

药品不良反应数据挖掘技术在药物警戒中的应用

周瑞珊¹, 卢佩雯¹, 陈君恒¹, 石艺杨¹, 何明秀¹, 韩芳芳^{1,2*}, 蔡永铭^{1,2,3*} (1.广东药科大学, 广州 510006; 2.国家药监局药物警戒技术与评价重点实验室, 广州 510006; 3.广东省中医药精准医学大数据工程技术研究中心, 广州 510006)

摘要: 随着信息技术的发展, 医药电子数据海量增长, 药品不良事件报告大幅增加, 给药物警戒研究带来了巨大的挑战。而数据挖掘技术可以自动从真实世界数据中提取药品不良反应风险信号。因此, 对海量不良事件报告数据进行高效数据挖掘是实现药品不良反应自动检测的必要措施。本研究通过介绍当前主要的大型药品不良事件报告数据库和相关数据挖掘方法, 对药品不良反应数据挖掘技术在药物警戒中的应用及其局限性进行综述, 为药物警戒相关机构和科研人员提供参考。

关键词: 数据挖掘; 不良事件报告; 药物警戒; 药品不良反应; 自动检测

中图分类号: R969.3 文献标志码: A 文章编号: 1007-7693(2024)06-0864-07

DOI: 10.13748/j.cnki.issn1007-7693.20224098

引用本文: 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.

Application of Adverse Drug Reaction of Data Mining in Pharmacovigilance

ZHOU Ruishan¹, LU Peiwen¹, CHEN Junheng¹, SHI Yiyang¹, HE Mingxiu¹, HAN Fangfang^{1,2*}, CAI Yongming^{1,2,3*} (1.Guangdong Pharmaceutical University, Guangzhou 510006, China; 2.NMPA Key Laboratory for Technology Research and Evaluation of Pharmacovigilance, Guangzhou 510006, China; 3.Guangdong Provincial Traditional Chinese Medicine Precision Medicine Big Data Engineering Technology Research Center, Guangzhou 510006, China)

ABSTRACT: With the development of information technology, the massive growth of pharmaceutical electronic data and the significant increase in the reports of drug adverse event reports have brought great challenges to pharmacovigilance research. Data mining techniques can automatically extract the risk signals of adverse drug reaction from real-world data. Therefore, efficient data mining of massive adverse event reporting is a necessary measure to realize the automatic detection of adverse drug reactions. By introducing the current major large-scale adverse drug event reporting databases and related data mining methods, this study reviews the application and limitations of adverse drug reaction data mining technology in pharmacovigilance, which provides reference for pharmacovigilance-related institutions and researchers.

KEYWORDS: data mining; adverse event report; pharmacovigilance; adverse drug reaction; automatic detection

21 世纪以来, 信息科学迅猛发展, 数据产生、收集能力及方式逐渐完善, 数据量大幅增加。大数据是指所涉及的资料数据规模巨大, 以至于使用传统的数据库管理工具难以提取、管理、处理和组织, 影响着各行各业。在整个医药卫生行业中, 各种类型的医药电子数据都在高速生成, 药品不良事件报告数据也大幅增加。

药品不良事件 (adverse event, AE) 是指患者在接受药品治疗过程中发生的不良临床事件。药品不良反应 (adverse drug reaction, ADR) 是指患者在医护人员正确指导下及正常用法用量下接受正规合格药品治疗后, 出现与用药目的无关的有害反应。虽然 ADR 与 AE 都是接受药品治

疗后出现的各种与治疗目的无关的反应或事件, 但两者存在一定的区别。ADR 与药品治疗有因果关联, 而 AE 不一定与药品治疗有因果关联 (即可能由其他原因导致 AE 发生)。药品 AE 包括了 ADR、品质标准问题、用药错误和滥用等。因此, 有必要从海量的药品 AE 报告中发现与药品治疗有因果关联的 ADR, 以评估药品的安全性。

真实世界数据 (real-world data, RWD) 是指在真实世界环境下, 日常收集的各种与健康状况、医疗诊疗及康复保健相关的临床数据。但并不是所有的 RWD 经过整理分析, 就能成为真实世界证据 (real-world evidence, RWE)。RWE 的前提条件是 RWD 满足适用性, 即应基于特定的临床研究目

基金项目: 广东省药品监督管理局科技创新项目 (2022ZDZ06)

作者简介: 周瑞珊, 女, 硕士生 E-mail: 1226133085@qq.com
蔡永铭, 男, 博士, 教授 E-mail: cym@gdpu.edu.cn

*通信作者: 韩芳芳, 女, 博士, 讲师 E-mail: hanff@gdpu.edu.cn

的和监管决策用途^[1]。真实世界研究 (real-world study, RWS) 是指在真实世界环境下, 基于临床研究问题及目的, 收集 RWD 或基于 RWD 衍生出来的汇总数据, 经过数据清洗、整理、分析, 获取药品的使用情况及潜在风险-收益的临床证据 (即 RWE) 的研究过程。2021 年 4 月, 国家药品监督管理局药品审评中心为进一步指导和规范申办者如何利用 RWD 产生 RWE 以支持药物研发, 发布了《用于产生真实世界证据的真实世界数据指导原则 (试行)》^[1]。对于大量的 RWD, 可以通过数据挖掘 (data mining, DM) 技术, 充分利用数据资源, 挖掘出有价值的信息。

DM 是指从大量有噪声的数据中, 通过算法自动搜索及提取潜隐的、未知的、但有用的信息及知识的过程^[2]。DM 是做 RWS 的一个良好工具。相较于传统方法, DM 是在假设不明确的前提条件下, 处理和提取 RWD, 获得事先未知、有效且实用的信息和知识^[2]。DM 过程包括了数据采集、数据清洗和预处理、DM、数据解释和利用等过程^[3]。ADR 数据挖掘是药物警戒的核心方法之一^[4]。世界卫生组织 (World Health Organization, WHO) 将药物警戒定义为“检测、评价、认识及预防各种 ADR 或其他一切与药物/疫苗相关问题的学科和活动”。当一个或多个来源的观察性或试验性报告提示某药品与某些 AE 之间存在新的潜在关联, 或某些已知关联 AE 的新信息, 该信息被认为需要进一步验证, 则称为药物警戒风险信号^[5]。药物警戒的关键包括风险信号的产生与分析, 即如何收集及分析上市后药品的相关数据, 并得出说服力强的结论^[6]。而 DM 技术可以检测并识别风险信号。

本研究的主要目的是介绍当前主要的大型 AE 报告数据库, ADR 数据挖掘方法, ADR 数据挖掘在药物警戒中的应用及其局限性, 为药物警戒相关科研人员提供参考。

1 AE 报告数据库

WHO 乌普萨拉监测中心 (Uppsala Monitoring Centre, UMC) 是 WHO 药物监测国际合作中心, 拥有全球最具规模的 AE 数据库。自 1968 年以来, 至今有 150 多个国家的药物警戒中心与 UMC 合作, 定期提交 AE 报告数据。中国于 1998 年成为成员国, 开始 AE 监测数据的上报^[7]。VigiBase 是报告药品潜在 ADR 的全球数据库, 是

WHO 药物警戒系统的核心。药品 AE 信息以结构化形式储存在 VigiBase, 方便用户检索和分析数据, 发现潜在的药品风险信号^[8]。VigiAccess 于 2015 年由 WHO 推出, 旨在向公众提供 VigiBase 数据库中 AE 报告信息, 可从 <http://www.vigiaccess.org> 访问。

美国食品药品监督管理局 (U.S. Food and Drug Administration, FDA) AE 报告系统 (FDA Adverse Event Reporting System, FAERS) 是世界上最知名的 AE 报告数据库之一。美国药品监测工作由 FDA 的药品审评与研究中心负责, 以主动和被动相结合的方式监测药品 AE, 主要收集药品制造商、医疗卫生专业人员及消费者上报的 AE 数据。FAERS 数据库包括了患者信息、用药错误信息及结局等, 每季度更新一次数据库。原来的系统始于 1969 年, 于 1997 年进行重大修订后, 报告量显著增加, 一直沿用至今。FAERS 数据集可从 <https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html> 下载。

EudraVigilance 是欧洲药品管理局自 2001 年以来维护的欧洲级 AE 数据库。报告包含有关患者的信息, 包括他的病史, 服用的药品, 所给予的治疗和结局, 及相关的 ADR 或可疑反应的详细信息。各利益相关方, 包括药品上市许可持有人、监管机构、科研人员、医疗卫生人员和公众, 根据该数据库的访问政策具有一定的访问权限。不同利益方拥有的访问权限不同, 以便该数据库向不同利益方提供尽可能多的信息, 同时也能保护患者隐私数据安全^[9]。数据库可从 <http://www.adrreports.eu> 访问。

自 1999 年以来, 中国逐步建立起覆盖全国的国家 ADR 监测系统, 并于 2010 年开始, 国家药品监督管理局每年发布上一年度国家 ADR 监测年度报告 (简称“年度报告”)^[10]。截止到 2021 年, 数据库累计收到 ADR/AE 报告 1883 万份^[11]。ADR/AE 报告是指发生于中国境内单个患者的药品 AE 报告, 由医疗机构、药品生产及经营企业等单位上报^[10]。该数据库也被中国学者用于 DM 药物警戒研究。年度报告可从 <https://www.nmpa.gov.cn/index.html> 获取。

2 ADR 数据挖掘常用方法

ADR 数据挖掘技术通过 DM 方法对 AE 报告数据库中的药品 AE 进行信号检测, 识别潜在的

ADR 风险信号^[12]。ADR 数据挖掘方法的研究最早见于 1974 年, 由 Finney 提出“反应比例信号”概念^[13]。经过多年的研究发展, 目前在药物警戒领域中主要采用比例失衡分析 (disproportionality analyses, DPA) 法。

DPA 是药品 AE 报告系统数据库广泛应用的药物警戒风险信号检测方法。DPA 的原理是利用经典四格表法 (表 1), 比较目标药品-AE 与其他药品-AE 比例失衡程度, 以评估药品与 AE 之间的关联强度。比例失衡是指某一药品-AE 组合频率超出了整个数据库的背景频率, 或达到了预先设定的判别标准, 提示该药品与该 AE 存在的统计学关联, 且这种关联不是由偶然因素引起, 可能为 ADR 风险信号^[12]。DPA 正是通过计算量化药品-AE 组合的比例失衡程度, 来发现潜在的风险信号^[14]。信号值越大, 则认为目标药品与目标 AE 存在统计学关联的可能性越大^[15]。根据统计原理的不同, DPA 可分为频数法和贝叶斯法。

表 1 药品不良事件四格表

Tab. 1 Fourfold table of adverse drug event

报告数	目标不良事件	其他不良事件	合计
目标药品	<i>a</i>	<i>b</i>	<i>a+b</i>
其他药品	<i>c</i>	<i>d</i>	<i>c+d</i>
合计	<i>a+c</i>	<i>c+d</i>	<i>N=a+c+c+d</i>

注: *a*、*b*、*c*、*d*和*N*表示满足表格指定条件下的报告数量。

Note: *a*, *b*, *c*, *d*, and *N* indicated the number of reports that meet the specified conditions in the table.

2.1 频数类 ADR 数据挖掘方法

频数法是基于 AE 监测系统的数据进行直接计算的统计方法。报告比值比 (reporting odds ratio, ROR) 法和比例报告比 (proportional reporting ratio, PRR) 法是最常用频数类 DPA 法。

2.1.1 ROR 法 1992 年, Stricker 等^[16]使用 ROR 法报道了头孢克洛引起的血清病样反应, 这是将 DPA 应用于药品安全领域的最早报道^[17]。ROR 法与流行病学研究中“比值比” (odds ratio, OR) 计算相似, 计算目标药品-AE 组合与其他药品-AE 组合的比值之比。ROR 的计算见公式 (1)(式中 *a*、*b*、*c*、*d* 的定义参见表 1, SE 表示标准误)^[18]。通常, 当 $a \geq 3$, ROR 的 95% 置信区间 (confidence intervals, CI) 下限 > 1 , 则提示检测到 ADR 风险信号^[18]。

$$\begin{aligned} \text{ROR} &= \frac{(a/c)}{(b/d)} = \frac{ad}{bc} \\ \text{SE}(\ln \text{ROR}) &= \sqrt{\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)} \\ 95\% \text{ CI} &= e^{\ln(\text{ROR}) \pm 1.96 \sqrt{\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)}} \end{aligned} \quad (1)$$

2.1.2 PRR 法 英国医药监管局的 Evans 等^[19]于 1998 年提出 PRR 法, 并于 2001 年应用 PRR 法对英国黄卡系统数据进行分析。PRR 法与流行病学研究中“相对危险度” (relative ration, RR) 计算相似, 计算目标药品-AE 组合与其他药品-AE 组合的构成比之比。PRR 的计算见公式 (2)(式中 *a*、*b*、*c*、*d* 的定义参见表 1)^[19]。通常, 当 $a \geq 3$, PRR 值 ≥ 2 , 且卡方值 (χ^2) ≥ 4 , 则提示检测到 ADR 风险信号^[19]。

$$\begin{aligned} \text{PRR} &= \frac{a/(a+b)}{c/(c+d)} \\ \text{SE}(\ln \text{PRR}) &= \sqrt{\left(\frac{1}{a} + \frac{1}{a+b} + \frac{1}{c} + \frac{1}{c+d}\right)} \\ 95\% \text{ CI} &= e^{\ln(\text{PRR}) \pm 1.96 \sqrt{\left(\frac{1}{a} + \frac{1}{a+b} + \frac{1}{c} + \frac{1}{c+d}\right)}} \end{aligned} \quad (2)$$

2.2 贝叶斯类 ADR 数据挖掘方法

贝叶斯学派是统计学主要学派之一。其核心思想在于得知事件先验分布或随意指定先验分布的基础上, 计算事件 A 发生的条件下, 事件 B 出现的概率^[20]。贝叶斯类 ADR 数据挖掘方法是在四格表的基础上, 应用贝叶斯判别定理, 使得计算模型能在早期发现 ADR 风险信号, 其本质是计算药品与 AE 关联概率。目前, 贝叶斯置信传播神经网络 (Bayesian confidence propagation neural network, BCPNN) 法和多项伽马-泊松分布缩减 (Muti-item Gamma Poisson Shrinker, MGPS) 法是应用最广泛的贝叶斯类 DPA 法。

2.2.1 BCPNN 法 1998 年, UMC 的 Bate 等^[21]提出 BCPNN 法。BCPNN 法主要基于信息成分 (information component, IC) 及其 CI 去评估药品与 AE 之间的关联强度。IC 的计算见公式 (3)(式中 *a*、*b*、*c*、*d*、*N* 的定义参见表 1)^[21-22]。通常, IC 的 95%CI 下限 > 0 时, 则提示检测到 ADR 风险信号^[21]。BCPNN 基于 Beta 分布确定先验分布, 并主观设置超参数的值为 1。

$$\begin{aligned}
IC &= \log_2 \frac{P(x,y)}{P(x)P(y)} = \log_2 \frac{a(a+b+c+d)}{(a+b)(a+c)} \\
V(IC) &\approx \left(\frac{1}{\log_2} \right)^2 \left[\frac{N-a+\gamma-\gamma_{11}}{(a+\gamma_{11})(1+N+\gamma)} + \right. \\
&\quad \left. \frac{N-a-b+\alpha-\alpha_1}{(a+b+\alpha_1)(1+N+\alpha)} + \frac{N-a-c+\beta-\beta_1}{(a+c+\beta_1)(1+N+\beta)} \right] \\
\gamma &= \gamma_{11} \frac{(a+b+\alpha_1)(a+b+\beta_1)}{(a+\gamma_{11})(N+\alpha)(N+\beta)} \\
95\%CI \text{ 下限} &= \log_2 \frac{(a+\gamma_{11})(N+\alpha)(N+\beta)}{(N+\gamma)(a+b+\alpha_1)(a+c+\beta_1)} - \\
&\quad 2\sqrt{V(IC)}
\end{aligned} \tag{3}$$

式中： $\gamma_{11}=1, \alpha_1=\beta_1=1, \alpha=\beta=2$ 。 $P(x)$ 表示药物“ x ”在四格表中的概率； $P(y)$ 表示不良事件“ y ”在四格表中的概率； $P(x,y)$ 表示药物-不良事件组合“ x ”和“ y ”在四格表中的概率； $V(IC)$ 表示信息成分的方差； $\alpha, \alpha_1, \beta, \beta_1, \gamma, \gamma_{11}$ 均为超参数。

2.2.2 MGPS 法 2001 年, Dumouchel 等^[23] 提出 MGPS 法, 其是对 FDA 原有风险信号检测方法“伽马-泊松分布缩减”(gamma Poisson shrinker, GPS) 的改进。经验贝叶斯几何平均数 (empirical Bayes geometric mean, EBGM) 的计算是 MGPS 法的核心, EBGM 的计算见公式 (4) (式中 a, b, c, d 的定义参见表 1)^[24]。通常, EBGM 的 95%CI 下限 ≥ 2 , 则提示检测到 ADR 风险信号^[24]。但也有学者认为, 应以 EBGM 的 95%CI 下限 ≥ 1 作为风险信号的阈值, 而以 EBGM 的 95%CI 下限 ≥ 2 作为风险信号的阈值过于保守, 容易造成风险信号的漏检^[25]。MGPS 是基于 Gamma 分布确定先验分布, 并结合自身数据, 通过最大似然法估算超参数^[12]。

$$EBGM = \frac{a(a+b+c+d)}{(a+b)(a+c)} \tag{4}$$

2.3 频数法和贝叶斯法的优缺点

尽管频数法和贝叶斯法的原理与方法不同, 但本质上均是在药品 AE 四格表的基础上, 量化目标药品-AE 的比例失衡程度, 发现潜在的药物警戒风险信号。

2.3.1 频数法 优点是灵敏度高, 原理简单, 算法简便, 计算速度快。缺点是特异度低, 信号不稳定, 极易受个别值影响。由于估计值严重依赖于样本量大小, 当目标 AE 的报告数量 (a 或 c 值) 较少时, 一旦发生数值变动, 信号值波动较大, 可能会造成 ADR 风险信号的偏差^[26-27]。其

中, 对于 ROR 法, 当药品 AE 四格表中 b 或 c 值为零时, 即当数据库中无目标药品发生其他 AE 的数据, 或没有其他药品导致目标 AE 发生的数据时, ROR 值难以计算^[18]。有研究表明 ROR 比 PRR 更具优势, 因为 ROR 可以估计相对危险度, 从而消除一些偏倚, 而 PRR 法不能^[28]。但也有研究否认了这个观点, 认为这没有实际的作用, 在实践中 2 种方法产生的结果差不多^[29]。此外, 频数法无法反映时间变化趋势, 且容易受到混杂偏倚 (如年龄、性别等) 的影响^[27]。

2.3.2 贝叶斯法 优点是特异度高, 信号稳定, 误判概率小, 是为一种更谨慎的方法^[27]。贝叶斯法依赖于先验分布, 对样本量有最小要求, 但随着观察值越来越多, 先验分布被逐步修正, 得到趋向于“真实”分布的后验分布^[27]。相对于频率法, BCPNN 法对数据质量的要求较低, 检测到的统计学相关性较强^[21]。而 MGPS 可以对单个人群特征因素 (如年龄、性别等) 进行分层分析, 进而探索某一人群特征与 ADR 之间是否存在关联, 但无法同时对多个人群特征进行分层^[26]。贝叶斯法的缺点是灵敏度一般, 计算复杂, 需要一定的贝叶斯统计知识基础^[21,23]。

各种方法各有优劣, 国际范围内并未建立信号检测的金标准。Sakaeda 等^[15] 基于 FAERS 数据库, 对 ROR、PRR、IC、EBGM 4 种方法进行比较, 发现 ROR 法灵敏度最高, 检测出的 ADR 风险信号最多; EBGM 法灵敏度最低, 更为保守, 检测出的 ADR 风险信号最少。

2.4 其他 ADR 数据挖掘方法

除了上述的 DPA 法, 还有其他 DM 方法应用于潜在 ADR 的挖掘^[30], 包括关联规则法^[31]、聚类分析法^[32]、决策树法^[33]、主成分分析法^[34] 和神经网络法^[35] 等。这些方法主要基于机器学习算法实现潜在 ADR 的挖掘。但由于这些方法复杂度较高, 临床上多采用 DPA 法进行 ADR 风险信号检测, 其中应用最广泛的是频数法。

3 ADR 数据挖掘技术在药物警戒中的应用

药物警戒贯穿于整个药物发展的生命周期, 包括药物研发设计、上市前临床试验研究、上市后安全监测及评估, 直至被淘汰。药物警戒制度注重对 ADR 风险信号进行检测、识别、分析和评估, 进而对药品采取科学的管理和控制措施, 以保证患者用药安全^[36]。RWE 可用于评价药物安全

性和有效性, 药物警戒需要充分利用真实世界药品 AE 数据产生 RWE 以支撑药物警戒的发展和建设^[37]。

国家药物警戒中心正是利用 DM 技术自动检测 ADR 风险信号, 识别出在药品上市前某些未能发现的或未充分了解的 ADR。DPA 是目前 AE 报告系统广泛应用的风险信号检测方法。UMC 采用 BCPNN 法来进一步统计分析药品与 AE 的关系, 根据“药品-AE”表现 IC 的大小和数据的地区分布特点对信号进行分析和判定^[7]。EudraVigilance 采用 PRR 法进行信号检测, 根据信号的优先顺序进行分析、评估和验证, 确认为风险信号的才会采取控制措施^[9]。FDA 采用 MGPS 法来检测 AE 报告数据库, 因为它能查看药品-药品相互作用^[38]。国家药品不良反应监测系统应用了 PRR 和 BCPNN 2 种方法, 核心是提升模型输入参数的数据质量, 减少假阳性信号, 建立筛选分类策略和工具, 在资源有限的情况下关注重点药品和重点不良反应, 在不良反应方面关注心、肝、肾、脑、等部位的严重伤害^[9]。

此外, 药物警戒相关科研人员通过 DM 技术可以主动检测感兴趣药品的 ADR 风险信号, 并对药品安全性进行分析和评价。通过研究 AE 报告数据库, 特别是 RWD, 可以更好地获得 ADR 的临床表征(如时间、结果和预后等), 更好地监测药品的安全性, 及时检测罕见和危及生命的 ADR, 为医药卫生人员和患者提供参考依据。基于收集的 RWD, 科研人员需先对 AE 报告数据进行清洗和预处理, 再通过 ADR 数据挖掘方法识别潜在风险信号, 并结合临床研究和专业知识对结果进行解释和综合评估。

由于 FAERS 数据量大, AE 报告信息多, 且易于获取, 最常用于 ADR 的 DM 研究。Khaleel 等^[39]分享了详细的 FAERS 数据集的清洗和预处理流程, 并提供了一个干净、无重复数据的标准化 FAERS 数据集。基于预处理好的 AE 报告数据, 可采取多种 DM 方法进行风险信号检测, 用于单种药品的 AE 特征分析, 多种药品的 AE 差异性比较, 以及药品与 AE 相关性分析等。Mikami 等^[40]使用 ROR 和 IC 法挖掘免疫检查点抑制剂诱导的神经系统 AE 的特征和危险因素。陈欢等^[22]使用 BCPNN 法挖掘来迪派韦/索磷布韦上市后的安全警戒信号。Caldito 等^[41]通过 ROR 法比较利妥昔单

抗和奥瑞珠单抗 AE 特征的差异。Wu 等^[42]通过 PRR 和 IC 法挖掘质子泵抑制剂与痴呆事件相关性, 以进一步验证临床研究中质子泵抑制剂与痴呆之间关系相互矛盾的结果。

4 ADR 数据挖掘的局限性

尽管 AE 报告数据库经常被国家药物警戒中心用于评估安全问题, 但由于 AE 报告数据库自身固有的限制, ADR 数据挖掘具有一定的局限性。①报告偏移: 存在漏报、少报、多报、重报、报告数据质量参差不齐等问题。②数据缺失: 相关人口统计学信息(包括年龄和性别等)和临床信息(包括疾病严重程度、治疗措施和 AE 发生时间等)缺失, 导致统计分析困难及结果的不准确。③混杂偏移: 既往病史、伴随治疗、药品给药的剂量、频率和药品-药品相互作用等可能影响药品与 AE 的关系, 信号检测结果存在混杂偏移。④报告结局的局限性: 死亡与药品的使用没有明确关系, 死亡可能与疾病进展相关^[43]。⑤无法获知药品-AE 发病率: 由于缺乏接受药品治疗的总人数和发生 AE 的总人数, 无法直接计算药品-AE 发病率, 只能计算报告率^[22]。此外, 由于 DPA 算法自身的局限性, 无法推断药品与 AE 之间的因果关系: DPA 只是提供了信号强度的估计, 仅具有统计学意义, 仍需要临床研究进一步验证^[44]。

5 小结

随着 DM 技术的不断发展, ADR 数据挖掘技术广泛用于药物警戒研究, 并发挥着重要的作用。DM 利用 AE 报告数据库, 挖掘出使用情况和潜在 ADR 风险信号, 将这些信息相互关联、相互融合, 帮助医药卫生人员及早发现相关 ADR, 为药物警戒研究提供参考。本研究介绍了 AE 报告数据库, ADR 数据挖掘方法, ADR 数据挖掘技术在药物警戒中的应用及其局限性。尽管使用 DPA 法挖掘 AE 报告数据库存在一定的局限性, 但其 DM 结果仍具有临床价值, 可供研究参考。随着药品 AE 报告数据库的扩充和药物警戒的发展, 相信未来能更充分地了解药品安全性。

REFERENCES

- [1] 国家药品监督管理局药品审评中心. 国家药监局药审中心关于发布《用于产生真实世界证据的真实世界数据指导原则(试行)》的通告(2021年第27号)[EB/OL]. (2021-04-13) [2022-11-24]. <https://www.cde.org.cn/main/news/viewInfoCommon/2a1c437ed54e7b838a7e86f4ac21c539>.

- [2] DONG S S, HUANG Z X. A brief theoretical overview of random forests[J]. *J Integr Technol(集成技术)*, 2013, 2(1): 1-7.
- [3] HELMA C, GOTTMANN E, KRAMER S. Knowledge discovery and data mining in toxicology[J]. *Stat Methods Med Res*, 2000, 9(4): 329-358.
- [4] 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.
- [5] SHI W H, BA L, ZHOU J, et al. Comparative study on pharmacovigilance signal management system among the European union, the United States and Japan[J]. *China Pharm(中国药房)*, 2021, 32(4): 406-412.
- [6] ZHAN Q, WANG D Y. Pharmacovigilance and its signal problems[J]. *Chin J Pharm Epidemiol(药物流行病学杂志)*, 2003, 12(4): 211-213.
- [7] CUI X, WANG L X, LIU G Y, et al. Enlightenment of international pharmacovigilance system on establishment of pharmacovigilance system of Chinese medicine[J]. *China J Chin Mater Med(中国中药杂志)*, 2021, 46(21): 5450-5455.
- [8] ZHANG W, WEI F Q, YANG Y. Study of vigi base database of Uppsala monitoring centre[J]. *Chin J Pharmacovigil(中国药物警戒)*, 2020, 17(10): 676-680.
- [9] HOU Y F, LIU H L, QI Y. An overview of EudraVigilance and implications for China[J]. *Chin J Pharmacovigil(中国药物警戒)*, 2021, 18(11): 1058-1061,1065.
- [10] WANG D, REN J T, DONG D, et al. Trend analysis of ten-year annual reports on adverse drug reaction monitoring[J]. *Chin J Pharmacovigil(中国药物警戒)*, 2020, 17(5): 276-283.
- [11] 国家药品监督管理局. 国家药品不良反应监测年度报告(2021年)[EB/OL]. (2022-03-30) [2022-11-24]. <https://www.nmpa.gov.cn/xxgk/fgwj/gzwj/gzwjyp/20220329161925106.html>.
- [12] BATE A, EVANS S J W. Quantitative signal detection using spontaneous ADR reporting[J]. *Pharmacoepidemiol Drug Saf*, 2009, 18(6): 427-436.
- [13] FINNEY D J. Systemic signalling of adverse reactions to drugs[J]. *Methods Inf Med*, 1974, 13(1): 1-10.
- [14] LI C J. Researches on theory and application of adverse drug reaction signal detection[D]. Xi'an: The Fourth Military Medical University, 2008.
- [15] SAKAEDA T, TAMON A, KADOYAMA K, et al. Data mining of the public version of the FDA Adverse Event Reporting System[J]. *Int J Med Sci*, 2013, 10(7): 796-803.
- [16] STRICKER B H C, TIJSSEN J G P. Serum sickness-like reactions to cefaclor[J]. *J Clin Epidemiol*, 1992, 45(10): 1177-1184.
- [17] MOORE N, THIESSARD F, BEGAUD B. The history of disproportionality measures (reporting odds ratio, proportional reporting rates) in spontaneous reporting of adverse drug reactions[J]. *Pharmacoepidemiol Drug Saf*, 2005, 14(4): 285-286.
- [18] VAN PUJENBROEK E P, BATE A, LEUFKENS H G, et al. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions[J]. *Pharmacoepidemiol Drug Saf*, 2002, 11(1): 3-10.
- [19] EVANS S J W, WALLER P C, DAVIS S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports[J]. *Pharmacoepidemiol Drug Saf*, 2001, 10(6): 483-486.
- [20] LEE. Bayesian statistics[M]. London: Arnold, 1997.
- [21] BATE A, LINDQUIST M, EDWARDS I R, et al. A Bayesian neural network method for adverse drug reaction signal generation[J]. *Eur J Clin Pharmacol*, 1998, 54(4): 315-321.
- [22] CHEN H, WEN P. Pharmacovigilance signal detection of ledipasvir/sofosbuvir based upon Bayesian confidence propagation neural network[J]. *Chin J Hosp Pharm(中国医院药学杂志)*, 2021, 41(16): 1631-1636.
- [23] DUMOUCHEL W, PREGIBON D. Empirical bayes screening for multi-item associations[J]. *KDD*, 2001: 67-76.
- [24] SZARFMAN A, MACHADO S G, O'NEILL R T. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database[J]. *Drug-Safety*, 2002, 25(6): 381-392.
- [25] HAUBEN M, MADIGAN D, GERRITS C M, et al. The role of data mining in pharmacovigilance[J]. *Expert Opin Drug Saf*, 2005, 4(5): 929-948.
- [26] YE X F, WANG H N, CHEN W, et al. Application of data mining algorithms in pharmacovigilance[J]. *Chin J Pharmacovigil(中国药物警戒)*, 2008, 5(1): 36-40.
- [27] XIANG Y Y, XIE Y M, YI D H. Pharmacovigilance of major pharmaceutical innovation[J]. *China J Chin Mater Med(中国中药杂志)*, 2011, 36(20): 2831-2835.
- [28] ROTHMAN K J, LANES S, SACKS S T. The reporting odds ratio and its advantages over the proportional reporting ratio[J]. *Pharmacoepidemiol Drug Saf*, 2004, 13(8): 519-523.
- [29] WALLER P, VAN PUJENBROEK E, EGBERTS A, et al. The reporting odds ratio versus the proportional reporting ratio: 'deuce'[J]. *Pharmacoepidemiol Drug Saf*, 2004, 13(8): 525-526, discussion 527-528.
- [30] GAO S, GAO Y J, ZHU R Y, et al. Application progress of data mining technology in adverse drug reaction monitoring[J]. *China Mod Med(中国当代医药)*, 2021, 28(26): 31-35.
- [31] JO H G, 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*, 2021(12): 800571.
- [32] WARD D, THORSEN N M, 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.
- [33] LI L, QIN H L, GAN J Z, et al. Epidemiological characteristics of adverse drug reactions induced by traditional Chinese medicine injections based on Bayesian network and decision tree[J]. *Eval Anal Drug Use Hosp China(中国医院用药评价与分析)*, 2021, 21(1): 101-104,108.
- [34] XU J, ZHONG C M, ZHOU J, et al. Exploration of principal component analysis for discovering the signal of adverse drug reactions[J]. *Chin J Hosp Pharm(中国医院药学杂志)*, 2017,

- 37(1): 13-16.
- [35] BAJŽELJ B, DRGAN V. Hepatotoxicity modeling using counter-propagation artificial neural networks: Handling an imbalanced classification problem[J]. *Molecules*, 2020, 25(3): 481.
- [36] ZHANG Y J, CHEN H, YANG Y. Preliminary thinking on the establishment of China's future pharmacovigilance system[J]. *Chin J New Drugs(中国新药杂志)*, 2021, 30(22): 2017-2023.
- [37] WANG G P. Research on knowledge system of pharmacovigilance system[J]. *Chin J Mod Appl Pharm(中国现代应用药学)*, 2022, 39(13): 1755-1761.
- [38] WILSON A M, THABANE L, HOLBROOK A. Application of data mining techniques in pharmacovigilance[J]. *Br J Clin Pharmacol*, 2004, 57(2): 127-134.
- [39] KHALEEL M A, KHAN A H, GHADZI S M S, et al. A standardized dataset of a spontaneous adverse event reporting system[J]. *Healthcare (Basel)*, 2022, 10(3): 420.
- [40] MIKAMI T, LIAW B, ASADA M, et al. Neuroimmunological adverse events associated with immune checkpoint inhibitor: A retrospective, pharmacovigilance study using FAERS database[J]. *J Neurooncol*, 2021, 152(1): 135-144.
- [41] CALDITO N G, SHIRANI A, SALTER A, et al. Adverse event profile differences between rituximab and ocrelizumab: Findings from the FDA Adverse Event Reporting Database[J]. *Mult Scler*, 2021, 27(7): 1066-1076.
- [42] WU B, HU Q Z, TIAN F Y, et al. A pharmacovigilance study of association between proton pump inhibitor and dementia event based on FDA adverse event reporting system data[J]. *Sci Rep*, 2021, 11(1): 10709.
- [43] CHEN C, WU B, ZHANG C Y, et al. Immune-related adverse events associated with immune checkpoint inhibitors: An updated comprehensive disproportionality analysis of the FDA adverse event reporting system[J]. *Int Immunopharmacol*, 2021(95): 107498.
- [44] ABDEL-WAHAB N, SHAH M, SUAREZ-ALMAZOR M E. Adverse events associated with immune checkpoint blockade in patients with cancer: A systematic review of case reports[J]. *PLoS One*, 2016, 11(7): e0160221.
- 收稿日期: 2022-12-06
(本文责编: 曹粤锋)