

ICS ****.*
C**



团体标准

T/CACM ****—202*

中药上市后安全性评价数据分析方法指南

Guidelines of data analysis methods for safety evaluation on post-marketing Chinese Patent Medicine

(文件类型：公示稿)

(完成时间：2024 年 10 月)

20**-**-**发布

20**-**-**实施

中华中医药学会发布

目 录

前 言	II
引 言	III
1. 范围	4
2. 规范性引用文件	4
3. 术语和定义	4
3.1 上市中药 Post-marketing Chinese Patent Medicine	4
3.2 中药上市后安全性评价 Post-marketing safety evaluation of Chinese Patent Medicine	4
3.3 药品不良反应 Adverse drug reaction, ADR	4
3.4 药品不良事件 Adverse drug event, ADE	4
3.5 信号 Signal	4
4. 中药上市后安全性评价数据基本要求	5
4.1 相关性	5
4.2 真实性	5
4.3 全面性	5
4.4 完整性	5
4.5 可溯源性	5
4.6 伦理性	5
5. 中药上市后安全性评价数据来源	5
5.1 ADR/ADE 监测等信息系统	5
5.2 临床研究	5
5.3 药物毒理研究	6
5.4 期刊文献	6
5.5 灰色文献	6
6. 中药上市后安全性评价数据分析方法	6
6.1 描述性统计分析	6
6.2 比值失衡法	6
6.3 贝叶斯法	7
6.4 关联规则	8
6.5 决策树	8
6.6 回归分析	9
6.7 Meta 分析	10
6.8 神经网络	10
6.9 随机森林	11
6.10 支持向量机	11
参 考 文 献	12

前 言

本文件按照 GB/T1.1—2020《标准化工作导则第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由北京中医药大学国家中医药发展与战略研究院中医药大数据研究中心提出。

本文件由中华中医药学会归口。

本文件起草单位：北京中医药大学、河南中医药大学、天津中医药大学、中国中医科学院、广东药科大学、华西医院中国循证医学中心、兰州大学、首都医科大学、中日友好医院、中国医学科学院药物研究所、黑龙江中医药大学、山东中医药大学、湖南中医药大学、广西中医药大学、广州中医药大学第七临床医学院、复旦大学华山医院、武汉大学中南医院、华中科技大学同济医学院附属同济医院、鲁南制药集团股份有限公司、普华和诚（北京）信息有限公司。

本文件主要起草人：吴嘉瑞、张俊华、荆志伟、蔡永铭、张晓滕。

本文件其他起草人（按姓氏笔画排序）：丁长松、于子凯、王连心、王伽伯、王辉、田金徽、付长庚、刘树民、刘鑫馗、孙鑫、李欣、李修洋、李海、李霞、宋学坤、邱晓燕、张程亮、张景媛、金敏、金锐、周唯、赵崇军、费宇彤、郭凤英、梅全喜、曹庆恒、韩爱庆、程虹、靳洪涛、赫军。

引 言

近年来，中药上市后安全问题引起业内广泛关注，安全性评价成为热点研究领域。特别是中药注射剂等重点品种的安全性再评价工作在国家药品监督管理局指导下持续推进，取得较大进展。然而，由于中药安全问题影响因素复杂、关联关系判断难、数据分析方法应用局限等瓶颈，中药上市后安全性评价的高质量研究尚为数不多，海量的中药安全相关数据没有被充分利用。如何高质量开展中药上市后安全性评价已成为我国药物警戒工作中最具挑战性的内容，特别是高效识别中药不良反应信号、判断不良反应与中药之间关联关系、分析诱发中药不良反应发生的关键因素等科学问题备受关注。通过梳理和明晰相关数据分析方法在中药上市后安全性评价中的适用性和应用场景，可以为上市后安全性评价的开展提供技术支撑，指导科学、系统地挖掘中药潜在风险，剖析不良反应发生的内在规律等，为阐释中药安全性问题提供科学依据，促进中药临床安全合理使用。因此，制定《中药上市后安全性评价数据分析方法指南》具有重要的科学意义和实践价值。

本文件以中药上市后安全性评价数据分析为科学问题切入点，通过文献整理与挖掘深入解析常用临床数据分析方法的特点，结合专家调研、访谈、会议论证等方式，明确中药上市后安全性评价相关数据分析方法的适用性和应用场景，规范中药安全性评价相关数据的统计挖掘，为药品监管部门、药品上市许可持有人、药品生产企业、医疗机构、高校与科研院所开展中药上市后安全性评价数据分析提供指南。

本文件受国家自然科学基金项目《基于三维立体数据库和多维数据挖掘的中药注射剂过敏反应关键影响因素研究》（编号：81473547）、《中药注射剂不良反应随机森林信号模型建立及免疫毒理学评价方法研究》（编号：81473514）、国家中医药传承创新团队子项目《基于大数据的中药安全性评价方法学研究》（编号：ZYYCXTD-C-202005-10）、中国民族医药学会科研立项项目《中药民族药临床数据分析方法学研究》（编号：2020MZ298-110101）和国家中医药管理局青年岐黄学者培养项目（文件号：国中医药人教函〔2022〕256号）联合资助。

中药上市后安全性评价数据分析方法指南

1. 范围

本文件提供了中药上市后安全性评价的数据基本要求、数据来源和分析方法的建议。

本文件适用于药品监管部门、药品上市许可持有人、药品生产及经营企业、医疗机构、高校与科研院所开展中药上市后安全性评价使用。

2. 规范性引用文件

本文件没有规范性引用文件。

3. 术语和定义

下列术语和定义适用于本文件。

3.1

上市中药 Post-marketing Chinese Patent Medicine

已经获得国家药品监督管理部门批准，并在市场上销售和使用的中药。

3.2

中药上市后安全性评价 Post-marketing safety evaluation of Chinese Patent Medicine

以上市中药为研究对象、以临床安全性为主要研究目的的医学研究。

3.3

药品不良反应 Adverse drug reaction, ADR

合格药品在正常用法用量下出现的与用药目的无关的有害反应。

3.4

药品不良事件 Adverse drug event, ADE

药品治疗过程中出现的不良临床事件，不一定与该药有因果关系。

3.5

信号 Signal

来自一个或多个来源的，提示药品与事件之间可能存在新的关联性或已知关联性出现变化，且有必要开展进一步评估的信息。

4. 中药上市后安全性评价数据基本要求

4.1 相关性

与中药上市后安全性评价的目的相关联,以确保研究结果能为上市中药的用药决策提供基础,如上市中药 ADR/ADE 案例等。

4.2 真实性

要求数据来源可靠,记录真实、准确,必须客观反映中药安全问题的真实情况。上市中药安全性评价的数据来源较多,针对同一问题尽可能选择相对质量较好的数据,以确保结果的可靠性,增强评估结论的可信度。

4.3 全面性

数据应尽量全面,尽可能涵盖上市中药 ADR/ADE 的患者信息(如性别、年龄、疾病、过敏史等)、用药信息(如药品名称、剂量、疗程、联合用药、药品上市许可持有人、生产批号等)、ADR/ADE 信息(如 ADR/ADE 表现、发生时间、严重程度、预后情况等),以确保全面认识上市中药 ADR/ADE。

4.4 完整性

数据应力求完整,需紧密围绕研究目标,避免存在数据大量缺失情况,以确保最终分析结果的精确性。

4.5 可溯源性

数据应可溯源,每一个数据都应能够提供明确的来源和更新记录,提高数据分析的透明度,确保数据分析结果的可信度。

4.6 伦理性

涉及患者数据使用需要符合医学伦理学规范,遵循数据安全与隐私保护,以及中国人类遗传资源管理条例等相关法规文件要求,尊重和保护受试者合法权益。

5. 中药上市后安全性评价数据来源

5.1 ADR/ADE 监测等信息系统

通过国家 ADR/ADE 监测系统、医院信息系统(HIS 系统)、医疗保险数据系统等收集可疑的上市中药 ADR/ADE。

5.2 临床研究

以药物临床评价为目标的研究,包括随机对照试验、生态学研究、横断面研究、病例对照研究、队列研究等。

5.3 药物毒理研究

指为评价上市中药安全性开展的单次给药毒性实验、重复给药毒性实验、局部毒性实验、致敏实验、致癌、致畸、致突变实验等，可从整体动物、器官、细胞、分子多层面开展。

5.4 期刊文献

指国内外期刊中刊载的中药上市后安全性评价相关原始文献，包括中药 ADR/ADE 个案报告、中药 ADR/ADE 群案报告、中药毒理研究等。

5.5 灰色文献

未公开发表的文献，包括学位论文、会议论文、产品报告等。

6. 中药上市后安全性评价数据分析方法

6.1 描述性统计分析

6.1.1 基本概念

描述性统计分析是指通过收集、整理、汇总、分析和展示数据，以了解数据的分布、中心趋势、离散程度和相关性等的过程。

6.1.2 分析路径

定性变量数据（包括二分类变量、顺序变量和名义变量），一般采用频数和百分比描述，顺序变量资料的集中趋势和离散程度可采用中位数和四分位间距描述。

定量资料数据，先进行正态性检验，如果变量近似正态分布一般采用平均数和标准差，偏态分布需采用中位数和四分位间距。

6.1.3 考察指标

描述性统计主要考察指标包括：

- 频数及百分比，用以描述数据的分布情况；
- 平均数、中位数，用以反映数据的集中趋势；
- 标准差、四分位数、变异系数，用以观察数据离散程度。

6.1.4 应用场景

描述性统计的主要应用场景包括：上市中药 ADR/ADE 发生率分析，上市中药 ADR/ADE 临床表现分布，上市中药 ADR/ADE 影响因素分析等，用以分析原患疾病、用药原因、ADR/ADE 表现、转归等与上市中药安全相关的整体情况。

6.2 比值失衡法

6.2.1 基本概念

比值失衡法是通过比较药物使用与不良反应报告的比例来识别风险信号的方法。假设上市中药与某些不良事件存在因果关联，那么该“药品—不良事件”组合的实际发生数将超过预期，即二者的比值出现失衡，当比值失衡的程度超过事先界定的阈值时，则可认为出现一个风险信号。

6.2.2 分析路径

“药品—不良事件”组合比值失衡程度的计算基于 2×2 四格表。安全信号的检测既可通过对比值失衡指标 95% 可信区间的下限是否大于 0 进行检验，也可进行卡方检验，还可使用贝叶斯可信递进神经网络法、伽马泊松分布缩减法等。

6.2.3 考察指标

比值失衡法的主要考察指标包括：报告比（reporting ratio, RR）、信息成分（information component, IC）、比例报告比（proportional reporting ratio, PRR）、报告比值比（reporting odds ratio, ROR）

6.2.4 应用场景

比值失衡法主要应用于上市中药 ADR/ADE 信号发现，推断上市中药与 ADR/ADE 间可能存在的因果关系。

6.3 贝叶斯法

6.3.1 基本概念

贝叶斯法是以贝叶斯定理为基础，基于流行病学背景知识和病例信息，进行药物与 ADR/ADE 因果关系判断的方法。

6.3.2 分析路径

朴素贝叶斯法假定诱发 ADR/ADE 的各种原因相互独立。以药物引发了 ADR/ADE 为假说，通过收集所有的事实和证据，定量描述这些资料对假说支持的程度。具体实施步骤包括：确定分析中需要考虑的参数、收集病例有关资料（病史、ADR/ADE 发生过程、去激发、再激发等）、估算先验比、估算似然率、估算后验比（即支持药物是 ADR/ADE 原因的后验比）。

贝叶斯网络可表示联合条件概率分布，允许在变量的子集间定义条件独立性，提供一种因果关系图形，可用于非独立诱因 ADR/ADE 分析。

6.3.3 考察指标

贝叶斯法的主要考察指标包括：贝叶斯概率（似然率、先验概率、后验概率），贝叶斯因子。

6.3.4 应用场景

贝叶斯法主要应用场景包括：

——ADR/ADE 与药物的关联评价，利用贝叶斯法计算 ADR/ADE 与药物之间的关联程度，从而识别可能引起 ADR/ADE 的上市中药；

——ADR/ADE 影响因素分析，分析 ADR/ADE 与用药剂量、疗程、联合用药等用药因素间的关系，分析诱发 ADR/ADE 的主要影响因素；

——ADR/ADE 预测，可利用贝叶斯法结合已有的病例数据，构建上市中药 ADR/ADE 预测模型。

6.4 关联规则

6.4.1 基本概念

关联规则是通过分析数据中属性值之间的关联关系，发现 ADR/ADE 与相关属性的关联模式的一种数据挖掘方法。

6.4.2 分析路径

关联规则形如 X（药物、患者、用药情况等相关信息） $\rightarrow$ Y（不良反应发生）的蕴涵式，分析路径包括以下步骤：

——识别所有的频繁项目集，即支持度高于研究预期设定的最低支持度的项目组合；

——从识别出的频繁项目集中构造关联规则，并确保置信度高于研究预期设定的最低置信度，可以通过 Apriori 算法、FP-growth 算法等分析 ADR/ADE 与相关属性间的关联关系。

6.4.3 考察指标

关联规则的主要考察指标包括：支持度、置信度、提升度、频繁度、熵等。

6.4.4 应用场景

关联规则主要应用于上市中药 ADR/ADE 影响因素分析，如分析不同因素（患者特征、药物自身特征、用药条件等）对 ADR/ADE 发生的影响，找到潜在关联规律为安全性评价和风险管理提供支持。

6.5 决策树

6.5.1 基本概念

决策树是一种基于树状结构进行决策分析的数据挖掘方法，可自上而下生成分类预测图模型。

6.5.2 分析路径

决策树的基本架构包括决策节点、分支和叶子。决策树的每个节点、子节点的个数与决

策树用的算法有关，常用算法包括 ID3、C4.5、CART 等。在沿着决策树从上到下遍历的过程中，在每个节点都会遇到一个问题，在每个节点上问题的不同回答导致不同的分支，最后达到一个叶子结点，这个过程就是利用决策树进行分类的过程。

6.5.3 考察指标

决策树的主要考察指标包括：信息增益、增益率、基尼指数、互信息，并以树状图模型来表示中药安全性评价的决策规则和关键因素。

6.5.4 应用场景

决策树主要应用于 ADR/ADE 影响因素分析、ADR/ADE 预测等，如可预测患者特征、药物成分、用药条件等不同信息条件下，出现 ADR/ADE 的概率。

6.6 回归分析

6.6.1 基本概念

回归分析是研究因变量（或反应变量）和一个或多个自变量（或解释变量）之间相互作用和相关性的一种分析方法。

6.6.2 分析路径

根据数据类型不同分为连续型变量、分类变量、计数型变量、生存时间数据等。

1) 连续型变量最常用线性回归分析。采用普通最小二乘法估计参数，建立自变量和因变量之间的线性回归方程。

2) 分类变量主要采用 Logistic 回归。根据分类因变量是否发生的可能性建立逻辑回归模型，估计各自变量的 Odds 比值，分析各预测变量与不良反应的关系。

3) 计数型变量常用 Poisson 回归。当因变量形式为数值计数时，假定其符合 Poisson 分布并建立 Poisson 回归模型。

4) 生存时间数据使用 Cox 比例风险回归模型。根据存活时间和状态建立回归模型，评估各因素对不良反应发生时间的影响。

6.6.3 考察指标

回归分析的主要考察指标包括：回归系数、OR 值、RR 值、95%可信区间（confidence interval, CI）等。

6.6.4 应用场景

回归分析主要应用于上市中药 ADR/ADE 影响因素分析，如分析患者特征（性别、年龄、疾病史等）、剂量、疗程等对 ADR/ADE 发生的影响。

6.7 Meta 分析

6.7.1 基本概念

Meta 分析是一种针对同一科学问题的多个研究结果进行总结、分析的方法。

6.7.2 分析路径

在安全性评价中可选用常规 Meta 分析、单率 Meta 分析、网状 Meta 分析等。

常规和网状 Meta 分析基于 RCT、非随机对照试验（nRCT）、队列研究和病例对照研究等开展。根据数据类型、研究间异质性大小，选择相应的指标合并方法。研究间异质性较小选用固定效用模型，反之则选用随机效用模型。

单率 Meta 分析基于单臂试验开展，用以探讨 ADR/ADE 发生率、死亡率以及 ADR/ADE 发生的相关因素等。

6.7.3 考察指标

Meta 分析的主要考察指标包括：

- 二分类变量，考察比值比（Odds ratio, OR）、相对危险度（Relative Risk, RR）；
- 连续变量，考察平均差（Mean difference, MD）、标准化平均差（Standardized mean difference, SMD）；
- 网状 Meta 分析，主要考察曲线下面积（Area Under Curve, AUC）。

6.7.4 应用场景

Meta 分析主要应用于分析上市中药 ADR/ADE 发生率等，如系统评价某上市中药 ADR/ADE 整体发生率、累及各个临床表现的发生率，比较不同中药在治疗同一疾病时 ADR/ADE 的发生率等。

6.8 神经网络

6.8.1 基本概念

神经网络是一种模仿生物神经网络工作机制的计算模型，是由大量处理单元互联组成的非线性、自适应信息处理系统。

6.8.2 分析路径

神经网络的计算方法基于神经元之间的连接和权重，通过不断的学习和调整，实现对输入数据的处理和输出。其分析过程主要包括输入处理、激活函数、权重更新、输出处理等步骤。常见的神经网络包括多层感知机、卷积神经网络、循环神经网络等。

6.8.3 考察指标

神经网络的主要考察指标包括：损失函数、准确率、AUC、相关系数和特征重要性等。

其能整合多源信息，有效处理中药安全性评价的不确定性，评估药物成分、患者特征与不良反应之间的关系，为安全性评价提供依据。

6.8.4 应用场景

神经网络主要应用于 ADR/ADE 影响因素分析、ADR/ADE 预测等。如通过收集已发生 ADR/ADE 病例数据，包括患者特征、药物成分、剂量等信息，利用神经网络构建模型，分析各因素对 ADR/ADE 发生的影响，预测未发生 ADR/ADE 的病例出现 ADR/ADE 的可能性，为上市中药 ADR/ADE 提供风险评估。

6.9 随机森林

6.9.1 基本概念

随机森林是一个用随机方式建立的包含多个决策树的分类器，其基本单元是决策树，通过集成多个模型，将各个结果进行投票或平均，以得到最终的预测。

6.9.2 分析路径

随机森林的建模过程为：

- 从原始数据集中使用自主采样法抽取样本，形成多个子数据集；
- 对每个子数据集，构建一个决策树。在每个节点，随机选择一部分特征进行分裂；
- 重复以上步骤，直到生成指定数量的决策树。

6.9.3 考察指标

随机森林法的主要考察指标包括：基尼指数、信息增益、增益率和袋外错误率等。

6.9.4 应用场景

随机森林法主要应用于上市中药 ADR/ADE 影响因素分析、ADR/ADE 预测等，探索各种变量与 ADR/ADE 之间的关联，评估在某种危险因素条件下发生 ADR/ADE 的概率等。

6.10 支持向量机

6.10.1 基本概念

支持向量机是一种寻找最优决策边界（或称为超平面）的监督学习方法，通过核函数将数据映射到高维特征空间，并在特征空间中寻找一个最优的超平面，以实现对新数据的分类和预测。

6.10.2 分析路径

建立训练样本，并且该训练样本至少涵盖 2 个及以上类别。根据训练样本类别，假设权重向量及偏置量。当需要对样本进行分类时，应用支持向量机评估样本的映射函数。如果映射函数满足一定条件，则可以完成分类样本的分配，并赋值为 1，将其分类为正类，否则将

其分类为负类，并赋值为-1。当遇到多分类问题时，一般建议转化成众多二分类问题组合，并获得相应的解。还可以与神经网络等算法集成，形成更强大的混合模型，提高预测性能。

6.10.3 考察指标

支持向量机法的主要考察指标包括：特征向量之间的距离、支持向量、超平面。

6.10.4 应用场景

支持向量机法主要应用于上市中药 ADR/ADE 预测。

参 考 文 献

- [1]赵军宁.中药监管科学[M].北京：中国医药科技出版社，2024.
- [2]王永炎，杜晓曦，吕爱平.中药上市后临床再评价设计与实施[M].北京：人民卫生出版社，2012.
- [3]吴嘉瑞.中医药临床大数据研究[M].北京：中国医药科技出版社，2020.
- [4]吴嘉瑞，张冰.中药注射剂临床应用系统评价研究[M].北京：人民卫生出版社，2018.
- [5]谢雁鸣，王志飞.中医药大数据与真实世界[M].北京：人民卫生出版社，2016.
- [6]张俊华，孙鑫.循证中医药学[M].上海：上海科学技术出版社，2018.
- [7]吴嘉瑞,张冰.试论数据挖掘决策树方法在药物警戒研究中的应用[J].中国药物警戒杂志, 2012, 9(1): 29-32.
- [8]吴嘉瑞,张冰.试论数据挖掘方法在药品不良反应评价领域的应用[J].中药新药与临床药理, 2007, 18(6): 485-487.
- [9]范明,孟小峰译. Jiawei Han, Micheline Kamber,Jian Pei 原著. Data Mining: Concepts and Techniques, Third Edition.机械工业出版社. 2012, 8.
- [10]崔盈盈,刘施,吴嘉瑞,等.基于贝叶斯网络与决策树的双黄连注射剂不良反应流行病学特点研究[J].药物流行病学杂志, 2017, 26(1): 19-26.
- [11]胥洋,丁呈怡,詹思延,等.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.
- [12]谢雁鸣,廖星,姜俊杰,等.中成药上市后安全性医院集中监测技术规范[J].中国中药杂志, 2019, 44(14): 2896-2901.
- [13]李振吉,翁维良.中成药上市后临床研究方法技术体系与应用[M].北京：人民卫生出版社, 2022, 10.
- [14]崔盈盈,吴嘉瑞,谭迪,等.基于关联规则的清开灵注射液致药品不良反应发生规律研究[J].中国医院用药评价与分析, 2017, 17(12): 1585-1590.
- [15]崔盈盈,吴嘉瑞,张丹,等.基于多维数据挖掘的穿琥宁注射液致药品不良反应发生规律研究[J].中国医院用药评价与分析, 2017, 17(12): 1591-1594.

[16]吴嘉瑞,张冰.双黄连注射剂不良反应文献的数据挖掘研究[J].中国药物警戒, 2008, 5(3): 139-143.

[17]张利丹,谢雁鸣,杨云云.真实世界中丹参川芎嗪注射液治疗脑梗死合并用药的复杂网络分析[J].中国中医基础医学杂志, 2021, 27(10): 1595-1601.

[18]艾青华,谢雁鸣,李霖,等.运用倾向评分法研究真实世界参芪扶正注射液对门冬氨酸氨基转移酶的影响[J].中医杂志, 2014, 55(18): 1596-1600.

[19]卢恩仕,韩明光,刘祖发.贝叶斯网络在中医药领域的研究现状及展望[J].中医学报, 2022, 37(2): 316-320.

中华中医药学会团体标准公示稿