

中医药随机对照试验中采用真实世界数据作为对照组的 研究设计及挑战

胡晶^{1,2}, 李博^{1,2}, 张会娜^{1,2}, 冯硕^{1,2}, 廖星^{3*}

(1. 首都医科大学附属北京中医医院, 北京市中医药研究所, 北京 100010;

2. 北京中医药循证医学中心, 北京 100010;

3. 中国中医科学院中医临床基础医学研究所, 循证医学基础研究室, 北京 100700)

[摘要] 在设计随机对照试验(RCT)时,如果对照组存在患者招募和入组困难的情况,就会影响试验整体实施。近年来,真实世界数据(RWD)作为除RCT之外的数据来源,在医疗领域中发挥着越来越重要的作用。中医药RCT中可以尝试采用将RWD作为对照组的试验设计,不仅可以有效解决中医药RCT西医对照组患者入组困难的问题,同时能提供有力证据来评价中医药的疗效。倾向评分法目前已广泛应用于真实世界研究中混杂因素的处理,该文对RCT采用RWD作为对照组的这类设计中,基于倾向评分法常见的4种研究设计形式以实例分别进行了介绍,包括不对称随机分配、基于倾向评分分层法的两阶段设计、倾向评分联合复合似然法及倾向评分多种方法的联合。同时,这种设计类型也存在着方法学的挑战,包括RWD数据源必须是高质量且关键信息需要规范收集、RCT和RWD患者基线特征应该具有可比性、协变量选择时需要把所有已知与干预措施和结局相关的协变量都纳入进行分析等。在中医药领域采用这种设计时,还存在着有些RWD中医证型信息缺失、中医结局指标缺失等问题,在使用RWD时,需要根据数据实际情况决定如何分析。该文对以RWD作为RCT对照组的设计类型及面临的方法学挑战进行了介绍,期望能为研究者今后使用这类设计提供方法学借鉴。

[关键词] 随机对照试验; 真实世界数据; 中医药; 倾向评分法; 疗效评价

[中图分类号] R242;R2-031;R965.3;R969.4 **[文献标识码]** A **[文章编号]** 1005-9903(2023)01-0181-07

[doi] 10.13422/j.cnki.syfjx.20230194

[网络出版地址] <https://kns.cnki.net/kcms/detail/11.3495.R.20220901.1048.002.html>

[网络出版日期] 2022-09-01 15:06

Design and Challenges of Real World Data as Control Group in A Randomized Controlled Trial of Chinese Medicine

HU Jing^{1,2}, LI Bo^{1,2}, ZHANG Huina^{1,2}, FENG Shuo^{1,2}, LIAO Xing^{3*}

(1. Beijing Hospital of Traditional Chinese Medicine, Capital Medical University, Beijing Institute of Chinese Medicine, Beijing 100010, China;

2. Beijing Evidence-based Chinese Medicine Center, Beijing 100010, China;

3. Center for Evidence-based Chinese Medicine, Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China)

[Abstract] In the design of randomized controlled trials (RCT), difficulties in patient recruitment and enrollment of control group would limit the overall implementation of the trials. In recent years, as a data source, real-world data (RWD) plays an increasingly important role in the medical field. In RCT of Chinese medicine, RWD could be designed as control group. This design can effectively solve the problem of inclusion

[收稿日期] 2022-04-26

[基金项目] 国家自然科学基金面上项目(81973694);中国中医科学院科技创新工程项目(CI2021A00701-3)

[第一作者] 胡晶,博士,副研究员,从事循证中医药及疗效评价方法学研究,E-mail:hujingebm@163.com

[通信作者] *廖星,博士,研究员,从事循证中医药临床评价研究,E-mail:okfrom2008@hotmail.com

difficulty for the patients in the western medicine control group of traditional Chinese medicine(TCM) RCT, it also can provide high quality evidence to evaluate the efficacy of TCM. In recent years, propensity score method has been widely used to deal with confounding factors in real world study. In this paper, four common research designs based on propensity score method were introduced with examples, including propensity score matching and data augmentation, two-stage design of propensity score stratification method, propensity score-integrated composite likelihood approach, and combination of different propensity score methods. However, there are some methodological challenges in this type of design, including the RWD data sources must be of high quality and the key information needs to be collected in a standardized method, the baseline characteristics of RCT and RWD patients should be comparable, and all known covariates related to the intervention and outcome need to be included for analysis. When adopting this design in the field of TCM, there are still some problems such as the lack of TCM syndrome information and TCM outcomes in RWD. When using RWD, it is necessary to decide how to analyze according to the data condition. This paper discussed the design types and methodological challenges of using RWD as control group in RCT, hoping to provide methodological ideas for researchers to use this type of design in the future.

[Keywords] randomized controlled trial; real-world data; traditional Chinese medicine; propensity score method; efficacy evaluation

随机对照试验(RCT)被认为是评价临床干预措施有效性的金标准,理论上,RCT通过随机的方法,使观察组和对照组已知或未知的混杂因素达到均衡,避免了患者的选择性偏倚,达到对治疗效应的无偏估计^[1]。因此,RCT中干预措施对疾病的疗效可以通过对观察组和对照组的结局指标进行直接比较来估计,因为两组间的差异不是由于混杂因素所引起的,而是仅归因于治疗^[2]。自1982年第1篇中药RCT报告以来^[3],RCT方法已广泛运用于评价中医药临床疗效。但中医药RCT在实施过程中,常常会面临对照组患者入组困难的情况,例如,观察组为中医药,对照组为西药或常规西医治疗,但由于来中医院就诊的患者大多是希望使用中医药,不愿意进入西医对照组,这会对RCT实施过程中患者的招募和入组造成困难,进而影响试验进度。另外,分配到西医对照组的患者,可能试验中途会使用中医药而造成沾染,也会造成RCT中对照组的脱落率更高。在这种情况下,可以考虑这样一类研究设计,即在中医药RCT中采用真实世界数据(RWD)作为对照组^[4-5],以评价中医药疗效。

近年来,RWD作为除RCT之外的数据来源,在医疗领域中发挥着越来越重要的作用。RWD是指来源于日常所收集的各种与患者健康状况和(或)诊疗及保健有关的数据,常见的RWD来源包括医疗信息系统(HIS)、医保系统、疾病登记系统、自然人群队列和专病队列数据库等^[6-7]。真实世界研究(RWS)是指针对预设的临床问题,在真实世界环境

下收集与研究对象健康有关的数据(即RWD)或基于这些数据衍生的汇总数据,通过分析,获得药物的使用情况及潜在获益-风险的临床证据,即真实世界证据(RWE)的研究过程。RWD通过恰当的分析 and 严格的质量控制,可以产生RWE,用于支持医疗决策。近几年来,国家先后制订了一系列支持RWE用于上市后安全监测、儿童药物研发与审评等决策支持的文件^[7-9]。而国内学术界对于RWS也掀起了热潮高浪。而实际上,中医药领域最早引入真实世界概念和研究理念的文章发表于2010年^[10]。目前,RWD已被广泛应用于中医药研究领域,除经典的实用性RCT和各类观察性研究,RWD还可以用于解决中医药RCT对照组招募困难的问题,即中医药RCT中采用RWD作为对照组。这类设计可以有效解决中医药RCT中西医对照组患者入组困难的问题,同时能提供有力证据来评价中医药的疗效。本文对这种基于RWD设置RCT对照组的4种研究设计形式进行全面介绍,并分析其存在的问题和挑战,以期为未来临床研究者提供方法学思路。

1 以RWD作为RCT对照组的设计类型介绍

对于控制混杂因素的方法,Rosenbaum和Rubin于1983年提出倾向评分法(PSM)^[11-12],这种方法能够有效地均衡非随机对照研究中各对比组间协变量的分布,在组间协变量均衡的基础上进一步评价处理因素的效应,从而得到接近RCT的结果。倾向评分(propensity score)是指在一定协变量条件下,一个观察对象可能接受某种处理因素的可能性,

是一个从0到1范围内连续分布的概率值。PSM控制混杂偏倚的方法主要有4种:匹配、分层、校正和加权。PSM实施时,首先以处理因素作为因变量,其他多个混杂变量作为自变量,通过建立1个回归模型(最为常用的是logistic回归模型)来估计每个研究对象接受处理因素的可能性,即计算每个研究对象的倾向评分,通过平衡各对比组间的倾向评分从而有效地均衡混杂变量的分布,达到控制混杂偏倚的目的^[13-14]。既往有关PSM的广泛应用,多是基于数据分析阶段,有关各种已知混杂因素的处理^[15-17],而在设计阶段引入PSM是近几年来一种备受真实世界研究者青睐的方法。最为典型的就是RCT中采用RWD作为对照组的研究设计,这种设计采用PSM的目的是均衡RCT中观察组患者和RWD中对照组患者临床重要的协变量。以下将介绍目前医学研究领域中,比较常见的基于PSM,采用RWD作为RCT对照组的研究设计。

1.1 不对称随机分配的研究设计 不对称随机分配的研究设计是指RCT中随机一小部分患者到对照组,而不是完全采用RWD中的患者作为对照组^[18],这样可以对随机到对照组的这部分患者进行分析,目的是减少非同对照的偏倚,例如由于医疗实践或疾病诊疗标准等改变引起的偏倚。研究以1个随机、双盲、双模拟、多中心、非劣效的RCT(观察组和对照组的比例为1:1)为实例,试验计划纳入600例急性细菌性皮肤感染患者,观察组为一种口服抗生素,对照组为一种阳性药,主要结局指标为患者就诊48~72h原发灶表面积减少20%或更多。在RCT设计时,观察组将随机分配300例患者,对照组随机分配150例,剩余150例对照组计划从RWD中获得。RWD经筛选后,共有450例患者符合RCT的纳入排除标准,可作为对照组。研究采用倾向评分匹配法的2种匹配方案:①1:1精确匹配(pair-matching)。将RCT中的观察组人群与RWD中筛选出来的对照组进行1:1匹配(即300:300),在匹配的人群中随机选择需要的患者例数(150例),将匹配的对照组和RCT中的对组合并构成最终需要的对照组。见图1A。②根据卡尺调整的最邻近匹配法(nearest-neighbor matching adjusted by caliper)。确定卡尺距离,在RCT的对照组人群和RWD中寻找距离观察组倾向评分最邻近的人群,如果从RWD中获得的对照组患者的数量少于或多于所需数量,增加或减少卡尺距离,重新进行最邻近匹配,并且持续这个过程直到获得期望的对照组

患者数量见图1B。

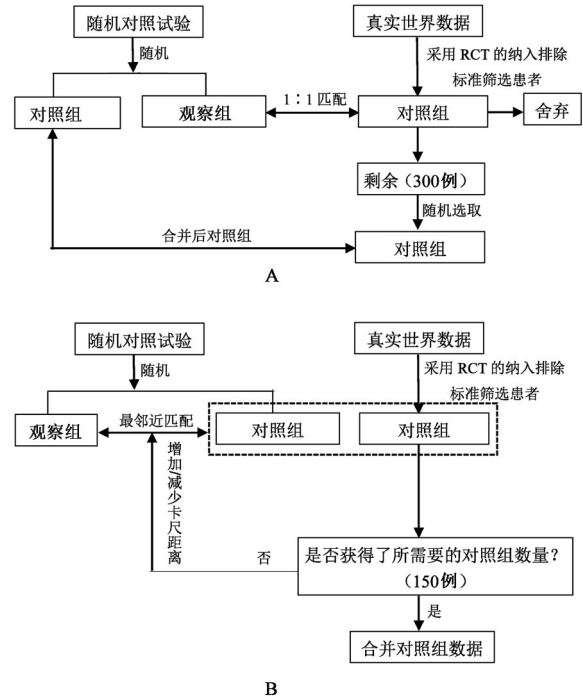


图1 不对称随机分配RCT中基于倾向评分匹配法的2种匹配方案
Fig. 1 Two matching methods based on propensity score matching method in an asymmetrically randomly assigned RCT

1.2 基于倾向评分分层法的两阶段设计 倾向评分分层法的原理与传统的分层分析方法基本相同,只是分层变量不是每个混杂变量,而是倾向评分。在实施时,首先将计算出来的倾向评分进行排序,然后按照倾向评分的百分位数将全部研究对象划分若干个层(一般5~10层)。研究者根据两组人群的倾向评分或某一组人群的倾向评分来确定每一层的临界值,最常用的方法是根据两组共同倾向评分等分为若干层。有研究者以1个两臂、非劣效RCT为实例对基于倾向评分分层法两阶段设计进行了具体应用介绍^[19],观察组为一种金属裸支架,用于改善有症状缺血性疾病的症状,阳性对照组为血管内近距离放射治疗,主要结局指标是靶血管失败率。该RCT以2:1的比例将患者随机分配至观察组和对照组,观察组148例,对照组74例。1个注册登记库中的历史数据已经收集了接受血管内近距离放射治疗的患者,因此该研究决定将从该库中选取的患者作为RCT的对照组。注册登记库经筛选后,共有334例患者符合RCT的纳入排除标准。研究确定了6个基线协变量作为混杂变量,采用倾向评分分层法分为5层,第一步是将RCT中222例患者与RWD中334例患者进行协变量的均衡性分析,如果未达到均衡,则重新进行此过程;如果达到

均衡,则进行第二步,将RCT中148例观察组患者和所有的408例对照组患者进行协变量的均衡性分析,如果未达到均衡,则重新进行前述过程,如果达到均衡,则对临床疗效进行统计分析。具体步骤见图2。

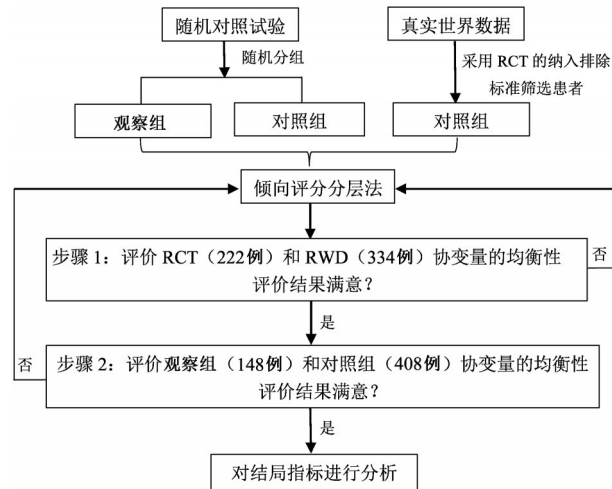


图2 基于倾向评分分层法的两阶段设计流程
Fig. 2 Two stage design flow chart based on propensity score stratification method

1.3 倾向评分联合复合似然法 倾向评分联合复合似然法(propensity score-integrated composite likelihood approach)是将PSM和复合似然函数联合使用的一种方法。有研究者以1个RCT作为实例对方法进行介绍^[20],试验目的是评价一种心血管设备在充血性心力衰竭患者中的有效性和安全性,在进行样本量估计后,该试验共需要354例患者(观察组和对照组的比例为1:1)。试验计划纳入267例患者,以2:1的比例将患者随机分配至心血管设备组(183例)和对照组(84例),剩余87例对照组患者计划从RWD中获得。RWD经筛选后,共有1570例患者符合RCT的纳入排除标准,进一步整理后,剩余1192例。研究确定了17个基线协变量作为混杂变量,首先计算每个患者的倾向评分,根据评分值将所有患者(267例+1192例)分为5层,并使每层内的患者数量基本相等。在每层内,检查和比较RCT患者和RWD患者协变量的分布情况。采用倾向评分分层的原因是因为在每个层内,来自RWD和RCT的患者在基线协变量方面往往比总体患者更相似。另外,由于RWD的患者数量比RCT的患者数量多,研究者提出在每一层内建立1个复合似然函数,用于对RWD的患者降低权重,以便在利用RWD患者数据时他们不会主导研究结果。然后通过最大化

复合似然函数对每一层的疗效指标进行分析,然后将每一层的分析结果合并起来以获得患者总体的疗效评价结果。具体步骤见图3。

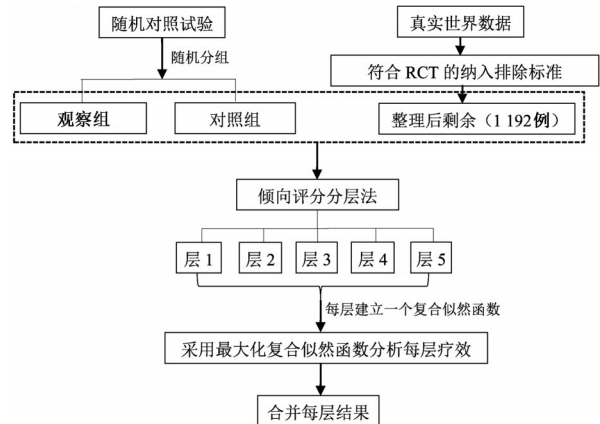


图3 倾向评分联合复合似然法的设计流程
Fig. 3 Design flow chart based on propensity score-integrated composite likelihood approach

1.4 倾向评分多种方法的联合 YUE等^[21]的研究以1个非劣效的RCT为实例,目的是评价一种心血管设备在治疗心血管疾病中的安全性和有效性,主要结局指标是治疗的成功率。经样本量估计,观察组需要250例患者,在筛选RWD后,共有1000例患者符合RCT的纳入排除标准,研究确定了15个基线协变量作为混杂变量。研究提出了3种设计方案:①设计1,倾向评分分层后再进行分层。首先采用倾向评分分层法将所有患者(观察组250例,对照组1000例)分为5层,并使每层内的患者例数相等,假如1个层内没有治疗组患者,说明患者在本层内不均衡,将层内的对照组患者去除。再对剩余患者采用倾向评分分层法分为5层,可以发现组间的可比性提高了。②设计2,1:1倾向评分匹配后进行分层。采用1:1最邻近匹配法,从RWD中对观察组的每个患者进行1:1倾向评分匹配,匹配后有250对患者。对这250对患者协变量的均衡性进行分析,如果不均衡,则采用倾向评分分层法分为5层。与设计1相比,设计2的观察组和对照组在各个倾向评分分层中的患者数量更加均衡,但设计2对照组患者例数为250例,比设计1的500例少。③设计3,2:1倾向评分匹配后进行分层。该设计按照2:1(对照组-观察组)的比例匹配两组患者,之后的过程与设计2类似。YUE等^[21]的研究指出,除了这3种设计方法之外,还可以有多种方法进行设计。另外,研究也对这3种设计方法的把握度进行了估计,结果显示这3种方法之间的把握度通常是可比的。

2 以RWD作为RCT对照组的挑战

2.1 RWD数据源的质量第一重要 美国食品药品监督管理局(FDA)在2017年发布的《使用真实世界证据支持医疗器械监管决策指南》^[22]中指出,只有当RCT和RWD这2种数据的合成从临床和监管的角度来看是合理的,并且RWD数据源是高质量时,才能作为RCT的对照组。我国药品监督管理局药品审评中心在2021年发布的《用于产生真实世界证据的真实世界数据指导原则(试行)》^[23]中,也对RWD的来源、评价、质量保障等方面,给出了具体要求和指导性建议。在以RWD作为RCT对照组的设计中,RCT中所包含的所有关键信息包括基线协变量和结局指标信息等,RWD中都需要准确的收集和记录,并且数据收集、汇总、管理和记录方法都需要是经过验证和标准化的^[24]。如果一些重要的关键信息在数据库中没有收集,或者数据收集不规范的话,用这个数据库做为对照组就存在着问题。另外,RWD中退出或失访的患者比例及缺失数据也不应过多。

2.2 RCT和RWD患者基线特征的可比性 RWD人群和RCT人群应具有同质性,换句话说,RWD和RCT中人群的基线协变量应该具有均衡性,如果不均衡,则可能出现选择性偏倚,此时将RWD人群作为RCT对照组的合理性可能就会受到质疑。研究者通常认为PSM是处理基线不均衡性的一种有效的统计分析方法,但在实际情况中,即使采用了PSM,RCT和RWD中人群的基线特征可能仍然不可比。YU等^[25]采用1个观察组和对照组非同期的双臂研究作为例子,两组间有20个基线协变量是不均衡的,另外,由于两组是非同期的,而此期间疾病治疗随着时间发生了很大变化,因此患者入组时间也是1个重要的协变量。由于两组患者基线特征不具有可比性,因此对临床疗效直接进行组间比较是不恰当的,为了试图解决这个问题,研究进行了PSM,但结果显示不论是将患者入组时间纳入倾向评分模型(图4A)或不纳入(图4B),两组间的倾向评分分布都无法达到很好的重叠。由于RCT和RWD中患者入组的时间不一样,患者治疗效果可能因医疗技术的发展或新指南的推出而不一致,再有,结局的定义或判定标准可能也会随着时间推移而改变。有研究者^[26]建议采用与RCT患者平行的数据作为对照组要优于历史对照,平行对照组可采用疾病注册登记模式,并且数据记录应尽可能完整和准确。另外,如果RWD和RCT的人群特征存在

着明显差异,例如种族差异,也不建议将RWD作为RCT的对照组。在这种情况下,针对这个RWD的任何继续分析都可能导致结果无法解释,为此,只能是放弃这个RWD数据来源,寻找另外一个与RCT人群具有同质性的数据源才是可行之举。

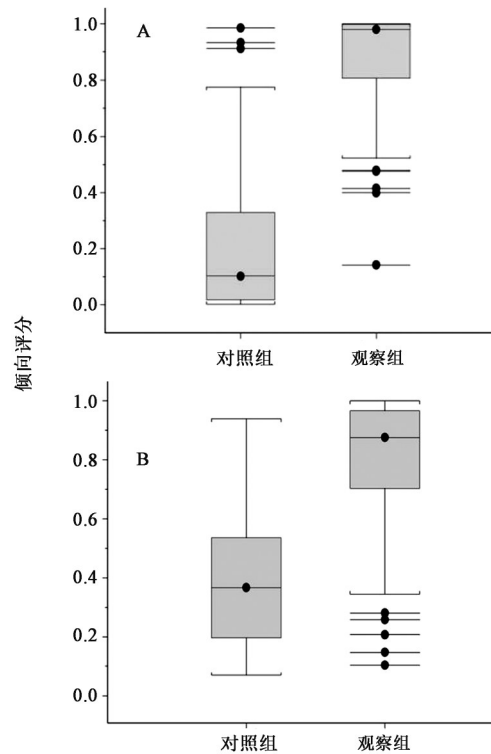


图4 观察组和对照组的倾向评分分布^[21]

Fig. 4 Distribution of propensity scores in the test group and the control group

2.3 协变量的选择 考虑混杂因素时,需要把所有已知的与干预措施和结局相关的协变量都纳入进来进行分析。但是在实践中,可能很难判断一个协变量是否是混杂因素。因此,研究中最好纳入尽可能多的已收集的变量。如果一个关键的混杂因素没有被纳入倾向评分模型中,可能会导致对治疗效果的错误估计^[25]。这些混杂因素的确定应该是一个基于科学推理和最新临床研究的深思熟虑的过程。确定混杂因素的这个过程不应该和结局指标进行分析,以避免潜在的偏倚,确保研究结论的可靠性^[26]。另外,还有一种做法需要避免,即基于逐步回归的结果或者最能预测治疗分配的模型选择协变量,因为这种方法也可能会排除一些关键的协变量,如果这样,所有协变量在组间达到均衡的最终目标可能就无法达成^[23]。

YUE等^[21]的研究指出,在提交给美国FDA的一些医疗器械案例中,为了实现组间所谓“更好”的

倾向评分平衡,从倾向评分建模中排除了一些关键协变量。这种做法是不可取的,因为排除一些协变量可能会导致“更好”的倾向评分分布平衡,但往往无法实现所有基线协变量达到均衡的研究目标。该研究者团队通过1个实例验证了这一问题^[27],即拟合了两个倾向评分模型的1个研究中,第一个模型纳入10个协变量(模型I),第二个模型纳入15个协变量(模型II)。研究选择了模型I作为分析模型,部分原因是模型I在治疗组和对照组之间的倾向评分分布有“更好”的重叠。模型II中纳入了一些组间不均衡的协变量,但模型I并未纳入。如果模型I不对这些协变量进行调整,那这些不均衡的协变量即使经过倾向评分法之后,可能仍然不均衡,模型I所谓的“更好”的重叠可能会导致错误的研究结论。

3 总结

近年来,RWD作为除RCT之外证据的数据来源,在医疗领域中的应用越来越广泛。2020年1月7日国家药品监督管理局发布了《真实世界证据支持药物研发与审评的指导原则(试行)》^[7]中提及RWD有3大类,其中一类为使用RWE作为外部对照的单臂试验,单臂临床试验也是验证研究药物有效性和安全性的一种方法。但如果试验中仅采用RWD作为外部对照,RCT和RWD中患者入组的时间不一样,患者治疗效果可能因医疗技术的发展或新指南的推出而不一致,另外,结局的定义或判定标准可能也会随着时间而改变。因此,在RCT中,一部分对照组患者采用RWD作为补充,这类设计在有效解决对照组患者入组困难的同时,也能减少非同期对照的偏倚,例如由于医疗实践或疾病诊疗标准等改变引起的偏倚。但这种设计类型也给研究设计、数据分析和结果解释带来了挑战。首先能作为RCT对照组的RWD数据源需要是高质量的,RCT中所包含的所有关键信息都需要在RWD中准确的收集和记录。RCT和RWD中患者的基线协变量也应该具有均衡性,如果不均衡,合并2种数据的合理性可能就会收到质疑。混杂因素的确定也应该是一个基于科学推理和最新临床研究的深思熟虑的过程,如果关键的混杂因素没有被纳入倾向评分模型中,可能会导致对治疗效果的错误估计。

在中医药RCT中采用RWD作为对照组的研究设计中,除了存在以上挑战外,在中医药临床研究领域还面临一些特殊的方法学挑战,例如,如果一个中医药RCT的目的是评价某种中药对特定中医

证型患者的疗效,对照组为西药,但在一些西医医院的HIS库或注册登记库中,可能并没有患者的中医证型信息,在这种情况下,选择这些数据库可能就存在着这种研究设计无法实施。另外,如果一个中医药RCT的结局指标是中医证候等,HIS库中可能也没有相关信息,此时可能也会导致这种研究设计无法实施。建议在RCT中采用RWD作为对照组的研究设计中,选择客观、公认的结局指标,例如病死率、住院率或实验室指标等。

[利益冲突] 本文不存在任何利益冲突。

[参考文献]

- [1] BYAR D P, SIMON R M, FRIEDEWALD W T, et al. Randomized clinical trials. Perspectives on some recent ideas[J]. N Engl J Med, 1976, 295(2): 74-80.
- [2] GREENLAND S, PEARL J, ROBINS J M. Causal diagrams for epidemiologic research [J]. Epidemiology, 1999, 10(1): 37-48.
- [3] 陈可冀, 钱振准, 张问渠, 等. 精制冠心病片对冠心病心绞痛双盲法治疗112例疗效分析[J]. 医学研究通讯, 1982, 11(2): 24-25.
- [4] HANNAN E L. Randomized clinical trials and observational studies: Guidelines for assessing respective strengths and limitations [J]. JACC Cardiovasc Interv, 2008, 1(3): 211-217.
- [5] SHERMAN R E, ANDERSON S A, DAL PAN G J, et al. Real-world evidence - What is it and what can it tell us?[J]. N Engl J Med, 2016, 375(23): 2293-2297.
- [6] U. S. Food and Drug Administration. Framework for FDA's real-world evidence program [EB/OL]. (2019-12-26) [2022-04-26]. <https://www.fda.gov/media/120060/download>.
- [7] 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]. (2021-5-16) [2022-04-26]. <https://www.nmpa.gov.cn/xxgk/ggtg/qtggtg/20200107151901190.html>.
- [8] 国家药品监督管理局. 真实世界研究支持儿童药物研发与审评的技术指导原则(试行)[EB/OL]. (2021-8-7) [2022-04-26]. <https://www.nmpa.gov.cn/directory/web/nmpa/images/1598928284781079702.pdf>.
- [9] 国家药品监督管理局. 真实世界数据用于医疗器械临床评价技术指导原则(试行). [EB/OL]. (2021-3-17) [2021-8-7]. <https://www.nmpa.gov.cn/xxgk/ggtg/qtggtg/20201126090030150.html>.
- [10] 田峰, 谢雁鸣. 真实世界研究: 中医干预措施效果评价的新理念[J]. 中西医结合学报, 2010, 8(4): 301-306.

- [11] ROSENBAUM P R, RUBIN D B. The central role of the propensity score in observational studies for causal effects[J]. *Biometrika*, 1983, 70(1):41-55.
- [12] ROSENBAUM P R, RUBIN D B. Reducing bias in observational studies using subclassification on the propensity score[J]. *JASA*, 1984, 79:516-524.
- [13] RUBIN D B. The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials[J]. *Stat Med*, 2007, 26(1):20-36.
- [14] RUBIN D B. For objective causal inference, design trumps analysis [J]. *Ann Appl Stat*, 2008, 2(3):808-840.
- [15] BHAVAN K P, BROWN L S, HALEY R W. Self-administered outpatient antimicrobial infusion by uninsured patients discharged from a safety-net hospital: A propensity-score-balanced retrospective cohort study[J]. *PLoS Med*, 2015, 12(12):e1001922.
- [16] FRANCESCA L, BEAUDOIN, ROEE G, et al. Persistent pain after motor vehicle collision: comparative effectiveness of opioids versus non-steroidal anti-inflammatory drugs prescribed from the emergency department—a propensity matched analysis [J]. *Pain*, 2017, 158(2):289-295.
- [17] HAN L, GOULET J L, SKANDERSON M, et al. Evaluation of complementary and integrative health approaches among US veterans with musculoskeletal pain using propensity score methods [J]. *Pain Med*, 2019, 20(1):90-102.
- [18] LIN J, GAMALO-SIEBERS M, TIWARI R. Propensity score matched augmented controls in randomized clinical trials: A case study [J]. *Pharm Stat*, 2018, 17(5):629-647.
- [19] XU Y, LU N, YUE L, et al. A study design for augmenting the control group in a randomized controlled trial: A quality process for interaction among stakeholders [J]. *Ther Innov Regul Sci*, 2020, 54(2):269-274.
- [20] CHEN W C, WANG C, LI H, et al. Propensity score-integrated composite likelihood approach for augmenting the control arm of a randomized controlled trial by incorporating real-world data [J]. *J Biopharm Stat*, 2020, 30(3):508-520.
- [21] YUE L Q, LU N, XU Y. Designing premarket observational comparative studies using existing data as controls: challenges and opportunities [J]. *J Biopharm Stat*, 2014, 24(5):994-1010.
- [22] U. S. Food and Drug Administration. Use of real-world evidence to support regulatory decision-making for medical devices [EB/OL]. (2020-9-13) [2022-04-26]. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices>.
- [23] 国家药品监督管理局药品审评中心. 2021.《用于产生真实世界证据的真实世界数据指导原则(试行)》. [EB/OL]. (2021-12-31) [2022-04-26]. <https://www.cde.org.cn/main/news/viewInfoCommon/2a1c437ed54e7b838a7e86f4ac21c539>.
- [24] HILL J L, REITER J P, ZANUTTO E L. Applied Bayesian Modeling and Causal Inference From an Incomplete-Data Perspective [M]. New York: Wiley, 2004:44-56.
- [25] YU L Q. Statistical and regulatory issues with the application of propensity score analysis to nonrandomized medical device clinical studies [J]. *J Biopharm Stat*, 2007, 17(1):1-13.
- [26] RUBIN D B. Using propensity scores to help design observational studies: Application to the tobacco litigation [J]. *HEOR*, 2001, 2:169-188.
- [27] YU L Q. Regulatory considerations in the design of comparative observational studies using propensity scores [J]. *J Biopharm Stat*, 2012, 22(6):1272-1279.

[责任编辑 王鑫]