

利用现实世界数据仿真临床试验： 目标试验的发展与应用

周佳薇¹ 黄丽红² 尤东方¹ 陈峰¹ 赵杨¹

¹南京医科大学公共卫生学院生物统计学系,南京 211166;²复旦大学附属中山医院生物统计室,上海 200032

通信作者:赵杨,Email:yzhao@njmu.edu.cn

【摘要】 临床试验是评价干预措施疗效和安全性的金标准,但存在花费大、耗时长等限制。现实世界数据可为比较性研究提供强大的数据基础,但研究质量参差不齐。本文介绍了仿真目标试验,其利用现实世界数据,按照临床试验的设计,事先定义暴露和结局、设立纳入排除标准、确定时间零点、估计样本量和制定统计分析计划等,以期提高观察性研究的证据等级,并初步讨论仿真目标试验的证据等级评价标准,通过案例解读仿真目标试验。

【关键词】 仿真目标试验; 观察性研究; 证据质量; 临床试验

基金项目: 国家自然科学基金(82173620, 82373690, 82273733, 82204156); 江苏省高校优势学科建设工程

The emulation of clinical trials with real-world data: development and application of target trial

Zhou Jiawei¹, Huang Lihong², You Dongfang¹, Chen Feng¹, Zhao Yang¹

¹Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, China; ²Department of Biostatistics, Zhongshan Hospital, Fudan University, Shanghai 200032, China

Corresponding author: Zhao Yang, Email: yzhao@njmu.edu.cn

【Abstract】 Clinical trial is the gold standard for evaluating the efficacy and safety of interventions; however, it is limited by high costs and long time. Real-world data (RWD) can provide a robust data basis for comparative research, but the quality is uneven. This review introduces the target trial emulation, in which researchers, using RWD and following the design of clinical trials, define exposure and outcome in advance, set eligibility criteria, determine the time zero, estimate sample size, and plan statistical analysis, to enhance the quality of evidence for observational studies. This review preliminarily discusses the standard of evidence quality evaluation in target trial emulation. Then, the target trial emulation is shown through case interpretation.

【Key words】 Target trial emulation; Observational study; Quality of evidence; Clinical trial

Fund programs: National Natural Science Foundation of China (82173620, 82373690, 82273733, 82204156); Priority Academic Program Development of Jiangsu Higher Education Institution

一、概念和背景

临床研究中,干预治疗的有效性和安全性通常

通过临床试验确定。临床试验作为比较性研究的金标准,采用随机化和盲法,避免或减少临床试验

DOI:10.3760/cma.j.cn112338-20230821-00081

收稿日期 2023-08-21 本文编辑 张婧

引用格式:周佳薇,黄丽红,尤东方,等.利用现实世界数据仿真临床试验:目标试验的发展与应用[J].中华流行病学杂志,2024,45(2):279-285. DOI:10.3760/cma.j.cn112338-20230821-00081.

Zhou JW, Huang LH, You DF, et al. The emulation of clinical trials with real-world data: development and application of target trial[J]. Chin J Epidemiol, 2024, 45(2):279-285. DOI: 10.3760/cma.j.cn112338-20230821-00081.



中的偏倚,并在实施过程中遵循事先计划的研究方案,降低测量误差对结果的影响,从而得出疗效和安全性的估计。但在实际的临床研究中,并非每一个临床干预问题都能从临床试验中得到答案。在干预措施效果评价的过程中,临床试验并不始终具有可及性。例如,评价急性心绞痛发作时,使用硝酸甘油缓解症状的疗效。显然,出于伦理考量,无法为患者随机分配药物。对于这一类紧急的临床情况,需要考虑其他途径寻求疗效的估计。欧美国家部分研究机构十分重视流行病学研究中的比较性研究,将其纳入法案,作为评价干预措施效果的有效手段^[1]。

对监管部门而言,更关心的是在真实诊疗情景中干预措施的疗效和安全性。临床试验的干预措施与现实世界中治疗情况存在差异;临床试验有着相对严苛的纳入和排除标准,对于被排除的群体,无法应用结果外推,可能需要借助其他研究。由此,观察性研究可以纳入考量。随着医疗数据量级的爆发式增长和医疗信息系统的不断升级和完善,观察性数据来源相较于临床试验更加丰富,范围更加广泛。因此,利用观察性数据进行比较性研究,探索干预治疗的有效性和安全性^[2],即为仿真目标试验最初的思想雏形^[1,3]。

目标试验这一概念由 Hernán 和 Robins^[4]于 2016 年正式提出,目标试验本意指一项假想的临床试验(或以真实的临床试验为蓝本),从顶层设计上打破临床试验和观察性研究的隔阂,明确将观察性研究同临床试验联系起来,主要用于使用观察性数据回答干预的疗效和安全性问题。首先假设利用临床试验探究该问题,撰写对应的临床试验方案和统计分析计划(或以已有的临床试验方案为依据),使用观察性数据,进行统计分析,最后解读研究结果等。利用观察性数据按照既定的方案完成目标试验的过程即为仿真目标试验。

仿真目标试验本质上仍然是观察性研究中的一类。将临床试验的设计和分析思想应用到仿真目标试验中^[5],根本目的在于提高观察性研究在疗效比较中的可靠性^[6],从而完成疗效因果推断。通过仿真目标试验,有效提升该类观察性研究的证据等级,将现实世界数据转变为现实世界证据。

二、目标试验的设计

1. 干预和结局的定义:首先,根据目标试验的目的,确定干预和预期结局终点的定义。干预的定义包括试验组和对照组中干预的实施方式、剂量、

时间及其他要求。仿真目标试验中,不建议使用安慰剂作对照。实际的诊疗中,一般不会使用安慰剂进行治疗,即使支持性治疗的患者,也会在病情、病程和临床诊疗记录频率上与干预患者存在较大的差异^[7]。建议选择对结局终点无影响的干预作为对照,或选择明确已知影响的干预作为阳性对照。结局终点的定义要尽可能客观准确,包括结局的测量时间、方式及对测量者或评估者的要求等。对于验证性仿真目标试验,干预和结局的定义应尽可能接近临床试验中的定义。

2. 纳入和排除标准:研究设计阶段,针对目标人群的选取,制定合理的纳入和排除标准。仿真目标试验中,纳入和排除标准的定义可能会根据实际数据情况进行调整,此时需事先评估这样的调整对比较性研究结局的影响,并在方案中说明^[8]。

3. 样本量估计:尽管与临床试验相比,现实世界数据来源丰富,可供选择的样本更多。但在仿真目标试验中,样本量估计仍有必要。对于验证性仿真目标试验,考虑到仿真目标试验与临床试验中疗效估计可能存在差异,仿真目标试验的样本量应至少超过临床试验。

4. 随机化与盲法:均为临床试验控制偏倚的重要手段^[9]。仿真目标试验中,通常为无实施盲法的临床研究。实效性临床试验的设计中,患者和临床医生均已知治疗^[10]。当目标试验的目的是比较现实治疗策略的疗效时,非盲的设计并非仿真目标试验的重要缺点^[4];但当仿真目标试验作为确证性的证据支持药物新适应证审批时,有必要考虑非盲带来的影响。

现实世界的干预已经产生,随机化无法实施。对于随机化的仿真通过调整所有混杂因素来实现,以确保不同干预开始时,人群之间的可比性^[4]。仿真目标试验通过匹配、回归法,或基于倾向性评分进行分层、匹配、回归或加权,或 G 方法、双稳健模型等达到基线可比性。若基线信息未包含所有混杂因素,则不可能完全成功仿真随机化。虽然尚不存在良好控制未观测混杂的方法,但可以通过 E 值和阴性对照等方法衡量未观测混杂的潜在影响^[11-12]。

5. 统计分析:传统临床试验中,通常估计意向性治疗人群的疗效作为主要分析结果。然而,真实的诊疗中,患者可能存在多种采取治疗的途径,且采取治疗的持续时间较短。这需要研究者充分收集现实世界数据,并考虑其覆盖程度。此外,仿真

目标试验的统计分析为治疗过程分析,与临床试验中的符合方案人群更接近。

目标试验中,与临床试验类似,需要事先计划结局指标的统计分析方法,并考虑期中分析和多重性避免假阳性结果,提高统计分析结果的可信度。此外,考虑到现实世界与临床试验的差异,特别是验证性目标试验,预先已经明确对疗效比较的结果产生影响甚至导致偏差的因素,应对其进行敏感性分析,进一步明确其影响程度,通常也需要事先制定敏感性分析计划。例如,肿瘤临床试验中,以生存为结局时,对患者的定期随访和生存终点的记录有详细的要求,但在现实世界研究中,若存在因缺乏有计划的随访造成生存终点数据删失较多的情况,应采用敏感性分析来探究这一情况对结局的影响。此外,对于观察性研究中常见的选择偏倚,也需要在制定统计分析计划时,选取适当的方法,控制其对结果的影响^[13-15]。

6. 基线时间的选择:临床试验中,通常认为给予干预前最后一次获取的临床信息为基线,在仿真目标试验中,需要确定一个合适的随访时间零点,以获取基线信息。患者在随访时间零点满足纳入和排除标准,并从该时间点开始接受干预措施,进入随访并观察结局^[4]。然而,在实际诊疗中,满足纳入排除标准/开始随访与开始干预的时间可能并不一致,若根据随访过程中,出现的干预进行患者分组,则可能导致非死亡时间偏倚,影响干预疗效估计。此时应选择满足纳入排除标准/开始随访的时间作为零点,并考虑 landmark 分析、时依法等敏感性分析方法探究非死亡时间偏倚对疗效估计的影响^[16]。

此外,基线定义时,可能会遇到 2 种情况:第一种为患者诊疗过程中仅有一个时间点满足纳入排除标准,应选择该时间点作为基线;第二种为患者诊疗过程中存在多个时间点满足纳入排除标准^[17],此时既可以选择一个符合标准的时间:随机选择或选择第一个时间点,也可以选择所有符合标准的时间点或其子集,用于仿真多个嵌套试验,每个试验随访开始时间不同^[3]。

三、数据的要求与处理

与临床试验相比,开展观察性研究的操作要求较小,观察性数据在日常诊疗和队列中均更常见。仿真目标试验的常见数据来源包括医院信息系统数据(电子病历数据)、医保支付数据、登记队列数据、药品安全性主动监测、自然人群队列数据等多

个方面。不同现实世界数据来源的数据重点和质量不同,例如电子病历数据的用药记录、实验室检查等信息更加全面,而这些临床信息可能是影响疗效比较的混杂因素^[6],登记数据在随访和结局中质量更高。利用现实世界数据进行仿真目标试验时,可以尝试将不同来源的数据结合起来,或在不同类型的数据中验证干预疗效。

现实世界数据质量参差不齐,大多存在数据不完整、记录不准确、收集不全面等问题。利用现实世界数据时,需充分考虑数据存在的问题,并制订全面的数据治理计划,保证研究质量。

四、分析与结果解读

1. 一致性评价:仿真目标试验的统计分析结果,既要考虑其统计学意义和临床意义的解释,又需解读仿真目标试验与临床试验的差异,一致性评价指标有 3 个:监管一致、估计一致和检验一致^[7]。

监管一致,2 个研究的点估计方向一致,且结果有统计学意义;估计一致,仿真目标试验的点估计值在临床试验的 95%CI 内;检验一致,2 个研究结果之间的标准化差异无统计学意义。检验的统

计量: $Z = \frac{\hat{\theta}_{TT} - \hat{\theta}_{RCT}}{\sqrt{\hat{\sigma}_{TT}^2 + \hat{\sigma}_{RCT}^2}}$ 。式中, $\hat{\theta}_{TT}$ 和 $\hat{\sigma}_{TT}^2$ 为仿真目标

试验疗效的点估计和方差估计, $\hat{\theta}_{RCT}$ 和 $\hat{\sigma}_{RCT}^2$ 为临床试验疗效的点估计和方差估计^[18]。

2. 差异来源:差异的来源可以分为仿真差异和偏倚^[19]。仿真差异主要围绕人群、干预、结局、统计分析等多个方面,偏倚是观察性研究中不可避免的问题。

(1) 人群差异来源:在人群中,仿真差异产生于纳入排除标准、人群特征和地区等。在仿真目标试验中,临床试验的纳入排除标准可能不适用。例如,在干预实施前,实验室检查可能无法在登记队列数据中获取;部分临床试验中患者筛选存在研究者的主观判断。人群来源上,仿真目标试验与临床试验也可能存在分布上的差异,这也会造成结果之间的差异。有研究提出,按照临床试验的人群分布特征对仿真目标试验人群进行加权,使 2 个研究人群分布尽可能一致,但也有学者认为现实世界数据是实际接受干预人群更普遍的代表,更能反映干预在人群中的真实疗效^[18,20]。实际上,加权和不加权的 2 种分析策略均应被考虑,并根据目的决定主次,这取决于仿真目标试验是为了尽可能地验证临床试验的结果,还是探究干预在人群中的疗效^[21]。

地区和人种的差异也需要考虑,选择相同地区和人种的临床试验结果作为仿真目标试验的参考。

(2) 干预差异来源: 临床试验中会明确干预实施方式和时间, 并且采取措施维持受试者的依从性, 控制失访率。但现实世界数据中, 患者是否获取干预治疗可以从历史诊疗数据中得知, 却并不完全表示患者真实接受干预, 而依从性因人而异。此外, 数据覆盖面也是干预差异的来源, 患者的干预获取渠道可能大于研究的数据覆盖面, 此时, 患者的依从性估计有偏差。

(3) 结局差异来源: 主动监测和因疾病需要而获取治疗之间存在差异, 反映出仿真目标试验中疾病程度往往更重。在检测时间上, 临床试验中制定随访策略, 既用于监督受试者的依从性, 也可获取结局和不良事件信息。现实世界数据依赖患者个人的医疗需求, 并受到地区经济和医疗资源的影响。现实世界数据中, 还会因数据覆盖面造成结局捕获不全, 例如使用部分医院的电子病历数据, 若患者更改居住地或就诊医院则缺失部分结局信息。因此解读结果时, 需考虑结局缺失的影响^[22]。

(4) 统计分析差异来源: 在疗效的统计分析方面, 一般选择同样的统计分析方法, 但临床试验与仿真目标试验的分析集不同。从这个角度上来看, 临床试验与现实世界研究之间互相补充, 作为严格标准和普遍情况的疗效估计^[23]。此外, 应采用敏感性分析探索仿真差异对疗效估计的潜在影响。

(5) 混杂因素: 在观察性研究中, 混杂偏倚是一个不可避免的问题, 混杂偏倚包括已知和未知混杂偏倚。尽管可以采用倾向性评分等方法控制已测

量的混杂偏倚, 但某些已知对疗效估计有影响却未测量或缺少来源的因素, 仍会影响一致性评价。已知的混杂偏倚应在研究设计时尽可能被考虑。未知混杂偏倚同样会对疗效估计产生影响, 且不应被忽视, 仿真目标试验中需要更多的研究来解决相关问题。

五、仿真目标试验的评价

当观察性研究与临床试验结果冲突时, 研究者往往怀疑观察性研究受到混杂偏倚的影响, 将其证据等级放在临床试验之后。通过目标试验设计, 利用现实世界数据仿真目标试验, 目的在于提升观察性研究的科学性等级。同样地, 也需要合理、有效、清晰的评价工具衡量仿真目标试验的结果, 为目标试验的设计、证据等级的划分与比较、目标试验结果的解读提供依据。

制定评价系统工具前, 充分借鉴既往评价方法。参考了非随机干预研究偏倚风险评估工具 ROBINS-I 和用于评价证据等级和治疗建议程度的 GRADE 工具^[24-25]。见表 1。

ROBINS-I 工具于 2016 年发表于英国医学杂志, 确定了非随机干预研究中 7 个重要的偏倚^[24]。该评价工具中, 每个偏倚中有一系列问题, 通过回答综合评估这类偏倚^[24]。GRADE 工具常用于制定治疗指导原则, 主要包括证据等级评价和治疗建议程度^[26-29]。在证据等级评价中, 根据其研究设计、质量和结果等划分证据等级^[25]。

对于仿真目标试验的证据等级的评价, 应将 ROBINS-I 和 GRADE 工具中证据等级评价部分结合起来。既应考虑仿真目标试验中可能存在的偏

表 1 ROBINS-I 和 GRADE 工具的主要内容及比较

工具	主要内容	优点	缺点
ROBINS-I	非随机干预研究中 7 个重要偏倚的衡量: 1. 混杂偏倚; 2. 受试者/患者选择偏倚; 3. 干预分类偏倚; 4. 干预偏差偏倚; 5. 缺失数据偏倚; 6. 结局测量偏倚; 7. 选择性结果报告偏倚。	<ul style="list-style-type: none"> 全面总结观察性研究可能出现的偏倚; 每个偏倚的衡量以系列问题为支撑。 	<ul style="list-style-type: none"> 仅关注偏倚, 未考虑设计方案、样本量等方面的评价。
GRADE	1. 证据等级评价: (1) 根据研究设计类型确定初始等级; (2) 考量 8 个影响因素; (3) 得到最终等级。 2. 治疗建议程度: (1) 证据等级; (2) 考虑治疗有效性和安全性计算净获益; (3) 不同人群的需求或期望治疗结果; (4) 资源有效利用的性价比。 最终给出不同治疗在不同情况的建议程度。	<ul style="list-style-type: none"> 包括证据等级评价和治疗建议程度 2 个方面, 临床使用价值更高。 	<ul style="list-style-type: none"> 证据等级评价考虑不全面; 证据等级划分较粗糙。

倚,还应考虑代表性、设计方案的合理性、样本量是否达到足够的检验效能、组间均衡性、敏感性分析的必要性等。

在各项研究证据等级评价的基础上,考虑将其转化为治疗建议程度。将临床试验和仿真目标试验证据结合起来,参考 GRADE 工具治疗建议程度部分,临床医生根据临床试验的效力和仿真目标试验的效果,决定采用何种干预措施。《中华人民共和国医师法》规定,在尚无有效或者更好治疗手段等特殊情况下,医师取得患者的知情同意后,可以采取具有循证医学证据的药品用法实施治疗^[30]。既反映了医疗领域日新月异的发展,临床证据体系的多样组成,又对临床医生治疗建议程度的判断产生了新的要求。监管部门的新药审批和适应证扩展中,有效的评价系统能够准确反映仿真目标试验的设计、实施和结果解读的质量,为审批提供有效支持。科学研究中可以结合治疗建议评价和精准治疗,在寻找目标人群的治疗方式上更进一步。因此,未来基于 ROBINS-I 和 GRADE 工具构建仿真目标试验评价体系和同临床试验组成证据链转化为治疗建议程度,均需要进一步的讨论和更恰当的评价工具。

六、案例介绍与解读

随着医疗数据信息收集系统的不断完善,仿真目标试验的开展拥有良好的基础。目前,仿真目标试验大多选择重复已经完成或正在进行的临床试验,即为验证性仿真目标试验。仿真目标试验尚在学习中,国内外在不同临床领域的仿真目标试验中仍存在空白。

仿真目标试验实例^[3,7,31-32]中最突出的为重复随机对照试验倡议^[21,33-38],由美国食品药品监督管理局、布莱根妇女医院等发起。该倡议包括 40 项仿真目标试验,其目的是摸索利用现实世界数据产生现实世界证据的透明过程,衡量如何得出与临床试验一致的结果,并讨论差异的影响因素^[7]。

目前,该项目公开了 32 项研究结果,主要围绕心血管疾病和相关的结局事件^[36]。在这些研究中,使用倾向性评分 1:1 最近邻匹配,设置卡钳值为 0.01,控制事先选定的 120 个协变量,以达到组间均衡。协变量包括人口学信息、日历时间、合并疾病、合并用药等信息。一致性评价则考虑了监管一致、估计一致和检验一致 3 个指标。

从目前已经发表的结果来看,24 项达到监管一致,21 项达到估计一致,24 项达到检验一致。监

管一致通常是最严格的,这也符合监管部门的期望;估计一致次之,对点估计有一定的要求,对于可信区间(统计学意义)则放宽标准,在这种情况下,探索影响因素对结果造成的影响是十分有必要的;最后为检验一致,通常是最宽松的评价标准。

值得探讨的是造成结果之间差异的原因,主要包括仿真差异、残余混杂和偶然性等。仿真差异,即利用现实世界数据仿真临床试验时,在设计上存在的差异,主要从 PICOT,也就是人群(population)、干预(intervention)、对照(comparator)、结局(outcome)和时间(time)5 个方面展开,这些差异可能会导致 2 个研究探索的不是同一个医学问题,而人群的差异甚至会使在设计上完全一致的 2 个临床试验得到不一致的结果,残余混杂可能是由不恰当的对照引起的,也可能由于研究中存在未观测混杂等。针对这些可能的因素,该倡议的研究者将继续通过敏感性分析进行补充和讨论。

七、讨论与展望

临床试验是药物疗效和安全性评价的金标准,但在现代流行病学的发展中,无论是可及性还是操作性,观察性研究都更有优势^[39]。因此,现实世界研究,在人群代表性、样本量大小、数据范围等方面,有着不容忽视的优势。然而,限制观察性研究发展的原因有数据质量差、易受混杂影响和结果选择性报告等。如何有效利用现实世界数据进行疗效的因果推断产生现实世界证据,并划分证据等级,是各方关注的重点。观察性研究文献中,背景和讨论部分均引用历史临床试验,既说明进行观察性研究的理由(临床试验实施的限制),又将两者结果进行比较,然而在方法部分,却几乎不曾提及临床试验^[40]。为打破局限,在仿真目标试验中,研究者按照临床试验的设计思路,事先准备研究方案和统计分析计划等,从而提高研究质量,并探讨结果差异的潜在来源。

仿真目标试验的提出和发展必将成为观察性研究乃至流行病学史上的重大事件,仿真目标试验从设计上有效结合传统观察性研究和临床试验的优势^[41-42],为生物医学研究带来重大改变。未来,仿真目标试验必将作为临床试验的有力辅助证据^[43],为新药研发、监督管理、审批决策带来新的发展^[44]。

自现实世界研究等概念被提出起,有关现实世界研究是否可以成为临床试验的有效替代,一直是热点讨论话题。毫无疑问,临床试验和现实世界研

究均存在各自的优势与劣势。仿真目标试验可以看作两者的桥梁,充分考虑现实世界数据的广泛优势,在设计和分析上借鉴临床试验控制偏倚,在结果解读中再将两者结合起来,互相支持补充,完善药物获益风险评估。因此,仿真目标试验有希望助力、重复,甚至部分替代随机对照试验,促进新药审评上市、扩展已有药物的适应证范围^[7]。

近年来,相关案例不断出现,例如 OPERAND 等,实际上对于仿真目标试验的探索和应用,国内外仍存在很大的进步空间。对于偏倚的控制,可以采用倾向性评分等方法控制已观测混杂,对于研究中部分观察甚至未观察的混杂因素则无能为力。加强仿真目标试验中控制混杂因素的方法研究,可以提升应用研究质量。同时,观察性研究中还包括选择偏倚,尽管仿真目标试验中已经采用了一些措施控制选择偏倚,但在结果解释和差异解读时,仍需考虑这个因素可能的影响。各临床领域内,关于仿真目标试验与临床试验结果差异的解读仍然十分有限,需要进一步地探索和积累^[7,13,45]。另外,结果一致性评价,可以借鉴临床试验中的非劣效检验,在设计时,考虑数据和仿真差异对结果的影响,结合已有研究和背景知识,确定非劣效界值,利用非劣效检验评价一致性,或可作为辅助性评价指标。

本文介绍了目标试验的概念和设计,说明仿真目标试验有效利用现实世界数据和临床试验设计优势,提高现实世界研究质量,未来在方法和医学领域内的空白仍需更多研究。

利益冲突 所有作者声明无利益冲突

作者贡献声明 周佳薇:文献整合、论文撰写;黄丽红、尤东方:论文修订、经费支持;陈峰:指导性支持;赵杨:选题设计、经费支持

参 考 文 献

- [1] Hernán MA. With great data comes great responsibility: publishing comparative effectiveness research in epidemiology[J]. *Epidemiology*, 2011, 22(3):290-291. DOI: 10.1097/EDE.0b013e3182114039.
- [2] Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease[J]. *Epidemiology*, 2008, 19(6):766-779. DOI:10.1097/EDE.0b013e3181875e61.
- [3] Danaei G, Rodriguez LA, Canero OF, et al. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease[J]. *Stat Methods Med Res*, 2013, 22(1):70-96. DOI:10.1177/0962280211403603.
- [4] 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.
- [5] García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening[J]. *Eur J Epidemiol*, 2017, 32(6):495-500. DOI: 10.1007/s10654-017-0287-2.
- [6] Admon AJ, Donnelly JP, Casey JD, et al. Emulating a novel clinical trial using existing observational data. predicting results of the PreVent study[J]. *Ann Am Thorac Soc*, 2019, 16(8):998-1007. DOI:10.1513/AnnalsATS.201903-241OC.
- [7] Franklin JM, Paterno 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.
- [8] Liu RS, Rizzo S, Whipple S, et al. Evaluating eligibility criteria of oncology trials using real-world data and AI[J]. *Nature*, 2021, 592(7855):629-633. DOI:10.1038/s41586-021-03430-5.
- [9] Fang YX, He WL, Wang HW, et al. Key considerations in the design of real-world studies[J]. *Contemp Clin Trials*, 2020, 96:106091. DOI:10.1016/j.cct.2020.106091.
- [10] Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy[J]. *JAMA*, 2003, 290(12):1624-1632. DOI:10.1001/jama.290.12.1624.
- [11] Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies[J]. *Epidemiology*, 2010, 21(3):383-388. DOI:10.1097/EDE.0b013e3181d61eeb.
- [12] VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-Value[J]. *Ann Intern Med*, 2017, 167(4):268-274. DOI: 10.7326/M16-2607.
- [13] Veronesi G, Grassi G, Savelli G, et al. Big data, observational research and P-value: a recipe for false-positive findings? A study of simulated and real prospective cohorts[J]. *Int J Epidemiol*, 2020, 49(3):876-884. DOI:10.1093/ije/dyz206.
- [14] 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.
- [15] Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias[J]. *Epidemiology*, 2004, 15(5):615-625. DOI:10.1097/01.ede.0000135174.63482.43.
- [16] Duchesneau ED, Jackson BE, Webster-Clark M, et al. The timing, the treatment, the question: comparison of epidemiologic approaches to minimize immortal time bias in real-world data using a surgical oncology example[J]. *Cancer Epidemiol Biomarkers Prev*, 2022, 31(11):2079-2086. DOI:10.1158/1055-9965.EPI-22-0495.
- [17] Cain LE, Robins JM, Lanoy E, et al. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data[J]. *Int J Biostat*, 2010, 6(2):18. DOI:10.2202/1557-4679.1212.
- [18] 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.
- [19] 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.
- [20] 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.

- [21] Sheffield KM, Dreyer NA, Murray JF, et al. Replication of randomized clinical trial results using real-world data: paving the way for effectiveness decisions[J]. *J Comp Eff Res*, 2020, 9(15): 1043-1050. DOI: 10.2217/cer-2020-0161.
- [22] Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials[J]. *N Engl J Med*, 2012, 367(14): 1355-1360. DOI: 10.1056/NEJMSr1203730.
- [23] Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research[J]. *Clin Trials*, 2012, 9(1):48-55. DOI:10.1177/1740774511420743.
- [24] Sterne JA, Hernan 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. DOI:10.1136/bmj.i4919.
- [25] Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations[J]. *BMJ*, 2004, 328(7454):1490. DOI:10.1136/bmj.328.7454.1490.
- [26] Guyatt GH, Oxman AD, Kunz R, et al. Going from evidence to recommendations[J]. *BMJ*, 2008, 336(7652): 1049-1051. DOI:10.1136/bmj.39493.646875.AE.
- [27] Guyatt GH, Oxman AD, Kunz R, et al. What is "quality of evidence" and why is it important to clinicians? [J]. *BMJ*, 2008, 336(7651): 995-998. DOI: 10.1136/bmj.39490.551019.BE.
- [28] Guyatt GH, Oxman AD, Vist GE, et al. GRADE:an emerging consensus on rating quality of evidence and strength of recommendations[J]. *BMJ*, 2008, 336(7650): 924-926. DOI:10.1136/bmj.39489.470347.AD.
- [29] Brożek JL, Akl EA, Alonso-Coello P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions[J]. *Allergy*, 2009, 64(5): 669-677. DOI: 10.1111/j.1398-9995.2009.01973.x.
- [30] 中国人大网. 中华人民共和国医师法(2021年8月20日第十三届全国人民代表大会常务委员会第三十次会议通过)[EB/OL]. (2021-08-20) [2023-10-26]. http://www.npc.gov.cn/npc/c2/c30834/202108/t20210820_313104.html.
- [31] Xie Y, Bowe B, Gibson AK, et al. Comparative effectiveness of SGLT2 inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and sulfonylureas on risk of kidney outcomes: emulation of a target trial using health care databases[J]. *Diabetes Care*, 2020, 43(11): 2859-2869. DOI: 10.2337/dc20-1890.
- [32] Rogawski McQuade ET, Benjamin-Chung J, Westreich D, et al. Population intervention effects in observational studies to emulate target trial results: reconciling the effects of improved sanitation on child growth[J]. *Int J Epidemiol*, 2022, 51(1): 279-290. DOI: 10.1093/ije/dyab070.
- [33] 石舒原, 赵厚宇, 周庆欣, 等. 真实世界证据与随机对照试验: RCT DUPLICATE 项目方法学介绍[J]. *药物流行病学杂志*, 2020, 29(3):198-205. DOI: 10.19960/j.cnki.issn1005-0698.2020.03.012.
- Shi SY, Zhao HY, Zhou QX, et al. Real-world evidence and randomized controlled trials: the methodology of RCT DUPLICATE[J]. *Chin J Pharmacoepidemiol*, 2020, 29(3): 198-205. DOI: 10.19960/j.cnki.issn1005-0698.2020.03.012.
- [34] 石舒原, 周庆欣, 孙凤, 等. 真实世界证据与随机对照试验: RCT DUPLICATE 项目成果[J]. *药物流行病学杂志*, 2019, 28(11): 757-762. DOI: 10.19960/j.cnki.issn1005-0698.2019.11.013.
- Shi SY, Zhou QX, Sun F, et al. Real-world evidence and randomized controlled trials: the outcome of RCT DUPLICATE[J]. *Chin J Pharmacoepidemiol*, 2019, 28(11): 757-762. DOI: 10.19960/j.cnki.issn1005-0698.2019.11.013.
- [35] 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. DOI: 10.19960/j.cnki.issn1005-0698.2019.08.002.
- 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. DOI:10.19960/j.cnki.issn1005-0698.2019.08.002.
- [36] Wang SV, Schneeweiss S, RCT-DUPLICATE Initiative, et al. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials[J]. *JAMA*, 2003, 329(16):1376-1385. DOI:10.1001/jama.2023.4221.
- [37] 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.
- [38] 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(二)[J]. *中华流行病学杂志*, 2022, 43(11): 1835-1841. DOI: 10.3760/cma.j.cn112338-20220513-00409.
- 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 2) [J]. *Chin J Epidemiol*, 2022, 43(11): 1835-1841. DOI: 10.3760/cma.j.cn112338-20220513-00409.
- [39] West SG. Alternatives to randomized experiments[J]. *Curr Direct Psychol Sci*, 2009, 18(5): 299-304. DOI: 10.1111/j.1467-8721.2009.01656.x.
- [40] Labrecque JA, Swanson SA. Target trial emulation: teaching epidemiology and beyond[J]. *Eur J Epidemiol*, 2017, 32(6):473-475. DOI:10.1007/s10654-017-0293-4.
- [41] 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.
- [42] Didelez V. Commentary: should the analysis of observational data always be preceded by specifying a target experimental trial? [J]. *Int J Epidemiol*, 2016, 45(6): 2049-2051. DOI:10.1093/ije/dyw032.
- [43] 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.
- [44] Eichler HG, Koenig F, Arlett P, et al. Are novel, nonrandomized analytic methods fit for decision making? The need for prospective, controlled, and transparent validation[J]. *Clin Pharmacol Ther*, 2020, 107(4):773-779. DOI:10.1002/cpt.1638.
- [45] Gibbons RJ, Gardner TJ, Anderson JL, et al. The American heart association's principles for comparative effectiveness research: a policy statement from the American heart association[J]. *Circulation*, 2009, 119(22): 2955-2962. DOI:10.1161/CIRCULATIONAHA.109.192518.