为核心,系统比较不同药物警戒信号检测方法应用的研究尚不多见。基于此,本文综述了当前药物警戒信号检测方法的进展以及能够进行信号检测的R软件包,以期为相关研究者提供参考。

1 药物警戒信号检测方法及其进展

近年来,数据挖掘技术在药物警戒领域的应用日趋成熟,信号检测方法也随之不断迭代,已从基于频数统计的比例失衡分析方法逐步演进为基于复杂算法的数据驱动的检测方法。了解这些方法是评估和选择相应R软件包的基础,表1对药物警戒领域信号检测方法进行了总结。

早期药物警戒信号定量检测以频数法为核心,其遵循比例失衡分析的原理,基于经典四格表进行直观计算,可快速判断某不良事件在特定药品报告中的比例是否显著高于其在所有其他药品报告中的比例,满足被动监测中批量筛查可疑信号的需求。但在处理报告数极少的罕见事件时结果波动剧烈^[6],为此,贝叶斯方法被引入,借助先验分布进行信息平滑,提高了在大规模数据中检测结果的稳健性和特异度^[14]。随着对药物风险识别时效性的要求不断提高,药物警戒的目标从“发现关联”向“及时预警”转变,检测逻辑从针对累积数据的静态分析转向了序贯监测^[15],通过在数据积累过程中进行周期性统计分析,捕捉药物与不良事件的时序关联,及早发现潜在安全性信号,实现药品/疫苗上市后的早期风险预警。然而,传统的比例失衡分析方法基于四格表,其核心是处理双变量关联,难以有效控制药物警戒数据中混杂因素多、异质性强等问题。随之演进的回归模型^[7, 16-17]将信号检测推向多变量分析阶段,使研究者能够在控制已知混杂因素(如年龄、合并用药等)的前提下,更准确地评价药品与风险的因果关联强度。

表 1 药物警戒信号检测方法总结
Table 1. Summary of pharmacovigilance signal detection methods

方法类别	核心原理	主要算法	适用场景	特点
频数法	遵循比例失衡分析 (disproportionality analysis, DPA) 的原理, 比较某不良事件在特定药品报告中比例和在其他药品报告中比例的失衡程度, 以评估药品与不良事件之间的关联强度	比例报告比值比 (proportional reporting ratio, PRR)、报告比值比 (reporting odds ratio, ROR) 等	广泛应用于自发呈报系统 (spontaneous reporting system, SRS) 数据的信号检测	计算简便、易于理解, 但易受个别值的影响, 结果波动较大 ^[6]
贝叶斯方法	同归属于 DPA 的范畴, 引入了先验分布, 对不稳定的比例估计值进行平滑, 使结果更稳健	贝叶斯置信传播神经网络 (Bayesian confidence propagation neural network, BCPNN)、多项伽马泊松收缩模型 (multi-item Gamma Poisson Shrinker, MGPS) 等	适用于大规模 SRS 数据的稳健的信号检测	特异度较高, 误判概率小, 但灵敏度不足 ^[14]
序贯方法	在数据积累过程中, 通过周期性地、反复地统计分析, 尽早发现潜在安全信号	序列对称分析 (prescription sequence symmetry analysis, PSSA) ^[19] 、序贯概率比检验 (sequential probability ratio test, SPRT) ^[15] 和树状扫描统计量 (TreeScan) ^[20-21] 等	适用于上市后药品/疫苗不良反应的实时监测, 尤其适配纵向诊疗数据	实时性强, 适配低发生率不良反应监测, 对数据质量要求高, 实施复杂
回归模型	采用多变量分析来量化药品与不良事件之间的关联强度, 并控制年龄、合并用药等混杂因素的影响	Logistic 回归, 可直接输出优势比 (odds ratio, OR) 及其置信区间 ^[7, 16] ; LASSO (least absolute shrinkage and selection operator) ^[17] 正则化回归方法等	验证信号时控制混杂因素, 可基于常规诊疗信息数据研究药物风险因素	可控制混杂因素, 因果推断能力强, 但无法控制未测量的混杂因素
机器学习算法	从海量数据中自动学习复杂的模式和特征, 进行探索、预测或分类, 分为无监督机器学习与有监督机器学习	无监督机器学习: 关联规则法 (association rule mining, ARM) ^[22] 、聚类分析法 (cluster analysis) ^[23] 和主成分分析法 (principal component analysis, PCA) ^[24] 等。有监督机器学习: K-近邻算法 (K-nearest neighbor, KNN) ^[25] 、随机森林算法 (random forest, RF) ^[26] 和梯度提升机 (gradient boosting machine, GBM) 算法 ^[27] 等	适用于探索性分析、信号优先级排序、个例报告自动评估等	能够整合复杂的高维特征, 提高检出率和效率, 但部分复杂模型存在“黑箱”问题, 可解释性差
其他方法	如利用分子结构、基因组学等生物学数据, 从机制层面预测风险; 不依赖外部本体知识库, 采用数据驱动的阈值确定方法来识别潜在信号等	模式发现算法 ^[10] 、功能富集分析 (enrichment analysis) ^[28] 等	专注于更广义的模式发现或特定应用场景	前者易于从生物学或临床角度解释, 难以探知与现有知识体系无关的新发现; 后者统计关联可能缺乏明确的生物学或临床意义

由于临床实践中多药联用极为常见, 单纯考察单一 DEC 可能低估或误判药物相互作用风险。联合用药信号检测旨在识别多重 DEC, 以判定特定不良事件在多种药物联用时的报告频率是否显著高于预期。Qian 等^[16] 基于自发报告系统建立了药物相互作用信号自动检测系统, 采用 Logistic 回归、 Ω 收缩测度、加性模型和乘性模型等方法识别潜在药物-药物相互作用信号。

Dijkstra 等^[13] 指出, 合并用药背景可能产生“无辜旁观者效应”, 即与真正致害药物共同使用的无害药物被错误识别为信号。因此, 联合用药信号检测不仅需要统计模型支持, 还应结合用药时序、适应证、药理机制和临床合理性进行综合判断。

随着医药领域数字化、智能化应用, 电子健康记录、医保数据与可穿戴设备等多源信息融

合, 真实世界数据成为具有潜力的医疗大数据载体^[18], 信号检测数据来源也不再局限于传统自发报告系统。与自发报告系统相比, 电子健康记录和医保数据通常包含处方、诊断、检验、就诊时间和随访信息, 可用于构建暴露窗口和结局观察窗口, 因此更适合开展主动监测、时序分析和混杂控制。常用方法包括处方序列对称分析、序贯概率比检验和树状扫描统计量等^[15, 19-21]。此外, 随着机器学习^[22-27]和模式发现算法^[10]的应用, 信号检测进入了高度自动化和数据驱动阶段。这些方法能够自动捕捉海量高维数据中的复杂非线性模式, 甚至不依赖外部知识库即可识别异常关联, 极大提升了对未知风险的探索效率。此外还有一些专注于特定应用场景的方法和工具, 如功能富集分析^[28], 借鉴基因组学中的概念, 检验一组有共同特征(如作用机制相同)的药品是否与某些特定的不良事件(如某个MedDRA分类下的所有事件)显著相关。

综上所述, 信号检测方法的发展进程显示出从简单直观到兼顾稳健、从静态回顾到动态监测、从单变量关联到多因素混杂控制、从预设阈值驱动到数据驱动的特点, 为公众用药安全提供更精准、及时的技术保障。

2 R语言在药物警戒信号检测中的应用

2021年12月1日《药物警戒质量管理规范》正式施行, 要求上市许可持有人需建立持续的信号检测、评价和风险控制流程。Sun等^[29]通过对我国上市后药品信号检测现状调研发现, 我国药物警戒体系仍处于发展完善阶段, 药品不良反应报告系统已覆盖持有人、医疗机构和经营企业, 但系统自动化水平、信号检测和验证能力仍需提升。我国已建成国家、省两级药物警戒监测网络, 覆盖全国97.9%县级区域, 2025年全年报告量达272.8万份, 监测数据体量持续攀升^[30]。当前监测数据仍以自发报告为主, 正逐步融合各类真实世界数据。在我国药物警戒制度不断完善与监测数据持续积累的背景下, R语言具有良好的本土实践价值。一方面, R语言降低了复杂算法的应用门槛并提高了研究的可复现性。它将似然比检验、高维回归等理论复杂、计算量密集模型封装为标准、易用的函数接口, 打破了高级统计方法在实际应用中的瓶颈。另一方面, 得益于

其开源特性, 研究人员能够灵活组合不同R包的功能, 构建符合特定需求的、可重复的自动化分析流程, 对于建立规范的信号管理流程具有重要的实践意义。

本文以“pharmacovigilance”“signal detection”“disproportionality”等关键词在R语言的官方仓库CRAN和主流代码托管平台GitHub检索核心功能与药物警戒信号检测直接相关的软件包(不排除已归档或长期未维护的软件包)。经筛选, 截至2025年10月31日, 共纳入12个可用于药物警戒信号检测的软件包。为系统比较其特点, 本文从核心方法、应用领域、主要功能、开发活跃度及学术论文支持这五个维度建立比较框架, 并进行梳理总结, 比较结果见表2。

整体来看, 当前R软件包已实现了从经典到前沿算法的高度集成, 不仅涵盖了早期的比例失衡分析法(如PhViD、pvda实现的基础频数法), 也支持旨在提升稳健性的贝叶斯模型(如openEBGM、pvEBayes), 更进一步拓展至针对混杂因素控制的复杂回归模型(如adapt4pv的自适应Lasso)以及大数据驱动的模式发现算法(如MDDC)。在应用场景上, 除通用检测工具外, R软件包针对特定场景提供精准支持, 例如, AEenrich和sglr针对疫苗安全分别能够实现富集分析和序贯监测, mederrRank则聚焦于用药错误报告的风险排序。这种多样性使得用户能根据具体研究背景选择合适的工具。此外, 全球主流上市后监测数据库所依赖的算法^[31](如VigiBase使用的BCPNN, EudraVigilance使用的ROR/PRR, FAERS曾使用的MGPS), 在R的多个软件包中都有涉及, 为跨数据库的方法学比较和评估提供了便利。

3 代表性R软件包介绍

综合表2中各软件包的核心方法和开发活跃度, 本文选取了pvm、pvLRT、adapt4pv和MDDC作为代表性软件包, 分别从核心功能、分析流程(输入数据、核心函数、输出结果)、优点和局限性、应用实例等方面进行介绍。其中, pvm侧重多信号检测方法横向比较; pvLRT代表基于似然比检验的统计推断方法; adapt4pv为面向个体报告层面的高维药物暴露变量筛选方法; MDDC则不依赖外部医学术语, 是完全由数据驱动的新型检测方法。

表2 药物警戒信号检测相关 R 软件包综合比较
Table 2. Comprehensive comparison of R packages related to pharmacovigilance signal detection

名称	核心方法	应用领域	主要功能	开发活跃度 (最后更新时间)	学术论文 支持情况*
adapt4pv ^[32]	自适应 Lasso	药品不良反应监测数据	信号检测	2023-05-30	有 ^[11]
AEenrich ^[33]	功能富集分析	疫苗不良反应监测数据	信号检测	2021-11-01	有 ^[10]
MDDC ^[34]	模式发现算法	药品不良反应监测数据	数据预处理、信号检测、 结果可视化	2025-04-07	有 ^[28]
mds ^[35]	不含特定检测算法	医疗器械不良事件监测数据	数据预处理	2020-06-15	无
mederrRank ^[36]	贝叶斯分层模型	用药错误监测数据	信号检测	2023-09-05	有 ^[37]
openEBGM ^[36]	贝叶斯 DPA (EBGM/MGPS)	药品不良反应监测数据	数据预处理、信号检测、 结果可视化	2023-09-14	有 ^[38]
PhViD ^[39]	频率学&贝叶斯 DPA	药品不良反应监测数据	信号检测	已归档 (Archived)	无
pvda ^[40]	频率学&贝叶斯 DPA (支持 分组计算)	药品不良反应监测数据	信号检测	2025-01-17	无
pvEBayes ^[41]	经验贝叶斯	药品不良反应监测数据	数据预处理、信号检测、 结果可视化	2025-10-01	无
pvLRT ^[31]	似然比检验 (LRT)	药物/疫苗不良反应监测数据	数据预处理、信号检测、 结果可视化	2023-03-06	有 ^[12]
pvm ^[42]	频率学&贝叶斯 DPA、Lasso 回归等 (侧重于方法比较)	药品不良反应监测数据	信号检测	2025-02-11	有 ^[13]
sglr ^[43]	序列广义似然比检验 (sequential GLR)	疫苗不良反应监测数据 (上市前试验)	信号检测	2022-04-19	有 ^[44]

注：*学术论文支持指有该软件包的开发或使用说明的相关学术论文发表。

3.1 pvm：全面的常用信号检测方法比较工具

pvm 是一个集成了多种方法的综合工具箱，适用于方法学比较、敏感性分析和候选信号稳健性评价，主要面向自发报告系统数据。该包涵盖了以下方法：频数法，如 ROR、PRR、卡方检验、Fisher 精确检验、二项似然比检验、泊松均值检验；贝叶斯方法，如 BCPNN 的两种变体和 MGPS，以及 Lasso 回归^[42]等。用户前期需完成药品名称规范化、不良事件术语统一、重复报告处理及 DEC 计数。通过输入标准 DEC 的 2x2 列联表^[13]，根据不同方法调用相应的函数来计算不同测度指标，可比较其排序、评分或变量筛选结果。pvm 的优势是便于在实际工作中进行方法学比较、敏感性分析以及交叉验证研究。但同时，它又受限于无法自动完成数据清洗和方法筛选，需用户凭借专业知识解读并比较不同方法的结果。

应用示例：Dijkstra 等^[13]围绕自发报告系统中常见的共用药偏倚开展模拟研究，比较 ROR、PRR、BCPNN 和 LASSO 等方法。结果提示，不同方法对小样本、效应大小和合并用药造成的“无

辜旁观者效应”敏感性不同。pvm 适合用于候选信号的稳健性评估，当某一组合被单一方法提示为信号时，可进一步考察其在多种方法下结果是否一致，再决定是否进入后续验证。

3.2 pvLRT：基于似然比检验的药物安全评价工具

pvLRT 可将异常报告识别置于假设检验框架下，通过比较零假设模型和备择模型发现报告数升高的 DEC。该包实现普通泊松模型和零膨胀泊松模型下的似然比检验方法，并支持单个或多个药品同时检验^[12, 31]，适合需要减少阈值任意性、强调统计推断或与传统 DPA 结果交叉验证的研究。pvLRT 的核心分析函数为 pvlrt()，该函数接收预处理好的 DEC 列联表数据。用户可以通过关键参数 zi_prob 来选择使用标准的泊松模型 (zi_prob=0) 或是由数据驱动估计的零膨胀泊松模型 (zi_prob=NULL)。分析完成后，可用 summary() 提取详细的检验统计量和 P 值；也可用泛型函数 plot() 生成气泡图、条形图和热图；还可用 AIC() 和 BIC() 对不同模型的拟合优度进行比较^[31]。pvLRT 的优势在于提供较严格的统计推断框架，可用 P 值评价显著性，并在多重检验框架下控制 I 类错

误和错误发现率。主要局限性为结果高度依赖于输入列联表的数据质量和模型设定,若原始数据存在重复、编码不一致或筛选不合理,结果可能偏倚。此外,pvLRT基于汇总表,不能像个体级回归模型那样直接纳入年龄、性别、适应证和合并用药等协变量。

应用示例:Chakraborty等^[12]使用FAERS数据库中他汀类药物的不良事件数据开展分析,在似然比检验框架下比较多个同类药品与不良事件之间的异常报告证据,并通过统计量、*P*值和图形展示结果。该示例解决了在不完全依赖经验阈值的情况下,对候选信号进行统计检验的问题,因此,pvLRT适用于需要较严格推断、存在较多零计数,或需要比较同类药品信号差异的场景。

3.3 adapt4pv: 基于自适应Lasso的变量选择工具

adapt4pv面向药物警戒中的高维变量筛选问题,主要实现自适应Lasso、标准Lasso、类别不平衡子抽样Lasso及倾向评分相关方法^[11, 32]。与比例失衡分析法不同,它以个体报告为分析单位,将目标不良事件作为二分类结局,将多个药物暴露作为候选变量。该包适用于多药暴露、高维药物协变量和目标事件明确的信号筛选任务。adapt4pv通常要求输入个体报告层面的药物暴露矩阵 x 和二分类结局 y 。 x 为 $N \times p$ 二值药物暴露矩阵, N 表示报告数量, p 表示候选药物数量^[11], y 表示目标不良事件是否出现。核心函数包括adapt_bic()、adapt_cisl()、lasso_bic()和cisl()等。输出结果主要为被筛选出的候选药物变量、回归系数、选择准则或校正后的*P*值,即变量筛选和信号优先级排序结果。adapt4pv的优点是能够处理复杂的相互作用和高维数据,通过引入数据驱动的自适应权重,克服了标准Lasso回归中所有变量被同等惩罚的局限,从而能够更准确地识别真实的信号,为解决传统DPA方法难以处理的混杂问题提供了新的可能性。主要局限性在于,作为一个多阶段的回归模型,其计算复杂度高于传统DPA方法,结果的可解释性不如传统方法直观;对计算资源的要求更高,且方法的稳定性和普适性仍需更多验证。

应用示例:Courtois等^[11]基于法国药物警戒数据库,以个体报告作为分析单元,以目标不良事件(药物性肝损伤)是否发生为结局,以多个

药物暴露为候选解释变量,比较自适应Lasso、标准Lasso、类别不平衡子抽样Lasso和高维倾向评分方法,目的是在大量候选药物中筛选更值得进一步关注的变量。总之,adapt4pv适合用于能获得个体报告数据、候选药物较多且共用药复杂的研究,其结果可作为后续评价优先级的依据。

3.4 MDDC: 不依赖医学术语的模式发现算法工具

MDDC将信号检测视为观测计数相对于期望计数异常升高的模式发现问题。与依赖MedDRA等医学本体或预设层级结构的方法不同,MDDC强调数据驱动阈值和不良事件相关性分析。它适合探索性模式发现,尤其适用于希望减少对外部本体和固定经验阈值依赖的场景。MDDC要求输入药品-不良事件列联表,使用前需规范药品名称和不良事件术语,并检查列联表格式。主要函数包括check_and_fix_contin_table()、mddc_boxplot()、mddc_mc()、report_drug_AE_pairs()和plot_heatmap()等。输出包括异常组合、*P*值或校正*P*值、观测计数、期望计数和标准化残差等。MDDC的优势是独立于外部知识库且阈值由数据驱动,并结合不良事件相关性辅助评价候选信号^[10]。其局限性是基于汇总列联表数据,无法直接控制个体层面的年龄、性别、适应证和合并用药等混杂因素;当列联表中零计数单元格比例较高时,错误发现率可能升高。

应用示例:Liu等^[10]基于FAERS数据库提取 β 受体阻滞剂相关数据,开展MDDC方法的应用研究。研究者先根据观测计数与期望计数的偏离识别异常单元格,再结合不良事件相关性筛选候选信号。其研究目的不是验证单个药物已知风险,而是在缺乏明确的医学本体或不希望依赖固定阈值时,如何从较大列联表中发现异常报告模式。因此,MDDC适合用于大规模自发报告数据的早期探索,但其结果仍需结合病例质量、药理机制和外部证据进一步验证。

4 R语言实际应用挑战与未来发展方向

尽管R语言是非常强大的统计工具,但在药物警戒实践中的应用仍面临诸多挑战。首先体现在数据层面。SRS数据固有的问题,如漏报、信息不完整、报告偏倚、混杂因素以及缺乏分母,是开展分析及结果解释的基础性障碍之一^[45];数

据质量也是一大挑战,新冠肺炎疫情期间的药物警戒实践凸显了数据量激增、区域报告差异、个案报告质量不一、重复报告以及 MedDRA 编码不一致等问题^[46];数据清洗较为复杂,当前许多 R 包要求输入预处理好的列联表,这需要用户具备强大的数据清洗、标准化和整合能力。其次体现在方法层面。信号检测方法没有金标准, DPA 简单但易受混杂影响, LRT 在控制错误率上有优势但依赖模型假设。实践中方法的选择、解释与验证面临极大困难,定量信号往往需要结合定性审查(病例评估、临床判断)才能进行有效验证和优先级排序^[29]。机器学习算法虽然能处理高维数据和交互作用,但常被视为“黑箱”,其决策过程缺乏透明度,引发了对可解释性和临床适用性的担忧^[23]。值得注意的是,美国 FDA 与欧洲 EMA 等机构正积极制定 AI 药物安全监测的监管框架,要求其决策过程透明、可重复且需人类监督^[47]。

结合当前药物警戒领域的技术进展和使用需求,未来信号检测相关 R 软件包的发展方向可聚焦以下方面:一是多源数据的融合利用,结合自发报告系统、电子健康记录、保险理赔、社交媒体、文献等多源数据,有望提供更全面、及时的信号证据,开发能够处理异构数据、进行融合分析的 R 包将是重要方向^[9]。二是自动化、智能化与实时监测,机器学习和人工智能将在信号管理流程的自动化中扮演核心角色,未来的软件应朝着更智能的信号优先级排序、接近实时的监测发展^[48]。三是可解释的因果推断,当前信号检测主要关注关联性而非因果性,未来亟须将因果推断框架整合到信号检测工具中,用于信号评价的分析方法包括传统回归模型(Logistic 回归、Poisson 回归等),还可通过引入倾向性评分、疾病风险评分、工具变量等对统计模型进行调整,从而进行因果效应估计。发展可解释性 AI 技术能够打开“黑箱”,增加用户对复杂模型输出的理解和信任^[49]。四是数据分析前端和后端的生态链完善,在提供更易用的数据预处理模块、更强大的交互式可视化和解释性工具方面, R 语言仍具备广阔的开发潜力。

5 结语

本文系统梳理了 R 语言在药物警戒信号检测

中的应用,归纳其涵盖的经典频数法、似然比检验、自适应 Lasso 及数据驱动模式发现等多类方法,可支持从数据预处理到信号验证的全流程实践。目前 R 语言在药物警戒的实际落地仍有多重阻碍,真实世界数据难以清洗、检测方法缺乏金标准、模型可解释性难以满足临床审查需求等,大幅限制了数据价值的转化。未来 R 语言将向多源数据融合、实时智能监测、可解释因果推断、完善分析生态等方向发展,从而实现药物警戒从关联信号识别向因果效应评估的升级。通过规范数据治理、优化算法框架、配套监管机制,多方协同破解现存痛点,将有利于充分发挥 R 语言生态优势,精准识别用药安全风险,为药品安全监管提供坚实的技术支撑。

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 王斌. 药物警戒与合理用药[M]. 上海: 上海交通大学出版社, 2025.
- 2 World Health Organization. The importance of pharmacovigilance [EB/OL]. (2022-01-01) [2025-11-14]. <https://www.who.int/publications/i/item/10665-42493>.
- 3 CIOMS Working Group. Practical aspects of signal detection in pharmacovigilance: report of CIOMS Working Group VIII[M]. Geneva: CIOMS, 2010.
- 4 Gill SK, Christopher AF, Gupta V, et al. Emerging role of bioinformatics tools and software in evolution of clinical research[J]. Perspect Clin Res, 2016, 7(3): 115-322. DOI: 10.4103/2229-3485.184782.
- 5 R. The R project for statistical computing[EB/OL]. [2025-11-14]. <https://www.r-project.org/>
- 6 Candore G, Juhlin K, Manlik K, et al. Comparison of statistical signal detection methods within and across spontaneous reporting databases[J]. Drug Saf, 2015, 38(6): 577-587. DOI: 10.1007/s40264-015-0289-5.
- 7 Harpaz R, Dumouchel W, Lependu P, et al. Performance of pharmacovigilance signal-detection algorithms for the FDA Adverse Event Reporting System[J]. Clin Pharmacol Ther, 2013, 93(6): 539-546. DOI: 10.1038/clpt.2013.24.
- 8 周瑞珊, 卢佩雯, 陈君恒, 等. 药品不良反应数据挖掘技术在药物警戒中的应用[J]. 中国现代应用药学, 2024, 41(6): 864-870. [Zhou RS, Lu PW, Chen JH, et al. Application of adverse drug reaction of data mining in pharmacovigilance[J]. Chinese Journal of Modern Applied Pharmacy, 2024, 41(6): 864-870.] DOI: 10.13748/j.cnki.issn1007-7693.20224098.
- 9 Lavertu A, Vora B, Giacomini KM, et al. A new era in

- pharmacovigilance: toward real-world data and digital monitoring[J]. *Clin Pharmacol Ther*, 2021, 109(5): 1197–1202. DOI: [10.1002/cpt.2172](https://doi.org/10.1002/cpt.2172).
- 10 Liu A, Mukhopadhyay R, Markatou M. MDDC: an R and Python package for adverse event identification in pharmacovigilance data[J]. *Sci Rep*, 2025, 15(1): 21317. DOI: [10.1038/s41598-025-00635-w](https://doi.org/10.1038/s41598-025-00635-w).
 - 11 Courtois É, Tubert-Bitter P, Ahmed I. New adaptive Lasso approaches for variable selection in automated pharmacovigilance signal detection[J]. *BMC Med Res Methodol*, 2021, 21(1): 271. DOI: [10.1186/s12874-021-01450-3](https://doi.org/10.1186/s12874-021-01450-3).
 - 12 Chakraborty S, Markatou M, Ball R. Likelihood ratio test-based drug safety assessment using R package pvLRT[J]. *R J*, 2023, 15(1): 101–121. DOI: [10.32614/RJ-2023-027](https://doi.org/10.32614/RJ-2023-027).
 - 13 Dijkstra L, Garling M, Foraita R, et al. Adverse drug reaction or innocent bystander? A systematic comparison of statistical discovery methods for spontaneous reporting systems[J]. *Pharmacoepidemiol Drug Saf*, 2020, 29(4): 396–403. DOI: [10.1002/pds.4970](https://doi.org/10.1002/pds.4970).
 - 14 向永洋, 谢雁鸣, 易丹辉. 药物警戒数据挖掘方法比较及其在中成药预警中的应用[J]. *中国中药杂志*, 2011, 36(20): 2831–2835. [Xiang YY, Xie YM, Yi DH. Pharmacovigilance of major pharmaceutical innovation[J]. *China Journal of Chinese Materia Medica*, 2011, 36(20): 2831–2835.] DOI: [10.4268/cjcm2011020](https://doi.org/10.4268/cjcm2011020).
 - 15 张新估, 郭晓晶, 叶小飞, 等. 药品不良反应监测中的序贯概率比检验方法[J]. *中国药物警戒*, 2018, 15(4): 221–225. [Zhang XJ, Guo XJ, Ye XF, et al. Sequential probability ratio analysis in post-market drug safety surveillance[J]. *Chinese Journal of Pharmacovigilance*, 2018, 15(4): 221–225.] DOI: [10.3969/j.issn.1672-8629.2018.04.007](https://doi.org/10.3969/j.issn.1672-8629.2018.04.007).
 - 16 Qian YF, Ye XF, Du WM, et al. A computerized system for detecting signals due to drug-drug interactions in spontaneous reporting systems[J]. *Br J Clin Pharmacol*, 2010, 69(1): 67–73. DOI: [10.1111/j.1365-2125.2009.03557.x](https://doi.org/10.1111/j.1365-2125.2009.03557.x).
 - 17 Ellis DE, Hubbard RA, Willis AW, et al. Comparing LASSO and random forest models for predicting neurological dysfunction among fluoroquinolone users[J]. *Pharmacoepidemiol Drug Saf*, 2022, 31(4): 393–403. DOI: [10.1002/pds.5391](https://doi.org/10.1002/pds.5391).
 - 18 郭美玲, 柴克燕, 刘蕴资, 等. 数据挖掘方法在真实世界研究中的应用进展[J]. *药物流行病学杂志*, 2026, 35(4): 451–461. [Guo ML, Chai KY, Liu YZ, et al. Application progress of data mining methods in real-world study[J]. *Chinese Journal of Pharmacoepidemiology*, 2026, 35(4): 451–461.] DOI: [10.12173/j.issn.1005-0698.202511032](https://doi.org/10.12173/j.issn.1005-0698.202511032).
 - 19 贺冰洁, 张明轩, 詹思延. 药物流行病学中处方序列对称分析研究的系统综述[J]. *中华流行病学杂志*, 2021, 42(9): 1641–1649. [He BJ, Zhang MX, Zhan SY. Prescription sequence symmetry analysis in pharmacoepidemiology: a systematic review[J]. *Chinese Journal of Epidemiology*, 2021, 42(9): 1641–1649.] DOI: [10.3760/cma.j.cn112338-20201208-01386](https://doi.org/10.3760/cma.j.cn112338-20201208-01386).
 - 20 孙一鑫, 王森, 杨明芳, 等. 树状扫描统计量用于药品和疫苗安全性监测的综述[J]. *中华流行病学杂志*, 2021, 42(7): 1286–1291. [Sun YX, Wang M, Yang MF, et al. Review on tree-based scan statistic in drug and vaccine safety monitoring[J]. *Chinese Journal of Epidemiology*, 2021, 42(7): 1286–1291.] DOI: [10.3760/cma.j.cn112338-20201103-01297](https://doi.org/10.3760/cma.j.cn112338-20201103-01297).
 - 21 Kulldorff M, Dashevsky I, Abery TR, et al. Drug safety data mining with a tree-based scan statistic[J]. *Pharmacoepidemiol Drug Saf*, 2013, 22(5): 517–523. DOI: [10.1002/pds.3423](https://doi.org/10.1002/pds.3423).
 - 22 Jo HG, Seo J, Choi S, et al. East Asian herbal medicine to reduce primary pain and adverse events in cancer patients: a systematic review and meta-analysis with association rule mining to identify core herb combination[J]. *Front Pharmacol*, 2022, 12: 800571. DOI: [10.3389/fphar.2021.800571](https://doi.org/10.3389/fphar.2021.800571).
 - 23 Ward D, Thorsen NM, Frisch M, et al. A cluster analysis of serious adverse event reports after human papillomavirus (HPV) vaccination in Danish girls and young women, September 2009 to August 2017[J]. *Euro Surveill*, 2019, 24(19): 1800380. DOI: [10.2807/1560-7917.ES.2019.24.19.1800380](https://doi.org/10.2807/1560-7917.ES.2019.24.19.1800380).
 - 24 徐瑾, 钟长鸣, 周鹃, 等. 使用主成分分析的药品不良反应风险信号挖掘方法探索[J]. *中国医院药学杂志*, 2017, 37(1): 13–16. [Xu J, Zhong CM, Zhou J, et al. Exploration of principal component analysis for discovering the signal of adverse drug reactions[J]. *Chinese Journal of Hospital Pharmacy*, 2017, 37(1): 13–16.] DOI: [10.13286/j.cnki.chinhosp-pharmacy.2017.01.03](https://doi.org/10.13286/j.cnki.chinhosp-pharmacy.2017.01.03).
 - 25 陈泉. 基于机器学习的自发呈报系统中药品不良反应信号检测研究[D]. 中国人民解放军海军军医大学, 2024. [Chen X. Signal detection of adverse drug reaction in spontaneous reporting system based on machine learning[D]. Naval Medical University, 2024.
 - 26 Pham M, Cheng F, Ramachandran K. A comparison study of algorithms to detect drug-adverse event associations: frequentist, Bayesian, and machine-learning approaches[J]. *Drug Saf*, 2019, 42(6): 743–750. DOI: [10.1007/s40264-018-00792-0](https://doi.org/10.1007/s40264-018-00792-0).
 - 27 Jang MG, Cha S, Kim S, et al. Application of tree-based machine learning classification methods to detect signals of fluoroquinolones using the Korea Adverse Event Reporting System (KAERS) database[J]. *Expert Opin Drug Saf*, 2023, 22(7): 1–8. DOI: [10.1080/14740338.2023.2181341](https://doi.org/10.1080/14740338.2023.2181341).
 - 28 Li SR, Zhao LL. Vaccine adverse event enrichment tests[J]. *Stat Med*, 2021, 40(19): 4269–4278. DOI: [10.1002/sim.9027](https://doi.org/10.1002/sim.9027).
 - 29 Sun X, Zhang Y, Qian D, et al. An investigation into the current status of post-marketing signal detection for pharmaceutical products in China: an interview-based study[J]. *J Chin Pharm Sci*, 2025, 34(2): 175–184. DOI: [10.5246/jcps.2025.02.014](https://doi.org/10.5246/jcps.2025.02.014).
 - 30 国家药品不良反应监测中心. 国家药品不良反应监测年度报告(2025年)[EB/OL]. (2026-04-27) [2026-05-26]. https://www.cdr-adr.org.cn/center_news/202604/t20260427_51496.html.
 - 31 Chakraborty S, Markatou M, Liu A. pvLRT: likelihood ratio test-based approaches to pharmacovigilance[EB/OL]. (2023-03-06) [2025-10-24]. <https://cran.r-project.org/web/packages/pvLRT/index.html>.
 - 32 Courtois E, Ahmed I, Perdrey H. adapt4pv: adaptive approaches for signal detection in pharmacovigilance[EB/OL]. (2023-05-30) [2025-10-16]. <https://CRAN.R-project.org/package=adapt4pv>.
 - 33 Li SR, Chen HF, Zhao LL, et al. AEEnrich: adverse event

- enrichment tests[EB/OL]. (2021-11-01) [2025-10-11]. <https://CRAN.R-project.org/package=AEenrich>.
- 34 Liu A, Mukhopadhyay R, Markatou MI. MDDC: modified detecting deviating cells algorithm in pharmacovigilance[EB/OL]. (2025-04-07) [2025-10-24]. <https://cran.wustl.edu/web/packages/MDDC/index.html>.
- 35 Chung G. mds: medical devices surveillance[EB/OL]. (2020-06-15) [2025-10-15]. <https://CRAN.R-project.org/package=mds>.
- 36 Ihrie J, Canida T, Ahmed I, et al. openEBGM: EBGM disproportionality scores for adverse event data mining[EB/OL]. (2023-09-14) [2025-10-24]. <https://cran.r-project.org/web/packages/openEBGM/index.html>.
- 37 Venturini S, Franklin JM, Morlock L, et al. Random effects models for identifying the most harmful medication errors in a large, voluntary reporting database[J]. *Ann Appl Stat*, 2017, 11(2): 504-526. DOI: [10.1214/16-AOAS974](https://doi.org/10.1214/16-AOAS974).
- 38 Canida T, Ihrie J. openEBGM: an R implementation of the Gamma-Poisson Shrinker data mining model[J]. *R J*, 2017, 9(2): 499-519. DOI: [10.32614/RJ-2017-063](https://doi.org/10.32614/RJ-2017-063).
- 39 PhViD: pharmacovigilance signal detection[EB/OL]. [2025-10-24]. <https://cran.r-project.org/web/packages/PhViD/index.html>.
- 40 Gauffin O, Fusaroli M. pvda: disproportionality functions for pharmacovigilance[EB/OL]. (2025-01-17) [2025-10-24]. <https://cran.r-project.org/web/packages/pvda/index.html>.
- 41 Tan Y, Markatou M, Chakraborty S, et al. pvEBayes: empirical Bayes methods for pharmacovigilance empirical[EB/OL]. (2025-10-01) [2025-10-24]. <https://CRAN.R-project.org/package=pvEBayes>.
- 42 Dijkstra L, Garling M, Foraita R, et al. bips-hb/pvm[EB/OL]. [2025-10-24]. <https://github.com/bips-hb/pvm>.
- 43 Narasimhan. sgldr: Sequential generalized likelihood ratio decision boundaries proposed by Shih, Lai, Heyse and Chen (2010, <doi:10.1002/Sim.4036>)[EB/OL]. (2022-04-19) [2025-10-24]. <https://cran.r-project.org/web/packages/sgldr/index.html>.
- 44 Shih MC, Lai TL, Heyse JF, et al. Sequential generalized likelihood ratio tests for vaccine safety evaluation[J]. *Stat Med*, 2010, 29(26): 2698-2708. DOI: [10.1002/sim.4036](https://doi.org/10.1002/sim.4036).
- 45 王蒙. 贝叶斯工具变量方法研究及其在药品不良反应主动监测中的应用[D]. 中国人民解放军海军军医大学, 2019. [Wang M. Research and application of Bayesian instrumental variable in active surveillance of adverse drug reaction[D]. Naval Medical University, 2019.
- 46 Lino MM, Mather S, Trani M, et al. Challenges and innovations in pharmacovigilance and signal management during the COVID-19 pandemic: an industry perspective[J]. *Vaccines (Basel)*, 2025, 13(5): 481. DOI: [10.3390/vaccines13050481](https://doi.org/10.3390/vaccines13050481).
- 47 Nagar A, Gobburu J, Chakravarty A. Artificial intelligence in pharmacovigilance: advancing drug safety monitoring and regulatory integration[J]. *Ther Adv Drug Saf*, 2025, 16: 20420986251361435. DOI: [10.1177/20420986251361435](https://doi.org/10.1177/20420986251361435).
- 48 Barbieri MA, Battini V, Carnovale C, et al. Artificial intelligence in pharmacovigilance signal management: a review of tools, implementations, research, and regulatory landscape[J]. *Expert Opin Drug Saf*, 2026, 25(2): 207-222. DOI: [10.1080/14740338.2025.2545926](https://doi.org/10.1080/14740338.2025.2545926).
- 49 Ferreira-Da-Silva R, Cruz-Correia R, Ribeiro I. Beyond black boxes: using explainable causal artificial intelligence to separate signal from noise in pharmacovigilance[J]. *Int J Clin Pharm*, 2026, 48(2): 677-681. DOI: [10.1007/s11096-025-02004-z](https://doi.org/10.1007/s11096-025-02004-z).

收稿日期: 2026 年 03 月 03 日 修回日期: 2026 年 05 月 26 日
 本文编辑: 李绪辉 桂裕亮