

真实世界证据与随机对照试验：RCT DUPLICATE 项目启动、实施、进展解读 与启示(一)

石舒原^{1,2} 刘佐相¹ 赵厚宇¹ 聂晓璐^{1,3} 符祝⁴ 宋海波^{5,6} 姚晨^{4,7} 詹思延^{1,8} 孙凤^{1,4}

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中国康复科学所,中国残联残疾预防与控制中心,北京 100068;³国家儿童医学中心,首都医科大学附属北京儿童医院临床流行病与循证医学中心,北京 100045;⁴海南省真实世界数据研究院,博鳌乐城国际医疗旅游先行区管理局,乐城 571437;⁵国家药品监督管理局药品评价中心,北京 100022;⁶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 100022;⁷北京大学临床研究所,北京 100191;⁸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】 近些年,医疗产品监管机构开始重新审视真实世界证据(RWE)对监管决策的潜在价值。RWE 能否代替金标准随机对照试验(RCT)产生的证据尚不确定。哈佛大学研究团队于 2018 年发起了 RCT DUPLICATE 项目,旨在利用医疗索赔数据库模拟 30 个 RCT,以探索效力-效果差距的量化方法并解释其潜在来源,增强 RWE 的可信度。本文回顾了 RCT DUPLICATE 项目的产生背景,重点介绍 RCT DUPLICATE 项目的研究目的、研究设计和实施流程,以期帮助国内学者更好地理解 RWE 的适用范围和应用价值。

【关键词】 真实世界证据; 真实世界数据; 模拟

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);海南省博鳌乐城先行区真实世界研究专项计划(HNLC2022RWS012)

Real-world evidence and randomized controlled trials: the initiation, implementation, progress interpretation and revelation of RCT DUPLICATE (part 1)

Shi Shuyuan^{1,2}, Liu Zuoxiang¹, Zhao Houyu¹, Nie Xiaolu^{1,3}, Fu Zhu⁴, Song Haibo^{5,6}, Yao Chen^{4,7}, Zhan Siyan^{1,8}, Sun Feng^{1,4}

¹Department of Epidemiology and Biostatistics, Peking University School of Public Health, Beijing 100191, China; ²China Institute of Rehabilitation Sciences, Center for Prevention and Control of Disability of China Disabled Persons Federation, Beijing 100068, China; ³Center for Clinical Epidemiology and Evidence-based Medicine, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China; ⁴Hainan Institute of Real World Data, the Administration of Boao Lecheng International Medical Tourism Pilot Zone, Lecheng 571437, China; ⁵Center for Drug Reevaluation, National Medical Products Administration, Beijing 100022, China; ⁶Key Laboratory for Research and Evaluation of Pharmacovigilance, National Medical Products

DOI:10.3760/cma.j.cn112338-20220513-00408

收稿日期 2022-05-13 本文编辑 万玉立

引用格式:石舒原,刘佐相,赵厚宇,等.真实世界证据与随机对照试验:RCT DUPLICATE 项目启动、实施、进展解读与启示(一)[J].中华流行病学杂志,2022,43(11):1828-1834. DOI:10.3760/cma.j.cn112338-20220513-00408.

Shi SY, Liu ZX, Zhao HY, et al. Real-world evidence and randomized controlled trials: the initiation, implementation, progress interpretation and revelation of RCT DUPLICATE (part 1)[J]. Chin J Epidemiol, 2022, 43(11):1828-1834. DOI:10.3760/cma.j.cn112338-20220513-00408.



Administration, Beijing 100022, China; ⁷Peking University Clinical Research Institute, Beijing 100191, China; ⁸Clinical Epidemiology Research Center, Peking University Third Hospital, Beijing 100191, China
Corresponding author: Sun Feng, Email: sunfeng@bjmu.edu.cn

【 Abstract 】 In recent years, researchers, pharmaceutical companies, and political makers gradually using more real-world data (RWD) to produce real-world evidence (RWE) for policy-making. A research team of Harvard University launched the RCT DUPLICATE project in 2018, aiming to replicate 30 randomized controlled trials using the medical claims database in order to explore methods for quantifying the efficacy-effectiveness gap and explain its potential sources, to enhance the credibility of the RWE. This paper reviews the background of RCT DUPLICATE Initiative, highlights the research purposes, research design and implementation process of the RCT DUPLICATE Initiative, to help domestic scholars better understand the scope and application value of RWE.

【 Key words 】 Real-world evidence; Real-world data; Emulation

Fund programs: National Natural Science Foundation of China (72074011); The Second Batch of Key Projects of Scientific Act for Drug Regulation of China ([2021]37-10); Real World Study Project of Hainan Boao Lecheng Pilot Zone (HNL2022RWS012)

一、概述

1. 研究背景：随机对照试验 (randomized controlled trial, RCT) 一直被视为评估药物疗效的金标准。传统 RCT 有诸多优点, 但需要严格的纳入排除标准、研究样本和随访时间有限, 很难用于研究药物相关的罕见不良事件, 真实世界证据 (real-world evidence, RWE) 被逐渐提出用于补充 RCT 证据, 涉及定义和监测疾病状况、卫生技术评估、为临床研究设计提供信息、提高试验效率、临床使用和批准后评估药物效果, 以及支持药品全生命周期监管决策等^[1-3]。2016 年美国《21 世纪治愈法案》的公布让各方利益相关者开始关注 RWE 用于监管决策的潜在用途^[1]。

真实世界数据 (real-world data, RWD) 是指从各种来源定期收集的, 与患者健康状况或医疗服务提供有关的数据^[2]。RWE 是指基于 RWD 通过严格的研究设计和统计分析获得医疗产品潜在获益或风险的临床应用证据^[2]。迄今发表的基于 RWD 开展的观察性研究大多依赖于医疗数据库, 这些数据库可持续收集患者的医疗卫生相关信息, 包括医疗保险索赔数据、电子健康记录 (electronic health record, EHR) 和一些注册登记信息。基于这些数据库开展的回顾性研究通常无法随机化分组, 因此, 面对根据自我意愿和用药指征接受不同治疗的患者, 当对各种干预措施进行比较时, 如何控制选择性偏倚和混杂偏倚以提高 RWE 结果的真实性是主要挑战。

2. RCT DUPLICATE (Randomized, Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of

Epidemiology) 项目: 该项目是 2018 年由美国食品药品监督管理局 (FDA)、哈佛大学布莱根妇女医院 (Brigham and Women's Hospital) 以及一家医疗保健技术 (Aetion) 公司共同牵头发起的全球较早的由监管部门资助的 RCT 模拟 (英文中的“replication、emulation 和 simulation”, 统一用模拟来表示) 的项目^[3], 旨在利用 3 个医疗索赔数据模拟 30 个 RCT。团队成员包括 FDA、Aetion 公司员工, 项目的专家顾问委员会核心小组成员来自哈佛大学布莱根妇女医院, 包括 RCT 和 RWD 研究设计、统计分析等法学方面的顶级专家, 涉及领域主要包括生物统计学、药物流行病学以及计算机应用。

该项目的目的是评估 RWE 和 RCT 的疗效评估是否存在差异, 如若存在差异, 进一步探索差异的大小以及产生的原因, 便于充分考虑潜在的相关因素, 以帮助决定 RWE 是否可以用于支持监管决策^[4]。主要包括 3 个部分: ①确定在新药审批以及药品上市后适应证扩展环节, RWE 是否可以补充甚至取代作为金标准的 RCT 以及如何起到充分补充作用; ②初步探讨何种类型的临床问题可以利用 RWD 进行分析, 以及如何利用 RWD 进行相关研究设计和统计分析, 进而协助 FDA 继续制定在监管决策中使用 RWE 的标准, 即确定 RWD 在何种使用条件下可为新药安全性和有效性评估提供证据补充; ③规范 RWE 研究流程步骤及相关统计分析策略的建立。

项目组基于医疗索赔数据库设计相应的 RWE 研究来匹配待模拟的 RCT 的关键研究特征, 并设计了一个实施流程细则, 以确保模拟 RCT 研究的过程具有一致性、透明化和可重复性。这是第一次系统

地评估基于 RWD 使用结构化过程设计 RWE 研究模拟多个 RCT 的能力。为了提高 RCT 的可重复性,项目组选择大样本且治疗方案相对简单的 RCT 开展相关的 RWE 研究^[5-6]。

RCT DUPLICATE 项目的主要研究内容为基于 RWD 分析,通过标准化、透明化、规范化、合理化的研究设计,产生 RWE,并与选取的相应 RCT 结果进行比较,为 RWD 的使用建立经验证据基础,具体包括 4 个子项目:①模拟已经完成 III/IV 期临床试验的研究;②预测正在进行 IV 期临床试验的研究结果;③基于纵向索赔数据库应用流行病学方法设计实施模拟心血管领域的 RCT;④其他试验的模拟验证。见表 1^[7]。

该项目的意义在于,如果基于医疗索赔数据库的 RWE 研究与已经发表的 RCT 研究结果相一致,同时能够准确地预测正在进行的 RCT 结果,将有助于增强大众尤其是监管机构对 RWE 的信任度,在未来缺少 RCT 证据的情况下,加快新药的上市审批流程,减少新疗法推向市场的时间和成本。进一步也可以基于 RWE 来拓展药物适应证范围,推动 RWE 在 FDA 决策中的应用,使得实时更新捕获风险-获益信息成为可能,实现对医疗产品全生命周期的监管。

二、研究方法

1. RCT 筛选过程:并非所有 RCT 都可利用 RWD 进行模拟,需要选择合适的 RCT 并寻找与之相匹配的数据源。RCT DUPLICATE 项目组考虑选

择与监管决策相关的 RCT,包括已发表或已注册正在进行的,而且可获取相匹配的 RWD 来源以确保这些 RCT 能被模拟。项目组尝试涵盖尽可能多的治疗领域的 RCT,但是发现在某些治疗领域并不适合进行模拟,特别是那些主要依靠症状体征或替代指标作为结局的 RCT,其结局指标无法在医疗索赔数据库中获取。此外,项目组选取的 RCT 多为大规模和高功效的 RCT,以保证研究结果的可靠性。

项目组通过两个步骤选出最终合适的 RCT:①初筛确定候选 RCT;②二次筛选确定最终符合项目要求的 RCT。

初筛旨在寻找多样化的 RCT,以考察其在 RWD 中的可重复性。这一阶段主要兼顾了 4 个方面^[8]:①对照类型(阳性对照或安慰剂+标准治疗对照);②结局效应(效应大小);③试验目的(优效或非劣效);④临床领域(尽可能多地覆盖不同临床领域)。初筛的 RCT 来源可分为两类:①从获批上市的新药研究、补充药物适应证的研究和 FDA 推荐的研究中选择已发表了研究结果的 RCT;②从临床试验注册网站(<https://www.clinicaltrials.gov/>)和 FDA 建议中选择正在进行且尚未发布结果的 RCT。

二次筛选依据项目组事先制定的标准进行,基于现有数据库选择待研究 RCT 的标准:①RCT 关注的结局变量在现有数据库中有诊断或测量;②能保证测量的准确性,如所关注的临床结局事件为心肌梗死、中风、骨折、大出血或死亡等诊断明确的临床事件,在数据库中一般会有清晰的记录且不会掺杂

表 1 RCT DUPLICATE 项目

| 项目 | 资助方 | 任务 | 意义 |
|-----------------------------------|-------|--|--|
| 模拟已经完成 III/IV 期临床试验的研究 | FDA | ①确定用于模拟的 III/IV 期临床试验; ②建立相应的数据库分析平台; ③基于 RWD 模拟 RCT,分析比较两者的结果一致性; ④组织研讨会,总结阶段性经验 | 回答何时何种情况可以有效应用 RWE 支持决策;增强监管信心,指导补充性新药审批 |
| 预测正在进行 IV 期临床试验的研究结果 | FDA | ①确定用于预测的 IV 期临床试验; ②建立相应的数据库分析平台; ③基于 RWD 模拟 RCT,分析比较两者的结果一致性; ④组织研讨会,总结阶段性经验 | 提供前瞻性经验证据;提供结构化、经过测试的分析模板,强化开展相关研究信心和透明度 |
| 基于纵向索赔数据库应用流行病学方法设计实施模拟心血管领域的 RCT | NHLBI | ①确定用于模拟的 RCT; ②尝试性探索多种设计和分析方法,比较各方法的结果异同; ③使用 Meta 分析探索影响因素; ④组织研讨会,总结阶段性经验 | 进一步扩展可应用的设计、分析方法;制定适用于此类研究的方法学指南 |
| 其他试验的模拟验证 | | 前瞻性模拟 4 个正在进行的 RCT: ①CAROLINA; ②DECLARE-TIMI 58; ③ENTRACTE; ④CANVAS | 确定 RCT 结果是否可以通过这种方法进行预测 |

注:NHLBI:美国国家心肺血液研究院;RWD:真实世界数据;RWE:真实世界证据;RCT:随机对照试验;FDA:美国食品药品监督管理局;CAROLINA:利格列汀与格列美脲对 2 型糖尿病患者心血管事件的研究;DECLARE-TIMI 58:达格列净对心血管事件发生率影响的多中心研究;ENTRACTE:托珠单抗与依那西普在类风湿关节炎和心血管疾病危险因素患者中的比较研究;CANVAS:卡格列净心血管评价研究

主观因素的干扰,这类 RCT 在 RWD 中的模拟将会得到更可靠的结果;③能准确地捕捉到暴露的信息;④对重要的混杂进行了准确测量并可用于调整。不满足上述任一标准的 RCT 即被排除。

项目组排除:①HIV 或癌症治疗相关的 RCT,因为这些疾病治疗药物的评估需要更详细的临床数据,而这些数据在医疗索赔数据库中是无法获得的;②疫苗相关的 RCT;③完全基于患者自我报告的指标(如疼痛、症状缓解、功能状态改善等)、替代指标(如实验室指标)以及索赔数据中无法获取相关指标的 RCT。总的来说,项目组主要依据主要结局的可重复性,排除难以模拟的 RCT。

结合上述两个步骤,项目组最终确定了 40 个 RCT,包括 33 个已完成并发表结果的 RCT 和 7 个仍在进行但尚未得到结果的 RCT^[8]。在进行初步的可行性和功效分析后,发现可能会有 RCT 无法进行模拟,因此项目组共选择了 40 个 RCT,即使项目初期一些 RCT 由于可行性较差被排除,也能确保项目组完成预设的任务目标,即模拟 30 个 RCT。在这 40 个 RCT 中,24 个为上市后研究且关注常规临床中患者可用的药物。有 25 个为优效性试验,其中 19 个成功显示出了优效性,2 个没有发现优效性但满足了预先规定的非劣效界值,4 个没有发现优效性,也没有达到预先规定的非劣效界值。在 15 个非劣效试验中,7 个在主要结局指标发现了优效性,其余的 8 个显示出非劣效结果。尽管项目组优先考虑选择阳性对照试验,但最终纳入的 RCT 中仅有 17 个为阳性对照组的试验;其余 23 个试验均以安慰剂联合标准治疗作为对照。

最终选择的 40 个 RCT 涵盖了多个治疗领域,其中包括降糖药(17.5%)、直接口服抗凝剂(15.0%)、抗血小板药物(17.5%)、抗高血压药物(10.0%)、抗骨质疏松药物(10.0%)、抗哮喘药物(10.0%)、慢性阻塞性肺疾病治疗药物(7.5%)、抗心力衰竭药物(7.5%)、降脂药物(2.5%)和抗心律失常药物(2.5%)。所有选定的 RCT 中,39 个(97.5%)的研究对象是成年人群,仅 1 个(2.5%)关注儿童^[8]。

2. RWD 来源:项目组选择了 3 个美国医疗索赔数据库用于开展本研究,包括两个商业健康保险索赔数据 Optum Clinformatics (2004–2019 年) 和 IBM MarketScan (2003–2017 年),以及 Medicare 的 A、B、D 3 个子集(2011–2017 年),Medicare 数据库没有用于抗血小板试验的模拟研究。此外受限于

数据可及性的问题,对 CAROLINA 研究结果的预测参考 RCT 研究的实施时间,即使用截至 2015 年 9 月的数据^[9]。在整个研究期间,每个数据库都会进行实时更新。

数据来源涵盖了参与医疗保险计划的患者的所有医疗服务相关信息,包括人口统计信息、登记开始和结束日期、配药日期、使用剂量和供应天数等。Medicare 数据库涵盖了所有行政记录的死亡,但在商业保险数据库中,无法获取到完整的院外死亡信息,且在所有数据来源中都没有记录死亡原因。

3. RWE 研究设计与实施过程:项目组设计了 RWE 研究的实施流程框架,旨在让 RWE 研究的设计和分析尽可能的结构化、透明化并具有可重复性^[8]。研究队列的创建和研究变量的选取是借助 Aetion 公司数据平台来实现的^[10-11],其中包括可行性和功效分析以及评估治疗效果并最终比较分析。该平台纳入了与索赔有关的海量数据,且平台可跟踪记录在线使用者的分析内容和时间。这一框架流程为研究发起者与 FDA 合作时,在进行正式的比较分析之前,为设计 RWE 研究提供了一个直观透明化的模式,提高了实施的可行性。

在明确研究方案并进行统计分析之前,项目组首先需评估研究的可行性和有效性,包括研究组之间的协变量平衡性和统计功效的估计,最终确定研究方案,对主要分析策略进行详细说明并在 <https://www.clinicaltrials.gov/> 上进行方案的注册登记。在此之前研究不会开展任何关于治疗特异性结果的分析^[12],如果研究者认为研究的预测功效不够,可以决定终止该项研究,但需要详细记录终止原因。整个流程旨在模拟常规临床试验方案的提交过程,确保特定的研究设计和统计分析方法的选择不会受 RWE 研究结果的影响^[3]。

在确定研究可行性之后,需撰写一份在医疗索赔数据库中进行 RWE 研究的方案,该方案会详细说明 RWE 研究的设计,包括操作细节、编码定义和判断模拟成功与否的相关标准细则。同时,报告可行性和功效分析的结果。这些结果和研究实施的更新以及最终的模拟研究结果会与 Aetion 公司数据平台报告一起发布在研究官方网站(www.rctduplicate.org),其中包括所有变量、结局的定义和具体实施细节。项目组也会适时地咨询临床和方法学专家,确保选择的研究设计、测量方法和统计分析方法的科学性。尽管具体的研究设计和分

析会有所不同,但项目组优先采用可得到有效估计的研究设计,比如新用药者和阳性对照队列设计^[4]。尽管项目组尽可能地与待模拟的 RCT 的纳入排除标准保持一致,但 RCT 和 RWE 的研究人群之间仍然会存在一些差异,在进行敏感性分析的时候可以通过加权或其他方法来保证 RWE 的研究人群的特征尽可能接近 RCT 人群的特征^[13]。考虑到未来在没有 RCT 证据的情况下开展 RWE 研究时,研究者无法获取到相应的 RCT 研究对象的基本特征,因此,基于 RWD 的主分析中不应包含上述敏感性分析所采取的方法以保证 RWE 研究过程不受已知的 RCT 研究结果的干扰。

大多数 RCT 采用意向性治疗(intention-to-treat)原则进行分析,即无论随机分组后的治疗计划是否改变,患者都认为按原随机化分组随访到了研究的预定结束时间。然而,在 RWD 中通常采用实际治疗分析(as treated analysis, AT)来估计治疗效果,将患者转换治疗计划或停止研究治疗时处理为删失来估计治疗效果。项目组预估 RWE 研究中患者对药物的依从性相对差于 RCT 中的患者,故选择 AT 来评估治疗效果,使其与 RCT 的结果更加一致。

4. RWE 的统计分析过程:在针对上述 RCT 开展 RWE 研究中,项目组均采用 1:1 的倾向性评分(propensity score, PS)来控制 120 余个潜在的混杂因素^[14],在药物暴露前 6 个月测量混杂因素。虽然 RCT 通常样本量较小也仅需要展示较少的患者特征,但在 RWE 研究中,为了平衡尽可能多的潜在混杂因素,并模拟基线随机化,需要更大的样本量和多维度的协变量集。协变量一般包括人口学基本信息、治疗开始时间、共患疾病以及合并使用的药物(如心血管和其他药物的使用)等特定变量。由于 Optum 和 MarketScan 数据库中只提供了一部分患者的实验室指标,并没有将其纳入 PS 分析中,但评估了暴露组之间指标在匹配后的平衡性。在基于年龄、性别和合并症等估计 PS 并进行 1:1 匹配后,会进行功效分析,以确定非随机 RWE 研究的预期功效^[15]。项目要求 RWE 研究至少与相应的 RCT 具有同等的功效。

研究随访起始时间定义为治疗开始后第二天,随访终止时间可包含以下几种情况:直至治疗停止并设置 30 d 拓展观察期、转换到对照组、发生结局事件、进入养老院、医疗保险到期或研究期结束。在 PS 匹配队列中使用 Cox 回归估计风险比(hazard

ratio, HR)及其 95%CI,在每个数据源中分别进行分析,然后使用固定效应 Meta 分析进行汇总。之所以选择 Meta 分析,是因为汇总的估计数量非常少,而且在 3 个数据库中使用了统一的研究设计。预先设定的敏感性分析包括“as-started”分析,在该分析中,患者不会因治疗改变而做删失处理,而是在随访的 365 d 时做删失处理。

所有 RWE 研究的主要结局均为复合主要不良心血管事件(major adverse cardiovascular events, MACEs),除 DECLARE-TIMI 研究(由于在开展 RWE 研究模拟该 RCT 时没有足够的统计功效进行 MACEs 的分析)外, MACEs 主要包括非致死性心肌梗死、非致死性中风或心血管死亡。此外 TECOS 研究中对 MACEs 的定义还包括了因不稳定型心绞痛住院治疗。DECLARE-TIMI 研究只分析以住院治疗心力衰竭和心血管死亡的主要复合结局。针对所有 RCT,如果排除了患有癌症和其他慢性病的患者,大部分研究对象的死亡可归因于心血管疾病,所以项目组将全因死亡作为心血管疾病死亡的替代结局指标。另外 5 项抗血小板药物研究根据纳入排除标准保留的合格研究对象数量有限,使得可重复性功效不足,加之这些研究评估的是患者住院期间的治疗,因此项目组考虑无法在 RWD 中获取相关信息而放弃去模拟这些 RCT。

在这一阶段中,将进行最终的可行性分析,以确定在 RWE 研究中重要的混杂因素能否达到预期的组间平衡。根据临床知识和平衡参数,如匹配后 C-统计量和绝对标准差,在不考虑结局数据的情况下,判断预先指定分析的组间平衡是否充分^[16]。此外,根据最终的协变量向量分布分别计算合并多个数据库的分析和基于每个数据库的单个分析的功效。如果单个数据库研究本身没有足够的功效,但在合并分析中有足够的功效,项目组仅报告合并分析的结果;否则,还将分别报告具有足够功效的单一数据库研究的结果。最后,项目组建议酌情进行敏感性分析,以确定预先指定的主要分析的稳定性和次要分析的真实性,例如评估研究人群一个子集中某些实验室指标的平衡性^[17]。

5. RCT-RWE 研究结果一致性评估方法:项目组的主要目的是评估 RCT 结果与 RWE 研究结果之间差异的大小和产生原因^[8]。预先制定了两个一致性度量指标:①监管一致性(regulatory agreement);②估计值一致性(estimate agreement)。此外,也建议通过计算 RCT 和 RWE 研究结果效应

估计值之间的标准化差异进行假设检验并评估结果是否存在差异。

(1) 监管一致性: 定义为 RWE 研究模拟 RCT 结果的方向和统计显著性的能力。例如, 模拟一个优效性 RCT 的 RWE 研究结果也应该是优效的。对于非劣效性 RCT, 监管一致性要求 RWE 研究显示的非劣效性应使用与 RCT 相同的界值。对于结果阴性的 RCT, RWE 研究结果也应该显示无统计学意义。监管一致性衡量的是 RWE 在多大程度上可以提供与 RCT 相同的监管决策证据, 这适用于典型的监管决策规则。因为当 RCT 显示出强有力的证据拒绝零假设时, 研究之间的一致性更可信, 所以将根据 P 值的不同分层 (<0.01 、 $0.01\sim$ 和 >0.05) 来进行敏感性分析以评估监管一致性。一般来说, 当 RCT 没有显示试验组间结局差异有统计学意义时, RWE 研究与 RCT 出现监管一致性的差异由于偶然性导致的可能性最高^[8]。

(2) 估计值一致性: 因为 RWE 研究可能比相应的 RCT 具有更高的功效, 会出现 RCT 结果无统计学意义而 RWE 研究结果有统计学意义, 在这种情况下, 即使 RWE 研究和 RCT 的估计值是接近的, 也不满足监管一致性。例如, 如果 RCT 报告的 $HR=0.85$ ($95\%CI: 0.71\sim 1.02$), RWE 研究的 $HR=0.85$ ($95\%CI: 0.73\sim 0.99$); 在这种情况下, 尽管估计值十分一致, 项目组会得出 RWE 研究没有与 RCT 达到监管一致性的结论。

出于这个原因, 项目组又制定了一个次要指标, 即估计值一致性, 定义为当 RWE 研究的治疗效果估计处于 RCT 治疗效力估计值的 $95\%CI$ 内, 即满足估计值的一致性^[8]。项目组将这个标准建立在 RCT 的 $95\%CI$ 基础上, 是因为这些 RCT 通常都具有相似的统计功效 ($80\%\sim 90\%$), 因此 $95\%CI$ 的宽度将处于一个适当的范围内。

此外, 根据 RCT 对 $95\%CI$ 标准的解释, 研究期望 95% 实际人群的效应都在 $95\%CI$ 内。因此, 估计值一致性的定义要求 RWE 效应估计值应在基于 RCT 证据的真实效应的 $95\%CI$ 范围内。假设 RWE 的估计没有偏倚, 估计值一致性的概率仅取决于 RCT 和 RWE 研究之间的方差比率。当方差相等时, RWE 研究的无偏估计值有 83% 的概率会达到与 RCT 估计值的一致性。作为敏感性分析, 研究会按照 $70\%CI$ 来确定估计值一致性。

(3) 标准化差异: 除了上述评价模拟研究的两种一致性度量指标, 项目组还利用效应估计值和 CI

的散点图以及标准化差异来探讨所有 RWE 模拟结果和相应 RCT 结果之间差异的大小和方向。计算公式:

$$Z = \frac{\hat{\theta}_{RWE} - \hat{\theta}_{RCT}}{\sqrt{\hat{\sigma}_{RWE}^2 + \hat{\sigma}_{RCT}^2}}$$

其中, $\hat{\theta}$ 表示治疗效果的估计值, 最常见的是对数风险比, $\hat{\sigma}^2$ 是相关方差。这一分析可以更好地理解治疗效果估计值的一致性以及导致估计值差异的因素。通过对 RWE 和 RCT 结果之间的统计学显著差异进行假设检验, 也可以在标准化差异的基础上产生关于一致性的二元判定。假设 $\alpha=0.05$, 且 RWE 和 RCT 结果都基于大样本, $|Z|>1.96$ 时拒绝无差异的原假设。

如果 RCT 采用的是阳性对照, 则评判为“良好”(good)的对照; 如果对照是与结局无关的替代药物, 且替代药物用于特征高度相似的患者, 判定为“中等”的对照; 如果对照是与结局无关的替代药物, 但用于不同特征的患者, 判定为“不好”的对照。如果 RCT 结局测量具有高特异性, 判定为“良好”的结局指标, 如果 RCT 结局测量具有中等特异性, 判定为“中等”的结局指标。

值得注意的是, 在根据选定的待模拟 RCT 去开展 RWE 研究时, 即使 RWE 研究可以通过专门设计来匹配 RCT 的特征, 但考虑到 RWD 的局限性, 这些潜在的差异, 如 RWE 中通常缺乏完整的疾病史和不同于 RCT 严格的对照和服药依从性等均可能发生, 并对结果的解释带来挑战。由于某些数据库中缺乏患者完整的病史记录, 部分患者可能无法获得特定的生物标记物或影像学资料, 以及患者症状体征记录不规范, 都可能会给 RCT 的匹配带来困难^[18]。

三、小结

使用 RWD 产生 RWE 来支持药物的监管决策已经受到广泛关注, 近几年研究者和政策制定者亦围绕如何更好地评估 RWE 的可信度和透明度展开了系列讨论并发起相关研究^[19-23]。由于 RCT 长期以来一直被视为评估干预措施与疗效之间因果关系的金标准, RCT DUPLICATE 项目寻求的解决方法是根据 RCT 的研究设计, 获取潜在的 RWD 进行研究设计, 通过执行与 RCT 相同的统计分析策略并辅以敏感性分析, 探讨该类 RWE 研究在多大程度上可以模拟 RCT 的结果。如果研究结果一致, 这将增强对潜在 RWD 的来源、质量及其产生的 RWE 的可信度, 从而为确定真实治疗效果提供监管决策支持。然而, 这些模拟 RCT 的 RWE 研究的最终目的

并非鼓励监管机构用 RWE 来替代 RCT,而是通过开发一个通用的模拟 RCT 的研究流程,评估 RWE 研究在提供药物疗效证据方面的潜在作用。通过使用严格设计的 RWE 研究来模拟 RCT 的结果,可以更全面地了解在监管决策中使用这些设计的优势和局限性。如果正确使用,未来 RWE 可以作为传统 RCT 的补充,以帮助提高医疗保健中的临床决策效率,并改善医疗服务人群覆盖不足的情况。

通过开展 RCT 和 RWE 研究结果之间差异比较的研究,可以从 RWE 模拟 RCT 研究中获得诸多启示。例如,在 RCT DUPLICATE 和 OPERAND (Observational Patient Evidence for Regulatory Approval and Understanding Disease) 等更多的团队开展独立的研究,能够总结何种设计和分析可以产生较为一致的结果,从而确定未来研究的最佳实践^[24]。比如医疗索赔数据可能缺乏临床症状或指标细节,但提供了个体连续的纵向数据;EHR 和患者登记信息通常有详细的临床信息,但缺乏个体在该系统外的其他医疗服务信息。这都可能导致不同的研究结果。随着研究者对 RWE 应用领域的不断深入探索,以 RCT DUPLICATE 项目为代表的系列研究将进一步探索与提升 RWE 在证据质量体系 and 临床决策中的作用。

利益冲突 所有作者声明无利益冲突

作者贡献声明 石舒原:文献整合、论文撰写;刘佐相:文献收集;赵厚宇、聂晓璐:论文修订;符祝、宋海波:论文修订、研究材料支持;姚晨、詹思延:指导性支持;孙凤:提出研究选题、获取研究经费

参 考 文 献

- Bonamici S. H. R. 34-21st century cures act[EB/OL]. [2022-03-15]. <https://www.congress.gov/bill/114th-congress/house-bill/34>.
- U. S. Food & Drug Administration. Framework for FDA's real-world evidence program[EB/OL]. [2022-03-15]. <https://www.fda.gov/media/120060/download>.
- Franklin JM, Glynn RJ, Martin D, et al. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making[J]. *Clin Pharmacol Ther*, 2019, 105(4):867-877. DOI:10.1002/cpt.1351.
- Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials? [J]. *Clin Pharmacol Ther*, 2017, 102(6): 924-933. DOI:10.1002/cpt.857.
- Hernán MA, Hernández-Díaz S, Robins JM. Randomized trials analyzed as observational studies[J]. *Ann Intern Med*, 2013, 159(8): 560-562. DOI: 10.7326/0003-4819-159-8-201310150-00709.
- 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.
- 姚晓莹,张靖雪,詹思延.真实世界证据与随机对照试验:RCT DUPLICATE 项目概述[J]. *药物流行病学杂志*, 2019, 28(8):495-497, 517.
- Yao XY, Zhang JX, Zhan SY. Real-world evidence and randomized controlled trials: the overview of RCT DUPLICATE[J]. *Chin J Pharmacoepidemiol*, 2019, 28(8): 495-497, 517.
- 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.
- Paterno 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.
- Wang SV, Verpillat P, Rassen JA, et al. Transparency and reproducibility of observational cohort studies using large healthcare databases[J]. *Clin Pharmacol Ther*, 2016, 99(3):325-332. DOI:10.1002/cpt.329.
- Kim SC, Solomon DH, Rogers JR, et al. Cardiovascular safety of tocilizumab versus tumor necrosis factor inhibitors in patients with rheumatoid arthritis: a multi-database cohort study[J]. *Arthritis Rheumatol*, 2017, 69(6):1154-1164. DOI:10.1002/art.40084.
- Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE special task force on real-world evidence in health care decision making[J]. *Value Health*, 2017, 20(8): 1003-1008. DOI:10.1016/j.jval.2017.08.3019.
- Deming WE, Stephan FF. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known[J]. *Ann Math Stat*, 1940, 11(4):427-444. DOI:10.1214/aoms/1177731829.
- Rassen JA, Shelat AA, Myers J, et al. One-to-many propensity score matching in cohort studies[J]. *Pharmacoepidemiol Drug Saf*, 2012, 21 Suppl 2: 69-80. DOI:10.1002/pds.3263.
- Gagne JJ, Glynn RJ, Avorn J, et al. A combined comorbidity score predicted mortality in elderly patients better than existing scores[J]. *J Clin Epidemiol*, 2011, 64(7):749-759. DOI:10.1016/j.jclinepi.2010.10.004.
- Franklin JM, Rassen JA, Ackermann D, et al. Metrics for covariate balance in cohort studies of causal effects[J]. *Stat Med*, 2014, 33(10): 1685-1699. DOI: 10.1002/sim.6058.
- Paterno E, Gopalakrishnan C, Franklin JM, et al. Claims-based studies of oral glucose-lowering medications can achieve balance in critical clinical variables only observed in electronic health records[J]. *Diabetes Obes Metab*, 2018, 20(4):974-984. DOI:10.1111/dom.13184.
- Franklin JM, Glynn RJ, Suissa S, et al. Emulation differences vs. biases when calibrating real-world evidence findings against randomized controlled trials[J]. *Clin Pharmacol Ther*, 2020, 107(4): 735-737. DOI: 10.1002/cpt.1793.
- U.S. Food & Drug Administration. Real-world data:assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products [EB/OL]. [2022-03-15]. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory>.
- U.S. Food & Drug Administration. Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological products[EB/OL]. [2022-03-15]. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug>.
- U.S. Food & Drug Administration. Real-world data:assessing registries to support regulatory decision-making for drug and biological products guidance for industry[EB/OL]. [2022-03-15]. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-registries-support-regulatory-decision-making-drug-and-biological-products>.
- Arlett P, Kjaer J, Broich K, et al. Real-world evidence in EU medicines regulation:enabling use and establishing value [J]. *Clin Pharmacol Ther*, 2022, 111(1): 21-23. DOI: 10.1002/cpt.2479.
- Flynn R, Plueschke K, Quinten C, et al. Marketing authorization applications made to the European medicines agency in 2018-2019: what was the contribution of real-world evidence? [J]. *Clin Pharmacol Ther*, 2022, 111(1):90-97. DOI:10.1002/cpt.2461.
- The MRCT Center of Brigham and Women's Hospital and Harvard. Real world evidence[EB/OL]. [2022-03-15]. <https://mrctcenter.org/blog/projects/real-world-evidence>.