

# 偏倚风险评估系列:(三)交叉设计随机对照试验

武珊珊 孙凤 詹思延

100050 北京,首都医科大学附属北京友谊医院国家消化系统疾病临床医学研究中心(武珊珊); 100191 北京大学公共卫生学院流行病与卫生统计学系(孙凤、詹思延); 100191 北京大学循证医学中心(孙凤、詹思延)

通信作者:詹思延, Email: siyan-zhan@bjmu.edu.cn

DOI: 10.3760/cma.j.issn.0254-6450.2017.10.028

**【摘要】** 针对交叉设计随机对照试验 Cochrane 偏倚评估工具 2.0 版本(RoB2.0)的主要内容进行详细介绍,主要阐述了与平行设计 RoB2.0 的不同之处,并举例说明交叉设计 RoB2.0 的使用方法和注意事项。交叉设计 RoB2.0 针对交叉设计的自身特点,设置了相应的信号问题,为交叉设计试验纳入系统综述进行证据整合提供偏倚风险信息。

**【关键词】** 偏倚风险; 评估工具; 随机对照试验; 交叉设计; 系统综述

**Bias risk assessment: (3) Revised Cochrane bias risk assessment tool for individual randomized, cross-over trials (RoB2.0)** Wu Shanshan, Sun Feng, Zhan Siyan

National Clinical Research Center of Digestive Diseases, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China (Wu SS); Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China (Sun F, Zhan SY); Center of Evidence-Based Medicine and Clinical Research, Peking University, Beijing 100191, China (Sun F, Zhan SY)

Corresponding author: Zhan Siyan, Email: siyan-zhan@bjmu.edu.cn

**【Abstract】** This paper introduces the revised Cochrane bias risk assessment tool for individual randomized, cross-over trials (RoB2.0), comparing RoB2.0 of cross-over design with RoB2.0 of parallel design, and illustrating the application of RoB2.0 for cross-over trials in a published clinical trial. As a comprehensive tool specially designed for cross-over design, it provides more information on risk of bias in systematic reviews for the evidence synthesis.

**【Key words】** Risk of bias; Tool for assessment; Randomized controlled trial; Cross-over design; Systematic review

## 一、交叉设计 RoB2.0 的制定背景

与平行设计试验为每位受试者分配一种干预措施并进行组间比较不同,交叉设计试验是按事先设计好的试验顺序,在各个时期对研究对象逐一实施各种干预,以比较各干预组间的差异,是一种将自身比较和组间比较设计思路综合应用的研究设计类型。一个简单的随机交叉设计就是“AB/BA”设计,在试验的第一阶段受试者先接受干预 A 或干预 B,然后第二阶段分别交叉到干预 B 或干预 A,在每个阶段相当于一个平行设计试验。更为复杂的交叉设计类型包括 3 种或 3 种以上的干预措施以及 3 个或 3 个以上的处理阶段。与平行设计试验相比,交叉设计的优势在于:①每个受试者都是其自身的对照,消除了受试者个体间的变异。②在达到同等检验效

能的情况下,样本量只有平行设计试验的一半。③每个受试者接受每种干预措施,从而有利于筛选出最佳干预措施。

交叉设计适合于能够暂时缓解的、稳定的慢性疾病的短期效果评估,如应用于缓解哮喘、癫痫及透析等疾病的疗效研究。但在以下情况中应用交叉设计则可能不妥:①疾病病情会随时间而进展,尤其是病情进展迅速时,如退行性病变,在试验过程中病情会逐步发生恶化。②主要结局为不可逆性指标,如死亡、怀孕率等,或干预措施彻底治愈了相应疾病。③在干预措施疗效持久时,如药物的半衰期较长,则在前一阶段给予的干预措施,其疗效会持续到后一阶段,即存在滞后效应。④前后两个处理阶段之间的洗脱期,引发了第二处理阶段的撤退效应。

因此,在对交叉设计试验进行偏倚风险评估时,需要考虑交叉设计自身的特点。因此,Cochrane方法学工作组在其2011年发表的偏倚评估工具1.0版本(Cochrane Collaboration's tool for assessing risk of bias in randomized trials, RoB1.0)基础上<sup>[1]</sup>,对该工具进行了更新,专门针对交叉设计制定了相应的偏倚评估工具,即交叉设计偏倚评估工具2.0版本(RoB2.0)。该工具于2016年10月20日在网站(<https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool>)发布<sup>[2]</sup>。本次介绍的交叉设计偏倚评估工具2.0版本(revised Cochrane risk of bias tool for cross-over trials, RoB 2.0)主要针对两种干预措施、两个处理阶段和两种顺序的交叉设计,即“AB/BA”设计。

二、工具解读

交叉设计 RoB2.0 在领域设置上与平行设计 RoB2.0 类似,包括5个领域:①随机化过程中的偏倚;②偏离既定干预的偏倚;③结局测量的偏倚;④结局数据缺失的偏倚;⑤结果选择性报告的偏倚。最后还有对这5个领域的整体评价。二者的不同之处主要围绕3个方面:①是否存在滞后效应;②是否存在阶段效应;③是否仅报告第一阶段的结果。即在上述领域信号问题的设置上有所调整(表1)。

1. 信号问题调整的解读:由于交叉设计 RoB2.0 与平行设计 RoB2.0 各领域的大部分信号问题类似,故本文将不再对类似内容逐一解读,仅针对表1中列出的调整信号问题进行详细阐述。

(1)随机化过程中的偏倚:该领域主要针对是否存在阶段效应这一问题,对相应的信号问题进行了调整。阶段效应是交叉设计所特有的,由于试验分两个阶段(或时期)进行,受试者在不同阶段接受不同的处理,若前后两个阶段的效应值估计存在系统性差异,则可能存在阶段效应,如受试者的病情随时间发生了不可逆的进展、社会背景因素发生了重大的改变(如卫生保健策略或制度的改变)等<sup>[3]</sup>。阶段效应可通过比较第一阶段期初与第二阶段期初受试

者的基本特征是否有差异来衡量。若前后两个阶段期初受试者的基本特征存在显著差异,则说明存在阶段效应,即受试者的基本特征随时间或社会大背景发生了趋势性的改变。阶段效应会导致前后两个阶段的可比性下降,从而引起效应值估计的偏倚。在对交叉设计进行偏倚风险评估时,若两个顺序组(AB组和BA组)研究对象的分配比例为1:1,则研究结局任何随时间变化的一般趋势(阶段效应)均会相互抵消,不影响处理效应的估计,此时偏倚风险为低风险。若两个顺序组研究对象的分配比例不等,研究结局随时间变化的一般趋势(阶段效应)无法完全抵消,从而可能对处理效应的估计带来偏倚。值得注意的是,在两个顺序组研究对象分配比例不等(即存在阶段效应)的情况下,若在统计分析时将阶段效应纳入了模型进行分析,则可认为对效应值的估计影响较小,此时偏倚风险为低风险。

(2)偏离既定干预的偏倚:该领域主要针对滞后效应这一问题,对相应的信号问题进行了调整。在交叉设计中,前一个阶段的处理效应可能会影响到后一阶段的处理效应。这种前一阶段的处理在后续的阶段中仍然存在的效应,称之为滞后效应。若存在滞后效应,则研究对象在第二阶段相当于接受了“伴随处理措施”,从而对处理效应的估计带来偏倚,通常会低估效应值,倾向于无效假设,即容易出现假阴性的结果<sup>[4]</sup>。滞后效应可通过分别对比两个阶段期初的两组(治疗组 and 对照组)基线特征是否平衡而检出。若在第二阶段期初,治疗组和对照组的基线特征存在显著不平衡,则说明很可能存在滞后效应。因此,交叉设计试验中,为避免前一阶段用药或处理因素对下一阶段结果的影响,在进入下一阶段之前,需要等待一段时间,直到药物在体内完全代谢,受试者完全恢复到试验开始前的生理和心理状态,以洗脱前一阶段处理对后续阶段的影响。这段介于两个阶段之间的间歇期即称为“洗脱期”。不同试验阶段之间需要设计足够长的洗脱期,以消除滞后效应。洗脱期的长短一般视药物在体内的半衰期

表1 交叉设计 RoB2.0 信号问题设置(与平行设计 RoB2.0 对比)

领域	信号问题调整(交叉设计 RoB2.0)
随机化过程中的偏倚	增加了两个信号问题,以评估阶段效应
偏离既定干预的偏倚	在评估干预分配的效果时,调整为对洗脱期时长的评价,以评估滞后效应;在评估干预依从的效果时,增加对洗脱期时长的评价,以评估滞后效应
结局测量的偏倚	与平行设计 RoB2.0 相同
结局数据缺失的偏倚	增加了“对交叉设计处理缺失数据方法是否恰当”的信号问题
结果选择性报告的偏倚	增加了“是否仅报告了第一阶段结果”的信号问题
整体偏倚	与平行设计 RoB2.0 相同



而定(通常至少为 5~6 个半衰期),并需结合生物作用的特点来确定。需要注意的是,2×2 交叉设计(AB/BA 设计)由于自由度为 3,如果其中 2 个自由度用于估计处理效应和阶段效应,则剩下的 1 个自由度就只能与滞后效应和处理×阶段的交互作用中的一个有关。因此,2×2 交叉设计是无法通过统计学方法来区分滞后效应与和处理×阶段的交互作用,只能在两个阶段之间设计一个足够长的洗脱期,以假设滞后效应不存在。与平行设计 RoB2.0 类似,交叉设计 RoB2.0 也按照系统综述作者(注意,这里指的是系统综述的作者,而非原始研究的作者)的研究目的分为两类,在评估时需要根据不同的目的选择性回答不同的问题,分别为评估干预分配效果和评估干预依从效果。由于交叉设计试验的周期较平行设计试验要长,因此在评价干预依从效果时,需要注意交叉设计中可能会存在不同处理阶段受试者依从性不同的情况,如受试者往往在第二阶段的依从性有所下降。此时,系统综述作者需要注意研究者是否采用了一些分析手段(如:逆概率加权或工具变量)对实际接受的干预与预计实施的干预的差异进行了调整。

(3)结局测量的偏倚:与平行设计 RoB2.0 相同,具体内容见前期系列讲座<sup>[5-6]</sup>。

(4)结局数据缺失的偏倚:该领域与平行设计 RoB2.0 的信号问题类似,评价结局数据缺失是否会导致效应值估计的偏倚。需要注意的是,若前两个信号问题的回答显示研究中有一定比例的失访,且两组结局数据缺失的比例和缺失原因存在系统性差异,则需进一步回答第三个信号问题,即“是否有证据支持,即便存在缺失数据,分析结果仍然是稳健的”。由于交叉设计本身的特殊性,交叉设计中可能存在滞后效应,因而平行设计中常用的缺失数据处理方法“末次结转法”在交叉设计中并不适用,尤其在每个阶段仅进行一次观察时,如果第二阶段受试者的数据缺失,用第一阶段的观察值来填充显然是不合理的。而如果剔除了有缺失值的受试者,则损失了样本量,降低了研究的把握度<sup>[7-8]</sup>。因此,对于结果是否稳健,需要根据原始研究处理缺失数据的方法及敏感性分析的结果等综合进行判断。此外,交叉设计中,将受试者效应作为固定效应还是随机效应进行分析,目前仍存在争议。若作为固定效应,则任何一阶段存在缺失数据的受试者均被剔除,从而损失了样本量,降低了检验效能;若作为随机效应,则只在某一阶段数据缺失的受试者亦可纳入分

析,样本量损失较小,从而提高了研究的精确度。已有研究证明<sup>[7]</sup>,在交叉设计中只要缺失值不是太多,且为随机缺失时,将受试者效应作为随机效应进行分析,所得到的估计就是稳健的。

(5)结果选择性报告的偏倚:该领域与平行设计 RoB2.0 的信号问题类似,除多重结局测量及多重数据分析外,还针对交叉设计自身的特点,增设了一个信号问题,即“是否基于滞后效应的统计学检验结果选择性仅报告第一阶段的结果”。若研究者发现试验中滞后效应有统计学意义,则很可能选择性只报告第一阶段的结果,放弃了第二阶段的数据。此时,交叉设计基于个体误差来比较处理因素效应的优势未被充分利用,且样本量变小,从而降低了研究的精确度。此外,在多重数据分析时,研究者很可能基于自己偏好的结果,选择性报告配对分析或非配对分析的结果(第一阶段结果)。交叉设计的统计分析应体现自身配对的优势,使用配对分析的方法。非配对分析通常会降低研究的检验效能,使效应值倾向于无效假设。若研究者希望得到两种干预措施等效或效应值无差别的结论,则很可能选择性报告非配对分析的结果,当然通常这种情况不太多见。选择性报告常见的原因是,选择性报告的那些结果在组间差异具有统计学意义,或较为新颖,或可以使文章更容易发表。这种选择性的报告结局很可能给决策带来误导。对结果选择性报告的判断,应依据研究计划书、统计分析计划书、试验注册平台等信息来源进行判断。“事前”研究计划的发表必须先于试验的开始,在试验已经开始实施后进行的补救注册或计划书发表等不能排除选择性报告存在的可能。

2. 评估标准及整体偏倚评估:在完成信号问题的回答与评估后,评估者需要根据回答的情况对相应领域的偏倚风险按照相应标准细则作出“整体偏倚风险低”、“整体偏倚风险高”或“可能存在偏倚风险”的评估。最后,根据所有单个领域的评估结果对“整体偏倚”这一综合领域作出评价。本文不再列举评估标准细则,感兴趣的研究者可参阅交叉设计 RoB2.0 官方指南文件了解<sup>[9]</sup>。关于整体偏倚的评估标准,与平行设计 RoB2.0 类似,若 5 个领域评估结果均为低风险,则整体偏倚风险低;若其中任意一个领域的评估结果为高风险,或多项领域内的评估结果是可能存在风险,则整体偏倚风险高;除外上述两种情况的交叉设计试验则可能存在偏倚风险。需要强调的是,RoB2.0 整体偏倚这一领域是对单个研究单个结局水平的评价,即对同一个交叉设计试验,不同

的结局可能会有不同的整体评估结果。

3. 评估相关基本信息的添加:与平行设计RoB2.0类似,在对交叉设计试验使用该工具进行偏倚风险评估时,评估者需先填写一些与工具使用的相关信息,包括待评估的结局、评估所需的信息来源及评估者评估的目的(干预分配的效果或干预依从的效果)。在此基础上,再选择相应的信号问题进行偏倚风险评估。

### 三、实例分析

以2016年8月18日发表在新英格兰医学杂志上的交叉设计随机对照试验研究举例说明交叉设计RoB2.0的实际应用。该研究者希望评估妊娠期I型糖尿病患者中,与传感器增强型胰岛素泵(SAP)比较,自动胰岛素泵的疗效和安全性<sup>[10]</sup>。该研究除最终正式发表的试验结果文章外,还附有研究方案和补充结果作为附件,并在Current Controlled Trials网站进行了注册(ISRCTN71510001),这些资料作为支持信息来源在偏倚评估过程中被使用。为了覆盖交叉设计RoB2.0的全部内容并提供更多实例信息供使用者参考,本实例分析中既按照评估分配效果进行了评价,同时也按照评估依从效果进行了评价。在工具实际使用中,综述作者只需要根据自

己的研究目的选择其中一种目的进行偏倚评估即可。篇幅所限,对评估前准备、研究基本信息以及偏倚评估具体内容,本文未按照交叉设计RoB2.0中提供的表格进行完整展示。由于交叉设计RoB2.0与平行设计RoB2.0基本类似,主要针对其中4个领域的7个条目进行了调整,因此该实例重点对此7个条目进行偏倚评估和解读(表2)。

## 讨 论

总的来说,交叉设计RoB2.0在平行设计RoB2.0的基础上,针对交叉设计自身的特点,进一步对各个领域偏倚评估的信号问题进行了调整。该工具是目前国内外唯一针对交叉设计的随机对照试验制定的偏倚评估工具。该工具主要围绕交叉设计中容易出现偏倚风险的3个方面进行了信号问题的设置,包括:①是否存在滞后效应;②是否存在阶段效应;③是否仅报告第一阶段的结果。

需要注意的是,对交叉设计试验的偏倚风险评估在很大程度上依赖于交叉试验报告的规范程度。遗憾的是,交叉试验的报告方式各种各样,部分情况下很难或无法从交叉试验中获得相应信号问题的答案。其次,交叉设计融合了平行设计和配对设计的

表2 RoB2.0评估实例:偏倚评估

领域	信号问题	回答	支持信息/理由
随机化过程中的偏倚	1.4 两个顺序组(AB/BA)研究对象的分配比例是否相等	不是	见补充结果11页Figure S2 随机化后两个顺序组的研究对象人数分别为7人和9人,可认为两组研究对象的分配比例不等
	1.5 如在1.4问题中回答不是/可能不是/不清楚则回答:阶段效应是否纳入统计分析	是	发表文献统计分析部分 对夜间血糖正常范围时间比例及持续血糖监测的数据分析,研究者均将阶段效应纳入了统计分析
综述作者的目的是评估干预依从的效果			
偏离既定干预的偏倚	2.5 研究对象是否依从了分配的干预措施	是	该研究欲评估的干预措施为自动胰岛素泵和SAP泵,除第一处理阶段中1名研究对象因胎儿为13体综合征终止妊娠退出外,其他所有16名研究对象均完成了两个处理阶段并纳入了分析。研究对象在两个处理阶段的依从性均较高
	2.6 如在2.3,2.4或2.5问题中回答不是/可能不是/不清楚则回答:研究是否采用了恰当的方法估计干预的依从效果	不适用	
	2.7 在第二阶段研究结局评估之前,是否有足够长的洗脱期以消除滞后效应的影响	可能是	发表文献方法部分 第一处理阶段结束后,所有研究对象均经过2周的洗脱期,才进入第二处理阶段。据此,可推测在第二处理阶段研究结局评估之前,有足够长的洗脱期以洗脱第一处理阶段治疗的影响,消除滞后效应
缺失结局数据的偏倚	3.3 如在3.2问题中回答不是/可能不是/不清楚则回答:是否有证据支持,即便存在缺失数据,分析结果仍然是稳健的	不适用	
结果选择性报告的偏倚	5.3 基于滞后效应的统计学检验结果选择性报告结局	不是	发表文献方法和结果部分及研究计划书中的统计分析部分 由于该研究设计了为时2周的洗脱期,据此可推测该研究中不存在滞后效应,且在统计分析计划及结果部分均说明了该研究结果是基于两个处理阶段的分析

注:支持信息/理由的内容均为研究发表文献、计划书、注册网站等来源的原文摘录

优点,是一种能够基于受试者个体内水平上进行效应值估计的设计,因此在对交叉设计进行统计分析时应体现自身配对设计的优势,使用一些配对分析的方法。某些情况下,尽管试验研究者采用了配对分析方法,但结果未报告或报告不充分,从而导致系统评价者无法提取配对数据。虽然系统评价者可以获取一些非配对数据,但其统计学效能大大降低,在Meta分析中通常会导致其权重比其应有的权重低的多。因此,尽管这算不上偏倚的来源,未体现在交叉设计RoB2.0工具中,但也是我们进行证据整合时不可忽视的一个问题。此外,该工具许多条目涉及大量方法学专业知识和专业词汇,非方法学专业的综述作者使用时可能存在困难,可能需要开发团队后续制定手册进一步明确一些方法学词汇定义及使用方法,并举例说明其含义,以使交叉设计RoB2.0能够更好的在证据整合及循证医学工作中体现其意义和价值。

利益冲突 无

#### 参 考 文 献

- [1] Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] [Z]. The Cochrane Collaboration, 2011.
- [2] Higgins JP, Savović J, Page MJ, et al. A revised tool for assessing risk of bias in randomized trials//Chandler J, McKenzie J, Boutron I, et al. Cochrane Methods[M]. Cochrane Database of Systematic Reviews, 2016, 10(S1):29-31.
- [3] Senn S. Cross-over trials in clinical research [M]. 2<sup>nd</sup> ed. Chichester: John Wiley & Sons, 2003.
- [4] Cummings P. Carryover bias in crossover trials[J]. Arch Pediatr Adolesc Med, 2010, 164(8): 703-705. DOI: 10.1001/archpediatrics.2010.126.
- [5] 杨智荣,孙凤,詹思延. 偏倚风险评估系列:(一)概述[J]. 中华流行病学杂志, 2017, 38(7): 983-987. DOI: 10.3760/cma.j.issn.0254-6450.2017.07.027.
- [6] 杨智荣,孙凤,詹思延. 偏倚风险评估系列:(二)平行设计随机对照试验偏倚评估工具2.0介绍[J]. 中华流行病学杂志, 2017, 38(9): 1285-1291. DOI: 10.3760/cma.j.issn.0254-6450.2017.09.028.
- [7] Yang ZR, Sun F, Zhan SY. Risk on bias assessment: (2) Revised Cochrane risk of bias tool for individually randomized, parallel group trials (RoB2.0) [J]. Chin J Epidemiol, 2017, 38(9): 1285-1291. DOI: 10.3760/cma.j.issn.0254-6450.2017.09.028.
- [7] 赵杨,于浩,娄冬华,等. 交叉设计多中心临床试验的混合效应模型[J]. 中国临床药理学与治疗学, 2005, 10(1): 116-120. DOI: 10.3969/j.issn.1009-2501.2005.01.026.
- [8] Zhao Y, Yu H, Lou DH, et al. Generalized linear mixed models in multi-center crossover clinical trails [J]. Chin J Clin Pharmacol Ther, 2005, 10(1): 116-120. DOI: 10.3969/j.issn.1009-2501.2005.01.026.
- [8] 陈家慧,富春枫,赵杨,等. 多水平Poisson回归模型在交叉设计临床试验中的应用[J]. 中国临床药理学与治疗学, 2007, 12(11): 1306-1310. DOI: 10.3969/j.issn.1009-2501.2007.11.023.
- [9] Chen JH, Fu CF, Zhao Y, et al. Multilevel Poisson regression models in crossover clinical trails [J]. Chin J Clin Pharmacol Ther, 2007, 12(11): 1306-1310. DOI: 10.3969/j.issn.1009-2501.2007.11.023.
- [9] 陈峰. 非独立数据的统计分析方法//方积乾,陆盈. 现代医学统计学[M]. 北京:人民卫生出版社,2002:45.
- [10] Chen F. Statistical analysis of non-independent data //Fang JQ, Lu Y. Modern medical statistics [M]. Beijing: People's Health Publishing House, 2002:45.
- [10] Stewart ZA, Wilinska ME, Hartnell S, et al. Closed-loop insulin delivery during pregnancy in women with type 1 diabetes [J]. N Engl J Med, 2016, 375(7): 644-654. DOI: 10.1056/NEJMoa1602494.

(收稿日期:2017-04-01)

(本文编辑:王岚)