

# 环境暴露对健康影响的非随机化研究 偏倚评价工具 ROBINS-E 介绍

王巍巍<sup>1</sup> 周庆欣<sup>2,3</sup> 马蕾<sup>4</sup> 冯世豪<sup>5</sup> 杨智荣<sup>6</sup> 孙凤<sup>2,3</sup> 詹思延<sup>2,3,7</sup>

<sup>1</sup>首都医科大学附属北京安定医院,国家精神心理疾病临床医学研究中心,精神疾病诊断与治疗北京市重点实验室,北京 100088;<sup>2</sup>北京大学公共卫生学院流行病与卫生统计学系,北京 100191;<sup>3</sup>北京大学循证医学中心,北京 100191;<sup>4</sup>哈尔滨医科大学附属第一医院精神卫生中心,哈尔滨 150001;<sup>5</sup>首都医科大学基础医学院,北京 100069;<sup>6</sup>英国剑桥大学临床医学院初级医疗中心 CB1 8RN;<sup>7</sup>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】** 针对非随机化环境暴露研究 Cochrane 偏倚评估工具 ROBINS-E 的主要内容进行详细介绍,并举例说明 ROBINS-E 的使用方法和注意事项。ROBINS-E 针对非随机化暴露研究(non-randomized studies of exposures, NRSE)的特点,设置了相应的评估领域和信号问题,为将 NRSE 纳入到系统综述,在 GRADE 框架下进行证据整合提供偏倚风险信息。

**【关键词】** 偏倚风险; 评价工具; 非随机化研究; 环境暴露; 系统综述

**基金项目:**国家自然科学基金(72074011);北京市属医院科研培育项目(PX2019071)

## Introduction of a tool to assess Risk of Bias in Non-randomized Studies-of Environmental Exposure (ROBINS-E)

Wang Weiwei<sup>1</sup>, Zhou Qingxin<sup>2,3</sup>, Ma Lei<sup>4</sup>, Feng Shihao<sup>5</sup>, Yang Zhirong<sup>6</sup>, Sun Feng<sup>2,3</sup>, Zhan Siyan<sup>2,3,7</sup>

<sup>1</sup>The National Clinical Research Center for Mental Disorders & Beijing Key Laboratory of Mental Disorders, Beijing An Ding Hospital, Capital Medical University, Beijing 100088, China; <sup>2</sup>Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; <sup>3</sup>Center of Evidence-based Medicine and Clinical Research, Peking University, Beijing 100191, China; <sup>4</sup>Department of Psychiatry, The First Affiliated Hospital of Harbin Medical University, Harbin 150001, China; <sup>5</sup>School of Basic Medical Sciences, Capital Medical University, Beijing 100069, China; <sup>6</sup>Primary Care Unit, School of Clinical Medicine, University of Cambridge, Cambridge CB1 8RN, UK; <sup>7</sup>Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing 100191, China  
Corresponding author: Sun Feng, Email: sunfeng@bjmu.edu.cn

**【Abstract】** This paper summaries the Risk of Bias in Non-randomized Studies-of Environmental Exposure (ROBINS-E), a tool for evaluating risk of bias about non-randomized studies of exposures (NRSE), and introduces the application of ROBINS-E in a published NRSE. According to the characteristics of NRSE, evaluation fields and signaling questions were designed in ROBINS-E to provide essential information about risk of bias for NRSE included in systematic reviews and GRADE. ROBINS-E is the tool in assessment of risk of bias in observational studies and quasi-randomized 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-randomized studies; Environmental exposure; Systematic review

DOI: 10.3760/cma.j.cn112338-20201112-01324

收稿日期 2020-11-12 本文编辑 李银鸽

引用格式:王巍巍,周庆欣,马蕾,等.环境暴露对健康影响的非随机化研究偏倚评价工具 ROBINS-E 介绍[J].中华流行病学杂志,2022,43(1):98-104. DOI: 10.3760/cma.j.cn112338-20201112-01324.

Wang WW, Zhou QX, Ma L, et al. Introduction of a tool to assess Risk of Bias in Non-randomized Studies-of Environmental Exposure (ROBINS-E)[J]. Chin J Epidemiol, 2022, 43(1):98-104. DOI: 10.3760/cma.j.cn112338-20201112-01324.



**Fund programs:** National Natural Science Foundation of China (72074011); Beijing Municipal Administration of Hospital Medical Research Cultivation Program (PX2019071)

环境或职业暴露对人体健康结局影响的相关证据通常来自非随机化研究 (non-randomized studies, NRS), 进行环境暴露决策时, 往往需要采用系统综述和证据质量分级等手段对相关研究证据进行评估和整合<sup>[1]</sup>。由于 NRS 中干预措施或暴露因素不是随机分配的, 其结果更容易受到各种潜在偏倚的影响。因此, 利用标准评估工具对原始研究和证据体的偏倚风险 (risk of bias) 进行评价, 对后续的证据评估和循证决策而言就显得至关重要<sup>[2-3]</sup>。

近期一项研究对环境健康危害评估 (environmental health hazard assessments) 中经常使用的 5 种偏倚风险评价方法进行了评价<sup>[4]</sup>, 结果发现, 虽然这 5 种方法选用的评估领域基本一致, 但不同方法中同一评估领域的相对重要性却各不相同, 因此有必要对这些评价方法进行整合和改进<sup>[4-5]</sup>。

#### 一、制定背景

1965 年, Cochran 和 Chambers<sup>[6]</sup> 建议参考随机对照试验 (randomized controlled trial, RCT) 设计的考量要素设计和评价 NRS。2016 年, Hernán 和 Robins<sup>[7]</sup> 提出使用 NRS 回答研究问题时应仿效理想化的 RCT 来进行因果推断。同年 10 月, Sterne 等<sup>[8]</sup> 在 *BMJ* 上发表了 ROBINS-I (risk of bias in non-randomized studies of interventions)。ROBINS-I 是一个领域评估式工具, 用于多种类型的非随机干预性研究的偏倚风险评价, 其核心理念是根据研究问题构造出一个无偏的随机化目标试验作为参考, 通过与目标试验的标准化比较来评价偏倚风险, 而不是通过 NRS 的设计类型 (例如病例对照设计) 直观判断偏倚风险是否存在<sup>[9]</sup>。在 ROBINS-I 中, intervention 指的是自愿接受的医疗干预或治疗措施, 而在环境或职业暴露对健康结局影响的研究中, 暴露通常是在无意的情况下发生, 能够获取的信息相对有限, 例如暴露开始时间和持续时间、暴露测量准确性以及不同暴露水平的区分等。因此, 在对环境或职业暴露非随机化研究进行偏倚风险评估时, 需要考虑暴露的特点。基于这样的考虑, Rebecca 博士和证据推荐分级的评估、制订与评价 (GRADE) 工作组在 ROBINS-I 的基础上开发了新的评价工具 ROBINS-E, 主要用于环境暴露或职业暴露观察性研究的偏倚风险评估。该工具于

2019 年发布<sup>[10]</sup>。

#### 二、工具解读

1. 评价领域的设置: ROBINS-E 在领域设置上与 ROBINS-I 相似, 由 7 个评价领域构成, 分别是混杂偏倚、研究对象选择偏倚、暴露分类的偏倚、偏离既定暴露的偏倚、数据缺失的偏倚、结局测量的偏倚以及结果选择性报告的偏倚。

2. 信号问题调整的解读: ROBINS-E 为每个领域明确设置了一个或多个信号问题, 要求评价员根据非随机化暴露研究 (NRSE) 研究提供的信息对信号问题做出“是 (Yes)”“可能是 (Probably yes)”“可能不是 (Probably no)”或“不是 (No)”的回答。由于 ROBINS-E 与 ROBINS-I 在研究对象选择的偏倚、数据缺失的偏倚、结局测量的偏倚、结果选择性报告的偏倚等领域的信号问题设置相似, 具体内容可以参见系列讲座<sup>[8-11]</sup>。本文不再对相似内容逐一解读, 仅针对调整的信号问题进行阐述<sup>[5]</sup>。见表 1。

(1) 混杂偏倚 (confounding bias): 评估暴露混杂导致的偏倚, 需要特别注意 3 种情况: 横断面研究、大效应值 (large magnitude of effect, 相对危险度  $>5$  或者  $<0.2$ )<sup>[12]</sup> 或反向残余混杂 (opposing residual confounding)、未测量的混杂。残余混杂是指有些混杂因素已经被识别和调整, 但由于测量不准确和误差导致其未被调整完全, 仍然可能歪曲真实因果效应。反向残余混杂指的是残余混杂因素会明显低估暴露的效应。未测量的混杂是指未识别和调整的混杂, 通常是由于研究条件或认识所限而导致的。

对于横断面研究, ROBINS-I 提到的时依混杂造成的偏倚无法进行评估<sup>[11]</sup>, 不能确定横断面观察到的短期混杂因素是否与长期效应相关, 对测量已知混杂因素对真实效应的影响造成一定的困难。

对于有大效应值或反向残余混杂效应的研究, 在进行偏倚风险评价时可以适当放宽标准<sup>[12]</sup>。通常认为, 大效应值的真实关联效应可能远远大于偏倚的影响, 反向残余混杂会低估效应值或令结果更趋于阴性。以上 2 种情况除了对研究水平的偏倚风险评估产生影响, 在 GRADE 评估中也会提高 NRSE 证据体的确定性。以吸烟和肺癌相关死亡率的前瞻性队列为例<sup>[13-14]</sup>, 虽然存在残余混杂和未测量混杂, 例如职业暴露或空气污染等, 但是由于吸

表 1 ROBINS-E 与 ROBINS-I 信号问题设置调整

领域	ROBINS-I	领域	ROBINS-E
混杂偏倚	1.1 研究中是否存在可能影响干预效果的混杂因素? 1.2 是否根据所接受干预的不同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? 1.3 干预中断或变更是否可能与结局的预后因素相关? 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 研究中的变量是否可以真实可靠地测量需要控制的混杂? 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? 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和时依协变量引起的混杂? 1.8 研究中的变量是否可以真实可靠地测量需要控制的混杂?	混杂偏倚	1.1 研究中是否存在影响暴露因素效应的混杂因素? 1.2 是否根据暴露因素的不同,对研究对象的随访时间进行了分段分析? 1.3 暴露中断或变更是否可能与结局的预后因素相关? 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需要控制的混杂? 1.6 对于受暴露影响的因素,作者是否避免进行控制? 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和时依协变量引起的混杂? 1.8 需要控制的混杂因素是否可以在研究中真实可靠地测量?
干预分类偏倚	3.1 是否对各个干预组进行了清晰地界定? 3.2 是否在干预起始时就对各组干预信息进行了记录? 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?	暴露错分偏倚	3.1 是否对暴露状态进行了明确定义? 3.2 是否在暴露起始时进入研究? 3.3 暴露状态记录是否早于结局评估? 3.4 暴露状态的分类是否受到结局知识或结局风险的影响? 3.5 危险人群暴露的水平、持续时间或范围是否足以检测暴露的效应? 3.5 危险人群暴露的水平、持续时间或范围是否足以检测暴露的效应? 3.6 随访期是否足以观察到目标结局? 3.7 暴露评估方法是否可靠(包括用于输入数据的方法)?
偏离既定干预的偏倚	4.1 是否存在非常规的干预改变? 4.2 干预变化是否在组间不均衡,并可能对研究结局造成影响? 4.3 重要的伴随干预措施组间是否均衡? 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? 4.5 研究对象是否依从了分配的干预措施? 4.6 研究是否采用了恰当的方法估计干预的依从效果?	偏离既定暴露的偏倚	4.1 是否存在暴露状态的改变? 4.2 是否许多研究对象的暴露因素发生了转换? 4.3 重要的伴随暴露因素在组间是否平衡? 4.4 是否采用了恰当的控制方法对上述问题进行校正?

烟(超过 25 支)的效应估计值达到 30,偏倚产生虚假关联的影响几乎可以忽略<sup>[15]</sup>,因此对混杂偏倚风险等级的评价为“低”或“中等”<sup>[14]</sup>。

对暴露与目标结局之间的关联进行探索性分析可以帮助识别其他职业暴露或空气污染等潜在残余混杂因素。一些探索性研究显示,最常见的职业暴露因素包括二氧化硫、焊接烟尘、引擎排放物、汽油、润滑油、溶剂、油漆、粘合剂、挖掘粉尘、木屑和吸烟史之间不存在关联<sup>[16]</sup>。在评价以吸烟为暴露因素的观察性研究中,上述提到的职业暴露因素的混杂偏倚可以排除。

(2) 暴露错分偏倚(exposure misclassification bias):在 NRSE 研究中,与暴露测量相关的偏倚是环境暴露研究中系统误差的主要来源<sup>[17]</sup>。识别错分偏倚的来源和类型非常重要,通常非差异性错分使研究效应的估计值偏低,而差异性错分对效应值的影响方向难以确定。对于在体内停留时间很短

的暴露因素,则需要多次重复测量,保证暴露因素水平测量的准确性<sup>[18]</sup>。研究中报告的检测样本数量以及测量方法的准确性若与暴露水平有关,则考虑可能存在随机误差,造成非差异性错分。对于错分偏倚如何处理,可以参考关于二噁英、饮用水消毒副产物与癌症关联的研究<sup>[19]</sup>。

评估各种暴露测量方法造成错分偏倚的风险十分关键,可以参考噪声暴露的划分方法<sup>[20]</sup>,风险由大到小顺序:①自我报告调查表:在工作场所,与同事距离 2 m 左右,是否提高声音才能进行正常对话(表示噪音水平>80 分贝);②建模:在职业环境中,可以考虑职业暴露矩阵,职业卫生学家根据职称或者专业级别对可能的暴露范围进行分类;③环境监测:在工作场所,使用噪声监测仪对噪声进行连续测量,传感器的测量误差在特定的暴露范围内;④个体监测:使用噪声监测仪对个体暴露进行测量,传感器测量误差在特定的暴露范围内;⑤个体

剂量:个体监测,考虑使用耳罩、测量听觉敏锐度等。

(3) 偏离既定暴露的偏倚 (bias due to departures from intended exposures):该领域主要针对暴露因素在研究过程中发生变化这一问题。此外,若伴随暴露因素在组间存在系统性差异,也会产生偏离暴露的偏倚。例如,当研究对象意识到空气污染对健康的不良影响后搬离当地,不仅空气污染暴露因素发生了变化,甚至水源作为伴随暴露因素也会发生改变。

3. 评估标准及整体偏倚评估:在完成信号问题的回答和评估后,基于信号问题的判断将每个领域和研究整体偏倚风险等级划分为:“低”“中”“高”或“极高”。关于整体偏倚风险等级的评估标准,与 ROBINS-I 类似,若 7 个领域评估结果均为低风险,则整体偏倚风险低;若 7 个领域的评估结果是低或中,则整体偏倚风险为中;若其中任意一个领域的评估结果为高风险但无任何评价领域为极高风险,则整体偏倚风险高;若其中任意一个领域的评估结果为极高风险,则整体偏倚风险极高;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)。参考 ROBINS-I 指南,评价员应根据最高偏倚风险领域来判断单个研究的偏倚风险级别,或者综合考虑所有领域的评级后,认为该研究可能发生更高的偏倚风险,则应上调单个研究的偏倚风险级别<sup>[8]</sup>。

### 三、ROBINS-E 在系统综述中的应用

基于环境暴露观察性研究制定系统综述时,推荐 ROBINS-E 工具对原始研究进行偏倚风险评估。操作流程主要包括 3 个阶段:第一阶段,定义系统综述研究问题,PECO 的 4 要素分别是目标人群 (population)、暴露因素 (exposure)、参照因素 (comparator) 和结局 (outcome);第二阶段,构造目标实验,明确混杂因素和联合暴露因素;第三阶段,

按照不同的结局评估偏倚风险的 7 个领域,包括环境暴露研究的优势和局限性。

1. 第一阶段:基于 NRSE 研究定义系统综述的研究问题时,应明确暴露因素是什么,是否涉及不同暴露水平,对照组是什么。例如,在探讨噪声与听力损伤的研究中,可以选择  $\geq 80$  分贝噪声水平作为暴露组,  $< 80$  分贝作为对照组;也可以选择长期稳定暴露在 80 分贝稳态噪声水平以上的工人,与初始暴露水平相比,每增加 5 分贝对听力水平产生的影响。研究人员可以参考 Morgan 等<sup>[21]</sup>提出的方法确定 PECO 框架下的暴露定义。见表 2。

评价小组需要对暴露和结局测量的准确性进行评估。不同研究间暴露水平的定义是否一致。纳入的研究采用直接测量指标还是采用间接测量指标。例如,肥胖作为结局指标时,如果不是通过测量身高、体重进行判断,而是通过测量腰围对肥胖结局进行定义,那么腰围就是肥胖结局的间接测量指标。在进行偏倚风险评价时,应识别和区分不同来源的间接测量指标<sup>[5]</sup>。此外,评价员可以参考测量环境暴露和健康结局混杂因素的公共数据库 (PhenX Toolkit; <https://www.phenxtoolkit.org/>) 定义暴露因素。

2. 第二阶段:评价员根据 PECO 问题、暴露和对照的暴露阈值、与结局相关的混杂因素和结局测量方法,构建本研究的目标实验。目标实验的目的是提供与参照实验的结构化对比,通常默认为低风险偏倚,假定目标实验的设计已经均衡了所有的预后因素和混杂因素。目标实验不仅可以在单个研究水平进行偏倚风险评价,也可以作为研究问题的最低偏倚水平参考进行研究间评估。

3. 第三阶段:对于合格的单个研究,必须探讨混杂因素存在的可能性,研究层面和结局层面都会受到不同混杂因素的影响。例如,研究的目标人群

表 2 PECO 框架下的暴露定义方法

序号	系统综述研究背景	方法	PECO 示例
1	计算暴露的健康效应,描述暴露与风险表征结局之间的量效关系	探索系统综述中暴露与结局关联的形态与分布	在新生儿中,妊娠期暴露每增加 10 分贝对出生后听力损害有何影响?
2	当临界值被多个系统综述结果重复时,可评估暴露临界值对健康结局的影响	采用基于系统综述纳入的研究中的分布定义的多个临界值	在新生儿中,孕期最高分贝暴露相比最低分贝暴露(如确定的三分位数、四分位数或五分位数)对出生后听力损害有何影响?
3	当暴露组临界值与对照组临界值能够确定或可从其他人群获得时,可评估两者之间的关联	采用外部或其他人群(可能来自其他研究)的平均临界值	在商业飞行员中,职业噪声暴露相比其他职业的噪声暴露对听力损害有何影响?
4	确定可以减轻对健康结局影响的暴露临界值	采用已知的与健康结局相关的暴露临界值	在工人中,暴露水平 $< 80$ 分贝相比 $\geq 80$ 分贝对听力损害有何影响?
5	评估可以通过干预措施改善暴露对结局影响的临界值水平的潜在效应	根据干预措施可以达到的暴露临界值选择对照组	在一般人群中,与不采取干预措施相比,将噪声水平降低 20 分贝对听力损害有何影响?

是一般人群,但某项研究仅仅纳入了工厂中的工人,并且有暴露于其他化学品的可能,那么就会影响间接性或选择偏倚的判断。

在系统综述中,敏感性分析通过对证据体进行探索性分析,从而确定纳入或排除有特定偏倚风险的研究后结果是否依然稳健<sup>[22]</sup>。敏感性分析方法应该在系统综述方案中提前指定,或者经过初步分析后确定。例如,由于未调整分析协变量造成的混杂会使研究有极高的偏倚风险。当证据体将未进行调整和进行调整的研究全部纳入时,若敏感性分析发现调整后的效应估计值与总体合并效应估计值存在差异,则认为效应估计值不稳健,受到了混杂因素的影响,可以考虑排除未进行调整分析的研究。如果两者效应估计没有差异,则考虑混杂因素影响不大,可以将未调整的研究保留在分析中。当研究间的效应估计基本一致,敏感性分析显示偏倚风险对结果没有影响,则可以考虑更新单个研究水平的偏倚风险评级,降低该领域的偏倚风险等级。

在进行整体证据质量评价时,需要将单个结局证据体的总体偏倚风险评价整合到 GRADE 中,具

体方法可以参考 Guyatt 等<sup>[23]</sup>制定的针对 RCT 和观察性研究的 GRADE 指南。对于 NRSE 而言,证据体评价起点为“高”级证据,降级因素包括:纳入原始研究的偏倚风险、原始研究之间的一致性、证据的间接性、系统综述结果的不精确性和发表偏倚<sup>[24]</sup>,可能的升级因素包括效应量大小、剂量反应梯度和反向残余混杂<sup>[12]</sup>。

#### 四、实例分析

以 2013 年 4 月发表在 *Environmental Health Perspectives* 上的 NRSE 举例说明 ROBINS-E 的实际应用<sup>[25]</sup>。该研究来自加州大学伯克利分校和美国 CDC 建立的环境因素与儿童生长发育的出生队列,目的是研究产前和产后双酚 A (bisphenol A, BPA) 暴露与 BMI、腰围、体脂百分比和肥胖的关系。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未进行完整展示。由于 ROBINS-E 与 ROBINS-I 结构类似,主要区别在混杂偏倚、暴露错分偏倚、偏离既定暴露偏倚 3 个领域,因此实例重点对此 3 个领域进行偏倚评估和解读(表 3)。结局指标是超重,定义为 5~9 岁儿童的 BMI $\geq P_{85}$ 。

表 3 ROBINS-E 的应用实例

领域	信号问题	回答	支持信息/理由
混杂偏倚	1.1 研究中是否存在影响暴露因素效应的混杂因素?	可能是	可能存在未测量的残余混杂
	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(问题 1.2~1.3)。		
	1.2 是否根据暴露因素的不同,对研究对象的随访时间进行了分段分析?	是	研究评估了在孕期、5 岁和 9 岁时,儿童接触 BPA 的情况,并对每个阶段进行了分析
	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		
1.3 暴露中断或变更是否可能与结局的预后因素相关?	可能是	BPA 暴露可能发生变化,与随着时间推移出现的食物消费的变化有关,例如通过罐头或加工食品、苏打水等暴露于 BPA,这些因素也可能与超重有关	
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:则继续回答与时依协变量引起的混杂相关的问题(问题 1.7~1.8)。			
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?	不适用		
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需要控制的混杂?	不适用		
1.6 对于受暴露影响的因素,作者是否避免进行控制?	不适用		
关于时依混杂的信号问题			

续表 3

领域	信号问题	回答	支持信息/理由
暴露分类偏倚	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和时依协变量引起的混杂?	是	潜在的混杂因素包括孕妇孕前体质指数、年龄、教育程度、在美国居住的年限、孕期吸烟、孕期汽水消费和家庭收入。考虑的时依协变量是儿童对汽水、快餐和糖果的消费、看电视、环境烟草烟雾暴露和户外活动时间。采用了广义估计方程,用年龄和 BPA 作为交互项进行了调整,对于协变量的纳入标准进行了明确规定:与暴露和结局均相关, $P < 0.2$ 或者移除任何一项协变量导致 BPA 的系数增加 10%
	1.8 若问题 1.7 回答为是/可能是:需要控制的混杂因素是否可以在研究中真实可靠地测量?	可能是	研究中的人口统计特征、饮食和行为信息均采用结构化问卷进行调查,但未提供信效度的信息
	偏倚风险评价 可选答问题:混杂偏倚的方向	中	
	3.1 是否对暴露状态进行了明确定义?	是	本研究是一个纵向队列,BPA 的暴露状态通过测量尿液中的 BPA 浓度测定
	3.2 是否在暴露起始时进入研究?	是	暴露测量从怀孕的前三个月开始,可以看做婴儿暴露的起始
	3.3 暴露状态记录是否早于结局评估?	是	孕期和 5 岁的暴露测量是在 9 岁评估体质指数之前进行的
	3.4 暴露状态的分类是否受到结局知识或结局风险的影响?	否	4 次暴露测量中只有 1 次是与结局在相同时间进行,因此考虑暴露分类不会受到结局的影响
	3.5 危险人群暴露的水平、持续时间或范围是否足以检测暴露的效应?	可能是	暴露的范围/变异性是足够的(妊娠期为 0.5~4.6 ng/ml,儿童期为 0.9~16.3 ng/ml)。BPA 水平可能随着时间变化,虽然无法确定受试者是否长时间暴露在同一水平浓度下,但研究者认为 4 次测量可以反映暴露随时间变化的情况
	3.6 随访期是否足以观察到目标结局?	是	对研究对象进行 9 年随访,认为随访期足够观察到结局
	3.7 暴露评估方法是否可靠(包括用于输入数据的方法)?	可能不是	尿样的采集、保存、运输、检测仪型号、检测方法均有详细的记录。但由于 BPA 的测定样本在 5~13 个样本时,诊断的灵敏度和特异度才能达到 0.8 以上,因此考虑存在非差异性错分风险
偏离既定暴露的偏倚	偏倚风险评价 可选答问题:偏离既定暴露的偏倚方向?	高 可能存在暴露非差异性错分导致趋零效应	
	4.1 是否存在暴露状态的改变?	否	研究者认为在不同阶段测量 4 次 BPA 可以捕捉到暴露状态的真实变化情况,在不同阶段按照当时的暴露状态进行了评估
	如果研究目的是评价暴露依从效应(类似符合方案集分析),请继续回答问题 4.2 和 4.3,如果评价暴露的分组效应,如在 4.1 问题中回答是/可能是,请回答 4.4		
	4.2 是否许多研究对象的暴露因素发生了转换?	不适用	
	4.3 重要的伴随暴露因素在组间是否平衡?	不适用	
4.4 如在 4.2 问题中回答是/可能是,或者在 4.3 问题中回答否/可能不是,则回答:是否采用了恰当的控制方法对上述问题进行校正?	不适用		
偏倚风险评价 可选答问题:偏离既定暴露的偏倚方向?	低		

注: BPA: 双酚 A

### 五、讨论

ROBINS-E 是由 ROBINS-I 衍生而来,主要用于环境暴露或职业暴露观察性研究的偏倚风险评价,前者聚焦于无意识的环境或职业暴露,而

ROBINS-I 涉及的措施主要指临床诊疗中的干预措施(如药物、器械、心理干预、康复理疗等)。ROBINS-E 制定小组通过严谨的定性研究方法确定相应的领域和条目,对环境暴露研究进行风险偏倚

评估,评价证据的确定性,最终目的是促进环境健康领域的决策。整体而言,ROBINS-E 是一个制定方法科学、纳入条目详尽、应用目标明确的 NRSE 的偏倚风险评价工具。

ROBINS-E 目前尚处于初步应用阶段,在使用过程中,尚有值得商榷的问题。首先,在可操作性方面,与 National Toxicology Programs' Office of Health Assessment and Translation、Office of the Report of Carcinogens 和 Newcastle-Ottawa Scale 这些暴露因素偏倚风险评估工具比较,采用 ROBINS-E 完成单个研究的评估时间显著增加,虽然从详细的信号问题中可以更好地识别偏倚风险,但在大型系统综述实施过程中,偏倚风险判断的准确性可能会受到一定的影响,这个问题有望在开发团队公布程序化工具后有所改善。其次,定义目标实验可能遇到困难,许多环境暴露因素造成的健康危害是潜在关联,没有足够的证据明确区分暴露组和对照组。在这种情况下,建议使用其他的可比的暴露数据来源,例如历史对照(引入暴露来源之前的数据)。最后,剂量反应关系在偏倚风险评价和 GRADE 中的作用未进行说明,有待后续工具进行更新。

对于暴露因素的研究者而言,还需要明确一点,关于动物的环境暴露效应,更多的是采用随机化分组的实验设计,应根据实验性研究的框架进行评估,而不是 ROBINS-E。

综上所述,建议环境健康领域的综述作者持续关注 ROBINS-E 的更新,在使用过程中积极与开发团队沟通,反馈遇到的问题和建议,以便 ROBINS-E 能够尽快推广应用,在环境健康领域的证据评估中发挥重要作用。

**利益冲突** 所有作者声明无利益冲突

**作者贡献声明** 王巍巍:实验设计、论文撰写、行政、技术或材料支持;周庆欣:数据分析、文章修改;马蕾、冯世豪、杨智荣:文章修改、提供建议;孙凤、詹思延:提供论文设计、行政、技术或材料支持

### 参 考 文 献

- Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes[J]. *Environ Health Perspect*, 2014, 122(10):1007-1014. DOI:10.1289/ehp.1307175.
- Balslem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence[J]. *J Clin Epidemiol*, 2011, 64(4):401-406. DOI:10.1016/j.jclinepi.2010.07.015.
- Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration[J]. *PLoS Med*, 2009, 6(7):e1000100. DOI:10.1371/journal.pmed.1000100.
- Rooney AA, Cooper GS, Jahnke GD, et al. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards[J]. *Environ Int*, 2016, 92-93:617-629. DOI:10.1016/j.envint.2016.01.005.
- Morgan RL, Thayer KA, Santesso N, et al. Evaluation of the risk of bias in non-randomized studies of interventions (ROBINS-I) and the 'target experiment' concept in studies of exposures: Rationale and preliminary instrument development[J]. *Environ Int*, 2018, 120:382-387. DOI:10.1016/j.envint.2018.08.018.
- Cochran WG, Chambers SP. The planning of observational studies of human populations[J]. *J Roy Stat Soc Ser A (General)*, 1965, 128(2):234-266. DOI:10.2307/2344179.
- 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.
- Sterne JA, Hernán 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.
- Schünemann HJ, Cuello C, Akl EA, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence[J]. *J Clin Epidemiol*, 2019, 111:105-114. DOI:10.1016/j.jclinepi.2018.01.012.
- Morgan RL, Thayer KA, Santesso N, et al. A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE[J]. *Environ Int*, 2019, 122:168-184. DOI:10.1016/j.envint.2018.11.004.
- 孙凤, 高乐, 杨智荣, 等. 偏倚风险评估系列: (五) 非随机干预性研究[J]. *中华流行病学杂志*, 2018, 39(33):374-381. DOI:10.3760/cma.j.issn.0254-6450.2018.03.024.
- Sun F, Gao L, Yang ZR, et al. Series of risk of bias assessment (5): Risk of bias in Non-randomized Studies of Interventions (ROBINS-I) [J]. *Chin J Epidemiol*, 2018, 39(33):374-381. DOI:10.3760/cma.j.issn.0254-6450.2018.03.024.
- Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence[J]. *J Clin Epidemiol*, 2011, 64(12):1311-1316. DOI:10.1016/j.jclinepi.2011.06.004.
- Doll R, Hill AB. Smoking and carcinoma of the lung[J]. *Br Med J*, 1950, 2(4682):739-748. DOI:10.1136/bmj.2.4682.739.
- Doll R, Hill AB. Mortality in relation to smoking: ten years' observations of british doctors[J]. *Br Med J*, 1964, 1(5395):1399-1410. DOI:10.1136/bmj.1.5395.1399.
- Bross IDJ. Spurious effects from an extraneous variable[J]. *J Chronic Dis*, 1966, 19(6):637-647. DOI:10.1016/0021-9681(66)90062-2.
- Blair A, Stewart P, Lubin JH, et al. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures[J]. *Am J Ind Med*, 2007, 50(3):199-207. DOI:10.1002/ajim.20281.
- Steenland K, Savitz DA. Topics in environmental epidemiology[M]. New York: Oxford University Press, 1997.
- Cox KJ, Porucznik CA, Anderson DJ, et al. Exposure classification and temporal variability in urinary bisphenol A concentrations among couples in utah-the HOPE study[J]. *Environ Health Perspect*, 2016, 124(4):498-506. DOI:10.1289/ehp.1509752.
- Kogevinas M. Epidemiological approaches in the investigation of environmental causes of cancer: the case of dioxins and water disinfection by-products[J]. *Environ Health*, 2011, 10 Suppl 1 (Suppl 1):S3. DOI:10.1186/1476-069X-10-S1-S3.
- Nieuwenhuijsen MJ. Exposure assessment in environmental epidemiology[M]. Oxford:Oxford University Press, 2015.
- Morgan RL, Whaley P, Thayer KA, et al. Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes[J]. *Environ Int*, 2018, 121:1027-1031. DOI:10.1016/j.envint.2018.07.015.
- Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials[J]. *Cochr Datab Systemat Rev*, 2016, 10 Suppl 1:29-31.
- Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias) [J]. *J Clin Epidemiol*, 2011, 64(4):407-415. DOI:10.1016/j.jclinepi.2010.07.017.
- 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.
- Harley KG, Schall RA, Chevrier J, et al. Prenatal and postnatal bisphenol A exposure and body mass index in childhood in the CHAMACOS cohort[J]. *Environ Health Perspect*, 2013, 121(4):514-520. DOI:10.1289/ehp.1205548.