

偏倚风险评估系列:(五)非随机干预性研究

孙凤 高乐 杨智荣 詹思延

100191 北京大学公共卫生学院流行病与卫生统计学系(孙凤、高乐、詹思延); 100191 北京大学循证医学中心(孙凤、詹思延); CB1 8RN 英国剑桥大学临床医学院初级医疗中心(杨智荣)

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

DOI: 10.3760/cma.j.issn.0254-6450.2018.03.024

【摘要】 对非随机干预性研究(NRSI)偏倚评估工具ROBINS-I(Risk Of Bias In Non-randomised Studies-of Interventions)的主要内容进行详细介绍,并举例说明ROBINS-I的使用方法和注意事项。ROBINS-I针对NRSI的特点,设置了相应的评估领域和信号问题,为NRSI纳入系统综述进行证据整合提供偏倚风险信息。ROBINS-I为观察性研究和类实验领域新开发的评估工具,现已得到了一定程度的应用,但仍在进一步完善中,使用者可持续关注其后续更新与进展。

【关键词】 偏倚风险; 评估工具; 非随机干预性研究; 观察性临床研究; 系统综述

基金项目:国家自然科学基金(71673003)

Series of risk of bias assessment (5): Risk Of Bias In Non-randomized Studies-of Interventions (ROBINS-I)

Sun Feng, Gao Le, Yang Zhirong, Zhan Siyan
Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China (Sun F, Gao L, Zhan SY); Center of Evidence-based Medicine and Clinical Research, Peking University, Beijing 100191, China (Sun F, Zhan SY); Primary Care Unit, Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridgeshire CB1 8RN, UK (Yang ZR)

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

【Abstract】 This paper summaries the Risk Of Bias In Non-randomized Studies-of Interventions (ROBINS-I), a tool for evaluating risk of bias about Non-randomized Studies-of Interventions (NRSI), and introduces the application of ROBINS-I in a published NRSI. According to the characteristics of NRSI, evaluation field and signaling question were designed in ROBINS-I to provide essential information about risk of bias for NRSI included in systematic reviews. ROBINS-I is the tool in assessment of risk of bias in observational studies and quasi-randomised studies. Although the tool has been used in practice to some extent, but it still needs further improvement. Attention should be paid to its update and progress.

【Key words】 Risk of bias; Tool for assessment; Non-randomised Studies-of Interventions; Observational clinical study; Systematic review

Fund program: National Natural Science Foundation of China (71673003)

随机对照试验(randomized controlled trials, RCT)是评价干预效果的金标准。但由于伦理、可行性等原因,RCT有时难以开展,此时非随机干预性研究(Non-randomised studies of the effects of interventions, NRSI),可以作为RCT的有效补充,而且NRSI在某些方面有其独特优势,人群特征更接近于真实世界,尤其适合研究长期的结局指标、不良反应等,近些年在医药卫生领域的应用日益广泛^[1]。NRSI包括观察性研究以及类实验,前者主要有队列研究和病例对照研究等。在这些研究设计中,由于

干预措施不是随机分配的,研究结果更容易受到各种潜在偏倚的影响。因此,利用评估工具对NRSI的偏倚风险进行评价显得尤为重要,它可帮助使用者科学谨慎地筛选出高质量NRSI研究,再进一步推广应用或者进行证据整合,从而有效促进循证决策。

一、制定背景

目前,观察性研究的质量评价工具应用最广泛的是2008年Newcastle-Ottawa Scale(NOS)量表^[2]和1998年Downs-Black清单^[3],其中前者多是针对经典的队列研究和病例对照研究设计,但是对这两类

研究的多数衍生设计并不适用;后者可同时评价RCT与NRSI,但其涵盖方法学质量和报告质量共计五个方面27个问题,使用过于繁琐。两个工具的评价内容既包含了内部真实性也涉及了外部真实性。在ROBINS-I (Risk of Bias In Non-randomised Studies-of Interventions)之前,也有不少研究人员采用Cochrane协作网研发的偏倚风险工具RoB1.0^[4]从随机序列产生、分配隐藏、盲法、结果数据不完整、选择性报告等方面评价NRSI研究的偏倚风险,但由于RoB1.0是针对RCT开发的,并不能完全适用于NRSI。因此,学者们一直在致力于制定一个更加令人满意的工具,以更好地完成NRSI的质量评价。

2016年10月Sterne等^[5]在BMJ上发表了关于NRSI的更普适的偏倚风险评估工具ROBINS-I,适用干预效果评价的多种非随机研究类型,主要包括:队列研究、病例对照研究、类实验等,是一个领域评估式的工具。该工具的发布对于NRSI领域的质量评价具有里程碑式的意义。ROBINS-I工具由“Cochrane Bias Methods Group”和“Cochrane Non-Randomised Studies Methods Group”两个学组成员在Cochrane协作网研发的偏倚风险工具RoB1.0^[4]基础上共同开发,从2011年的立题会议到最终发布历时5年,经历了数次专家论证,期间曾在2014年9月推出了1.0.0网络版本^[6](即ACROBAT-NRSI, A Cochrane Risk of Bias Assessment Tool: for Non-Randomized Studies of Interventions),随后根据专家和用户的反馈建议完成了进一步的修订。2016年ROBINS-I制作组将所有相关资料发布在其官方网站<http://www.riskofbias.info>,包括PDF格式的模板、Docx格式的模板、Access表单形式的评价模板及ROBINS-I详细使用指南。

二、工具解读

ROBINS-I工具的解读归纳为4个方面:

1. 评价领域的设置:ROBINS-I工具共包括7个评价领域,在领域设置上与平行设计RCT的RoB2.0工具基本相似,但ROBINS-I进一步把评价领域分别归入到了干预前、干预中和干预后3部分,即:(1)干预前(①混杂偏倚、②研究对象选择的偏倚);(2)干预中(③干预分类的偏倚);(3)干预后(④偏离既定干预的偏倚、⑤缺失数据的偏倚、⑥结局测量的偏倚、⑦结果选择性报告的偏倚)。

每个评价领域都由多个信号问题(Signaling question)组成,共计34个信号问题。具体与RoB2.0工具内容对比:①使用者先对各领域的信号问题作

出回答,继而汇总形成各领域偏倚风险的判断,最后根据7个领域的评价结果对单个NRSI的特定结局作出整体的偏倚风险评价;②从内容上来看,ROBINS-I工具关于干预后的4个评价领域与平行设计RCT的RoB2.0工具的评价领域多有重合,因此理解这后4个评价领域时,也可以参考该工具;③此外,在ROBINS-I中无“其他偏倚”这一领域,即不再考虑增删偏倚来源的相关领域,但可以适当调整每个领域里的信号问题。见表1。

表1 ROBINS-I和RoB2.0在领域设置上的对比

划分方式	评价领域	ROBINS-I	RoB2.0
干预前	1	混杂偏倚	随机化过程中的偏倚
	2	研究对象选择的偏倚	
干预中	3	干预分类的偏倚	(无)
干预后	4	偏离既定干预的偏倚	偏离既定干预的偏倚
	5	缺失数据的偏倚	结局数据缺失的偏倚
	6	结局测量的偏倚	结局测量的偏倚
	7	结果选择性报告的偏倚	结果选择性报告的偏倚
	整体	整体偏倚	整体偏倚

2. 评价流程:ROBINS-I工具操作流程主要包括3个阶段:①第I阶段:定义研究问题;②第II阶段:首先构造出一个假想的实用性随机化“目标”试验(“target” pragmatic randomized trial),将所研究的NRSI看作这个“目标”试验的模拟和近似;在评定每个评价领域时,通过信号问题帮助判定者进行判断;③第III阶段:每个领域的偏倚风险等级分为“低(low)、中(moderate)、高(serious)、极高(critical)和未获得评估信息(NI,即不清楚)”5个等级。见图1。

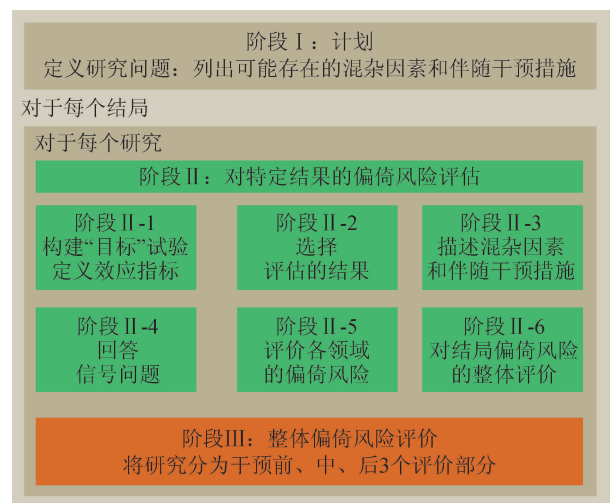


图1 ROBINS-I工具操作流程

基于NRSI研究制定系统综述时,若推荐用ROBINS-I工具对原始研究进行偏倚评价,则每个

NRSI研究均需要完成第II阶段的6个步骤(图1),其中步骤3~6是针对具体结局的,即对不同的结局需要分别完成步骤3~6的评价。步骤1:通过构建随机化“目标”试验,定义效应指标。步骤2:定义待评价的结局和结果。步骤3:对每个研究结局,分别描述混杂因素和伴随干预。步骤4:回答7个偏倚领域的信号问题。步骤5:对各信号问题的偏倚风险给出评判,继而形成各个领域的偏倚风险的判断。步骤6:对单个NRSI的特定结局作出整体的偏倚风险评价。

3. 各领域的评估标准及其信号问题的设置:ROBINS-I明确地为每个领域设置了数个信号问题,要求评估者先对信号问题作出“NA/Y/PY/PN/N/NI”的判断,分别表示“Not applicable(不适用)”、“Yes(是)”、“Probably yes(可能是)”、“Probably no(可能不是)”、“No(不是)”或“No information(未获得评估信息,即不清楚)”的回答。通过对信号问题的回答,作者可了解偏倚来源的信息,即相关“信号”,然后依据这些“信号”进一步对各领域的偏倚风险进行评估。需要强调的是,ROBINS-I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NRSI研究,不同的结局可能会有不同的整体评估结果。

(1)混杂偏倚:相关术语:临床试验研究中的选择偏倚(Selection bias as it is sometimes used in relation to clinical trials);分配偏倚(Allocation bias);病例混合偏倚(Case-mix bias);引导偏倚(Channelling bias)。

在NRSI中,不同干预组间研究对象的特征很可能有较大的差别。当一个或多个重要的预后因素可能影响个体是否接受某干预措施时,干预效应的估计就会产生偏倚。因此对尚未控制的混杂引起的偏倚进行风险评价是ROBINS-I的重要部分。

基线混杂可能是大多数NRSI普遍存在的问题,当某个或多个预后变量影响到患者试验开始时接受何种干预措施的选择时就会发生。例如一个比较罗格列酮和吡格列酮药物疗效的非随机试验,应当考虑药物治疗前的基线血糖,因为该指标是与心血管病结局发生的很强的预后因素,且可能会影响药物选择。控制可测量的混杂的适用方法^[6]包括:分层、回归、匹配、标化、G谱估计(g-estimation)和逆概率加权(inverse probability weighting)。这些方法可以对个体变量或倾向评分值加以控制。

ROBINS-I也适用于评价时依混杂造成的偏

倚。当NRSI研究者根据个体随访时间对不同干预进行划分时,即评价干预的依从效应时,时依混杂的控制尤为重要。若所接受的干预措施会改变某种预后因素,而变化后的这种预后因素又会反过来影响干预措施的选择,这种随时间改变的预后因素则会产生时依混杂。例如,降糖药物治疗开始后测量的血糖水平(一个基线调查后的预后因素)的升降可能会引起药物种类的选择。当基线调查后的预后因素本身就受到干预措施的影响时(如降糖治疗会影响基线后血糖水平),在统计分析中用常规的调整方法来控制时依混杂显然是不恰当的。但应注意的是,当研究对象的干预措施发生改变时,应当进行比较的是试验中所依从于干预的效应而非初始分配干预的效应。

(2)研究对象选择的偏倚:相关术语:观察性研究中的选择偏倚(Selection bias as usually used in relation to observational studies and sometimes used in relation to clinical trials);初始偏倚(Inception bias);领先时间偏倚(Lead-time bias);恒定时间偏倚(Immortal time bias)。

当排除的部分合格研究对象、初始随访时间或部分结局事件与干预措施和结局有关时,即使各干预措施的效应实际并无差别,此时所得到的结果仍会提示干预措施与结局之间存在联系。这就称为选择偏倚,不同于混杂。例如,关于补充叶酸防止神经管畸形的研究就存在偏倚,因为研究只局限于活产儿,由于死产和治疗性流产(从样本中排除)同时与干预和结局有关。另一个例子是NRSI证实的新型口服避孕药激素会显著增加静脉血栓的风险,与服用旧剂型的受试者相比,服用新剂型的受试者开始治疗的时间更近,而在治疗过程中,早期发生静脉血栓的风险最大。需要强调的是,在非随机研究中应纳入某医疗措施的新近应用者,随访应从决定使用治疗措施时开始,而不应该从治疗措施实施时开始,这是为了避免恒定时间偏倚。

我们所指的选择偏倚仅针对研究内部真实性,不涉及外部真实性的问题(外推性、适用性或可传递性等)。例如,在一项仅纳入无合并症患者的研究中,其结果不一定能外推到有合并症的人群中。但是这样并不会影响对在无合并症人群中的干预效应的估计。

(3)干预分类的偏倚:相关术语:错分偏倚(Misclassification bias);信息偏倚(Information bias);回忆偏倚(Recall bias);测量偏倚

(Measurement bias); 观察者偏倚 (Observer bias)。

在RCT中干预措施的错分问题较小,但在NRSI中很容易出现。例如降糖药处方记录缺失不能代表未服用降糖药。无差异错分与结局无关,通常会使效应估计值偏向无效假设。当干预状态的错分与结局或结局风险有关时,就会发生差异性错分。因此,干预的分组应该在不知晓结局的情况下进行,这一点尤其重要。典型的例子是病例对照研究中的回忆偏倚,已知病例或对照的状态可产生对暴露情况的回忆偏倚。如果结局对干预状态的信息有影响时,错分偏倚同样也会出现在队列研究中。例如,一个老年人的队列,研究结局为痴呆,一些观察对象在最初可能有中等程度的认知损害,因此在研究开始时对先前暴露情况的回忆可能会受到影响。

(4) 偏离既定干预的偏倚: 相关术语: 实施偏倚 (Performance bias); 时依混杂 (Time-varying confounding)。

偏离既定干预的偏倚 (有时称为实施偏倚), 除了试验组和对照组所研究的干预措施以外, 若其他伴随干预在组间存在系统性差异, 此时会产生偏离既定干预的偏倚。例如, 当研究对象知道他们所接受的干预措施, 或研究者对试验组和对照组间的差异有所假设时, 偏离既定干预的偏倚就有可能发生。此外, 出现以下情况时, 也可能产生偏离既定干预的偏倚: 各干预措施并不一定能成功实施 (例如若实验室检查出现错误, 那么就无法按既定方式给药), 或研究对象不依从干预措施, 或是重要的共同干预措施在各组间不均衡等。

依从性差或共同干预是否会导致偏倚取决于所研究的效应。对于干预分配效应来说, 这两者不会引入偏倚。相反, 若研究关注的是干预措施的依从效应 (如调查药物不良反应), 就会产生偏倚。例如, 一个比较食管癌低侵袭性和开放性术后呼吸道感染率的研究, 两组间在共同干预措施上存在巨大差异。首先, 单肺机械性通气应用于开放性手术组, 而低侵袭性治疗组采用的是双肺通气。其次, 硬膜外麻醉在开放性手术组运用的更多: 通常用硬膜外麻醉的患者的移动性较差, 因此也更容易有呼吸道感染。

需要注意的是, 一些常规的临床护理也会造成干预措施差异 (例如由于药物急性毒性而停止用药), 可以被看作是既定干预的一部分。

(5) 缺失数据偏倚: 相关术语: 失访偏倚 (Attrition bias); 观察性研究中的选择偏倚 (Selection bias as it is sometimes used in relation to observational

studies)。

数据缺失的原因包括数据损失 (失访)、错过预约、数据收集的不完整、进行分析时排除部分研究对象。在NRSI中, 所接受的干预措施、混杂或结局的测量过程都可能发生数据缺失。各组间数据缺失的程度和原因很重要。若缺失数据的比例较低, 且各组缺失的原因相近, 则偏倚的风险较小。反之, 偏倚风险会随之增大。

(6) 结局测量的偏倚: 相关术语: 检出偏倚 (Detection bias); 回忆偏倚 (Recall bias); 信息偏倚 (Information bias); 错分偏倚 (Misclassification bias); 观察者偏倚 (Observer bias); 测量偏倚 (Measurement bias)。

结局测量的偏倚是由于结局错分或测量误差导致的。若错分和测量误差与所接受的干预无关时, 则为无差异错分。与干预或结局无关的随机误差 (无差异测量误差) 不太可能产生偏倚。有差异的测量误差 (与干预状态有关) 会使得效应估计值存在偏倚。当结局评估者知晓干预措施时, 若对各组采用不同的结局测量方法, 或测量误差与干预状态和效应有关, 就会产生检出偏倚。该问题在主观性结局的测量中更为明显。对结局测量采用盲法可降低各组间的系统性差异。但在实际操作中, 相比起RCT, 盲法在非随机研究中通常很难实施。

(7) 结果选择性报告的偏倚: 相关术语: 结果报告偏倚 (Outcome reporting bias); 分析报告偏倚 (Analysis reporting bias)。

若基于结局效应的方向、大小和统计学显著性进行报告, 则容易产生选择性报告偏倚。该偏倚有3种常见类型: ①从多个测量结果中选择某种测量进行报告, 例如从多个时间点测量数据选择某个时间点的测量值, 或从多个疼痛量表中选择其中一个量表的得分; ②从多种方法估计的效应值中选择一个进行报告, 例如, 某研究同时计算了变化值以及调整基线后的终点值, 或用多种分析方法调整各类混杂, 研究者只选择了其中一个结果进行报告; ③选择某个亚组人群的结果进行报告, 这种情况在大型队列数据中较为常见。

4. 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事先制定的标准给出“低、中、高、极高或不清楚” (low, moderate, serious, critical, or no information) 的风险评估。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。

本文的实例部分将详细说明每个信号问题和领域如何评价。关于评估标准细则,感兴趣的研究者可进一步参阅ROBINS-I官方指南文件了解^[7]。

整体偏倚评估原则为:所有7个评价领域为低偏倚风险则整体偏倚风险为“低”(Low risk of bias),所有7个评价领域为风险低或中则总体偏倚风险为“中”(Moderate risk of bias),至少一个评价领域风险高但无任何评价领域为极高风险则整体偏倚风险为“高”(High risk of bias),若至少一个评价领域风险极高则整体偏倚风险为“极高”(Critical risk of bias),若缺乏关键评价领域的相关信息,则整体偏倚风险为“未获得评估信息(不清楚)”(No information, NI)。

三、实例分析

以下介绍1篇已经发表的研究举例说明ROBINS-I工具早期1.0.0版本使用(即ACROBAT-NRSI^[7],该工具的评估领域与ROBINS-I的相同)评

价NRSI并与既往偏倚评价方法进行了对比的方法学文献^[8],该文作者根据临床重要话题,选择了两个既往发表的权威的系统综述作为案例^[9-10],对其中纳入的NRSI原始研究进行了偏倚评价。本讲座选择该文^[7]中的一个有关降糖药的案例^[10]作为示例说明ROBINS-I在系统综述的实际使用。原综述作者纳入了16项NRSI研究,目的是评估与口服比格列酮相比,口服罗格列酮的2型糖尿病患者是否可以降低心血管病结局的风险^[10]。篇幅所限,本文并未按照图1的评估流程展示ROBINS-I中所提供的全套相关表格,此处仅对其中一项队列研究^[11]的偏倚评价结果进行示例(表2)。对于ROBINS-I的详细评估流程及该案例的相关表格,读者可登录网站(<http://cpse.bjmu.edu.cn/>)获取,包括:该案例的评估前准备与研究基本信息表、完整的ROBINS-I偏倚评估解读表及该案例最终纳入的16项NRSI研究的整体偏倚风险评价结果汇总表。

表2 ROBINS-I评估实例

领域	信号问题	回答	支持信息/理由
混杂偏倚	1.1 研究中是否可能存在影响干预效果的混杂因素? 若问题1.1回答为否/可能不是:可视研究由于混杂因素导致偏倚的风险低,因此不需要额外的指示性问题进行评估。 若问题1.1回答为是/可能是:需要进一步评价是否存在时依混杂。	可能是	缺失一些重要的可能的混杂信息:糖尿病病程、吸烟、BMI。
	1.2 是否根据所接受干预的不同对研究对象的随访时间进行了分段,并且以此时间分段进行分析? 若回答为否/可能不是:回答与基线混杂有关的问题(问题1.4~1.6)。 若回答为是/可能是:则回答问题1.3。	是	该研究共有14个时间段,每个时间段是6个月,对于每位研究者,均以6个月内始终用单种药纳入分析,因此在单独的每个6个月的观察期内,干预未发生变更,但6个月之后允许变更干预。分析时以此时间分段进行分析。
	1.3 干预中断或变更是否可能与结局的预后因素相关? 若回答为否/可能不是:继续回答与基线混杂有关的问题(问题1.4~1.6)。 若回答为是/可能是:则继续回答与基线混杂和时依协变量引起的混杂均相关的问题(问题1.7~1.8)。	是	患者有涉及治疗方案的变更。作者定义了14个周期,每个周期含6个月,从每个周期中定义所用药物,6个月之后允许变更干预,变更可能与结局的预后因素相关(比如,血糖控制情况和自身临床特征),后面研究结果也显示有较大比例患者进行了药物变更。但是,作者没有说明什么具体原因引起变更,只是用泊松回归来控制干预变更带来的影响。
	关于基线混杂的信号问题 1.4 作者是否采用了恰当的分析方法控制所有重要的混杂?		
	1.5 若问题1.4回答为是/可能是:研究中的变量是否可以真实可靠地测量了所需要控制的混杂?	不适用	
	1.6 对于受干预影响的因素,作者是否予以控制基线/时依协变量引起的混杂	不适用 不适用	
	1.7 作者是否采用了恰当的分析方法以控制所有重要的混杂因素以及时依协变量引起的混杂?	可能是	作者收集的测量指标都是一些客观指标(比如年龄,性别,入组前心血管疾病状况,研究期间降压药、降脂药、门诊病人胰岛素的使用情况)。并且运用了广义线性模型,并用年龄调整的Charlson评分对潜在并发症进行了调整控制。
	1.8 若问题1.7回答为是/可能是:研究中的变量是否可以真实可靠地测量了所需要控制的混杂?	可能是	作者对已收集到的诸多重要混杂因素进行了调整控制。但仍存在未被测量的混杂:糖尿病患病时间、吸烟、BMI,尚不确定这些因素对结局效应的影响。
偏倚风险评估		中	
可选答问题:	混杂偏倚的方向		

续表

领域	信号问题	回答	支持信息/理由
研究对象选择的偏倚	2.1 研究对象的选择(纳入研究或分析)是否根据干预开始后观察到的个体特征进行? 若问题2.1回答为可能不是/否:继续回答问题2.4。	可能不是	研究是一个回顾性队列研究,收集了所有符合上述标准的对象的基线和结局资料。
	2.2 若问题2.1回答为是/可能是:在干预开始后,是否有因素既影响研究对象的纳入,同时也与干预有关?	不适用	
	2.3 若问题2.2回答为是/可能是:在干预开始后,是否有因素既影响研究对象的纳入,同时也受结局的发生或与结局的原因影响?	不适用	
	2.4 大多数受试者的随访与干预的起始时间是否一致?	可能是	研究对象为新使用降糖药物的患者。对每个对象都有随访,研究结局为因心梗住院、全死因死亡、观察期无符合标准的研究对象的时间间隔、观察终点。
	2.5 若问题2.2和2.3回答为是/可能是,或者若问题2.4回答为否/可能不是:所运用的校正技术是否能校正选择偏倚? 偏倚风险评估 可选答问题:研究对象选择偏倚的方向?	不适用 低	
干预分类的偏倚	3.1 是否对各个干预组进行了清晰地界定?	可能是	虽然将干预措施分为3类:磺脲类,二甲双胍类,噻唑烷二酮类(又分为罗格列酮和吡格列酮),并且将各受试者的随访期按每6个月为一个时间段,该时间段内仅纳入使用其中一类干预措施的受试者,同时有多个干预措施的被排除。但是干预信息来自处方而非配药。
	3.2 是否在干预起始时就对各组干预信息进行了记录?	是	采用受试者数据注册系统,包括人口统计学资料、接受医疗服务时间、药物、诊断、实验室检查结果以及排除记录。数据不全者还利用了门诊和住院记录。这些信息不易受到结局的影响。
	3.3 干预组的划分是否因知晓结局或结局相关风险而受到影响? 偏倚风险评估 可选答问题:干预分类偏倚的方向?	否 中	按每6个月为1个时间段进行观察,在该时间段内研究对象的干预措施仅有1个,未发生改变,因此不存在干预措施受结局影响的情况。
偏离既定干预的偏倚	如果您的研究目的是评价干预的分配效应,请继续回答问题4.1和4.2。		
	4.1 是否存在非常规的干预改变?	不适用	
	4.2 如在4.1问题中回答是/可能是,则回答:干预变化是否在组间不平衡,并可能对研究结局造成影响? 如果研究目的是评价干预的依从效应,请继续回答问题4.3~4.6。	不适用	
	4.3 重要的伴随干预措施组间是否均衡?	否	未排除使用胰岛素以及调脂治疗等伴随干预的患者,表1显示多种伴随干预组间有差异。后期统计分析进行了多变量调整和分层分析。
	4.4 计划实施的干预措施在大多数受试者中是否成功执行?	可能是	可能成功实施,因为是以处方为干预判定方式,并未以配药为基准进行判定,因此不能确定是否真正实施。
	4.5 研究对象是否依从了分配的干预措施?	可能是	根据至少1次的门诊随访(包括心理治疗或营养调查),或住院记录,或许可以确定研究对象的依从情况。
4.6 如在4.3,4.4或4.5问题中回答不是/可能不是则回答:研究是否采用了恰当的方法估计干预的依从效果? 偏倚风险评估	不适用 中		
可选:偏离既定干预的偏倚方向?			该研究中的患者最长观察期为7年,每6个月为1个纳入间隔,6个月内服用4种药物任何一种的即纳入合格对象。最终,磺脲类、二甲双胍、罗格列酮和吡格列酮的患者数分别为11 200、12 490、1 879和806人;全程7年内始终单一使用上述4种药物之一的患者数分别为7 152、8 798、1 028和418人,即人数分别减少了36.1%、29.6%、45.3%和48.1%。因此,认为可能存在偏离既定干预的偏倚,尤其罗格列酮和吡格列酮组,但是,作者用泊松回归来控制了干预的变更。

领域	信号问题	回答	支持信息/理由
缺失数据的偏倚			
	5.1 全部或绝大多数研究对象的研究结局数据是否可以获得?	是	利用住院和门诊记录补充不完全的数据。
	5.2 研究对象是否因干预状态信息的缺失而被排除?	否	研究对象均有至少1次的降糖药处方记录。
	5.3 研究对象是否因分析中其他变量的信息缺失而被排除?	否	研究对象的信息用不同的数据来源进行了补充。
	5.4 若问题5.1回答否/可能不是,或问题5.2/5.3回答为是/可能是:缺失数据的比例和缺失数据的原因在组间是否相似?	不适用	
	5.5 若问题5.1、5.2或5.3回答为否/可能不是:是否应用了恰当的统计学方法处理缺失数据?	不适用	
	偏倚风险评估	低	
	可选:数据缺失的偏倚方向?		
结局测量的偏倚			
	6.1 结局测量是否受到已分配的干预的影响?	否	研究结局为因心梗住院、全死因死亡、观察期无符合标准的研究对象的时间间隔、观察终点,较为客观。
	6.2 结局评估者是否知道研究对象接受的干预?	可能是	没有提及结局评估者的信息,但是结局是因心梗住院、全死因死亡等,结局都较为客观,受结局评估者影响的可能性不是很大。
	6.3 各组间结局评估方法是否具有可比性?	是	各组间结局评估方法一致:数据都来自注册系统;结局的测量方法都是因心梗住院、全死因死亡等;时间段为6个月。
	6.4 结局测量中的任何系统性误差是否与所接受的干预措施相关?	否	结局较为客观,发生错分偏倚的可能性不大。
	偏倚风险评估	低	
	可选:结局测量偏倚的方向?		
结果选择性报告的偏倚			
	所报告的效果估计值是否从以下结果中选择出来进行?		
	7.1 某个结局的多重测量	否	未提及某结局是否采用多种测量方式。但是结局为因心梗住院、全死因死亡等,不太可能存在多种测量方式。
	7.2 干预-结局相关性多重数据分析结果	是	分别估计了粗的和调整后的率比。
	7.3 不同亚组分析	是	进行了分层分析,各层结果相似。
	偏倚风险评估	低	
	可选:结果选择性报告偏倚的方向?		
总体偏倚			
	偏倚风险评估	中	
	可选:该结局的整体偏倚方向?		

四、讨论

整体上来说,ROBINS-I内容丰富、翔实,完善了NRSI研究各个偏倚评估重要组成部分的细节,强调了偏倚评估在证据整合与评价过程中的重要地位。相比于NOS等之前常用的观察性研究偏倚风险评估工具,ROBINS-I有一些明显的优势:①针对性较强,ROBINS-I所针对的观察性研究主要是以评价干预效果为目的的;②评估内容更全面,ROBINS-I所设置的评估领域和信号问题涵盖了这类观察性研究常见的偏倚来源;③评估过程程序化,ROBINS-I细化了具体领域的评估过程,将评估中易于产生困扰和分歧的项目,均使用信号问题进行明确指示,有助于提高评估结果的可重复性和一致

性;④ROBINS-I不仅提供了偏倚风险评估,还提供了可能的偏倚方向评估,后者将有助于综述读者判断干预的真实效果方向。

然而,ROBINS-I尚处于初步使用^[12-18]和不断完善阶段,目前仍存在一些问题有待开发团队后续改进。首先,ROBINS-I的可操作性相比NOS等量表明显下降。根据我们的测试,完成ROBINS-I的时间相比NOS、RoB1.0等都有显著增加,判断各领域的偏倚风险时流程复杂。好在这个问题在开发团队公布程序化工具后已有所改善,目前已有规范的指导手册展示评估步骤和结果解读^[7]。其次,与RoB2.0新工具一样,ROBINS-I许多条目涉及到大量方法学专业知识和专业词汇,非方法学专业的综

述作者可以参考开发团队已制定的使用说明手册^[7],详细了解一些方法学词汇定义及使用方法。再次,工具中一些细节可能需要进一步统一,某些条目中,回答“是”或“可能是”对应的是“低风险”,而另一些条目中同样的回答却是“高风险”,这很容易给评估者造成困扰并增加出错的机会。这些不一致完全可以通过转变信号问题的提问方式进行统一。此外,现在还没有能够方便进行ROBINS-I评估的软件工具,但ROBINS-I团队正在努力,目前已经开发了ACCESS表单填写模板,挂在其官方网站(<http://www.riskofbias.info>)可供下载使用。期待未来能像RoB1.0那样,将ROBINS-I整合至下个版本的Review Manager软件中,或者开发一个界面友好的专用ROBINS-I评估软件。相信在开发这些专门的软件以后,评估过程将更加快捷,并会有更好的结果展示方法。最后,ROBINS-I中一些条目的设置从方法学角度来说可能仍有待商榷。特别是对于两种不同综述目的(干预的分配效应和依从效应)下分别有不同的信号问题,如何提高这两类问题的区分度,可能需要开发团队进一步说明并讨论。

此外,对ROBINS-I的使用者来说,还应明确该工具与另一个简称为TREND的清单之间的区别。后者是2003年由美国CDC组织制定的关于非随机对照设计的报告规范^[19]。该报告规范强调,非随机对照试验的研究报告要详细报告研究的假设(理论基础)、干预措施和组间比较的条件,研究设计以及为调整可能的偏倚所采用的方法等。TREND针对的是报告质量,而ROBINS-I针对的是偏倚风险。

综上所述,我们建议综述作者持续关注ROBINS-I的更新,以及Cochrane手册和Review Manager的相应更改的内容。在有条件的情况下,可积极地与开发团队沟通,及时反馈在理解、使用、方法学等方面遇到的问题与建议,使得ROBINS-I能够更好的在证据整合、乃至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] Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence—what is it and what can it tell us[J]. *N Engl J Med*, 2016, 375(23):2293–2297. DOI:10.1056/NEJMs1609216.
- [2] Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in Meta-analyses [EB/OL]. [2017-07-12]. http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm.
- [3] Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions[J]. *J Epidemiol Comm Health*, 1998, 52(6):377–384.
- [4] Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane collaboration's tool for assessing risk of bias in randomised trials [J]. *BMJ*, 2011, 343:d5928.
- [5] Sterne JAC, 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.
- [6] Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology* [M]. 3rd ed. Philadelphia: Lippincott Williams & Wilkins Publisher, 2008:345–380.
- [7] Sterne JAC, Higgins JPT, Reeves BC on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0, 24 September 2014 [EB/OL]. (2015-11-28) [2017-06-20]. <http://www.riskofbias.info>.
- [8] Bilandzic A, Fitzpatrick T, Rosella L, et al. Risk of bias in systematic reviews of non-randomized studies of adverse cardiovascular effects of thiazolidinediones and Cyclooxygenase-2 inhibitors: application of a new Cochrane risk of bias tool [J]. *PLoS Med*, 2016, 13(4):e1001987. DOI:10.1371/journal.pmed.1001987.
- [9] McGettigan P, Henry D. Cardiovascular risk and inhibition of cyclooxygenase: a systematic review of the observational studies of selective and nonselective inhibitors of cyclooxygenase 2 [J]. *JAMA*, 2006, 296(13):1633–1644. DOI:10.1001/jama.296.13.jrv60011.
- [10] Loke YK, Kwok CS, Singh S. Comparative cardiovascular effects of thiazolidinediones: systematic review and Meta-analysis of observational studies [J]. *BMJ*, 2011, 342:d1309.
- [11] Brownstein JS, Murphy SN, Goldfine AB, et al. Rapid identification of myocardial infarction risk associated with diabetes medications using electronic medical records [J]. *Diabetes Care*, 2010, 33(3):526–531. DOI:10.2337/dc09-1506.
- [12] Farrow LS, Smith TO, Ashcroft GP, et al. A systematic review of tranexamic acid in hip fracture surgery [J]. *Br J Clin Pharmacol*, 2016, 82(6):1458–1470. DOI:10.1111/bcp.13079.
- [13] Al-Moghrabi D, Salazar FC, Pandis N, et al. Compliance with removable orthodontic appliances and adjuncts: A systematic review and Meta-analysis [J]. *Am J Orthod Dentofac Orthop*, 2017, 152(1):17–32. DOI:10.1016/j.ajodo.2017.03.019.
- [14] Evans S, Ling M, Hill B, et al. Systematic review of meditation-based interventions for children with ADHD [J]. *Eur Child Adolesc Psychiatry*, 2017. DOI:10.1007/s00787-017-1008-9.
- [15] Health Quality Ontario. Effect of early follow-up after hospital discharge on outcomes in patients with heart failure or chronic obstructive pulmonary disease: a systematic review [Z]. Toronto: Health Quality Ontario (HQO). Ontario Health Technology Assessment Series, 2017.
- [16] Larivee N, Suissa S, Khosrow-Khavar F, et al. Drospirenone-containing oral contraceptive pills and the risk of venous thromboembolism: a systematic review of observational studies [J]. *BJOG*, 2017, 124(10):1490–1499. DOI:10.1111/1471-0528.14623.
- [17] Marin AC, Nyssen OP, McNicholl AG, et al. Efficacy and safety of quinolone-containing rescue therapies after the failure of non-bismuth quadruple treatments for *Helicobacter pylori* eradication: systematic review and Meta-analysis [J]. *Drugs*, 2017, 77(7):765–776. DOI:10.1007/s40265-017-0730-4.
- [18] Vlek SL, van Dam DA, Rubinstein SM, et al. Biliary tract visualization using near-infrared imaging with indocyanine green during laparoscopic cholecystectomy: results of a systematic review [J]. *Surg Endosc*, 2017, 31(7):2731–2742. DOI:10.1007/s00464-016-5318-7.
- [19] Des Jarlais DC, Lyles C, Crepaz N, et al. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement [J]. *Am J Public Health*, 2004, 94(3):361–366. DOI:10.2105/AJPH.94.3.361.

(收稿日期:2017-07-04)

(本文编辑:王岚)