

## 第二章 药品有效性评价指南

何娜<sup>1#</sup>, 李晓曦<sup>2#</sup>, 侯正尧<sup>3#</sup>, 张兰<sup>2</sup>, 刘芳<sup>1</sup>, 赵荣生<sup>1</sup>, 沈浩<sup>3</sup>, 吴行伟<sup>3\*</sup>, 董宪喆<sup>2\*</sup>, 翟所迪<sup>1\*</sup>

1. 北京大学第三医院 药学部, 北京 100191; 2. 首都医科大学宣武医院 药学部, 北京 100053; 3. 四川省医学科学院·四川省人民医院 药学部, 成都 610072

**【摘要】** 本章主要介绍了药品有效性评价的定义、流程和具体评价方法。药品有效性评价通过对比待评价药品与参比药品在治疗特定疾病时的临床治疗效果, 评估待评价药品是否能够显著改善患者状况并获得重要的健康收益。开展药品有效性评价首先要明确评价问题, 即按照PICO原则对目标人群、待评价药品、参比药品、结局指标等进行具化, 然后根据药品特征、疾病特征、文献资料的充分程度和数据的可及性选择适宜的评价方法, 包括二次研究、真实世界研究和RCT。二次研究系统收集和分析现有文献; 真实世界研究基于实际医疗环境中的数据, 反映药品在真实条件下的治疗效果; RCT则通过随机化分组, 阐明干预措施与结局之间的因果关系。此外, 本章还讨论了质量控制的重要性, 确保评价结果的真实性、正确性及可靠性。

**【关键词】** 药品有效性; 指南; 二次研究; 真实世界研究

**【中图分类号】** R95

**【文献标识码】** A

**【文章编号】** 1672-3384(2024)Z1-0009-13

**Doi:** 10.3969/j.issn.1672-3384.2024.Z1.002

### Chapter 2 Guideline for Drug Effectiveness Evaluation

HE Na<sup>1#</sup>, LI Xiaoxi<sup>2#</sup>, HOU Zhengyao<sup>3#</sup>, ZHANG Lan<sup>2</sup>, LIU Fang<sup>1</sup>, ZHAO Rongsheng<sup>1</sup>, SHEN Hao<sup>3</sup>, WU Xingwei<sup>3\*</sup>,  
DONG Xianzhe<sup>2\*</sup>, ZHAI Suodi<sup>1\*</sup>

1. Department of Pharmacy, Peking University Third Hospital, Beijing 100191, China; 2. Department of Pharmacy, Xuanwu Hospital, Capital Medical University, Beijing 100053, China; 3. Department of Pharmacy, Sichuan Academy of Medical Sciences and Sichuan People's Hospital, Chengdu 610072, China

**【Abstract】** This chapter primarily introduces the definition, process, and specific evaluation methods of drug effectiveness evaluation. Drug effectiveness evaluation compares the clinical therapeutic effects of a subject drug with a reference drug in treating specific diseases, to assess whether the subject drug can significantly improve patient outcomes and achieve important health benefits. The evaluation process begins by clarifying the evaluation question, which involves defining the target population, the drug under evaluation, the reference drug, and outcome measures according to the PICO principle. Based on the characteristics of the drug, the disease, the sufficiency of existing literature, and the availability of data, appropriate evaluation methods are selected, including secondary research, real-world studies, and RCTs. Secondary research systematically collects and analyzes existing literature; real-world studies use data from actual medical environments to reflect the drug's therapeutic effects; RCTs use randomization to clarify the causal relationship between the intervention and outcomes. Additionally, this chapter discusses the importance of quality control to ensure the authenticity, accuracy, and reliability of the evaluation results.

**【Key words】** drug effectiveness; guideline; secondary research; real world study

\*通信作者: 吴行伟, 博士, 副研究员, 研究方向: 临床药理学。E-mail: wuxingwei@med.uestc.edu.cn; 董宪喆, 博士, 副主任药师, 研究方向: 临床药理学与药事管理。E-mail: dongxianzhe@xwhosp.org; 翟所迪, 本科, 主任药师, 研究方向: 药物使用科学。E-mail: zhaisuodi@163.com

#何娜、李晓曦、侯正尧为共同第一作者

药品有效性评价是指待评价药品及参比药品在治疗特定疾病时,对其临床治疗效果进行对比和分析,判断待评价药品是否可显著改善患者疾病状况,并获得重要的健康收益。有效性评价是药品临床综合评价的核心维度之一,药品疗效明确(通常指待评价药品可解决当前疾病临床治疗未满足的需求,或较现有治疗措施有明显的治疗优势)是开展药品安全性和经济性等评价的基础。

本章内容包括开展药品有效性评价的流程设计和具体评价方法。

## 1 评价流程

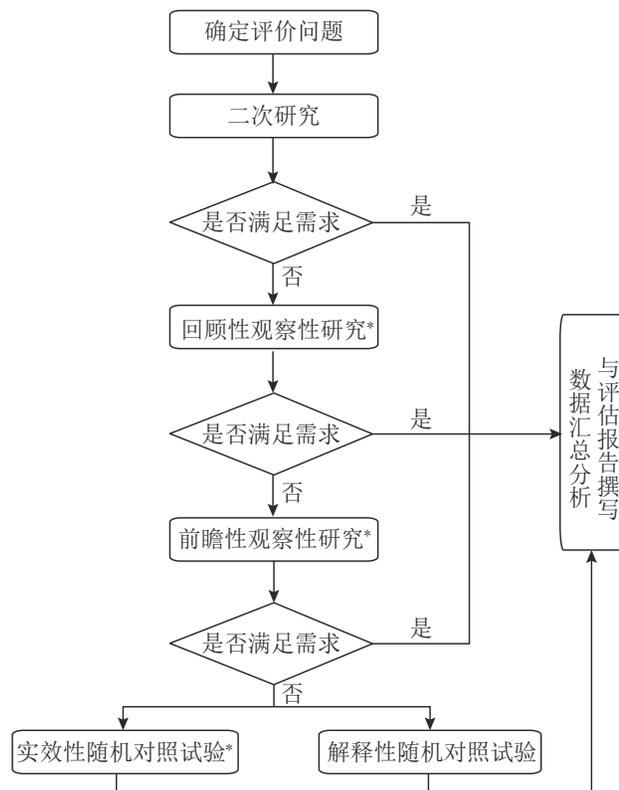
开展药品有效性评价首先要明确评价问题,即按照 PICO 原则对目标人群(population, P)、待评价药品(intervention, I)、参比药品(comparison, C; 优先选择国内外临床指南推荐的标准治疗或临床最常用的药品)、结局指标(outcome, O; 优先选择与所关注临床问题相关性更强的“硬”终点,其次选择替代结局指标)等进行具化。

明确评价问题后,根据待评价药品特征、疾病特征、既往文献资料的充分程度和数据可及性确定拟采用的研究方法。通常首先采取二次研究的方法对现有的证据进行梳理,若现有循证证据无法满足评价需求时,可考虑收集历史数据开展回顾性观察性研究;若回顾性数据不能满足需求时,可考虑开展前瞻性观察性研究或 RCT 等。有效性评价整体流程图见图 1。

## 2 评价方法

### 2.1 二次研究

二次研究是指针对某一具体临床问题,系统、全面地收集相关研究,采用恰当的文献评价方法和流程,筛选出符合标准的研究,对其进行定量或定性合



注:\*表示属于真实世界研究

图1 药品有效性评价整体流程图

成,得出当前最佳的综合结论。二次研究包括综述、系统评价和 meta 分析、快速卫生技术评估等研究类型,可提供证据级别较高的决策依据。

开展二次研究首先应纳入临床实践指南,尤其是基于系统性评价证据、多学科专家经验和患者意愿给出推荐意见的指南,直观地呈现待评价药品是否为特定适应证的一线推荐。其次是基于临床证据的评价,若已经有新近发表的质量良好的系统评价/meta 分析,可考虑开展快速评价<sup>[1-2]</sup>,即仅基于已发表的系统评价/meta 分析开展评价。如果上述条件不满足,下一步可同时纳入系统评价/meta 分析及原始研究开展完整评价。开展二次研究的总体思路见图 2。

开展二次研究的步骤可分为:①确定纳入/排除

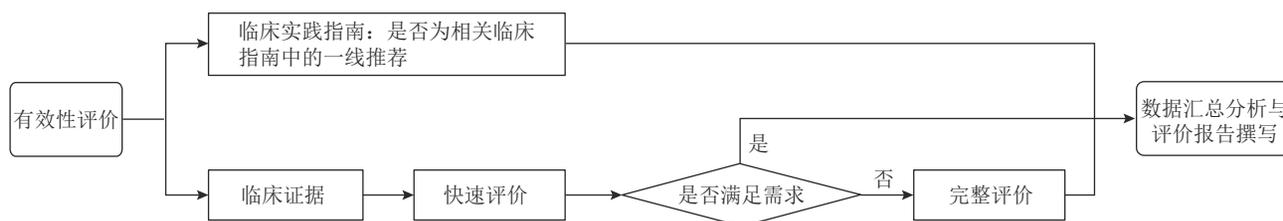


图2 基于二次研究的有效性评价总体思路

标准;②文献检索;③文献筛选和数据提取;④单篇研究的质量评价;⑤数据的汇总合并;⑥证据体的质量评价——评价结果的确定性;⑦评估结果的报告。

**2.1.1 确定纳入/排除标准** 基于待评价药品的特征及预评估的结果,确定评估的类型(快速评价或完整评价)。在PICO原则的基础上进一步完善文献的纳入排除标准:具体描述目标人群的临床特征,补充对年龄、合并症等的要求;具体描述待评价药品和参比药品的细节,补充对剂量和疗程的要求等;优化结局指标,明确重要、主要和次要结局指标;补充对纳入研究类型(study design, S)的要求。①快速评价主要纳入临床实践指南、系统评价/meta分析(含网状meta分析)、卫生技术评估(health technology assessment, HTA)。②完整评价一般在快速评价的基础上纳入RCT及队列研究、病例对照研究等观察性研究。RCT是有效性评价的金标准,通常而言有效性评价应优先纳入RCT,但对于随访周期长、罕见的结局事件,或急症抢救等RCT不可行或不符合伦理的情况,观察性研究也应被纳入并作为重要的证据来源。此外,还可基于文献发表语言、研究报告信息的完整性,以及同一篇研究重复报告等情况设定必要的排除标准。

**2.1.2 文献检索** 通常需检索的数据库包括:①综合性文献数据库,例如PubMed、Embase、the Cochrane Library等英文数据库,中国知网、万方数据知识服务平台、中国生物医学文献服务系统等中文数据库;②专科文献数据库,例如针对行为科学和心理健康领域的PsycINFO数据库等;③临床试验注册平台,例如ClinicalTrials.gov、http://www.chictr.org.cn/、https://trialsearch.who.int/等;④国内外卫生技术评估机构网站,例如英国国家卫生与服务优化研究院(National Institute for Health and Care Excellence, NICE)、加拿大药物和卫生技术局(Canadian Agency for Drugs and Technologies in Health, CADTH)等。

**检索策略:**通常以待评价药品名称、目标人群作为检索词,采用主题词与自由词结合的方法制定检索式。必要时可用经过验证的“search filters”辅助检索,例如对研究类型的限制、对人体研究的限制等。检索时应记录本次检索时间及数据库检索起止年限。

**补充检索:**是在数据库的文献检索基础上,必要时需辅以人工检索进行补充,尽可能避免遗漏。可采

用的补充检索方法包括但不限于以下几种:①对最终纳入研究的参考文献进行筛选;②检索最终纳入研究的相似研究,例如PubMed的“similar articles”;③检索引用纳入研究的文献;④检索UpToDate、Trip数据库和医脉通等相关知识型数据库;⑤检索相关学术会议进展。

**2.1.3 文献筛选和数据提取** 为避免研究者在文献筛选/收集数据过程引入偏倚,文献的筛选过程建议由2名研究者独立完成,研究结果的数据提取必须由2名研究者独立完成。2名研究者在正式文献筛选/数据提取前要有预试验,确保文献筛选/数据提取标准的一致性。应提前设定遇到不一致时的解决办法,例如讨论或征求第3人意见等。

**文献筛选** 文献筛选的流程通常为:①使用参考文献管理软件合并检索结果,删除相同报告的重复记录;②基于标题和摘要,删除明显不相关的报告;③获取潜在相关报告的全文;④查阅报告全文,判断研究是否符合纳入排除标准;⑤必要时,联系作者以确认研究是否符合纳入排除标准;⑥最后决定研究的纳入,着手数据收集。

文献筛选过程需记录文献纳入排除原因及纳入排除文献数量,以在结果部分展示文献筛选过程。除人工筛选外,需要时可辅以经过验证的文献筛选软件(例如RCT Classifier、Screen4Me、EBM AI-Reviewer)以提高效率。

**数据提取** 数据提取通常分为研究特征的提取和研究结果的提取。数据提取表示例见图3~7。

研究特征可能会影响待评价药品效果的有无或大小,提取研究特征有助于评价研究结果的适用性,提取的内容通常包括:①资料来源(作者、年份、发表时间);②方法(研究设计、研究开展的时间、研究开展的国家/地区);③PICOS [P为纳入标准(含诊断标准等)、样本量、患者特征(年龄、性别、原发疾病特征、并存疾病等),I与C为药品名称、给药途径、剂量、治疗时间等细节,O为报告的结局指标(含定义)、随访时间等,S为研究设计类型]。

除研究特征的提取外,还要提取各研究的研究结果,为后续定性、定量汇总做准备。提取的信息主要包括:①针对二分类变量的结局指标,干预组与对照组的样本量、干预组与对照组发生相关结局的患者

数、风险比(risk ratio, *RR*)、比值比(odds ratio, *OR*)、风险差(risk difference, *RD*)等;②针对连续型变量的结局指标,干预组与对照组的样本量、干预组与对照组的测量结果及标准差(standard deviation, *SD*)、组间均值差(mean difference, *MD*)、置信区间(confidence interval, *CI*)等。

此外,根据纳入研究类型的不同,某些研究类型有需特异性提取的信息,例如观察性研究需额外提取混杂因素校正的方法,同时提取粗发生率和校正后的效应值等。

年份	国家及机构	推荐意见	是否推荐待评价药物	推荐适应证	证据级别	推荐强度	是否首选
----	-------	------	-----------	-------	------	------	------

图3 纳入指南的基本特征

研究	疾病	药品	有效性评价结果	证据级别
----	----	----	---------	------

图4 纳入卫生技术评估的基本特征

研究	检索时间	疾病	拟评价药品	参比药品	纳入研究数	样本量	结局指标	效应值	结论
----	------	----	-------	------	-------	-----	------	-----	----

图5 纳入系统评价/meta分析的基本特征

**2.1.4 单篇研究的质量评价** 由2名研究者独立评价每一篇纳入研究的质量。对其评价结果不一致的

研究	国家/地区	患者基本特征				干预措施			结局指标	效应值
		疾病	男/女	年龄	样本量	拟评价药品	参比药品	疗程		

图6 纳入RCT的基本特征

研究	国家/地区	数据来源	样本量	患者基本特征	拟评价药品	参比药品	随访时长	结局指标	混杂偏倚控制方法	校正后效应值
----	-------	------	-----	--------	-------	------	------	------	----------	--------

图7 纳入观察性研究的基本特征

地方进行复核,必要时进行讨论或征求第3人意见。

纳入的指南,采用AGREE II<sup>[3]</sup>、AGREE-China<sup>[4]</sup>、“指南科学性、透明性和适用性评级”(scientificity, transparency, applicability rankings, STAR)工具<sup>[5]</sup>等评价其质量。纳入的HTA报告,采用国际卫生技术评估机构协作网(International Network of Agencies for Health Technology Assessment, INAHTA)制定的HTA checklist(报告清单)<sup>[6]</sup>评价。纳入的系统评价与meta分析,采用AMSTAR<sup>[7]</sup>或AMSTAR-2量表<sup>[8-9]</sup>进行方法学质量评价,或采用Risk of Bias in Systematic Reviews (ROBIS)工具<sup>[10]</sup>进行偏倚风险评价。纳入的RCT,采用Revised Cochrane Risk-of-Bias Tool for Randomized Trials (RoB 2.0)工具<sup>[11]</sup>进行偏倚风险评价。纳入的观察性研究,采用Risk of Bias in Non-Randomized Studies-of Interventions(ROBINS-I)量表<sup>[12]</sup>、纽卡斯尔-渥太华量表(Newcastle-Ottawa-Scale, NOS)<sup>[13]</sup>等进行偏倚风险评价。

**2.1.5 数据的汇总合并** 对纳入的临床实践指南、HTA、系统评价/meta分析(含网状meta分析)进行描述性评价和分析。在进行定性描述和分析时,应评估和判断主要的证据来源。

针对临床实践指南,可将方法学质量较高的指南或国内外权威学会(中华医学会,中国医师协会,中国药学会,美国、欧洲、日本特定疾病领域学会等)制定的指南作为主要证据来源。

针对系统评价/meta分析,可基于各系统评价的临床相关性、检索日期、纳入研究全面性、方法学质量等确定主要证据来源<sup>[14]</sup>。

针对RCT及观察性研究等原始研究,可考虑将低

偏倚风险的研究作为主要证据来源。

必要时对纳入的RCT或观察性研究进行定量合并(针对完整评估)。定量合并通常包括以下分析:

- ① meta分析, 比较多种干预措施, 或2种干预措施之间缺乏头对头研究时, 可考虑开展网状meta分析;
- ② 异质性分析;
- ③ 亚组分析;
- ④ 敏感性分析;
- ⑤ 发表偏倚评价。

**2.1.6 证据体质量评价—评价结果的确定性** 对于特定评价问题, 通过系统检索、筛选, 对符合纳入排除标准的研究总结形成的“证据体”进行基于证据推荐分级的评估、制订与评价(the Grading of Recommendations Assessment, Development and Evaluation, GRADE)系统的质量评价<sup>[15-16]</sup>, 以确定评价结果的确定性(把握度), 并形成临床研究证据总结表。GRADE证据分级的过程见图8, 临床研究证据总结表见图9。

**2.1.7 评估结果的报告** 应遵循相关报告规范, 例如PRISMA2020、MOOSE、ENTREQ等, 具体可参考EQUATOR协作网(<https://www.equator-network.org/>)。

证据的呈现形式亦可以参考一些可视化的方法, 例如辅以气泡图等证据图谱的形式, 更清晰、简明、系统地展示当前证据的质量、有效性、数量和人群特征等各种信息。

## 2.2 真实世界研究

真实世界研究即基于真实世界数据的研究, 研究数据来自真实医疗环境, 反映实际诊疗过程和真实条件下的患者健康状况<sup>[17]</sup>。基于真实世界数据的药品有效性评价是指在真实世界环境下收集与待评价药品疗效相关的患者诊疗数据或基于这些数据衍生的

汇总数据, 通过分析获得药品的使用情况及潜在获益-风险的临床证据的研究过程<sup>[18]</sup>。

基于真实世界数据的药品有效性评价设计类型主要包括基于主动收集患者医疗数据的干预性研究和观察性研究<sup>[19-20]</sup>。其中, 干预性研究一般包括实效性随机对照试验(pragmatic randomized controlled trial, pRCT)、单病例随机对照试验、单臂研究等; 观察性研究一般包括队列研究、病例对照研究、横断面研究等。pRCT研究设计见2.3.2部分。

**2.2.1 试验设计** 试验设计包括研究人群筛选与结局指标确定。真实世界研究通常设置较宽泛的纳入排除标准来筛选研究人群, 使得研究的外推性更好。由于真实世界研究是基于真实世界数据的研究, 因此数据的质量和数量是研究结果可靠性的基础。对于回顾性的真实世界数据分析, 常需开展系统的数据治理。由于真实世界环境下, 可能存在某些指标的缺失, 因此在试验设计时, 应时时关注数据是否支持等问题。

**结局指标确定** 结局指标的选择应结合研究目的、疾病特征、药品特性, 参考相关国内外专病指南、诊疗规范、临床实践专家共识, 兼顾研究设计的科学性、评价指标的可获得性、可分析性。在设计阶段研究者应明确主要及次要结局指标, 常见的主要和次要结局指标见表1。当评价指标不易获取而选择其他可测量的替代指标或中间指标时, 需要注意其与主要结局指标的相关性并作出详细说明。

**2.2.2 数据采集和治理** 真实世界数据可采用前瞻性或回顾性收集方式获取。对于回顾性研究, 在数据

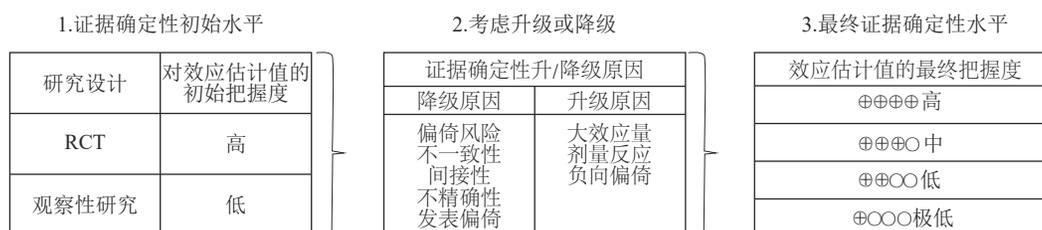


图8 GRADE证据分级的过程

疾病	拟评价药品	对照药品	结局指标	纳入研究数	样本量	效应值	GRADE分级	结论

图9 临床研究证据总结表

表1 药品有效性评价常见结局指标

有效性结局指标	内容
主要结局指标	①生命时长指标;②病情转归指标;③终点事件及复合终点事件
次要结局指标	①疾病控制指标;②症状体征变化指标;③实验室检查指标;④影像学检查指标;⑤病理学检查指标;⑥生命质量评价指标

收集前,研究者应预先制定数据治理及分析方案,包括数据提取策略、提取字段、重要变量缺失的评估及数据治理和分析方案等。

**数据采集** 药品有效性评价数据是指在真实医疗环境下,患者诊疗过程所记录的医疗数据及各种观察性研究数据等。其中前者一般包括医院信息系统数据、医保支付数据、死亡登记数据等;后者一般包括研究登记数据、自然人群队列观察性数据等,可从医院信息系统、专病数据库、死亡登记数据库、公共数据库中获取。常用公共数据库信息见表2。

研究者数据收集过程中,应对数据提取全流程进行严格质量控制。明确数据提取流程与规范,数据提取者在实施过程中对其流程做详细记录以保证过程可追溯。统一相关研究指标记录和描述的标准,必要时构建数据字典以保证各数据源资料信息规范性及可用性。同时,研究者可通过构建完善统筹协调机制及信息共享机制开展多来源、多类型、多中心的省域或区域药品有效性评价研究以解决数据孤岛问题。

**数据质量评价** 数据质量评价应贯穿于研究始终。研究者在数据收集、数据治理、数据分析过程中应积极开展数据质量评价,以保证研究科学性及其研究

结果的可靠性。数据质量评价流程见图10。

当数据收集完成后,研究者需对获取的源数据进行适用性评价。研究所获取数据及其使用过程应满足伦理审查法相关需求,严格执行患者隐私保护及数据安全相关要求。研究所涉及的关键变量在单个数据源中应有较好的覆盖度及数据完整性,以保证研究分析需求。数据样本量应兼顾统计学要求及临床可信度,同时注意避免经数据治理后符合研究需求的样本量明显减少。此外,数据源应保证在目标研究年限内采集完整、连续、可追溯。

在数据治理后,研究者应对研究数据相关性及其可靠性进行评价。前者包括研究关键变量信息覆盖度、药物治疗措施及临床结局定义准确性、研究样本目标人群代表性;后者包括经治理后研究变量及其变量值的完整性、数据准确性、数据治理过程的透明性。经数据质量评价合格后,则可按图进行下一步分析。若数据质量较低,难以保障药品有效性评价的开展,研究者可首先考虑完善已有相关数据或扩大数据来源,并重新进行数据质量评价<sup>[21]</sup>。

**数据治理** 数据治理是指针对特定的研究分析目的,对提取源数据进行治理以使其易于研究分析,其治理内容一般涉及数据安全性处理、数据清洗、数据转化等。

数据安全性处理是指基于真实世界数据的药品有效性评价要求严格按照《药品临床综合评价管理指南(试行)》的数据信息安全条款执行,坚持“谁主管谁负责、谁授权谁负责、谁使用谁负责”,加强评价过程中的数据采集、存储、挖掘、应用、运营、传输等环节的安全和管理。对患者个人敏感信息进行统一去标识

表2 常用公共数据库信息表

数据库名称	简介
美国国家癌症数据库	该数据库数据来源于1500多家癌症委员会认可的医院注册数据,这些数据用于分析和跟踪恶性肿瘤疾病患者及其治疗和结局
SEER 数据库	该数据库为美国癌症统计的权威来源,主要针对癌症预防和控制,以及协助与老年人有关的流行病学和卫生服务研究
Orphanet数据库	该数据库是收集罕见病相关知识,以提高罕见病患者的诊断、护理和治疗效果的数据,旨在提供有关罕见病的高质量信息
MIMIC重症医学数据库	该数据库是一个大型的、可免费获取的数据库,具有数据集时间跨度长(12年)以及个人健康数据完整等特点,使其成为重症医学界著名的科研数据集
BioLINCC 公共网站	该数据网站提供国家心肺血液研究所收集的110多个研究机构的临床和流行病学研究数据和生物样本
TCGA 数据库	该数据库目前包含了33种癌症的数据,每种癌症都涉及关键基因组变化的全面、多维的图谱

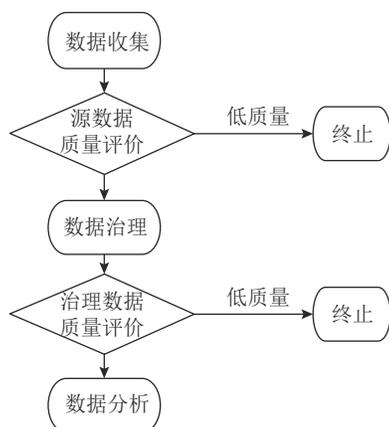


图10 数据质量评价流程图

化处理,以防止患者个人信息泄露。

数据清洗是指对源数据进行审查和校验,其内容一般包括删除重复冗余数据、逻辑核查、异常值及缺失值处理等。研究者在数据清洗过程中应首先对不同数据来源(例如不同信息系统)的数据在合并过程中产生的重复数据及无关数据进行识别。在保证数据完整性的前提下去除重复冗余数据,以减少后续数据的处理难度,提高研究分析效率。其后,研究者应对数据内容进行逻辑核查,以确保原始数据的可靠性与准确性。例如,患者年龄与出生年月推算值不同、相关检查检验结果不符合临床实际等。对于核查中发现的异常值,研究者应谨慎处理以避免引入偏倚,保证数据的真实性。数据更改要求对其内容及原因进行完整记录,以保证数据内容可追溯。最后,针对原始数据中存在的数据缺失问题可视数据缺失原因、缺失程度、数据指标重要性,对其缺失数据进行填补或对相关病例数据进行删除。

数据转化是指对清洗后的数据进行集中批量操作,使数据在形式及内容上更适用于研究分析,其内容一般包括数据格式标准化、医学术语归一化、编码标准统一化、衍生变量计算等。

数据格式归一化是指对不同数据源的数据格式进行标准化处理,使其合并在同一通用数据库中,以便对收集到的全样本进行数据整合。非结构化数据(自由文本数据)可利用可靠的自然语言处理技术处理,使其转化为结构化数据。医学术语归一化是指针对关键变量可按照研究目的,参照现有医学术语规范进行标准化处理。例如,由于信息记录缺乏统一标

准,患者诊断一般存在术语不规范、内容错误等问题。研究者可制作变量字典对相关术语进行归一化处理,必要时可参照国际疾病分类标准进行编码标准统一化。

回顾性研究设计所收集数据并非按照研究目的进行记录,其一般需要研究者进行基于已有数据的衍生变量计算以使其更易于统计分析。例如,医院信息系统数据中的肌酐值通常需要人为计算转化为肌酐清除率。计算衍生变量值需提前对相关变量及计算方法进行明确定义,同时保留原始数据及操作过程,以保证数据可追溯。

**2.2.3 统计学考虑** 统计学需考虑样本量估算和偏倚控制方法。

**样本量估算** 基于真实世界数据的药品有效性评价样本量大小取决于现有资料中有多少病例数符合研究纳入标准。真实世界研究设计通常使用能获得的全样本,研究者一般无需提前估算样本量。研究者更加关注在已有样本量的基础上,有多大把握能发现评价药品的阳性结果。核心在于检验效能( $1-\beta$ )的计算<sup>[22]</sup>,检验效能越大,所需的样本量就越大。研究设计一般要求在给定统计学检验水平下(例如 $\alpha=0.05$ ),检验效能要达到80%或90%,即要求Ⅱ型错误发生概率( $\beta$ )低于20%或10%。值得注意的是,当研究设计包含多个药品比较及亚组分析时,研究者仍需考虑各组间样本量是否能同时兼顾统计学要求。

**偏倚控制方法** 基于真实世界数据的药品安全性及有效性评价研究主要偏倚类型包括选择偏倚、信息偏倚、混杂偏倚。

选择偏倚是指在患者纳入排除阶段,纳入对象与排除对象在某些特征上存在差异而引起的系统误差。常见的选择偏倚包括入院率偏倚、幸存者偏倚、罹患率偏倚等。在研究设计阶段,研究者应制定明确的纳入排除标准。在研究实施阶段,实施者绘制筛选流程图并说明相应病例剔除原因。此外,研究者可利用统计学方法比较纳入研究各组、纳入病例与剔除病例在关键基线特征上的差异以控制选择偏倚。

信息偏倚是指在数据收集或研究实施过程中,因测量方法缺陷导致获取所需信息时所产生的系统误差。常见的信息偏倚包括回忆偏倚、测量偏倚、观察者偏倚等。建议研究者预先制定完善的信息收集计

划,尽可能多地收集研究所涉及的重要信息以评估潜在的信息偏倚影响。为保证数据真实性及研究结果可靠性,研究需识别并妥善处理数据资料中存在的异常值。建议研究者与相关领域临床专家对数值的真实性进行共同研判。异常值的修改要求保留修改痕迹并对其修改原因及依据进行说明记录,以确保研究可追溯性。

混杂偏倚是指药物暴露因素与疾病发生转归相关程度受到其他因素的干扰而导致的药物暴露效应估计的系统误差。常见的混杂偏倚如适应证偏倚。混杂偏倚控制常用统计学方法包括倾向性评分、多变量分析,见表3。

**2.2.4 研究结果报告** 基于真实世界数据的药品有效性评价结果报告形式可参照 STROBE 清单拓展的 RECORD 清单。

### 2.3 随机对照试验

当二次研究及观察性研究证据均无法满足评价需求时,可进一步开展RCT研究。

RCT就是从目标人群中选出合适的研究对象(P),进行随机化分组,对不同组实施不同的干预(I),随访一段时间,然后比较干预组和对照组(C)之间的结局(O),以比较干预效果的不同<sup>[23-24]</sup>。RCT具备阐明某个特殊干预措施和结局之间因果关系的能力,通过随机分配干预措施降低混杂因素的影响,并通过设盲降

低在结局确认和疗效判定时产生的偏倚。

RCT根据研究的目的可分为解释性RCT(即常见的RCT)和实效性RCT(pRCT)。解释性RCT,是评价某种干预措施在理想的、严格控制环境下的效力,pRCT则是衡量某治疗方法在真实世界环境下的疗效。

**2.3.1 解释性随机对照试验** 解释性RCT研究一般可分为方案设计、方案实施、数据分析3个阶段,见图11。

**研究人群(P)** 应确定研究的纳入和排除标准,其目的是为了确定一组用于研究干预措施对结局影响的人群。制定纳入和排除标准时应考虑对象的代表性、依从性和伦理等问题<sup>[25]</sup>。纳入标准是根据研究问题和研究效率确定目标总体,常见的纳入标准包括疾病特征、人口学特征、入组前疾病状态等。应根据研究目标确定研究对象的诊断标准,一般依据临床诊疗指南和规范或依据科学共同体制定的标准制定。纳入标准应产生足够数量的受试者,同时使试验发现的可靠性和招募的容易程度最大化。排除标准是依据研究问题所定义的不能参与研究的人群,常为潜在高风险人群及失访可能性较高的人群,此外,还应明确因不良反应、依从性等因素造成的剔除与脱落标准。

样本量计算是RCT研究中的重要环节,样本量的

表3 混杂偏倚控制常用统计分析方法

统计分析方法	适用条件
倾向性评分	
倾向性评分匹配法	倾向性分析中应用最为广泛的一种方法。要求组间的倾向性评分(PS)值满足重叠分布、PS值计算的条件独立性2个假设
逆概率加权法	当对照组和治疗组PS值分布重合较小、匹配后样本损失多、代表性降低时可使用
倾向性评分分层法	当倾向性评分匹配法及逆概率加权法均不宜使用时可考虑使用。分层时须保证每层均有足够数量的样本量以保证层内效应的稳定性
倾向性评分回归法	当倾向性评分匹配法及逆概率加权法均不宜使用时可考虑使用
多变量分析	
logistic 回归/多水平 logistic 回归	独立样本中单次测量二分类变量/聚集性样本(多中心研究)中单次测量二分类变量
线性回归/多水平线性回归	独立样本中单次测量连续型变量/聚集性样本中单次测量的连续型变量
Cox 比例风险回归/方差-校正 Cox 模型	独立样本中单次测量的生存变量/聚集性样本中单次测量的生存变量
泊松回归/多水平泊松回归	独立样本中单次测量的累积或计数数据/聚集性样本中单次测量的累积或计数数据
广义线性混合模型(GLMM)	聚集性样本中单次测量二分类变量、连续型变量;重复测量二分类变量、连续型变量
广义估计方差(GEE)	聚集性样本中单次测量二分类变量、连续型变量;重复测量二分类变量、连续型变量

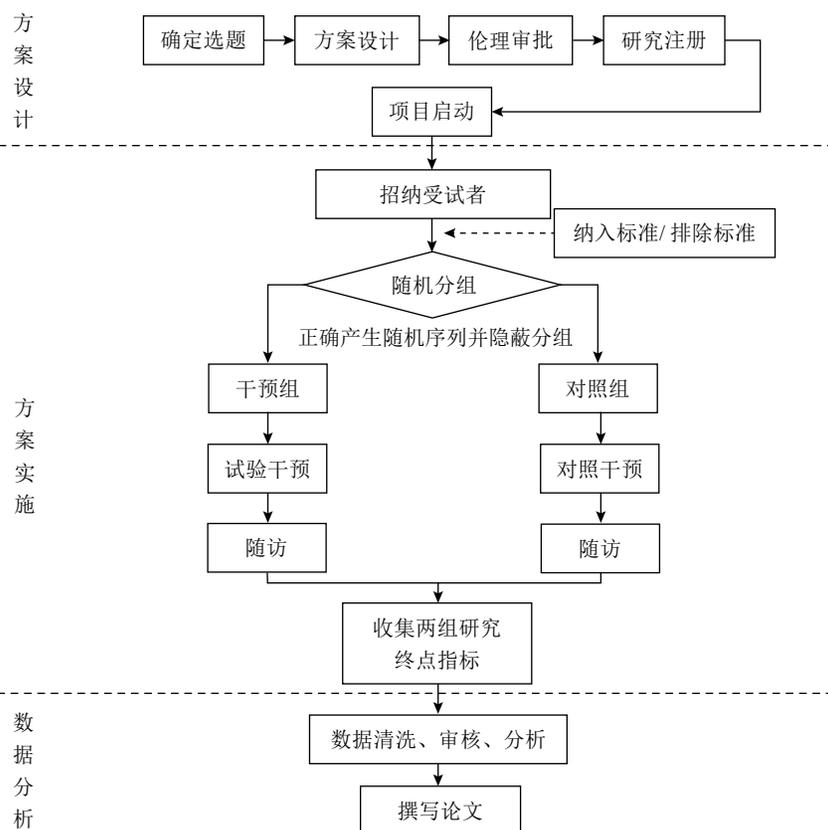


图 11 解释性 RCT 研究流程图

确定要综合考量研究目的、统计效能、伦理合规性、研究可行性等因素,其结果会直接影响到分析结果的可信度和研究效率。样本量计算需要考虑的因素有:

- ① 各组预期的结局(可以揭示组间结局指标的差异);
- ② I 类错误概率 $\alpha$ (通常取双侧 $\alpha=0.05$ ), $\alpha$ 越小样本量越大;
- ③ II 类错误概率 $\beta$ (通常取 $\beta=0.20$ ,检验效能 $\text{Power}=1-\beta=0.80$ ), $\beta$ 越小样本量越大;
- ④ 当结局为连续性变量时指标的标准差。其应在主要结局变量的基础上计算样本量,样本量估算时,可优先选择课题组的预实验结果,其次是通过类似研究的参考文献获取。还应考虑失访和退出等情况,将样本量进行相应扩大。

**干预措施(I)** 药品有效性评价研究中的干预措施是指干预组的治疗方案。设计干预措施时应充分考虑药物的剂量、疗程、给药频次、使用单一还是联合干预措施、具体操作方法、受试者的可接受性和注意事项等内容,以及治疗方式在实践中的可外推性等。要明确药品通用名、商品名、生产厂家、批号等信息及干预措施的具体内容,给出明确详细的定义或规定,

保证干预措施的规范性及可测量性,以及测量过程的可理解、可重复。

**对照措施(C)** 通过设置对照组控制混杂因素。对照组可分为安慰剂对照、剂量对照、空白对照和阳性对照等。最佳的对照是接受无疗效的治疗,比如接受与试验药物外观上无区别的安慰剂,可抵消因暗示或期望所导致的安慰剂效应,使研究分组之间的任何结局差异都可归因于干预产生的特定效应。从伦理学上讲,在实际研究中通常不能禁止或劝阻患者正在服用的常规药物来减少干预措施之外的治疗影响,此时可给予干预组和对照组标准治疗,在标准治疗的基础上检验干预组干预措施是否可以改善结局指标,但是该方法降低了整体事件发生率,从而增加了所需要的样本量。

**结局指标(O)** RCT 研究中的结局指标(也被称为终点指标)包括主要终点指标和次要终点指标<sup>[26]</sup>。一般情况下,一个临床研究仅设一个主要终点指标,该指标与临床试验的主要目的直接相关,往往包括疗效指标和其发生的时间点 2 个方面。

有效性终点指标即疗效指标,可以是疾病临床终点,可以是评价生活能力、临床症状等内容的相关量表,也可以是实验室检查结果等,可以反映临床获益,评价药物有效性。采用连续变量评价结局(如血糖、血压、生活质量等),通常比二分类变量(如糖尿病发生、死亡等)的结局所需要的样本量更少、干预时间更短、所需的费用也更低,但是不能够直接反映和回答所提出的科学问题。此外,能用定量变量评价的结局一般不要选择定性变量评价,能用多分类变量将数据分类做得更细一般不选择二分类变量采集数据<sup>[27]</sup>。

许多不同事件或测量组成的主要结局为复合结局,由于有更多的事件发生,复合结局通常比单一结局具有更高的效能。在经费和人力充足的情况下,可适当纳入行为学、社会学、生存质量、卫生经济学等指标。

**随机化分组 (randomization)** 随机化分组是RCT控制偏倚的一种方法。常见的随机策略包括简单随机化、区组随机化、分层随机化等。通常情况下,简单随机化是首选,将纳入的受试者个体采用简单的随机化方法分配到不同组。区组随机化,是先根据患者的某些特征进行排序,并划分为固定人数的“区组”,对每个区组内的人员进行随机化分配,是一种可以平衡对照组和干预组受试者人数的常用方法。分层随机化,是根据可能影响结局的因素进行分层,确保与结局紧密相关的变量水平在不同组之间均匀分布,也能提高小样本试验的效能。

随机化的精髓在于其不可预测性,只有同时采用分配隐藏才可以彻底避免选择偏倚,在最大限度上使各项已知和未知的预后因素在组间达到可比,最终可以将组间结局的差异归因为干预措施效果的差异<sup>[28]</sup>。分配隐藏要求在诊断患者、审核患者合格性、征求患者加入、随机分配等各个阶段,患者、医师和分组人员都不知道随机序列以及相对应的组别。中心随机分组的方法既符合分配隐藏又符合随机分组,该方法要求确定了一个患者的合格性并征得患者书面同意参与研究后,再向中心分组人员或部门索取随机数字,并给予对应的干预。通过互联网进行中心随机分组是目前最常用和最佳的方法。

**盲法设置 (blinding)** 盲法是指为了使患者、医师或资料搜集人员不知道患者接受的干预措施而采

用的一种程序或方法,同样是RCT控制偏倚的一种方法。与控制选择性偏倚的分配隐藏不同,盲法除控制选择性偏倚外,还可以控制信息偏倚。常用的盲法包括单盲、双盲和三盲。研究者应根据研究的目的、干预措施的属性等方面综合考虑,合理选择盲法设置。

**统计学分析** RCT的数据分析通常包括意向治疗(intent-to-treat, ITT)分析、改良意向治疗(modified ITT)分析、符合方案(per-protocol, PP)分析和按实际治疗(as-treated, AT)分析4种类型<sup>[29]</sup>。其中,ITT分析是RCT数据分析首选且最保守的方法,是指不管受试者是否按照计划完成试验过程,全部按其所分到的组别进行随访、评价和分析。ITT分析可避免由于破坏原始随机化而造成的偏倚,但是可能存在受试者脱落、改变治疗方案等情况。临床研究数据一般按照ITT原则将数据集分为全分析集(full analysis set, FAS)和符合方案集(per protocol set, PPS)。FAS是指尽可能按照ITT原则,以合理的方法尽可能少地排除受试者(如排除不符合纳入与排除标准的入组者、未服药者、无任何数据者),可以通过末次观察值结转法对部分由于受试者退出或剔除导致的数据缺失进行数据填补,并在数据统计分析中说明。PPS是FAS的一个子集,是更加符合研究方案的受试者数据集,一般由完成了预先确定的治疗量、主要变量可测定、基线变量没有缺失、无重大方案违背的受试者组成。

在统计学分析中,应明确3个方面的内容:①进行统计学分析的软件一般为SAS、Epiinfo、SPSS和R软件等。②统计学描述是应用统计参数和指标对变量数据进行合理描述和展示的过程<sup>[30]</sup>。描述性统计分析的主要任务是描述其集中趋势和离散趋势。在进行统计分析时,首先应对研究对象的一般资料进行分析描述。对于定量变量,需要考虑数据的分布特征,选择合适的指标来进行统计学描述。对于定性变量,可选择构成比、百分比、比值和率来进行统计学描述。③临床研究数据的统计学推断内容包括参数估计和假设检验2个部分。参数估计分为点估计和区间估计,其中点估计不考虑抽样误差,直接利用样本统计量的一个数值来估计总体参数;而区间估计是将样本统计量与标准误结合起来,以确定一个具有较大置信度的包含总体参数的范围,称为总体参数的置信

区间,通常设置为95%或99%。假设检验是临床研究数据统计分析的核心,是说明主要结局以及次要结局的分析方法<sup>[31]</sup>。比如符合正态分布且方差齐的定量变量,两组间比较采用 $t$ 检验,多组间比较采用方差分析,两个变量间的相关性分析可采用线性回归模型。对于定性变量,可选择卡方检验等探讨两组或多组之间的差异,使用logistic回归模型进行多因素分析探索独立危险因素。

**2.3.2 实效性随机对照试验** pRCT是指在真实或接近真实的医疗环境下,采用随机、对照的方式,比较不同干预措施的治疗结果(包括实际效果、安全性和成本等)的研究<sup>[32]</sup>。与传统RCT相比,pRCT的试验设计环境更偏向于真实医疗环境,纳入标准相对宽松,允许采用复杂干预措施评价整体疗效并观测多结局指标,研究结果更贴近临床诊疗实际,可更好地为医疗决策提供科学依据,是上市后药品、医疗器械等效果评价的最佳证据<sup>[33]</sup>。在开展pRCT时,可使用PRECIS-2工具辅助进行不同研究要素的设计,总体把握试验的实效性程度<sup>[34]</sup>。由于pRCT与RCT研究方法大致相同,下面主要对pRCT不同于RCT的研究设计进行补充说明。

**设计类型** pRCT常见的设计类型包括但不限于以下的类型:①个体实效性RCT(individual pRCT, ipRCT),即以个体为随机分组单位的RCT;②群组RCT(cluster RCT, cRCT),以群组为随机分组单位(干预单元)的RCT,此类型设计的试验在数据分析时应个体和群组两方面的干预效果进行评价和分析;③阶梯楔形RCT(stepwedged RCT, swRCT),一种特殊的群组RCT,常用于评价医疗卫生服务、卫生政策性的干预。研究中需根据研究目的、研究问题和研究条件来选择合适的类型。

**研究环境** 通常应考虑从研究结果可能被应用的场所中选取研究环境。比如恶性肿瘤、重度抑郁症等疾病对医护人员或医疗设备有特殊要求,只能在特定医院开展诊疗活动。

**研究人群** pRCT的纳入标准较宽泛、排除标准较少,并切合临床实际情况,允许不同研究对象间存在临床异质性。样本量估算考虑的因素除了干预措施、对照的效应大小及统计学检验参数外,还包括组内相关系数、群组大小、群组数目和结局测量次数

等<sup>[35]</sup>。为保证统计分析具有足够的检验效能,pRCT所需样本量通常较大。

**干预措施** pRCT的重要特征之一是干预方案的低标准化。pRCT允许干预实施者基于患者疾病特征、自身专业技能和执业经验等实际情况,灵活决定干预措施的实施细节,不强调所有受试者必须按照分配方案完成试验,在鼓励患者积极参与以提高数据收集的完整性与尽量减少研究本身对患者治疗依从性的影响之间进行综合权衡。但由于pRCT中干预措施的执行通常灵活性和可变性更大,研究者应对试验中允许医师或患者合并用药、改变剂量、换药或停药等条件或时间等作出必要的限定,也应描述试验中比较的各干预措施。

**对照措施** pRCT对照的选择要考虑临床实际情况,通常以常规治疗、标准治疗或公认有效的治疗措施为对照。

**结局指标** 因pRCT的研究目的在于评价干预措施在真实世界环境下的效果,结局指标应覆盖面更广,通常包含与患者日常生活相关的整体健康获益结局,一般测量干预的远期疗效、功能变化、生存质量、卫生经济学指标及远期终点事件等<sup>[36]</sup>。pRCT一般设定多个终点结局,包括主要结局和次要结局。主要结局强调选择对患者(或研究结果的运用者)有重要临床意义的指标,即能直接反映健康变化的测量指标。

**随机化分组** pRCT可采用整群随机或个体随机方法,还可结合医师或患者的意愿和偏好进行分组。在能实现的情况下,个体随机是优选。整群随机通常是评价一个项目干预的结果,当这些项目干预措施只能施加于群体(如医院、医师或社区)而非个体时,应基于群体进行干预。例如,针对社区开展的健康教育及针对个体的干预措施会影响群体里的其他人时(“沾染”现象),应当采用整群随机方法。

**盲法设置** 与传统RCT一致,pRCT通过随机分组和实施盲法最大程度地避免实施偏倚和测量偏倚。但对于pRCT而言,由于在临床实践中医师和患者对治疗的了解本身就是治疗的环节之一,对治疗结局产生的影响正是真实世界环境下治疗结局的一部分,因此,未对试验实施者和受试者实施盲法不一定对研究产生不利影响。为了尽量克服因知晓随机分组情况而导致的报告偏倚,pRCT也强调尽量对结果测量者

和统计分析人员施盲。

**统计学分析** pRCT统计分析的原则、内容及方法与RCT的基本原则一致,可参考《实效性随机对照试验的技术规范》<sup>[33]</sup>中相关内容,根据设计类型的不同选择适宜的分析方法。pRCT的主要统计分析需要基于ITT,即参与随机分组的对象,无论其是否接受该组的治疗,最终都应纳入所分配的组中进行统计分析。

**2.3.3 报告规范** 研究的报告可按照RCT报告标准CONSORT (consolidated standards for reporting of trials)2010修订版<sup>[37]</sup>进行撰写。CONSORT清单主要针对最常用的两组平行随机对照试验,包含七大部分<sup>[38]</sup>。此外,CONSORT小组还提供了CONSORT扩展(CONSORT-Extension)申明其他类型的报告声明,例如,针对整群随机试验、非劣效和等效性随机试验、适应性设计随机试验等。更多信息可在CONSORT官网(<http://www.consort-statement.org/>)进行查询及下载。

## 2.4 伦理学

伦理申请是对研究对象权益、安全和健康的保障,具体措施包括知情同意、资料保密、减少患者伤害和保障研究对象权益的措施。由于任何在人体上进行的研究都可能给研究对象带来或多或少的不便或伤害,而且消耗资源,因而研究设计的科学性也是伦理要求的一个部分。

伦理学要求临床上应使用疗效明确的干预,在不影响研究目的的情况下,应尽可能给予对照组现有已知最好的治疗(而不是安慰剂治疗),同时各组可以给予常规或现有最好的治疗作为基础治疗。当发现组间干预措施的获益-风险比出现明显差异时,应及时中止研究,然后给每一个参与研究的患者提供更为有效的治疗<sup>[36]</sup>。

大多数情况下,pRCT存在诸多伦理学问题的挑战。如医师灵活调整医疗干预措施(超剂量用药)可能带来临床安全性隐患;pRCT的对照组多采用常规或目前公认最佳的临床治疗方法,可能违反了临床均势原则,使干预组的受试对象错过了最佳治疗时机;如受试对象为急诊患者的pRCT,尚无证据表明使用干预和对照措施后患者的临床结果有差别,当符合纳入标准的患者病情危重、必须立刻采取治疗措施时,允许紧急处理后补办知情同意,但这种做法是否违背

了医学伦理,仍值得商榷<sup>[32]</sup>。

伦理委员会审批和受试者知情同意依然是保障pRCT受试者权益的重要措施。在某些特殊情况下(如比较不同医院管理政策对医院感染防控的影响),经伦理委员会审批,pRCT可以豁免患者知情同意。当pRCT的实施(包括干预和对照措施、数据收集等)完全符合实际临床医疗实践,通常认为也可申请豁免患者知情同意。

## 3 质量控制

为保障有效性评价的真实性、正确性及可靠性,须在项目正式开始前,根据事先制定的研究计划对参与项目的所有人员进行统一培训,并建议成立质量控制专家委员会,其成员应具备本课题所属专业领域的工作经验,包含临床医学、药学、流行病学等多学科专家。质量控制应贯穿于有效性评价研究过程始终:研究设计时,质量控制专家委员会对研究方案设计、数据治理计划书等进行论证和讨论,并针对其存在的相关问题提出意见和建议;研究过程中,质量控制专家委员会对研究过程进行考察和监督,确保研究按照研究设计严格执行;研究结束后,质量控制专家委员会对本研究的科学性、可靠性及应用价值进行再评价。

**作者贡献声明:**何娜(摘要、评价流程与二次研究部分的撰写/论文修改),李晓曦(随机对照试验、伦理学与质量控制部分的撰写/论文修改),侯正尧(真实世界研究部分的撰写/论文修改),张兰(论文修改/审阅),刘芳(论文修改/审阅),赵荣生(论文修改/审阅),沈浩(真实世界研究部分的撰写/论文修改),吴行伟(真实世界研究部分的撰写/论文修改),董宪喆(随机对照试验、伦理学与质量控制部分的撰写/论文修改),翟所迪(论文修改/审阅)。

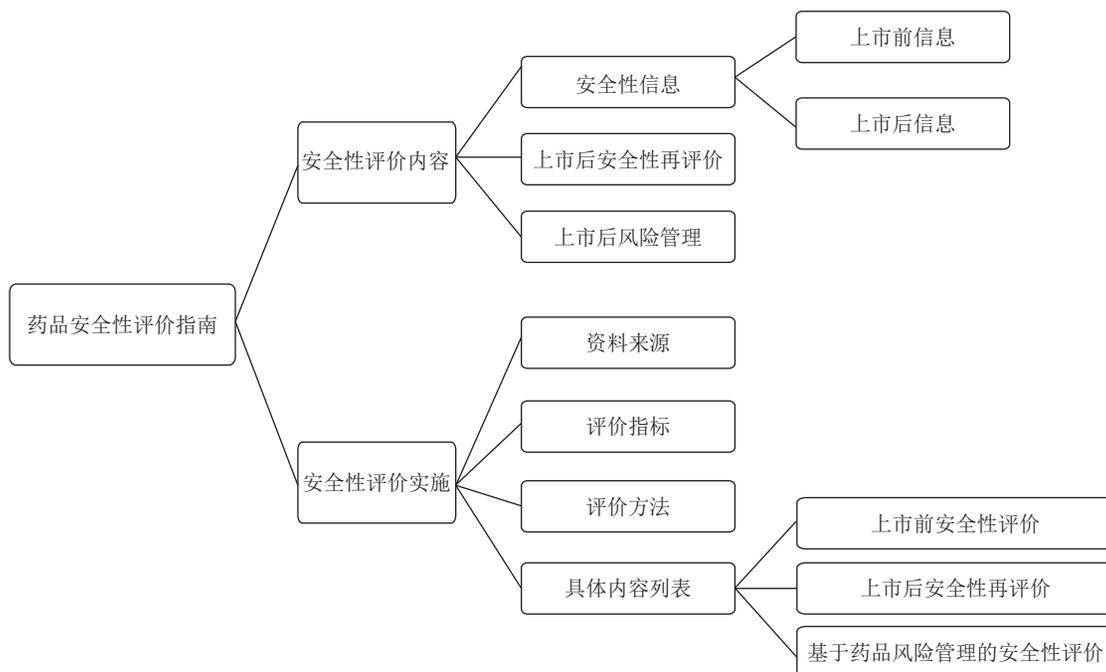
**附:**本章框架图见附图2。

## 【参考文献】

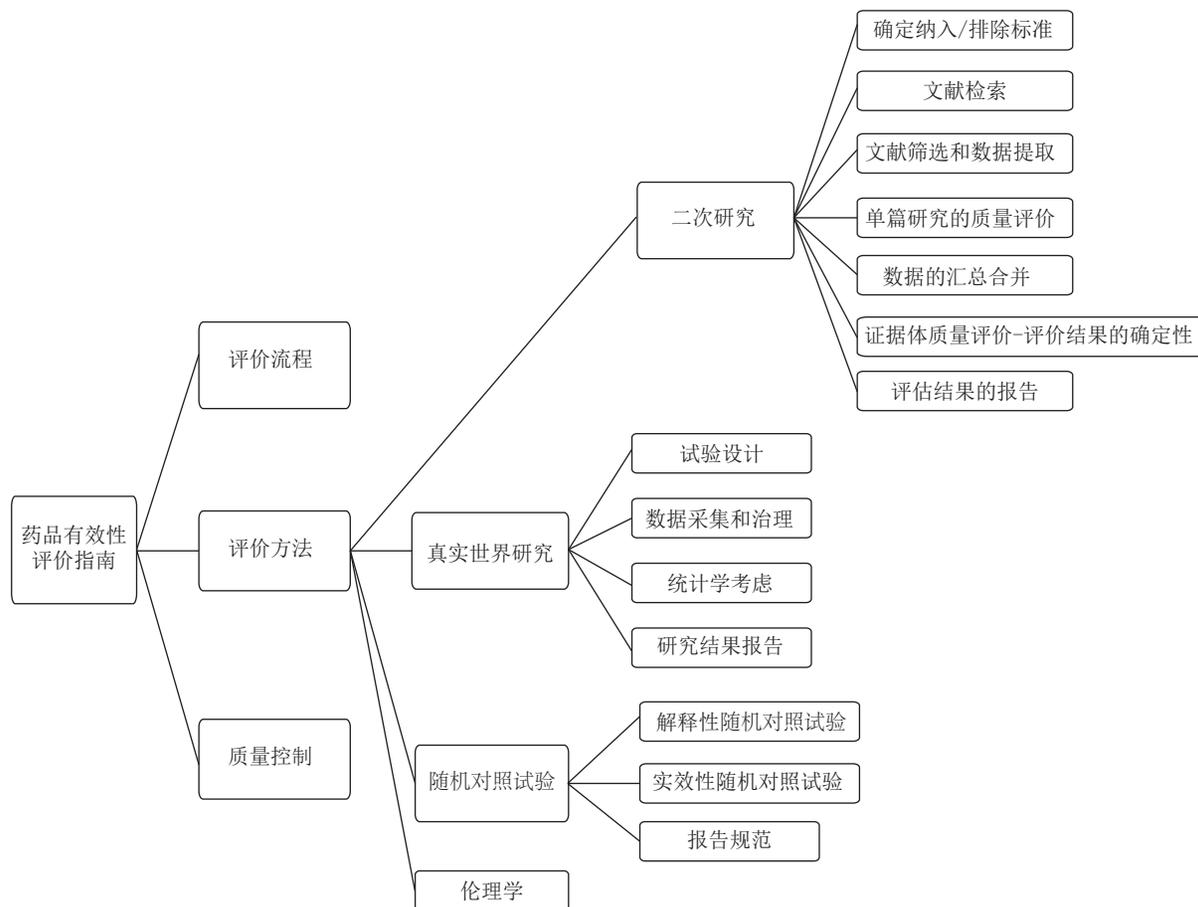
- [1] 李幼平,喻佳洁,孙鑫.快速评估方法与流程的探索[J].中国循证医学杂志,2014,14(5):497-500.
- [2] 唐惠林,门鹏,翟所迪.药物快速卫生技术评估方法及应用[J].临床药物治疗杂志,2016,14(2):1-4.
- [3] 韦当,王聪尧,肖晓娟,等.指南研究与评价(AGREE II)工具实例解读[J].中国循证儿科杂志,2013,8(4):316-319.
- [4] 王吉耀,王强,王小钦,等.中国临床实践指南评价体系的制定与初步验证[J].上海医学,2018,41(6):321-326.
- [5] 杨楠,赵巍,潘昉,等.针对临床实践指南科学性、透明性和适

- 用性的评级工具研发[J]. 中华医学杂志, 2022, 102(30): 2329-2337.
- [6] 嵇承栋, 朱琳懿, 万悦竹, 等. 国际卫生技术评估机构协作网卫生技术评估报告清单解读[J]. 中国循证医学杂志, 2016, 16(3): 369-372.
- [7] 熊俊, 陈日新. 系统评价/meta分析方法学质量的评价工具AMSTAR[J]. 中国循证医学杂志, 2011, 11(9): 1084-1089.
- [8] Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both[J]. BMJ, 2017, 358: j4008.
- [9] 陶欢, 杨乐天, 平安, 等. 随机或非随机防治性研究系统评价的质量评价工具AMSTAR 2解读[J]. 中国循证医学杂志, 2018, 18(1): 101-108.
- [10] Whiting P, Savović J, Higgins JP, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed[J]. J Clin Epidemiol, 2016, 69: 225-234.
- [11] Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials[J]. BMJ, 2019, 366: l4898.
- [12] Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions[J]. BMJ, 2016, 355: i4919.
- [13] 曾宪涛, 刘慧, 陈曦, 等. Meta分析系列之四: 观察性研究的质量评价工具[J]. 中国循证心血管医学杂志, 2012, 4(4): 297-299.
- [14] Lunny C, Pieper D, Thabet P, et al. Managing overlap of primary study results across systematic reviews: practical considerations for authors of overviews of reviews[J]. BMC Med Res Methodol, 2021, 21(1): 140.
- [15] 杨珂璐, 葛龙, 刘明, 等. GRADE在卫生技术评估中的应用[J]. 中国药物评价, 2020, 37(6): 411-416.
- [16] 杨楠, 肖淑君, 周奇, 等. GRADE在网状Meta分析中应用的基本原理和方法介绍[J]. 中国循证医学杂志, 2016, 16(5): 598-603.
- [17] 孙鑫, 谭婧, 唐立, 等. 重新认识真实世界研究[J]. 中国循证医学杂志, 2017, 17(2): 126-130.
- [18] 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]. (2020-07-17) [2023-12-12]. <http://www.nmpa.gov.cn/WS04/CL2138/373175.html>.
- [19] 中华中医药学会《中医药真实世界研究技术规范》制订组. 中医药真实世界研究技术规范——证据质量评价与报告[J]. 中医杂志, 2022, 63(3): 293-300.
- [20] 中华中医药学会. T/CACM 1371.5-2021 中医药真实世界研究技术规范-基于证据的中药有效性及安全性评价[S]. 北京: 中华中医药学会, 2021.
- [21] 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]. (2020-01-07) [2023-04-25]. <https://www.nmpa.gov.cn/xxgk/ggtg/qtggtg/20200107151901190.htm>.
- [22] 高培, 王杨, 罗剑锋, 等. 基于真实世界数据评价治疗结局研究的统计分析技术规范[J]. 中国循证医学杂志, 2019, 19(7): 787-793.
- [23] 彭晓霞, 唐讯. 临床研究设计[M]. 4版. 北京: 北京大学医学出版社, 2017.
- [24] 陈茜, 杨雁. 随机对照试验设计要点[J]. 中华糖尿病杂志, 2021, 13(6): 645-648.
- [25] 王瑞平, 肇晖, 李斌. 随机对照临床试验设计要点和规范[J]. 上海医药, 2022, 43(7): 72-77.
- [26] 王瑞平. 临床研究规范设计PICO原则[J]. 上海医药, 2022, 43(3): 67-72.
- [27] 王瑞平, 李斌. 临床医学研究数据分类浅谈[J]. 上海医药, 2022, 43(1): 3-6.
- [28] 唐金陵, Glasziou P. 循证医学基础[M]. 2版. 北京: 北京大学医学出版社, 2016.
- [29] 朱云鹏, 张恒, 郑哲, 等. 2022《美国心脏协会(AHA)心脏外科随机对照试验设计、执行和分析方法学标准的科学声明》解读[J]. 中国胸心血管外科临床杂志, 2023, 30(1): 10-16.
- [30] 王瑞平, 李斌. 临床医学研究数据统计分析思路概述[J]. 上海医药, 2022, 43(1): 7-9.
- [31] Laffel LM, Kanapka LG, Beck RW, et al. Effect of continuous glucose monitoring on glycemic control in adolescents and young adults with type 1 diabetes: a randomized clinical trial[J]. JAMA. 2020, 323(23): 2388-2396.
- [32] 唐立, 康德英, 喻佳洁, 等. 实效性随机对照试验: 真实世界研究的重要设计[J]. 中国循证医学杂志, 2017, 17(9): 999-1004.
- [33] 温泽淮, 李玲, 刘艳梅, 等. 实效性随机对照试验的技术规范[J]. 中国循证医学杂志, 2019, 19(7): 794-802.
- [34] Loudon K, Treweek S, Sullivan F, et al. The PRECIS-2 tool: designing trials that are fit for purpose[J]. BMJ, 2015, 350: h2147.
- [35] Hemming K, Haines TP, Chilton PJ, et al. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting[J]. BMJ, 2015, 350: h391.
- [36] 李立明. 流行病学[M]. 6版. 北京: 人民卫生出版社, 2007: 128-163.
- [37] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials[J]. BMJ, 2010, 340: c332.
- [38] 王瑞平, 李斌. 随机对照临床试验 CONSORT 声明解读[J]. 上海医药, 2022, 43(5): 58-62.

附件:中国药品综合评价指南参考大纲(第三版)附图1~10



附图1 第一章 药品安全性评价指南大纲框架图



附图2 第二章 药品有效性评价指南大纲框架图