

临床干预措施效力-效果差距弥合的方法学研究进展(一):改善真实世界研究的效果估计

刘佐相^{1,2} 龙子临^{1,2} 杨智荣³ 石舒原⁴ 许欣然^{1,2} 赵厚宇^{1,2} 杨祖耀⁵ 符祝⁶
宋海波^{7,8} 林腾飞⁹ 詹思延^{1,2,10} 孙凤^{1,2,6}

¹北京大学公共卫生学院流行病与卫生统计学系,北京 100191;²重大疾病流行病学教育部重点实验室(北京大学),北京 100191;³中国科学院深圳先进技术研究院计算机科学与控制工程学院,深圳 518055;⁴中国康复科学所/中国残疾人联合会残疾预防与控制中心,北京 100068;⁵香港中文大学公共卫生及基层医疗学院,香港 999077;⁶海南省博鳌乐城国际医疗旅游先行区管理局,海南省真实世界数据研究院,海口 571437;⁷国家药品监督管理局药品评价中心中药监测和评价部,北京 100076;⁸国家药品监督管理局药物警戒研究与评价重点实验室,北京 100076;⁹中国科学院深圳先进技术研究院先进计算与数字工程研究所生物医学信息技术研究中心,深圳 518055;¹⁰北京大学第三医院临床流行病学研究中心,北京 100191

通信作者:孙凤,Email: sunfeng@bjmu.edu.cn

【摘要】目的 干预措施在临床实践中的实际干预效果与随机对照试验(RCT)中表现的效力存在差异,即效力-效果差距。RCT结果与真实世界研究(RWS)结果的差异可能无法代表真实的效力-效果差距,这是因为当RWS与RCT在研究设计上有较大差异,或RWS结果估计存在偏倚时,效力-效果的估计可能是有偏的。其次,当发现干预措施存在效力-效果差距,不能对所有患者实行一刀切的临床决策,而需要进一步评估影响干预措施效果的真实世界因素,识别可能取得期望效用的患者群体。**方法** 检索PubMed、Embase、Web of Science、万方数据知识服务平台、维普数据库、中国知网6个数据库从建库至2022年12月31日的中英文文献,采用概括性综述的方法,对如何改进RWS设计从而弥合效力-效果差距的方法进行归纳整合和定性描述。**结果** 共纳入10篇文献,探讨如何以RCT研究方案为模板,制定相应的RWS方案,在正确估计效力-效果差距的基础上,进一步评估干预措施在患者亚群中的效果,选取能获得预期收益风险比的患者亚群,从而弥合效力-效果差距。**结论** 使用医疗大数据,模拟目标试验方案关键特征,可以提高研究结果的真实性和有效性,弥合效力-效果差距。

【关键词】 随机对照试验; 真实世界证据; 目标试验; 效力; 效果

基金项目:国家自然科学基金(72074011);中国药品监管科学行动计划第二批重点项目([2021]37-10);海南省博鳌乐城国际医疗旅游先行区管理局真实世界研究专项计划(HNLC2022RWS012)

DOI:10.3760/cma.j.cn112338-20230925-00189

收稿日期 2023-09-25 本文编辑 万玉立

引用格式:刘佐相,龙子临,杨智荣,等.临床干预措施效力-效果差距弥合的方法学研究进展(一):改善真实世界研究的效果估计[J].中华流行病学杂志,2024,45(2):286-293. DOI: 10.3760/cma.j.cn112338-20230925-00189.

Liu ZX, Long ZL, Yang ZR, et al. Progress in methodological research on bridging the efficacy-effectiveness gap of clinical interventions (1): to improve the validity of real-world evidence[J]. Chin J Epidemiol, 2024, 45(2):286-293. DOI: 10.3760/cma.j.cn112338-20230925-00189.



Progress in methodological research on bridging the efficacy-effectiveness gap of clinical interventions (1): to improve the validity of real-world evidence

Liu Zuoxiang^{1,2}, Long Zilin^{1,2}, Yang Zhirong³, Shi Shuyuan⁴, Xu Xinran^{1,2}, Zhao Houyu^{1,2}, Yang Zuyao⁵, Fu Zhu⁶, Song Haibo^{7,8}, Lin Tengfei⁹, Zhan Siyan^{1,2,10}, Sun Feng^{1,2,6}

¹Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; ²Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China; ³School of Computer Science and Control Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; ⁴China Rehabilitation Science Institute, China Disability Control and Prevention Center, China Disable Persons' Federation, Beijing 100068, China; ⁵School of Public Health and Primary Care, the Chinese University of Hong Kong, Hong Kong 999077, China; ⁶Administration of Hainan Boao Lecheng International Medical Tourism Pilot Zone, Hainan Institute of Real World Data, Haikou 571437, China; ⁷Department of Traditional Chinese Medicine Monitoring and Evaluation, Center for Drug Reevaluation, National Medical Products Administration, Beijing 100076, China; ⁸Key Laboratory for Research and Evaluation of Pharmacovigilance, National Medical Products Administration, Beijing 100076, China; ⁹Biomedical Information Technology Research Center, Institute of Advanced Computing and Digital Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; ¹⁰Clinical Epidemiology Research Center, Peking University Third Hospital, Beijing 100191, China

Corresponding author: Sun Feng, Email: sunfeng@bjmu.edu.cn

[Abstract] Objective Differences between randomized controlled trial (RCT) results and real world study (RWS) results may not represent a true efficacy-effectiveness gap because efficacy-effectiveness gap estimates may be biased when RWS and RCT differ significantly in study design or when there is bias in RWS result estimation. Secondly, when there is an efficacy-effectiveness gap, it should not treat every patient the same way but assess the real-world factors influencing the intervention's effectiveness and identify the subgroup likely to achieve the desired effect. **Methods** Six databases (PubMed, Embase, Web of Science, CNKI, Wanfang Data, and VIP) were searched up to 31st December 2022 with detailed search strategies. A scoping review method was used to integrate and qualitatively describe the included literature inductively. **Results** Ten articles were included to discuss how to use the RCT research protocol as a template to develop the corresponding RWS research protocol. Moreover, based on correctly estimating the efficacy-effectiveness gap, evaluate the intervention effect in the patient subgroup to confirm the subgroup that can achieve the expected benefit-risk ratio to bridge the efficacy-effectiveness gap. **Conclusion** Using real-world data to simulate key features of randomized controlled clinical trial study design can improve the authenticity and effectiveness of study results and bridge the efficacy-effectiveness gap.

[Key words] Randomized controlled trial; Real-world evidence; Target trial; Efficacy; Effectiveness

Fund programs: National Natural Science Foundation of China (72074011); The Second Batch of Key Projects of China Drug Regulatory Scientific Action Plan ([2021]37-10); Key Project of Real World Research Project of Administration of Hainan Boao Lecheng International Medical Tourism Pilot Zone (HNLC2022RWS012)

疗效评价是药物上市前和上市后的必然要求^[1],也是医学研究的重点内容,常涉及到“效力”(efficacy)和“效果”(effectiveness)这两个重要概念。效力指干预措施在理想条件下所产生的最大期望作用,主要来自于严格设计的随机对照试验(RCT);效果指干预措施获准上市后在真实医疗环境下所能达到的作用大小^[2-4]。基于真实世界数据(RWD)开展的真实世界研究(RWS)所产生的真实世界证据(RWE)^[5]是干预效果证据的重要来源。但是许多RWS发现干预措施在真实世界中的效果与上市前RCT中表现出的效力存在差异,这种潜在

差异被称为效力-效果差距^[2],是制药研发企业、监管部门、卫生技术评估和医疗保险机构等共同关注的问题^[6]。

切实的效力-效果差距主要来源于^[2,7-8]:①患者间生物学相关因素差异:患者基因组构成差异(导致药物作用相关靶点的差异化表达)以及身高、体重、年龄、合并症及基线疾病严重程度等个体差异;②患者和医护人员的行为差异:RWS中患者依从性不同以及不合理用药或者存在与干预措施有交互作用的共同处方等。

但是RCT与RWS之间结果的差异可能并不代

表真实的效力-效果差距。首先相较于 RCT, RWS 中可能存在^[9]: 纳入排除标准不详细、干预措施定义不明确、结局指标选择不规范的问题, 即研究间存在临床异质性, 则必然会导致结果的异质性。其次观察性 RWS 中使用既往收集临床数据进行分析, 缺乏随机化和盲法, 若缺少数据质量控制和统计学校正^[9], 可能会导致 RWS 结果的有偏估计。因此弥合效力-效果差距的首先需要规范和改进传统 RWS, 正确估计干预措施效果。

此外 RCT 具有严格的纳入排除标准以及标准化的试验流程, 能够保证患者的同质性和依从性, 组间差异可完全归因于干预措施作用, 而真实世界中患者异质性高, 依从性差, 直接使用组间平均差异, 无法代表干预措施的真实效果, 因此要划分具有不同效益-风险的患者亚群, 并评估影响干预措施效果的因素, 从而弥合效力-效果差距, 帮助临床决策。

在开展干预措施 RWS 时必须考虑如何帮助各利益相关方正确评估干预措施的效力-效果差距, 帮助精细化的临床决策。本研究通过采用概括性综述的方法对国内外关于改进 RWS、弥合效力-效果差距的经验进行系统梳理, 总结相关经验, 为我国药品等相关研究领域产品的研发、应用和监管等提供参考。

资料与方法

本研究是基于文献资料开展的 ScR 研究, 系统检索 PubMed、Embase、Web of Science、万方数据知识服务平台、维普数据库、中国知网 6 个数据库, 从建库至 2022 年 12 月 31 日的相关文献。

使用主题词和自由词相结合的检索方式。英文检索式(以 PubMed 为例): (("Observational Studies as Topic" [Mesh]) OR ("Real world" [Title/Abstract]) OR ("non-interventional study" [Title/Abstract])) AND ("Randomized Controlled Trials as Topic" [Mesh]) AND ((effectiveness [Title/Abstract]) OR (efficacy [Title/Abstract])), 中文检索式(以万方数据知识服务平台为例): 题名或关键词: (("随机对照试验" OR "随机临床研究" OR "随机对照临床试验" OR "随机对照临床研究") OR ("观察性研究" OR "真实世界" OR "非干预性研究" OR "疗效比较研究")) AND ("效力" OR "效果"))。

由 2 名研究者独立对文献进行筛选, 对于出现

分歧的文献与第三名研究者讨论后决定。使用定性方法来汇总结果, 将方法相似的文章进行归纳, 对其提出的效力-效果差距弥合方法进行整合综述。

结 果

1. 文献检索: 共检索到 9 943 篇文献, 删除重复文献后, 剩余文献 7 180 篇。通过阅读标题和摘要进行初步筛选, 排除不符合纳入标准的文献 7 098 篇, 阅读初筛纳入的 82 篇文献的全文, 依据纳入排除标准进行第二轮筛选, 并进行手动补充, 最终纳入 10 篇文献, 文献发表时间为 2008-2022 年。

2. RCT 模拟研究思路: 弥合效力-效果差距的思路分为两步, 改进 RWS 方法和规范研究流程, 最大程度减少研究设计差异对结果估计的影响, 得到与 RCT 级别相近的高质量 RWE^[2], 以及探究真实世界中造成效力-效果差距的因素, 并评估其影响, 预测干预措施在不同特征人群中的期望作用。使用 RWD 开展模拟目标试验的观察性研究(RCT 模拟研究)^[8], 越来越受到国内外专家学者的认可和广泛应用。

临床研究结果是为了回答具体的临床问题, 即在什么样的人群中, 哪种干预措施相较于哪种对照措施, 在哪个结局指标方面有什么样的效果, 可简化为 P (Population)、I (Intervention)、C (Comparison)、O (Outcome) 四部分(即 PICO)。当 RCT 与 RWS 在 PICO 定义等方面存在差异, 则其必定会导致结果估计的统计学异质性。其次 RCT 中通过随机化和盲法来最大限度减少混杂偏倚和信息偏倚等对结果估计的影响, 因此设计良好的 RCT 结果被认为是无偏估计。而观察性 RWS 使用的数据往往来自常规收集的医疗数据, 无法遵循随机化和盲法, 因此需要采用统计学方法控制和评估相应混杂对结果估计影响。如果能以 RCT 设计为模板, RWS 与其对照, 可以方便人们更直观地了解 RWS 开展过程中可能存在偏差的部分, 评估仿真模拟差异。最后在正确估计效力-效果差距基础之上, 识别真实世界中影响效力-效果差距的患者或其他因素, 并确定所有用药患者群体中最可能获益的人群, 从而弥合效力-效果差距。因此 RCT 模拟研究第一步也是最为重要的一步是根据目标试验方案结合 RWS 特点构建 RCT 模拟研究方案, 目标试验可以是已发表或注册的 RCT, 也可以是经过专家讨

论形成的理论上的 RCT。然后按照研究方案对 RWD 进行分析获得相应结果,并进行结果间一致性的评估。

3. RCT 模拟研究方案设计框架及关键点:

(1) 纳入排除标准: RCT 研究方案会设置详细的纳入排除标准,以选择预期反应好且干预对其无害的同质性患者。RWS 中通常会使用宽泛的纳入排除标准,以包含所有的真实世界患者,从而保证研究结果的外推性。当使用不同的纳入排除标准时,即使是相同的源人群,也会产生研究对象总体特征的差异,从而导致效力-效果差距。当 RCT 模拟研究纳入排除标准贴近目标试验纳入排除标准时,可以减少因人群特征差异而导致的效力-效果差距。但是受制于原始数据质量, RCT 模拟研究可能无法模拟目标试验的全部纳入排除标准而存在模拟差异,如缺乏相应的实验室检查数据等。此时应对相关情况进行详细解释和说明,可以使用两种研究间因每项纳入排除标准而被排除的患者比例、RWS 中患者平均可测量的基线前期时间长短,以及 RWS 中可获得实验室检查的患者比例等指标评估研究人群模拟差异^[10]。

此外,纳入排除标准的制定还需要考虑观察性研究中可能存在的相关选择偏倚,如当用药者均是治疗早期阶段的幸存者,而排除了因用药早期事件而停止用药的患者,可能会产生相应的现用药者偏倚,新用药者队列设计可以极大程度上避免这一问题^[11-12]。当然,当感兴趣的结果是极其罕见或在长期用药后发生时,使用现用药者也是合理的^[13]。还要注意纳入排除标准中不能包括对基线后事件的限制^[14],如为了提高患者的随访应答率而将患者入组后可用数据时间长度作为纳入排除标准也可能导致偏倚,但是可以对患者入组前的可用数据时间长度进行限定来尽可能选择更容易获得随访数据的患者。

(2) 治疗策略、结局定义: 治疗策略以及结局定义是决定研究间临床异质性的部分,若定义标准存在差异则会反映为研究间结果的统计学异质性。因此当开展 RCT 模拟研究时,治疗策略以及结局的定义必须尽可能贴近目标试验研究方案。

关于治疗策略, RCT 中往往对干预具有明确的定义,除了药品种类还对剂型、剂量、给药方式、用药时长等进行明确规定,而在很多 RWS 中,受制于数据源,干预组与对照组的划分仅考虑了是否使用研究药物,从而会存在可能的模拟差异,产生效

力-效果差距。干预措施定义的模拟应当考虑暴露与结局间的生物学作用机制。暴露与所研究结局之间的生物学作用机制可大体分为 3 种情况: 第一,对于存在明显阈值效应的药物,即药物的收益或风险需要在特定的暴露剂量点出现明显的改变,并在该剂量之后维持相应水平,此时暴露的定义需要考虑累积暴露量与阈值之间的关系; 第二,对于单时点发生的暴露,如外科手术和疫苗接种等,暴露的定义就比较简单,无需考虑剂量相关因素,但是需要考虑暴露后暴露状态随时间的改变,如疫苗接种后人体的抗体滴度会在短时间内达峰,并随着时间的推移而逐渐下降,此时定义暴露期需要考虑疫苗作用产生的时间和作用能够维持的时间; 第三,存在剂量-反应关系的干预措施时,除了考虑暴露是否发生还需要考虑剂量,根据暴露剂量选择研究对象^[13]。

对照组的选择也十分重要,不同对照的选择也可能导致不同的结果估计^[13]。RCT 中最常用的对照为安慰剂对照,干预组与对照组结果的差异可完全归因于暴露。但是 RWS 中很难设置安慰剂对照,因此组间结果差异并非完全为干预措施的作用。最好能通过确定接受标准护理或治疗的患者添加或不添加感兴趣的干预措施来模拟安慰剂对照试验。然而,在 RWD 中很难定义标准的护理或治疗^[10]。如果直接使用空白对照代替安慰剂对照,则干预组与对照组患者在疾病严重程度与不良事件风险方面会具有很大差异,可能导致混杂偏倚。相关文献仍推荐在评估干预措施的有效性与安全性时,使用阳性对照可能更稳妥^[13],如在 RCT Duplicate 项目中,对于安慰剂对照的 RCT,在 RWS 中均选用了既往文献未发现对心血管疾病有风险或具有保护作用的阳性对照药物^[10]。而当有必要选择空白对照时,需要更加重视随访零时刻的选取以控制时间相关的偏倚。

RCT 与 RWS 结果的差异还可能是由于结局指标定义的差异所导致。比如降糖药与低血糖等不良事件安全性的 RCT 模拟研究^[15],相较于 RCT,使用医疗索赔数据库只能捕捉到因低血糖就诊的较严重事件,而不太可能捕捉到不太严重的低血糖事件,这就会导致模拟差异,进而产生结果估计的不一致性。而心血管不良事件由于均需要有明确的就诊或诊断记录,结局定义可完全模拟,因此研究间结果的估计值置信区间重叠度较大。但是 RWS 仍受限于数据本身的特性而存在模拟差异,对于结

果的模拟差异,可以使用并排 Kaplan-Meier 图来进行评估^[10]。

(3) 随访零时刻确定: 随访期的定义, 对结果估计也有很大影响。研究对象开始随访的时间称为零时刻, 零时刻的选取对于控制时间相关性偏倚十分重要^[16]。在 RCT 中, 患者随机化分组和干预给予时点往往是统一的, 很少产生时间偏倚。而在观察性研究中, 通常入组时点和干预时点是不同的, 比如以满足入组条件的时点(通常是诊断时点)作为零时刻, 而处方时点往往会晚于相关诊断时点, 因此通常需要设置一定的宽限期, 根据零时刻后宽限期内相关处方记录将患者划分为干预组或对照组。但是干预组患者若在处方时间前发生了结局事件便会被错误划分至空白对照组而导致永恒时间偏倚和时间滞后偏倚^[16]。为了减少时间相关偏倚, 可以在新用药者队列设计基础上研究能度过永恒时间的患者, 即零时刻设置为干预给予时点^[17], 但是这种方法仅适用于阳性对照研究。而对于空白对照, 有研究证明, 使用基于时间的匹配可能获得更为准确的结果估计^[18], 包括以入组时点作为零时刻, 为每个干预组个体匹配空白对照个体, 或者计算干预组个体入组时点和接受干预措施时点间的间隔, 并以此为限制, 为干预组匹配具有相同时间间隔且基线特征可比的空白对照组。此外还需要考虑研究对象在研究窗内有多个满足条件的零时刻, 比如药物干预研究中, 患者在满足相关诊断和其他纳入排除标准后进入队列, 若一开始未被分配相应的干预则可能被分入对照组, 而在其后的随访过程中又使用研究药物则又作为干预组, 此时可能会有几个零时刻。针对这种情况可以仅纳入第一个或随机的一个零时刻, 或者是模拟多个嵌套试验, 即在多个合适时间点进行多次试验模拟, 这种方法虽然有更好的统计效能, 但是一个人可能被多次纳入分析而需要对方差进行调整^[14]。更为认可的方法是使用“模拟-删失-加权”方法^[14], 即为每一个人组患者(真实患者), 均产生一个克隆患者, 并将双方分配到不同的干预组, 在指定的宽限期内当真实患者接受的干预措施偏离既定分组干预措施时, 则将克隆患者做删失处理, 若在宽限期内发生结局事件或死亡等删失事件, 则对两个患者均做删失处理。然而, 由于这种删失不是随机的, 可能会导致选择偏倚, 这种潜在偏倚可以在分析中使用逆删失概率加权来解决。

(4) 因果分析策略与统计分析: RCT 中常用的

因果对比包括意向治疗(ITT)分析、符合方案(PP)分析或者实际治疗(AT)分析策略。采用不同的因果对比也会产生不同的结果。因此 RCT 模拟研究需要根据 RCT 所采用的因果对比策略而选择不同的统计分析方法。

ITT 分析策略是无论患者是否接受既定干预均保留在原组内进行的分析, 并不考虑实际治疗过程中的不依从, 比较的是治疗分配的效应, 其优势在于保持组间基线的可比。ITT 分析策略是建立在 RCT 良好随机化和依从性基础之上的一种分析策略, 而在观察性研究中无法进行随机化分组, 因此需采用相关的方法来减少组间基线的差异。在研究设计阶段可以使用限制和匹配的方法, 在数据分析阶段可以使用标准化、分层分析和多因素分析方法^[12-14]。但是越来越多的 RWS 依托于数据电子医疗大数据而展开, 具有更广的数据维度, 需要校正的协变量更加复杂, 此时基于倾向性评分的相关方法可能更加合适^[15], 如使用基于倾向性评分的个体化匹配或加权, 可以调整组间基线的不可比, 很好地模拟 RCT 中的随机化分组。但是, 倾向性评分仅能校正预先确定的基线协变量, 协变量的选择多依靠既往经验, 因此通常纳入较多的基线协变量。然而当包含过多仅与暴露相关协变量或无关的变量, 可能会使方差膨胀, 此时使用高维倾向性评分可能是一个不错的选择^[16]。已有研究探讨在 RCT 模拟研究中使用高维倾向性评分方法, 根据协变量可能造成的偏倚对协变量进行排序, 随后选取前 k 个协变量构建倾向性评分模型。

采用 PP 分析或 AT 分析策略时, 应当采用时依性变量校正的相关方法, 以调整依从性。当仅考虑暴露的时变性, 而认为各时点内协变量独立时, 可以使用时间依赖 Cox 比例风险回归模型以进行相关的生存分析^[19], 计算各时间段内的生存函数, 最后根据时间长短对效应值进行加权计算总效应值。当同时考虑暴露和混杂的时变性, 则应采用相应的 g 方法^[20], 包括 G 公式、逆概率加权的边际结构模型(IPTW-MSM), 以及 G-估计。其中 IPTW-MSM 是最常用和最便于计算的方法, 该方法基于反事实模型, 给每个观测者分配权重来重新获得一个混杂在组间均衡分布的伪人群。则两组间结局事件发生率的差异为干预措施的效应值。其分为两步, 首先根据各随访点的暴露和混杂因素计算权重。然后代入到边缘结构模型拟合回归, 从而估计干预措施的效应值。 g 方法应用的一个重要前提是假定不存

在未测量的混杂因素,目前对于估计未观测混杂的潜在影响方法尚不成熟,寻找合适的工具变量进行定量评估较为困难,但可以通过 E 值^[18]和阴性对照^[19]等方法进行矫正或定性评估剩余混杂。

此外还要详细描述除主分析以外的亚组分析和敏感性分析策略。开展 RWS 的目的之一是为医疗决策提供 RWE 支撑,而证据决策应当面向全部患者群体。在 RCT 模拟研究主分析中,仅验证了干预措施在总人群的平均效果,而未考虑人群内亚组收益风险比的异质性,很难预测个体的收益风险曲线^[13]。因此主分析后,还需要考虑可能的异质性因素进行充分的亚组分析^[21-22],探索干预措施在不同特征患者亚组之间的疗效差异,也是弥合效力-效果差距的重要手段。通常亚组分析可以分为验证性亚组分析、支持性亚组分析、探索性亚组分析,其中验证性亚组分析和支持性亚组分析通常需要在研究方案中进行明确的定义(事先定义),而探索性亚组分析既可以事先定义也可以在数据分析阶段定义(事后定义)。事先定义的亚组分析检验效能强,因果论证强度高。

进行亚组分析的重要前提是确定分组变量,分组因素应该是对暴露具有效应修饰作用的协变量,亚组分析分组变量的选择往往是基于既往研究和相关文献确定,通常包括人口学特征(如年龄、性别等)、实验室检查指标、基因组相关标志物、疾病的严重程度或分型、临床状况(如合并症、伴随用药)、地区(如国家、试验中心)和环境因素等。2021 年 Sagkriotis 等^[23]以新生血管性年龄相关黄斑变性的玻璃体内抗血管内皮生长因子(anti-VEGF)治疗为例,提出了一种基于机器学习方法识别影响干预措施真实世界效果患者特征的方法。主要步骤:①确定真实世界中使用 anti-VEGF 治疗且达到与 RCT 中干预组患者平均治疗效果的患者集;②使用机器学习的随机森林方法识别真实世界中影响 anti-VEGF 治疗效果的患者特征;③基于上一步骤选择的变量进一步建立决策树模型,并使用总数据集对决策树模型进行训练;④最终使用外部数据对该模型进行验证。此方法可以为确定 RCT 模拟研究的亚组分析分组变量确定提供思路。

开展亚组分析时需要注意以下几点问题:①样本量:使用事先定义亚组分析时,需要在研究设计中进行亚组样本量的计算,只有达到预先计算的样本量,亚组分析的结果才具有稳健性,因此为了保证亚组分析的样本量,主分析人群的样本量通常为

理论样本量的数倍^[24]。②亚组分析结果的可信度:包括亚组是否预先定义,事后定义亚组分析可能存在为了获得阳性结果而过度分析的嫌疑。以及亚组变量的选择是否具备生物学上的合理性等问题^[22]。③多重性问题:亚组分析往往包括一系列分析,涉及多次检验,因此如果不进行多重性调整,会增加总 I 类错误率^[22]。④报告的透明性:在文章撰写和结局报告时,需要对预先定义的亚组分析结果进行全面的汇报,避免报告偏倚^[25]。

由于观察性 RWS 中因果推断结论具有不确定性,因此结论的稳健性尤其重要,充分的敏感性分析有助于判断结论的稳健性。敏感性分析包括但不限于:不同数据处理方法;不同数据集定义;不同分析方法;模型中不同的协变量组合;未知或不可测的混杂因素影响等^[22-23]。

(5)初步可行性分析及研究方案注册:制定研究方案后,应进行初步的可行性分析。这是因为应用纳入排除标准进行人群筛选以及使用匹配或加权等方法校正基线混杂后,可能有大量个体被排除从而影响结果的稳健性和外推性。可以在筛选完研究对象后,使用简单的协变量进行匹配或加权,判断 RCT 模拟研究最终纳入的研究对象数量能否满足统计效能,若样本量过小,则应终止研究。

若初步可行性分析支持该 RCT 模拟研究的开展,为了避免研究结果的选择性报告,增加研究透明度,应在研究开始前在相关网站上进行注册并公布相应的研究方案,研究方案的撰写可参照临床研究方案的标准条目 SPIRIT 声明^[26]。

4. RCT-RWE 结果一致性判断及提高研究报告质量:按照研究方案完成 RCT 模拟研究后,需要对相应的结果进行发表,以供医护人员、患者、医疗决策部门的利益相关方进行参考。当选择的目标试验是已经开展和发表过的相关研究,可以将目标试验结果和 RCT 模拟研究结果联合起来进行报告,根据统计学异质性去寻找 RCT 模拟研究与目标试验之间存在的临床异质性以及方法学异质性,以帮助人们更好地了解 RCT 模拟研究中需要考虑的因素。

由于存在随机误差,即使不存在模拟差异时,RCT 模拟研究结果也无法与目标试验结果完全相等,因此仅定量判断 RCT-RWE 结果差异缺乏现实意义。RCT Duplicate 项目组提出了 RCT-RWE 结果一致性的 3 个定性评估指标:①监管一致性:RCT 模拟研究的方向和统计显著性;②估计值一致性:RWS 效应估计值是否在 RCT 结果的 95%CI 范围

内;③标准化差异:使用 Z 检验判断 RWS 结果与 RCT 结果间是否存在显著性差异^[27]。其中监管一致性是最严格的标准,标准化差异的标准最容易满足^[27]。

撰写稿件时,为了提升研究的报告质量,使研究更加完整、透明和清晰,应参照相应的医学研究报告规范。目前主要包括常规收集医疗卫生数据开展观察性研究的 STROBE 报告规范^[28]、加强流行病学观察性研究的 RECORD 报告规范^[29]或专门为观察性药物流行病学研究改进的 RECORD-PE 报告规范^[30]。此外,当进行同行评议评估 RCT 模拟研究质量时,可以参考已被广泛应用的非随机干预研究偏倚风险评估工具 ROBINS-I 评价其方法学质量^[31]。

讨 论

RCT 与 RWS 间效力-效果的差距日渐受到各利益相关方的重视,本研究采用 ScR 对此话题进行整合分析。

RCT 与 RWS 往往在研究设计的关键特征上存在较大差异,可能导致相应的统计学异质性。RCT 模拟研究以 RCT 为蓝本,制定与 RCT 相近的研究方案,减少了与 RCT 在研究设计上的差异,规范了观察性 RWS 的方法和步骤,从而有利于增强 RCT 与 RWS 间结果的可比性,弥合效力-效果差距。

但是 RCT 模拟研究本质仍然是观察性研究,仍然无法完全避免观察性研究本身存在的问题,可能会存在混杂和模拟差异。因此,需要进一步改进 RCT 模拟研究相关方法,提高 RCT 模拟研究质量。并在发表相关研究结果时,注意报告的规范性与透明性,方便各利益相关方在利用 RWE 时评估证据的质量。

利益冲突 所有作者声明无利益冲突

作者贡献声明 刘佐相:研究设计、论文撰写;龙子临、许欣然:论文撰写;杨智荣、石舒原、赵厚宇、杨祖耀、符祝、宋海波、林腾飞、詹思延:研究指导、论文修改;孙凤:研究设计/指导、论文修改、经费支持

参 考 文 献

- [1] 王拥军. 积极开展真实世界的药物疗效评价[J]. 中国药物评价, 2012, 29(1): 1-2. DOI: 10.3969/j.issn.2095-3593.2012.01.001.
- [2] Wang YJ. Active evaluation of drug efficacy in the real world[J]. Chin J Drug Eval, 2012, 29(1): 1-2. DOI:10.3969/j.issn.2095-3593.2012.01.001.

- [2] Eichler HG, Abadie E, Breckenridge A, et al. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response[J]. Nat Rev Drug Discov, 2011, 10(7):495-506. DOI:10.1038/nrd3501.
- [3] Sites BD, Masaracchia MM, Davis M. Distinguishing between efficacy and real-world effectiveness:the case for thinking beyond classic randomized controlled trial design[J]. Reg Anesth Pain Med, 2017, 42(2): 131-132. DOI:10.1097/AAP.0000000000000563.
- [4] Grimberg F, Asprion PM, Schneider B, et al. The real-world data challenges radar: a review on the challenges and risks regarding the use of real-world data [J]. Digit Biomark, 2021, 5(2): 148-157. DOI: 10.1159/000516178.
- [5] Schad F, Thronicke A. Real-world evidence-current developments and perspectives[J]. Int J Environ Res Public Health, 2022, 19(16): 10159. DOI: 10.3390/ijerph191610159.
- [6] Bothwell LE, Podolsky SH. The emergence of the randomized, controlled trial[J]. N Engl J Med, 2016, 375(6):501-504. DOI:10.1056/NEJMp1604635.
- [7] Nordon C, Karcher H, Groenwold RHH, et al. The "efficacy-effectiveness gap": historical background and current conceptualization[J]. Value Health, 2016, 19(1): 75-81. DOI:10.1016/j.jval.2015.09.2938.
- [8] Thompson D. Replication of randomized, controlled trials using real-world data: what could go wrong? [J]. Value Health, 2021, 24(1): 112-115. DOI: 10.1016/j.jval.2020.09.015.
- [9] 王靖雯, 邵明义, 符宇, 等. 国内真实世界临床研究方案设计的分析与思考 [J]. 中华中医药杂志, 2022, 37(4): 2125-2128.
- [9] Wang JW, Shao MY, Fu Y, et al. Analysis and reflection on the design of real world study scheme in China[J]. China J Tradit Chin Med Pharm, 2022, 37(4):2125-2128.
- [10] Franklin JM, Patorno E, Desai RJ, et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE initiative[J]. Circulation, 2021, 143(10): 1002-1013. DOI:10.1161/CIRCULATIONAHA.120.051718.
- [11] Ray WA. Evaluating medication effects outside of clinical trials:new-user designs[J]. Am J Epidemiol, 2003, 158(9): 915-920. DOI:10.1093/aje/kwg231.
- [12] Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application[J]. Curr Epidemiol Rep, 2015, 2(4):221-228. DOI:10.1007/s40471-015-0053-5.
- [13] Velentgas P, Dreyer NA, Nourjah P, et al. Developing a protocol for observational comparative effectiveness research: a user's guide[M]//Velentgas P, Dreyer NA, Nourjah P, et al. Developing a protocol for observational comparative effectiveness research: a user's guide[M]. Rockville: Agency for Healthcare Research and Quality, 2013.
- [14] Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available[J]. Am J Epidemiol, 2016, 183(8): 758-764. DOI: 10.1093/aje/kwv254.
- [15] Patorno E, Schneeweiss S, Gopalakrishnan C, et al. Using real-world data to predict findings of an ongoing phase iv

- cardiovascular outcome trial: cardiovascular safety of Linagliptin versus glimepiride[J]. *Diabetes Care*, 2019, 42(12):2204-2210. DOI:10.2337/dc19-0069.
- [16] Hernán MA, Sauer BC, Hernández-Díaz S, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses[J]. *J Clin Epidemiol*, 2016, 79:70-75. DOI:10.1016/j.jclinepi.2016.04.014.
- [17] 曾琳, 卓琳, 赵一鸣. 永恒时间偏倚[J]. *中华儿科杂志*, 2021, 59(2): 149. DOI: 10.3760/cma.j.cn112140-20201206-01082.
- Zeng L, Zhuo L, Zhao YM. Immortal time bias[J]. *Chin J Pediatr*, 2021, 59(2):149. DOI:10.3760/cma.j.cn112140-20201206-01082.
- [18] Wakabayashi R, Hirano T, Laurent T, et al. Impact of "time zero" of follow-up settings in a comparative effectiveness study using real-world data with a non-user comparator: comparison of six different settings[J]. *Drugs Real World Outcomes*, 2023, 10(1): 107-117. DOI: 10.1007/s40801-022-00343-1.
- [19] 王振宇, 陈朔华, 赵欣宇, 等. Cox 及其拓展模型在基于队列的依时暴露因素效应估计中的应用[J]. *中华流行病学杂志*, 2020, 41(6):957-961. DOI:10.3760/cma.j.cn112338-20200119-00046.
- Wang ZY, Chen SH, Zhao XY, et al. Application of Cox and extended regression models on modeling the effect of time-updated exposures in cohort studies[J]. *Chin J Epidemiol*, 2020, 41(6): 957-961. DOI: 10.3760/cma.j.cn112338-20200119-00046.
- [20] Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposures[M]//Fitzmaurice G, Davidian M, Verbeke G, et al. *Longitudinal data analysis*[M]. Boca Raton: CRC Press, 2009:553-599.
- [21] 国家药品监督管理局药品审评中心. 药物真实世界研究设计与方案框架指导原则(试行)[R]. 北京:国家药品监督管理局药品审评中心, 2023.
- Center for Drug Evaluation, NMPA. Guidelines for drug real-world research design and protocol framework (trial) [R]. Beijing:Center for Drug Evaluation, NMPA, 2023.
- [22] 国家药品监督管理局药品审评中心. 药物临床试验亚组分析指导原则(试行)[R]. 北京:国家药品监督管理局药品审评中心, 2020.
- Center for Drug Evaluation, NMPA. Guidelines for subgroup analysis of drug clinical trials (trial)[R]. Beijing: Center for Drug Evaluation, NMPA, 2020.
- [23] Sagkriotis A, Chakravarthy U, Griner R, et al. Application of machine learning methods to bridge the gap between non-interventional studies and randomized controlled trials in ophthalmic patients with neovascular age-related macular degeneration[J]. *Contemp Clin Trials*, 2021, 104: 106364. DOI:10.1016/j.cct.2021.106364.
- [24] Dijkman B, Kooistra B, Bhandari M. How to work with a subgroup analysis[J]. *Can J Surg*, 2009, 52(6):515-522.
- [25] Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation[J]. *Lancet*, 2005, 365(9454): 176-186. DOI: 10.1016/S0140-6736(05)17709-5.
- [26] 钟丽丹, 郑颂华, 吴泰相, 等. SPIRIT 2013 声明:定义临床研究方案的标准条目[J]. *中国中西医结合杂志*, 2014, 34(1): 115-122. DOI:10.7661/CJIM.2014.01.0115.
- Zhong LD, Zheng SH, Wu TX, et al. Standard protocol items: recommendations for interventional Trials 2013[J]. *Chin J Integr Tradit West Med*, 2014, 34(1):115-122. DOI: 10.7661/CJIM.2014.01.0115.
- [27] Franklin JM, Pawar A, Martin D, et al. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project [J]. *Clin Pharmacol Ther*, 2020, 107(4): 817-826. DOI: 10.1002/cpt.1633.
- [28] 王家莹, 付姝菲, 段玉婷, 等. 卫生保健仿真研究报告规范: 基于 CONSORT 和 STROBE 声明的扩展[J]. *中国循证医学杂志*, 2021, 21(2): 241-248. DOI: 10.7507/1672-2531.202008176.
- Wang JY, Fu SF, Duan YT, et al. Translated title of the contribution: reporting guideline for health care simulation research: extensions to the CONSORT and STROBE statements[J]. *Chin J Evidence Based Med*, 2021, 21(2):241-248. DOI:10.7507/1672-2531.202008176.
- [29] 廖星, 章轶立, 谢雁鸣. 真实世界研究标准: RECORD 清单和 GRACE 清单的解读[J]. *中国中药杂志*, 2015, 40(24): 4734-4738. DOI:10.4268/cjcmm20152403.
- Liao X, Zhang YL, Xie YM. An introduction of RECORD and GRACE checklist for studies in real world[J]. *China J Chin Mater Med*, 2015, 40(24): 4734-4738. DOI: 10.4268/cjcmm20152403.
- [30] 聂晓璐, 武泽昊, 赵厚宇, 等. 使用常规收集医疗卫生数据开展观察性研究的报告规范(药物流行病学版) RECORD-PE 规范中文版(上)[J]. *药物流行病学杂志*, 2019, 28(3):190-198, 212.
- Nie XL, Wu ZH, Zhao HY, et al. Chinese version for Reporting of studies conducted using observational routinely collected health data for pharmacoepidemiology, RECORD-PE (Part 1) [J]. *Chin J Pharmacoepidemiol*, 2019, 28(3):190-198, 212.
- [31] 王浩, 唐晓宇, 王和平, 等. ROBINS-I:评估非随机干预性研究偏倚风险的新工具[J]. *中国循证心血管医学杂志*, 2018, 10(7): 789-793. DOI: 10.3969/j. issn. 1674-4055.2018.07.05.
- Wang H, Tang XY, Wang HP, et al. ROBINS-I: A new tool for assessing the risk of bias in non-random intervention Studies[J]. *Chin J Evidence Based Cardiovasc Med*, 2018, 10(7): 789-793. DOI: 10.3969/j. issn. 1674-4055.2018.07.05.