

• 基础理论与方法 •

COSMOS-E: 病因学观察性研究的系统综述和Meta分析制作指南与案例解读

王巍巍¹ 杨智荣² 周庆欣^{3,4} 石舒原^{3,4} 张格⁵ 詹思延^{3,4} 孙凤^{3,4}

¹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; ²英国剑桥大学临床医学院初级医疗中心 CB1 8RN; ³北京大学公共卫生学院流行病与卫生统计学系 100191; ⁴北京大学循证医学中心 100191; ⁵掌番信息科技(上海)有限公司, 上海 200090

通信作者: 孙凤, Email: sunfeng@bjmu.edu.cn

【摘要】 本文针对病因学观察性研究的系统综述和Meta分析制作指南(Guidance on conducting systematic reviews and Meta-analyses of observational studies of etiology, COSMOS-E)的主要内容进行了介绍, 并举例说明COSMOS-E中的关键内容, 内容覆盖制作病因学观察性研究系统综述的每个步骤, 从提出研究问题、定义暴露和结局、评估偏倚风险到统计分析, 为研究人员开展和分析这类综述提供了指导。

【关键词】 病因学; 观察性研究; 系统综述; 指南

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)

DOI: 10.3760/cma.j.cn112338-20191024-00758

Introduction to COSMOS-E: Guidance on conducting systematic reviews and Meta-analyses on etiology of observational studies

Wang Weiwei¹, Yang Zhirong², Zhou Qingxin^{3,4}, Shi Shuyuan^{3,4}, Zhang Ge⁵, Zhan Siyan^{3,4}, Sun Feng^{3,4}

¹The National Clinical Research Center for Mental Disorders and Beijing Key Laboratory of Mental Disorders, Beijing Anding Hospital, Capital Medical University, Beijing 100088, China; ²Primary Care Unit, School of Clinical Medicine, University of Cambridge, Cambridge CB1 8RN, UK; ³Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; ⁴Center of Evidence-based Medicine and Clinical Research, Peking University, Beijing 100191, China; ⁵Zhangfan Information Technology (Shanghai) Co., Ltd, Shanghai 200090, China

Corresponding author: Sun Feng, Email: sunfeng@bjmu.edu.cn

【Abstract】 This paper introduces the conducting systematic reviews and Meta-analyses of observational studies of etiology (COSMOS-E) and illustrates the critical issues of COSMOS-E with a published systematic review. This document provides researchers with guidance on all steps in systematic reviews of observational studies of etiology, from shaping the research question, defining exposure and outcomes, to assessing the risk of bias and statistical analysis.

【Key words】 Etiology; Observational study; Systematic review; Guidance

Fund programs: National Natural Science Foundation of China (72074011); Beijing Municipal Administration of Hospitals Incubating Program (PX2019071)

DOI: 10.3760/cma.j.cn112338-20191024-00758

一、制作背景

系统综述的目的是评价和整合现有证据回答特定的研究问题, Meta分析是对相关研究结果的统计合并。系统综述通常是Meta分析的基础, 但Meta分析并不是系统综述的必要特征。目前, 基于观察性研究开展的系统综述逐渐增加, 在2014年发表的约8 000篇系统综述中, 36%是关于病因、预后或诊断的基于观察性研究的系统综述^[1]。病因学研究关

注暴露与疾病或健康相关结局之间的关联。潜在可致病的暴露也称为危险因素, 存在形式多种多样。危险因素或是不可改变, 如性别、遗传因素; 或随时间变化, 如代谢危险因素(如高胆固醇血症、胰岛素抵抗、高血压)、生活习惯(如吸烟、饮食)或环境因素(如空气污染、热浪)。从概念上讲, 暴露不同于干预措施, 后者是以改变健康结局为目的, 有明确的时间起点^[2]。对于难以在随机对照试验(randomized

controlled trials, RCTs)中研究的暴露,如空气污染、吸烟等,观察性研究有其独特优势。此外,观察性研究对于研究潜伏期较长的病因,如环境暴露或药物的致癌作用也有重要意义。

危险因素的流行病学研究通常依赖于比较分析(暴露与未暴露的组间比较),如队列研究对暴露与未暴露的人群进行随访追踪^[3-4]。其他研究设计如工具变量分析^[5-6]、孟德尔随机化^[7]、自身对照研究^[8-9]、病例对照研究^[3,10]、横断面研究^[11]和生态学研究^[12-13]也依赖比较分析方法。

二、COSMOS-E 的主要内容

1. 目的和适用范围:RCT 的系统综述制作^[14]和报告指南^[15]已经广泛应用。观察性研究系统综述的报告规范已经发表近 20 年^[16],但迄今尚未发布其制作指南。尽管通用结构相似,但观察性研究系统综述的制作流程标准化程度较低^[17],尤其针对研究设计问题缺少统一的指导意见^[18]。COSMOS-E 的目的是讨论并指导研究人员进行病因学观察性研究系统综述,但不要求必须按照该指南进行操作或报告综述。其步骤有些与医学干预的 RCTs 系统综述相似,内容涵盖了基本原理和更高级的主题,但不涉及非随机干预研究、诊断、预后或遗传研究的系统综述。此外,对病因学观察性研究的系统综述面临的争议问题^[18],本文也给出了不同的观点和可能性。

2. 筹备系统综述

(1) 组建综述团队:在设计阶段,团队应涵盖多个学科和方法学专业。在 Meta 分析中,可能遇到各种各样的问题,例如不同的研究设计(例如,病例对照研究和队列研究)是否可以合并、如何合并分析、剂量反应 Meta 分析是否可行。如识别潜在的混杂变量或评估暴露测量需要专业学科知识,复杂的统计分析方法强调对统计专业知识的要求。邀请信息学专家参与可以保证文献检索全面、高效。

(2) 提出研究问题:观察性研究系统综述需要明确研究问题,最初的选题可以宽泛,但考虑到清晰性和可行性,应进一步缩小范围。相较于 RCT 的系统综述,观察性研究系统综述的研究问题可能需要反复调整。提出研究问题后,评价员应锁定关键论文,了解关于相关证据有哪些、已经开展了哪些类型的研究。探索阶段目的有两个,一是明确问题是否已经在最新的系统综述中得到了回答,二是提示研究问题是否需要调整、如何改进以及进一步明确,使之成为切实可行的系统综述主题。

(3) 定义人群、暴露、比较和结局:与 RCT 的人

群、干预、对照和结局(population, intervention, control and outcome, PICO)格式^[15]相对应,观察性研究的系统综述应定义人群、暴露、对照和结局(population, exposure, control and outcome, PECO)^[19]。研究人群应反映目标人群,即结果适用的人群^[20],可以是一般人群,如胰岛素样生长因子和死亡率关联的 Meta 分析^[21],也可以是限制性人群,如母乳喂养和儿童白血病关联的综述^[22]。研究人群必须明确定义,以便有效地研究暴露-结局的关系^[23]。例如,假设辐射暴露会损害生长中的组织,则儿童比成人更适合作为研究人群^[23]。

在理想状态下,危险因素研究纳入的人群在随访开始时未出现研究结局,但在基于人群的研究中往往无法证实。如果不明确排除出现研究结局的情况,可能无法检测亚临床或早期疾病。在一项关于胰岛素抵抗与心血管事件关联的综述中,纳入的部分研究未明确排除基线时患有心血管疾病的研究对象,因此认为有较高的偏倚风险^[24]。

暴露和结局应明确定义。在病因学观察性研究中,需注意暴露因素的定义和测量,如社会经济状况、饮食、运动或环境化学物质等,并且需要评估研究间测量的可比性。结局应考虑目标结局是什么、如何确定结局发生。结局可以用不同的方法进行定义、分类或测量,如疾病(乳腺癌、血栓、糖尿病等)或健康相关状态(如生活质量、危险因素水平),在已知的研究结局中只有死亡对测量或确定方法不敏感。

评价员可以制定精确的暴露、结局定义,但如果研究没有采用这些定义,评价员则无法回答研究问题。在这种情况下,需要根据实际情况进行调整。在脂肪量和心血管风险的研究中,复杂的暴露测量方法如磁共振成像可能只在少数小样本研究中使用,如果使用体重指数或腰臀比,则会纳入更多的大型研究。

(4) 考虑混杂因素和偏倚:混杂是影响观察性研究真实性的一个重要因素。由于暴露和结局的共同影响因素,使组间的风险效应估计超出了目标暴露对结局的效应时,意味着混杂因素出现。在综述设计阶段,研究人员应仔细考虑哪些混杂因素可能歪曲暴露-结局关联。混杂因素不仅是一个是或否的现象,而且涉及程度问题。例如,比较素食主义者和非素食主义者死亡率的研究中,预期会出现强混杂,因为这些群体在其他生活方式特征上也有所不同,而这些特征又与死亡原因有关。然而,在吸烟与肺癌的病因学研究中,则几乎不会有强混杂因素,因为

肺癌的其他强危险因素较为罕见,且这些危险因素也不太可能与吸烟密切相关。通常,对于研究中出现的非预期不良事件,与其相关的强混杂因素很少见,例如石棉与间皮瘤之间的关联^[25]。

影响效应估计真实性的其他因素还包括测量(错误分类)偏倚或选择偏倚。错分是环境流行病学和职业流行病学中一类重要的偏倚,尤其是长期暴露^[26]。预先考虑潜在的混杂和偏倚,有助于界定哪些因素可能歪曲研究的真实性,有助于判断哪些类型的研究可能提供最优的无偏估计。

(5)制定研究方案:每个系统综述都应该有详细的研究方案。包括背景和原理、研究问题、暴露、对照和结局的定义、可能影响研究结果的潜在混杂因素和偏倚、研究入选标准、文献检索策略、数据提取(研究特征和结果)、偏倚风险评估和研究敏感性、统计方法、分析计划、证据评价方法。研究方案的制定通常是一个反复的过程,需要根据文献的范围和预实验来确定。评价员应注意不能根据研究结果变更方案,但是可以根据所获得的研究数量、研究类型修订方案。在系统综述的国际前瞻性登记注册(international prospective register of systematic reviews, PROSPERO)^[27]网站上注册研究方案,可以提高透明度,使编辑、同行评审员和其他人能将研究方案与已发表的报告进行比较,便于识别结果是否一致、是否存在选择性报告。

3. 检索相关研究:检索合格的研究包括几个步骤:①选择电子数据库,例如 Medline、Embase、Toxicology Literature Online [Toxline]等专业数据库或监管机构数据库;②与信息科学家或图书管理员合作^[28],共同制定检索策略,并进行测试和优化;③考虑更多检索方法,例如追踪或审查重要论文的引用论文;④决定是否检索灰色文献,如:学术文摘、论文、预印本。部分相关研究可能并未保存在电子数据库^[29-30],因此应当考虑将检索范围扩大至标准电子数据库以外。

由于文献数据库中研究类型的索引不准确,导致对观察性研究设计的检索结果不全面,同时,没有任何数据库能够完全覆盖所有的相关文献^[31],因此应检索多个电子数据库。目前尚无有效的系统检索策略来识别流行病学研究^[32],使用关键研究列表可以测试检索策略的灵敏度。在不损失灵敏度的前提下,初步检索到的研究数量可能非常庞大,很难做到精确的检索。人类技术评估信息检索(information retrieval for human technology assessment, HTA,

www.sure-info.org)网站提供了关于检索资源的信息以及制定检索策略的指导信息,包括用于识别观察性研究的检索过滤词。对于是否检索非英语的观察性研究取决于研究问题。一般来说,语言限制可能引入偏倚,因此需要慎重。研究人员不仅应检索暴露和结局相关的研究,而且应留意阴性暴露和结局对照的研究、暴露或结局的生态学和趋势研究以及基础科学论文。

检索会生成包含作者、标题、期刊等题录信息。由于同一项研究可能报告了多次^[33],需要根据标题和摘要对检索到的报告进行筛选,剔除明显不相关的重复出版物和文章。部分研究需要通过全文来确定是否符合标准、研究人群是否重叠。即使有明确定义的纳入排除标准,确定某些研究是否入选也会遇到困难。例如,研究人员计划开展一项关于儿童的综述,一些研究可能包括了青年成人,并且没有提供儿童数据。在这种情况下,评价员需要决定青年成人的比例占多少可以纳入研究,也可以尝试联系作者获得儿童的数据。

对于是否应该根据研究设计和方法学质量来纳入研究尚有争议^[18]。对于特定的综述,如果设计的特征与高风险偏倚明确相关,且容易识别(例如,长期暴露的病例对照研究),这样的研究可以首先排除。但由于风险偏倚评估在一定程度上是主观的,采用这种方式可能导致错误排除一些研究。另一方面,纳入所有可能的研究有助于在探究研究间异质性时发现重要结果^[34]。

记录和报告排除原因,可以增加筛选过程的透明度,并可据此进行敏感性分析、探讨排除研究或纳入研究的影响。因此,评价员应记录研究筛选过程,并在流程图中进行描述^[15](图1)。目前,已有专用软件可以完成研究筛选过程(<http://systematicreviewtools.com/>),机器学习和文本挖掘等工具可部分实现自动化查找合格研究和提取文章数据。

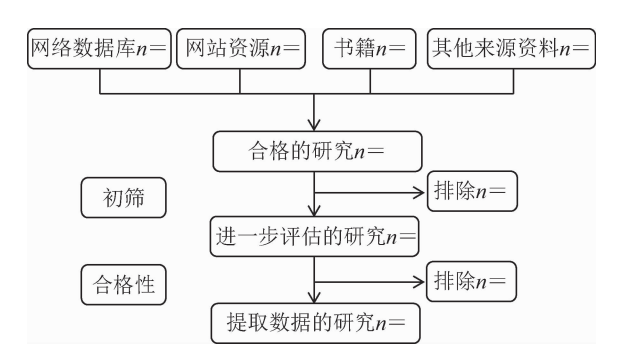


图1 研究筛选流程图

4. 数据提取:研究筛选、数据提取、风险偏倚评估过程建议由 2 位评价员独立平行进行,以减少错误和检测评价员间的理解偏差^[35-36],分歧可以通过讨论解决^[37]。每个综述应制作标准化的数据提取表单,可以采用典型的研究对表单进行预试验、优化,然后在通用软件(如 EpiData)或专用软件(如 Covidence, <http://systematicreviewtools.com>)中实现。对于所有纳入的研究,应提取核心数据:①文献目录信息,②研究设计,③风险偏倚评估,④暴露和结局以及相关定义,⑤研究对象的特征,⑥数据结果(每组研究对象的数量和结局的数量),⑦效应估计(调整后和未调整的)和标准误。

文献目录信息包括杂志名称、发表年份、卷期、页码和数字对象标识符(digital object identifier, doi)。研究设计的判断应基于研究实施过程的描述和评估,而非依据标题以及数据库中的索引^[3]。观察性研究设计的索引往往不充分,在专业期刊中,有 30%~50%索引为“病例对照研究”的研究并非真正的病例对照研究^[38-39]。

研究通常会报告效应值和测量精度(置信区间)或 *P* 值。效应值包括二分类变量结局的优势比、率比、相对危险度或危险度差,以及连续变量结局的均值差。标准误或标准差或许不能直接提取(标准误可能被错误描述为标准差),需要通过间接计算得出标准误。

调整混杂后的估计对于观察性研究来说是最重要的,但提取未校正的估计值和原始数据也很关键。比较调整前后的估计值可以深入了解混杂的重要性。许多研究对不同模型的效应值进行了报告,并根据不同的混杂因素进行了调整。在这种情况下,可以对最大调整估计值和最小调整估计值或粗略估计值分别进行 Meta 分析,可以参考胰岛素样生长因子和癌症风险的 Meta 分析^[40]。

5. 评价质量和偏倚:对研究设计质量的评价是系统综述的关键组成部分。观察性研究可能由于混杂或偏倚得出偏离真实关联的估计值。因此,观察性研究的 Meta 分析可能产生非常精确但虚假的结果^[41]。

研究质量一词经常用于此种情形,但是区分质量和偏倚风险很重要。一项高质量的研究可能存在极高的偏倚风险。例如,一项关于终生饮酒和子宫内膜癌风险的病例对照研究,采用了基于人群的设计以降低选择偏倚^[42],但只能依靠自我报告获得饮酒量,由于一些女性可能会低估饮酒量,从而导致社

会期望偏倚^[43]。研究敏感性是指研究发现真实效应的能力,与研究质量的关系比偏倚更为密切^[23]。如果研究结果为阴性,是否表明暴露与结局之间无因果关联?暴露人群的数量是否足够?暴露水平和持续时间是否足以检测到效应^[23]?随访时间是否足够长到可能发展为癌症?研究敏感性在职业流行病学和环境流行病学中尤为重要,而且在药物流行病学中也备受关注。评价员应该在观察性研究的综述中考虑评估偏倚风险和研究敏感性的问题。

(1) 单个研究的偏倚风险:偏倚通常来源于错误的信息收集或研究对象选择,导致发现偏离真实值的关联。通常情况下,偏倚是在研究的设计或实施过程中引入的,日后无法纠正。

与偏倚相反,混杂产生的关联是真实的,但不是因果关系,混杂是因为存在其他未考虑的与暴露和结局均相关的因素。时间依赖的混杂是一种特殊的混杂形式。对相关的混杂变量进行测量,可以通过统计分析进行调整,但调整后可能仍有残余的混杂。混杂经常与选择偏倚混淆。研究对象、随访时间或结局事件等预后变量在组间分布不均,通常称为选择偏倚。有向无环图(图 2)有助于阐明混杂和选择偏倚的关系^[44]。另一类重要的偏倚是信息偏倚,其中暴露或结局数据准确性的系统性差异可能导致个体暴露或结局的差异性错分。另外,偏倚还需要与随机误差区分开,随机误差是由于数据中的偶然变化而导致的与真实值的偏差。需要注意的是,混杂和选择偏倚指的是研究内部的偏差(内部有效性),而不是一般性或适用性问题(外部有效性)^[20]。对于观察研究中的偏倚风险评价目前尚没有一个公认的工具^[45]。考虑到观察性研究中研究设计、背景和研究问题的巨大异质性,应该根据下述的一般原则,为每个观察性研究系统综述和 Meta 分析制定一套标准。

(2) 通用原则:在评估偏倚风险时,有 7 项通用原则尤为重要^[46-48]。

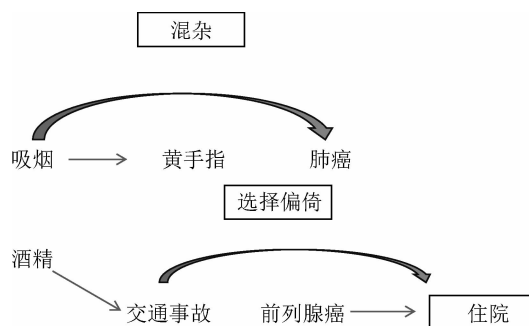


图 2 混杂和选择偏倚的因果关系

①应为每个综述问题和不同的研究设计分别定义相关领域的偏倚。相关领域的偏倚风险应该包括混杂偏倚(时间依赖),选择研究对象产生的偏倚(选择偏倚),暴露或结局测量的偏倚(信息偏倚),数据缺失导致的偏倚(选择偏倚),选择研究或者选择性报告结局导致的偏倚(选择偏倚)^[47]。研究者应评估不同结局的不同领域的偏倚风险。

②应对偏倚风险进行定性评估。对于每一项研究和偏倚领域,定性地对偏倚风险评估,例如,低风险、中风险或高风险。分类和定义标准应在文中加以描述。应避免通过评分进行定量评估。

③信号问题有助于判断偏倚。在每个偏倚领域中,简单的信号问题有助于判断偏倚的风险(表1)。Cochrane 偏倚风险评估工具(Cochrane risk of bias assessment tool for non-randomised studies of interventions, ROBINS-I)的开发人员正在开发用于环境暴露非随机化研究的工具^[47,49]。Cooper等^[23]编制了关于研究敏感性的问题清单。

④不同的结局必须分别评估。偏倚风险在不同的结局间是有区别的。例如,全死因结局确定的偏倚风险比主观的结局指标低;生活质量或疼痛或肺炎等依赖临床判断的结局。

⑤评估结果需文件记录。将文章中偏倚风险的评估依据的文本复制和存档,可以增加透明度,便于在分歧时进行讨论和重复评估结果。

⑥应避免总分制。总分涉及偏倚领域权重,通常每个条目分权重是相等的(0或1分),但偏倚的重要性取决于研究背景^[50-51]。量表中可能包含与偏倚不相关的条目,例如,Newcastle-Ottawa量表关于真实性的评价条目中,纳入了病例和对照组之间的无应答有无可比性^[52],是值得商榷的。因此,建议根据单个领域确定的最高风险,划分偏倚的风险级别。

⑦假设一项没有偏倚的试验有助于偏倚评估。可以假定理想试验,以此回答系统综述中提出的问题^[47],使综述问题更加清晰,阐明观察性研究中潜在的偏倚。

(3)报告偏倚、P值操纵和分析选择:选择性发表是削弱系统综述或荟萃分析结论的重要偏倚来源。阳性结果研究比阴性结果研究更容易发表(如

统计显著效应),有充分的证据表明RCT存在发表偏倚和其他报告偏倚^[53-54]:阳性试验更有可能发表,更可能迅速和多次发表,并且被引用的概率更大,从而更有可能纳入系统综述。

当一项研究考虑了多种暴露和结局,如果只对具有统计意义的关联进行全面分析、撰写和发表,系统综述的结果将被歪曲。如果基于P值选择研究人群和统计模型(P值操纵)^[55],会引入偏倚风险。

(4)如何处理偏倚风险:偏倚风险分析结果的呈现形式应透明,列出每项研究的偏倚风险因素。重点考虑如何处理高偏倚风险的研究。如果综述的目的是提供关于医学干预有效性的最佳证据,则综述通常仅限于低偏倚风险的研究。对于病因学观察性研究的系统综述,一般不建议基于偏倚风险评估排除研究。纳入所有研究,采用分层分析和回归分析探索不同偏倚风险和研究敏感性对于结果的影响,通常会提供更多的信息。

6. 探讨异质性:纳入系统综述的研究通常在研究设计、研究人群和偏倚风险方面有所不同^[1]。了解研究间异质性可以对综述有整体性的把握^[56],而且有助于决定是否进行统计学合并分析。研究者可以从研究间的差异中发现和利用更多的信息^[57]。例如,在一项克罗恩病与无炎症性肠病或溃疡性结肠炎的病例对照研究的综述中,研究了鸟分枝杆菌亚种副结核杆菌病(*Mycobacterium avium subspecies paratuberculosis*, MAP)与克罗恩病的关系^[58]。结果显示具有很强的相关性,表明MAP与克罗恩病的相关性是特异性的,而不是炎症性肠病中的一种普遍现象。

研究环境的差异也可以提供一些思路。生活方式因素,如吸烟、体育活动、性行为或饮食是许多观察性研究的目标暴露,但这些因素通常高度相关,其独立作用很难阐明。纳入特殊人群的研究,如不同

表1 不同偏倚领域的信号问题

偏倚领域	研究设计	信号问题
混杂偏倚	所有类型	哪些重要变量可能对暴露产生混杂
	所有类型	变量是否在合适的时间节点进行精确测量
	所有类型	作者是否使用了合适的分析方法和设计方法调整重要的混杂变量
选择偏倚	队列研究/横断面研究	入选研究与暴露和结局不相关
	队列研究/横断面研究	数据缺失原因与暴露和结局不相关
	病例对照研究	对照是否从发生病例的人群中抽样
	病例对照研究	缺失数据是否与病例或对照的疾病状态相关
信息偏倚	队列研究/横断面研究	结局评估人员是否不知晓研究对象的暴露状态
	队列研究/横断面研究	结局的评价方法在暴露组间是否可比
	病例对照研究	病例或对照的定义是否在不知道暴露因素的情况下确定
	病例对照研究	暴露状态的收集不受结局和结局风险因素的影响

的宗教或地理区域造成的不同生活方式,有助于了解(残余)混杂。

因此,在设计或分析阶段对异质性来源进行深入探讨,是病因学观察性研究系统综述和 Meta 分析的重要组成部分。分析方法,主要包括亚组分析和 Meta 回归,需要在研究方案中预先设定,或按照探索性数据分析进行解释。在分析阶段,异质性的探讨通常始于对森林图和漏斗图进行检查。

7. 是否进行 Meta 分析:通过分析偏倚风险和异质性的其他来源,考虑统计合并所有研究或亚组分析的效应值是否合适。作者应该说明不合并分析的原因^[59],目前对于如何处理合并分析的问题,持有不同的观点^[18]。无论合并与否,应基于研究的种类、敏感性和偏倚风险考虑,而不仅仅基于异质性的统计测量指标。具体原因有 2 个,首先,在没有统计异质性的情况下,合并有偏倚的研究结果可能得到同样有偏倚的合并估计,其置信区间较窄,有可能被错误地解释为决定性证据。纳入偏倚风险较高的研究通常会引入异质性,但也有例外。例如,观察性研究显示,富含 β -胡萝卜素的饮食对心血管死亡率的保护作用在各研究中非常一致。然而,补充 β -胡萝卜素的随机对照试验没有显示出任何益处,可能观察性研究的结果在一定程度上是由健康饮食和生活方式的混杂造成的^[41]。其次,虽然存在统计异质性,但如果研究的偏倚风险较低,且定性结果一致,表明存在一定程度的获益或风险,那么合并分析研究也是合理的。如果作者决定不提供总体的合并估计,可以考虑按照研究设计或人群分层 Meta 分析,即亚组分析。即使没有 Meta 分析,记录异质性和偏倚风险的系统综述仍然可以提供有价值的证据信息。

8. 统计学分析:

(1)观察性研究中的固定效应模型和随机效应模型:评价员进行合并研究及 Meta 分析时,需要确定采用固定效应模型还是随机效应模型^[60]。固定效应模型的前提假设是所有研究有相同的真实效应,效应的点估计之间的差异是由抽样误差造成的。随机效应模型假设真实效应在研究之间是不同的。统计学异质性存在的情况下,2 种模型的效应估计有所不同,与固定效应模型相比,样本量较小的研究在随机效应模型中得到的权重更高,由于将研究间的变异考虑在内,随机效应的置信区间更宽。不存在统计学异质性的情况下,随机效应模型和固定效应模型估计的结果相同。

在不同的观察性研究中,人群特征、暴露和结局

的定义往往不同。所有研究估计的真实效应都相同的假设极少成立,因此采用随机效应模型合并分析研究更合理^[18]。在这种情况下,需要重点考虑的因素是,对于给定的研究问题,较小或较大的研究是否具有更大的偏倚风险。在临床研究中,大型多中心试验往往比小型单中心研究具有更低的偏倚风险,倾向使用固定效应模型。而在病因学观察性研究中,较小的研究可能收集了更好的暴露和混杂的数据。选择的模型需要提前指定,但在敏感性分析中同时呈现 2 种模型的结果更有说服力。值得注意的是,尽管随机效应模型考虑了研究间的异质性,但不能帮助了解异质性的来源^[61]。

(2)研究间异质性的统计测量:统计学异质性的评价方法包括统计量 I^2 和 Cochrane's Q 检验。 Q 检验评价的是效应估计值之间的变异是否仅由随机造成的, I^2 可以量化不能被随机解释的变异^[62]。测量方法应谨慎解释:当纳入研究的数量较少时,统计量 I^2 具有不确定性^[63],Cochrane's Q 检验的检测效能有限^[64]。由于纳入观察性研究综述的数量通常为 10~20^[11],统计效能较低。此外,是否存在异质性的统计推断不需要与研究是否存在差异或存在偏倚风险的判断一致,可能重要的研究差异并没有转化为统计学异质性。

(3)漏斗图对称性:漏斗图是一种图形化的工具,用于判断小型研究和大型研究的估计值是否不同。根据标准误差或估计精度绘制效应值大小图形^[65]。如果不同研究的估计值仅仅因为随机变化而不同,那么将对称地围绕一个中心值分布,随着精度的增加,变异会减小。因此,该图形像一个倒置的漏斗。漏斗图的不对称性意味着研究规模与效应估计值存在关联或存在“小样本效应”^[66],小样本研究通常显示出较大的效应。可能的原因有真实的异质性(即小样本研究在研究人群、暴露水平等方面不同于大型研究)、选择偏倚(即选择性发表显示效应的小样本研究)、设计或分析的偏倚或偶然事件^[67-68]。不对称不等同于发表偏倚。特别是在观察性研究的背景下,其他来源的异质性也会影响漏斗图的不对称性。

(4)Meta 回归:用于探讨研究特征是否与效应大小相关,以及特定的研究特征是否可以解释观察到的统计学异质性。异质性的存在是进行 Meta 回归分析的原因,因此建议采用随机效应 Meta 回归。固定效应 Meta 回归在概念上是无意义的,可能产生较高的假阳性率^[69]。

纳入Meta回归模型的变量可能是研究特征,如研究设计、发表年份或偏倚风险,以及研究对象的特征,如年龄、性别、疾病进程等潜在的效应修正因子。应事先规定纳入Meta回归分析的变量(每10个研究纳入1个研究变量),减小假阳性结果的风险^[70]。当纳入研究对象特征时需注意,在研究层面观察到的关联可能不能反映个体水平的关联,即生态学谬误^[71]。

(5)合并不同类型指标:Meta分析取决于数据在单个研究中如何呈现。特别是在观察性研究中,研究者面临的问题是,根据研究设计和统计模型采用不同的指标来反映相同的暴露和结局的关联。例如研究吸烟(暴露)和高血压(结局)之间的关联,在队列研究中,结局是二分类变量时,可以采用风险比、发病率比、相对危险度、优势比等指标,当结局是连续变量时,则采用均值差或标化均值差。

当研究报告了不同统计指标,如风险比、相对危险度、优势比,需要考虑概念和技术层面的问题,例如是否忽略指标间差异进行合并分析,取决于纳入研究的设计类型有哪些(队列研究或病例对照研究)以及研究对象在病例对照研究中如何抽样^[72-73]。通常,病因学研究中的结局若是罕见的($<5\%$),不同的比率可以合并。对于非罕见结局,研究者必须慎重,优势比通常会高估相对危险度。

相对危险度与标化均值差或相关系数等其他指标也可以合并分析,目前已有方法和技术可以将优势比转换为标化均值差^[74]或者相关系数^[75-76]再合并。例如,在一项纤维蛋白原水平与术后失血的Meta分析中,研究报告了优势比、回归系数、相关系数^[77],可以将所有效应指标转化为相关系数,然后进行Meta分析^[77]。

(6)剂量反应Meta分析:在流行病学研究中,比较暴露的不同水平是常见做法。例如,研究血糖对心血管结局的影响,以其中一个组别作为参考,对不同的血糖水平组进行比较。然而,不同的研究可能报告不同类型的暴露变量(三分位数、四分位数或五分位数)。一种方法是对剂量最低和最高组比较的估计值进行Meta分析,但由于不同研究中剂量最低组和最高组的定义不同,因此不推荐。更为推荐的一种方法是模拟暴露与结局之间的关系,估计暴露增加一个单位,相关风险随之增加(或减少)的单位^[78-79]。例如,一项稳态模型评估胰岛素抵抗指数(HOMA-IR)和心血管事件相关性的Meta分析,使用剂量反应模型来估计HOMA-IR每增加一个标准差,心血管风险随之增加46%^[24]。

9. 结果解释和讨论:评价员应对结果进行全面的讨论:虽然总体估计值的置信区间很窄,但纳入的研究可能不够全面^[41],研究者应该谨记统计学显著性并不代表存在真实的关联。大效应值并不能弥补偏倚。如果纳入的研究有较低的偏倚风险,且异质性不大,研究人员可以得出结论,主要结果提供了合理真实的估计。如果研究具有较高的偏倚风险,研究人员应该得出结论,真实的效应仍然无法确定。证据推荐分级的评估、制订与评价(the grades of recommendation, assessment, development, and evaluation, GRADE)系统有助于规范判断“有多少把握认为效应估计值足以支持特定决策或推荐”^[80],同时考虑到研究设计、偏倚风险、结果的不一致性、不精确性、间接性以及报告偏倚^[81]。

一项或几项研究可能足以说明存在偏倚,并且其他研究也都存在这种偏倚。例如,多项队列研究表明,较高的C反应蛋白(C-reactive protein, CRP)水平与心血管风险相关,但其他心血管危险因素,包括吸烟、肥胖和体力活动,也与较高的CRP水平相关,可能混杂真实关联^[82]。然而,在孟德尔随机化研究中未发现任何关联^[82],该研究使用的遗传变异与CRP水平相关,但独立于行为或环境风险因素等流行病学研究中的混杂因素^[83]。究其原因,孟德尔随机化研究和经典队列研究估计了不同的效应:孟德尔随机化中是终身暴露,而队列研究中是特定(通常不明确)时间点之后的暴露。值得注意的是,当孟德尔随机化研究纳入了老年研究对象时,可能会出现选择偏倚^[84]。在解释流行病学研究的系统综述结果时,应考虑是否存在孟德尔随机化研究的证据,可以参考孟德尔随机化研究的指南^[85]。

在评价因果关系时,整合不同来源(如生态学研究、基础机制研究)的证据有助于得出最终结论,研究者应尽量获取基于不同分析方法和流行病学研究设计得到的结果,进行综合推断,因为每一种方法都有不同且独立的潜在偏倚来源,称之为三角测量^[86]。如果不同的方法都指向相同的结论,可以增加真实因果关系的可信度^[86]。例如,在关于吸烟和肺癌的讨论中,肺癌的时间趋势是一个重要的论据,来反驳遗传特征会像吸烟一样导致肺癌的假设^[87]。系统地讨论竞争性的危险因素可以增强对结果的解释^[88]。特别是在毒理学领域,机制证据在因果推论中起着关键作用,应对这方面的文献作系统综述,而不仅仅是快速检索几篇支持这一假说的文章^[17]。关于胰岛素样生长因子或肥胖与癌症风险的系统综述综合考

虑了实验室、动物和人群证据,来判断不同机制的合理性^[40,89]。

最后,应讨论研究结果在临床和公共卫生方面的重要性。病因的识别并不一定能转化为干预措施的建议^[90]。例如,流行病学和其他证据表明肥胖可能会增加几种癌症的风险^[89,91],但并不意味着减肥可以降低癌症风险。肥胖可能已经产生了危害,采用不同的干预措施进行减重对癌症风险影响也不同^[92]。

系统综述的优势在于,可以对某一领域进行清晰概述,并确定证据空白和需要进一步研究的类型之间的差距。因此,推荐采用详细的特定研究建议

替代“需要更多的研究”的通用说法。此外,在评估了研究的优点和局限性后,评价员可以指出在考虑未来研究时需要避免的陷阱。

三、案例解读

本部分内容以 de Souza 等^[93]发表在国际权威综合性医学杂志 *BMJ* 上的有关“饱和与反式不饱和脂肪酸摄入与全因死亡率、发生心血管疾病和 2 型糖尿病的风险:对观察性研究的系统综述和 Meta 分析”研究为例进行分析,见表 2。

这项研究旨在系统评价饱和脂肪酸和反式不饱和脂肪酸摄入与全因死亡率、心血管疾病及相关死

表 2 COSMOS-E 案例解读

内 容	解 读
筹备系统综述	
组建综述团队	作者在志谢部分列出了参与系统综述的团队成员,包括信息专家、统计专家、流行病学专家在内的专业人员负责制定研究方案、制定和实施检索策略以及统计分析。
形成研究问题	在前言部分,作者首先介绍最新发表的系统综述的研究结果与现行的饮食推荐不一致,建议重新评估饱和脂肪酸摄入的影响。美国最近实施针对反式脂肪酸的政策行动,从食品供应中去除了部分氢化植物油,但系统综述研究结果和膳食指南关于反式脂肪酸摄入建议存在争议。围绕目前的争议主题,扩展和更新研究结果,因此开展本研究进一步综述和评估饱和脂肪酸和反式脂肪酸与健康结局之间的关联。
定义人群、暴露、比较和结局	作者在方法部分明确说明研究人群定义为一般人群,未加以限定。暴露定义为自我报告或生物标记物测量的饱和脂肪酸或反式脂肪酸摄入。动物反式脂肪酸通常是采用已知的乳制品和肉制品食品表中的营养成分构成测量。结局定义为通过自我报告或医疗记录、注册网站获得的全因死亡率、冠心病、卒中和 2 型糖尿病。
考虑混杂因素和偏倚	作者在方法部分描述了可能的混杂因素有吸烟和年龄,针对暴露的定义,作者认为采用前瞻性记录或 24 h 回忆食物频率问卷、在亚人群中直接测量脂肪组织反式脂肪酸含量、采用更新的食物数据库分析膳食摄入三种方法评估暴露测量的研究发生错分偏倚的风险最小,如果采用了上述三种方法中的两种则认为是中等风险,如果只采用一种方法则认为是高风险,如果没有采用任何一种测量方法则认为是非常高风险。
研究方案	作者在基金部分提到方案经过 WHO NUGAG Subgroup on Diet and Health 同意,但并未在 PROSPERO 或其他平台注册。
检索相关研究	本研究选择的电子数据库有 Medline、Embase、Cochrane Central Registry of Controlled Trials、Evidence Based Medicine Reviews 和 CINAHL,手动检索系统综述和描述性综述的参考文献。以 Ovid Medline 为例,观察性研究设计相关的检索词包括“epidemiology studies”“case-control studies”“control groups”“matched-pair analysis”“cohort”“cohort studies”“longitudinal”“prospective”“retrospective”“case* adj5 control*”“case adj3 comparison*”“control group*”。无语言限制。
研究筛选	作者描述了合格性标准和筛选过程,绘制了研究筛选流程图。纳入的研究类型为在人群中开展的观察性研究(如前瞻性队列研究、病例对照研究、巢式病例对照或病例队列设计)。本研究未将特定的研究设计和方法学质量列为排除标准。
数据提取	作者说明了双人独立提取研究信息,包括研究设计、样本量、暴露和结局评估、研究对象特征和统计分析、混杂因素调整。
提取效应估计值和标准误差	作者在附录 1 中列出了,提取的信息包括最大限度调整和最小程度调整的效应估计值 <i>RR</i> 和测量精度(置信区间)或 <i>P</i> 值。
评价质量和偏倚	作者在方法部分说明了采用 Newcastle-Ottawa 量表评估纳入研究的偏倚风险。但未具体说明每个条目的评价标准。
探讨异质性	作者在分析数据部分说明采用统计量 <i>I²</i> 、Cochrane’s <i>Q</i> 检验、Meta 回归评价异质性。
Meta 分析:是否合并分析?	作者针对每个结局指标的合并分析,均说明了不纳入个别研究进行定量合并的原因。
统计学分析	
观察性研究中的固定效应模型和随机效应模型	文章中明确说明至少有 2 项研究提供了数据,则进行随机效应 Meta 分析,如果有 3 项或更少的研究进行合并分析,也会同时考虑固定效应估计。
研究间异质性的统计测量	本研究采用统计量 <i>I²</i> 和 Cochrane’s <i>Q</i> 检验评价异质性,并规定 Cochrane’s <i>Q</i> 检验 <i>P</i> < 0.10,并且 <i>I²</i> > 60% 则认为异质性显著。
漏斗图对称性	作者在数据分析部分说明,10 项及以上研究则通过漏斗图、Egger’s 和 Begg’s 检验法检查是否存在发表偏倚(显著水平 <i>P</i> < 0.1)。
Meta 回归	文章中明确说明,如果有 10 项及以上的研究,则采用 Meta 回归探索异质性来源,考虑因素包括基线研究年份、研究地域、随访时间、受试者中位年龄、吸烟者比例、参考类别中饱和脂肪量、人群的平均饱和或反式脂肪酸摄入量、性别、 α -亚麻酸、多不饱和脂肪酸总量、调整总热量、暴露测量的方法和频率、偏倚风险得分、调整血脂或血压(因果中间变量)。
合并不同类型指标	作者在方法学部分明确说明,正文中呈现的结果仅限于合并分析前瞻性队列研究的结局指标。附录 4 中部分森林图(e figure 51-53)呈现了合并分析队列研究和病例对照研究的结果。
剂量反应 Meta 分析	作者在局限性部分说明了由于队列数量限制,无法进行剂量反应 Meta 分析。
结果解释和讨论	作者对结果进行了充分全面的讨论,不仅介绍了证据级别较高的前瞻性队列研究的合并分析结果,还介绍了由于指标类型无法转换合并的其他研究的结果,结果不一致的情况下,作者的结论较为保守和慎重;作者讨论了根据偏倚风险大小进行敏感性分析的结果;作者将研究发现与既往的随机对照研究和病例对照研究结果进行对比讨论,并且整合基础研究的结果综合推断研究发现的机制合理性;作者说明了本文的发现仅限于描述关联,无法提供因果关系推断;作者在文末说明了研究结果的意义、未解决的问题以及未来相关研究需要解决的问题。

亡率、冠心病及相关死亡率、缺血性脑卒中和2型糖尿病5个结局的关联。

总体而言,该研究涵盖了病因学观察性研究系统综述的关键要素,条理清楚,分析方法合理,研究结果丰富翔实,讨论全面深入。但文章中对于评价质量和偏倚的内容较少。由于该研究的方案未发表,仅根据文章报告的内容对于判断系统综述的设计阶段考虑的是否全面是不充分的。

COSMOS-E的发表,将有助于提高研究人员对病因学观察性研究系统综述的认识,指导系统综述作者认识并解决病因学观察性研究系统综述设计中的关键问题,提高研究设计的科学性,增强证据的可信度。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study [J]. *PLoS Med*, 2016, 13 (5) : e1002028. DOI: 10.1371/journal.pmed.1002028.
- [2] Mansournia MA, Higgins JPT, Sterne JA, et al. Biases in Randomized Trials: A Conversation Between Trialists and Epidemiologists [J]. *Epidemiology*, 2017, 28 (1) : 54–59. DOI: 10.1097/ede.0000000000000564.
- [3] Vandembroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration [J]. *PLoS Med*, 2007, 4 (10) : e297. DOI: 10.1371/journal.pmed.0040297.
- [4] Dekkers OM, Horváth-Puhó E, Jørgensen JOL, et al. Multisystem morbidity and mortality in Cushing's syndrome: a cohort study [J]. *J Clin Endocrinol Metab*, 2013, 98 (6) : 2277–2284. DOI: 10.1210/jc.2012–3582.
- [5] Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? [J]. *Epidemiology*, 2006, 17 (4) : 360–372. DOI: 10.1097/01.ede.0000222409.00878.37.
- [6] Rassen JA, Brookhart MA, Glynn RJ, et al. Instrumental variables I : instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships [J]. *J Clin Epidemiol*, 2009, 62 (12) : 1226–1232. DOI: 10.1016/j.jclinepi.2008.12.005.
- [7] Mountjoy E, Davies NM, Plotnikov D, et al. Education and myopia: assessing the direction of causality by mendelian randomisation [J]. *BMJ*, 2018, 361k2022. DOI: 10.1136/bmj.k2022.
- [8] Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs [J]. *BMJ*, 2016, 354i4515. DOI: 10.1136/bmj.i4515.
- [9] Ponjoan A, Blanch J, Alves-Cabreros L, et al. Effects of extreme temperatures on cardiovascular emergency hospitalizations in a Mediterranean region: a self-controlled case series study [J]. *Environ Health*, 2017, 16 (1) : 32. DOI: 10.1186/s12940–017–0238–0.
- [10] Coureau G, Bouvier G, Lebaillly P, et al. Mobile phone use and brain tumours in the CERENAT case-control study [J]. *Occup Environ Med*, 2014, 71(7) : 514–522. DOI: 10.1136/oemed–2013–101754.
- [11] Mason KE, Pearce N, Cummins S. Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank [J]. *Lancet Public Health*, 2018, 3 (1) : e24–33. DOI: 10.1016/s2468–2667(17)30212–8.
- [12] Moses S, Bradley JE, Nagelkerke NJ, et al. Geographical patterns of male circumcision practices in Africa: association with HIV seroprevalence [J]. *Int J Epidemiol*, 1990, 19(3) : 693–697. DOI: 10.1093/ije/19.3.693.
- [13] Siegfried N, Muller M, Deeks JJ, et al. Male circumcision for prevention of heterosexual acquisition of HIV in men [J]. *Cochrane Database Syst Rev*, 2009(2) : Cd003362. DOI: 10.1002/14651858.CD003362.pub2.
- [14] Higgins J, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.1 (updated September 2020) [J]. Cochrane, 2020. Available from www.training.cochrane.org/handbook.
- [15] 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.
- [16] Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group [J]. *JAMA*, 2000, 283 (15) : 2008–2012. DOI: 10.1001/jama.283.15.2008.
- [17] Hoffmann S, de Vries RBM, Stephens ML, et al. A primer on systematic reviews in toxicology [J]. *Arch Toxicol*, 2017, 91 (7) : 2551–2575. DOI: 10.1007/s00204–017–1980–3.
- [18] Mueller M, D'addario M, Egger M, et al. Methods to systematically review and Meta-analyse observational studies: a systematic scoping review of recommendations [J]. *BMC Med Res Methodol*, 2018, 18: 44. DOI: 10.1186/s12874–018–0495–9.
- [19] 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 (Pt 1) : 1027–1031. DOI: 10.1016/j.envint.2018.07.015.
- [20] Dekkers OM, von Elm E, Algra A, et al. How to assess the external validity of therapeutic trials: a conceptual approach [J]. *Int J Epidemiol*, 2010, 39(1) : 89–94. DOI: 10.1093/ije/dyp174.
- [21] Burgers AM, Biermasz NR, Schoones JW, et al. Meta-analysis and dose-response metaregression: circulating insulin-like growth factor I (IGF- I) and mortality [J]. *J Clin Endocrinol Metab*, 2011, 96(9) : 2912–2920. DOI: 10.1210/jc.2011–1377.
- [22] Amitay EL, Keinan-Boker L. Breastfeeding and Childhood Leukemia Incidence: A Meta-analysis and Systematic Review [J]. *JAMA Pediatr*, 2015, 169 (6) : e151025. DOI: 10.1001/jamapediatrics.2015.1025.
- [23] Cooper GS, Lunn RM, Agerstrand M, et al. Study sensitivity: Evaluating the ability to detect effects in systematic reviews of chemical exposures [J]. *Environ Int*, 2016, 92–93605–610. DOI: 10.1016/j.envint.2016.03.017.
- [24] Gast KB, Tjeerdema N, Stijnen T, et al. Insulin resistance and risk of incident cardiovascular events in adults without diabetes: Meta-analysis [J]. *PLoS One*, 2012, 7 (12) : e52036. DOI: 10.1371/journal.pone.0052036.
- [25] Vandembroucke JP. When are observational studies as credible as randomised trials? [J]. *Lancet*, 2004, 363 (9422) : 1728–1731. DOI: 10.1016/s0140–6736(04)16261–2.

- [26] 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.
- [27] Booth A, Clarke M, Dooley G, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews[J]. *Syst Rev*, 2012, 12. DOI: 10.1186/2046-4053-1-2.
- [28] McGowan J, Sampson M. Systematic reviews need systematic searchers[J]. *J Med Libr Assoc*, 2005, 93(1): 74–80.
- [29] Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources[J]. *BMJ*, 2005, 331(7524): 1064–1065. DOI: 10.1136/bmj.38636.593461.68.
- [30] Kuper H, Nicholson A, Hemingway H. Searching for observational studies: what does citation tracking add to PubMed? A case study in depression and coronary heart disease[J]. *BMC Med Res Methodol*, 2006, 64. DOI: 10.1186/1471-2288-6-4.
- [31] Limeshow AR, Blum RE, Berlin JA, et al. Searching one or two databases was insufficient for Meta-analysis of observational studies[J]. *J Clin Epidemiol*, 2005, 58(9): 867–873. DOI: 10.1016/j.jclinepi.2005.03.004.
- [32] Waffenschmidt S, Hermanns T, Gerber-Grote A, et al. No suitable precise or optimized epidemiologic search filters were available for bibliographic databases[J]. *J Clin Epidemiol*, 2017, 82112–82118. DOI: 10.1016/j.jclinepi.2016.08.008.
- [33] von Elm E, Poglia G, Walder B, et al. Different patterns of duplicate publication: an analysis of articles used in systematic reviews[J]. *JAMA*, 2004, 291(8): 974–980. DOI: 10.1001/jama.291.8.974.
- [34] Berlin JA. Invited commentary: benefits of heterogeneity in Meta-analysis of data from epidemiologic studies[J]. *Am J Epidemiol*, 1995, 142(4): 383–387. DOI: 10.1093/oxfordjournals.aje.a117645.
- [35] Tendal B, Higgins JPT, Juni P, et al. Disagreements in Meta-analyses using outcomes measured on continuous or rating scales: observer agreement study[J]. *BMJ*, 2009, 339b3128. DOI: 10.1136/bmj.b3128.
- [36] Gotzsche PC, Hrobjartsson A, Maric K, et al. Data extraction errors in Meta-analyses that use standardized mean differences[J]. *JAMA*, 2007, 298(4): 430–437. DOI: 10.1001/jama.298.4.430.
- [37] Rohner EBJ, da Costa RR, Trelle S. Managing people and data in systematic review [M] // M Egger, Davey Smith G. *Systematic Reviews in Health Care: Meta-analysis in Context* Chichester. John Wiley & Sons; England.
- [38] Grimes DA. “Case-control” confusion: mislabeled reports in obstetrics and gynecology journals[J]. *Obstet Gynecol*, 2009, 114(6): 1284–1286. DOI: 10.1097/AOG.0b013e3181c03421.
- [39] Nesvick CL, Thompson CJ, Boop FA, et al. Case-control studies in neurosurgery[J]. *J Neurosurg*, 2014, 121(2): 285–296. DOI: 10.3171/2014.5.jns132329.
- [40] Renehan AG, Zwahlen M, Minder C, et al. Insulin-like growth factor (IGF) – I, IGF binding protein-3, and cancer risk: systematic review and Meta-regression analysis[J]. *Lancet*, 2004, 363(9418): 1346–1353. DOI: 10.1016/s0140-6736(04)16044-3.
- [41] Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies[J]. *BMJ*, 1998, 316(7125): 140–144. DOI: 10.1136/bmj.316.7125.140.
- [42] Friedenreich CM, Speidel TP, Neilson HK, et al. Case-control study of lifetime alcohol consumption and endometrial cancer risk[J]. *Cancer Causes Control*, 2013, 24(11): 1995–2003. DOI: 10.1007/s10552-013-0275-0.
- [43] Tourangeau R, Yan T. Sensitive questions in surveys[J]. *Psychol Bull*, 2007, 133(5): 859–883. DOI: 10.1037/0033-2909.133.5.859.
- [44] Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias[J]. *Epidemiology*, 2004, 15(5): 615–625.
- [45] Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography[J]. *Int J Epidemiol*, 2007, 36(3): 666–676. DOI: 10.1093/ije/dym018.
- [46] Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials[J]. *BMJ*, 2011, 343d5928. DOI: 10.1136/bmj.d5928.
- [47] 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, 355i4919. DOI: 10.1136/bmj.i4919.
- [48] Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies[J]. *Ann Intern Med*, 2011, 155(8): 529–536. DOI: 10.7326/0003-4819-155-8-201110180-00009.
- [49] 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.
- [50] da Costa BR, Hifiker R, Egger M. PEDro’s bias: summary quality scores should not be used in Meta-analysis[J]. *J Clin Epidemiol*, 2013, 66(1): 75–77. DOI: 10.1016/j.jclinepi.2012.08.003.
- [51] Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for Meta-analysis[J]. *JAMA*, 1999, 282(11): 1054–1060. DOI: 10.1001/jama.282.11.1054.
- [52] Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in Meta-analyses[J]. *Eur J Epidemiol*, 2010, 25(9): 603–605. DOI: 10.1007/s10654-010-9491-z.
- [53] Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy[J]. *N Engl J Med*, 2008, 358(3): 252–260. DOI: 10.1056/NEJMs065779.
- [54] Roest AM, de Jonge P, Williams CD, et al. Reporting Bias in Clinical Trials Investigating the Efficacy of Second-Generation Antidepressants in the Treatment of Anxiety Disorders: A Report of 2 Meta-analyses[J]. *JAMA Psychiatry*, 2015, 72(5): 500–510. DOI: 10.1001/jamapsychiatry.2015.15.
- [55] Head ML, Holman L, Lanfear R, et al. The extent and consequences of p-hacking in science[J]. *PLoS Biol*, 2015, 13(3): e1002106. DOI: 10.1371/journal.pbio.1002106.
- [56] Althuis MD, Weed DL, Frankenfeld CL. Evidence-based mapping of design heterogeneity prior to Meta-analysis: a systematic review and evidence synthesis[J]. *Syst Rev*, 2014, 380. DOI: 10.1186/2046-4053-3-80.
- [57] Davey Smith G, Egger M, Phillips AN. Meta-analysis. Beyond the grand mean? [J]. *BMJ*, 1997, 315(7122): 1610–1614. DOI: 10.1136/bmj.315.7122.1610.
- [58] Feller M, Huwiler K, Stephan R, et al. *Mycobacterium avium* subspecies paratuberculosis and Crohn’s disease: a systematic review and Meta-analysis[J]. *Lancet Infect Dis*, 2007, 7(9): 607–613. DOI: 10.1016/s1473-3099(07)70211-6.
- [59] Ioannidis JP, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding Meta-analysis in forest plots[J]. *BMJ*, 2008, 336

- (7658):1413–1415. DOI:10.1136/bmj.a117.
- [60] Borenstein M, Hedges LV, Higgins JP, et al. A basic introduction to fixed-effect and random-effects models for Meta-analysis[J]. *Res Synth Methods*, 2010, 1(2):97–111. DOI:10.1002/jrsm.12.
- [61] Greenland S. Invited commentary: a critical look at some popular Meta-analytic methods[J]. *Am J Epidemiol*, 1994, 140(3):290–296. DOI:10.1093/oxfordjournals.aje.a117248.
- [62] Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in Meta-analyses[J]. *BMJ*, 2003, 327(7414):557–560. DOI:10.1136/bmj.327.7414.557.
- [63] Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in Meta-analyses[J]. *BMJ*, 2007, 335(7626):914–916. DOI:10.1136/bmj.39343.408449.80.
- [64] Takkouche B, Cadarso-Suarez C, Spiegelman D. Evaluation of old and new tests of heterogeneity in epidemiologic Meta-analysis[J]. *Am J Epidemiol*, 1999, 150(2):206–215. DOI:10.1093/oxfordjournals.aje.a009981.
- [65] Sterne JA, Egger M. Funnel plots for detecting bias in Meta-analysis: guidelines on choice of axis[J]. *J Clin Epidemiol*, 2001, 54(10):1046–1055.
- [66] Sterne JA, Gavaghan D, Egger M. Publication and related bias in Meta-analysis: power of statistical tests and prevalence in the literature[J]. *J Clin Epidemiol*, 2000, 53(11):1119–1129.
- [67] Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in Meta-analyses of randomised controlled trials[J]. *BMJ*, 2011, 343d4002. DOI:10.1136/bmj.d4002.
- [68] Egger M, Davey Smith G, Schneider M, et al. Bias in Meta-analysis detected by a simple, graphical test[J]. *BMJ*, 1997, 315(7109):629–634. DOI:10.1136/bmj.315.7109.629.
- [69] Higgins JP, Thompson SG. Controlling the risk of spurious findings from Meta-regression[J]. *Stat Med*, 2004, 23(11):1663–1682. DOI:10.1002/sim.1752.
- [70] Thompson SG, Higgins JP. How should Meta-regression analyses be undertaken and interpreted? [J]. *Stat Med*, 2002, 21(11):1559–1573. DOI:10.1002/sim.1187.
- [71] Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting [J]. *BMJ*, 2010, 340c221. DOI:10.1136/bmj.c221.
- [72] Vandenbroucke JP, Pearce N. Case-control studies: basic concepts[J]. *Int J Epidemiol*, 2012, 41(5):1480–1489. DOI:10.1093/ije/dys147.
- [73] Knol MJ, Vandenbroucke JP, Scott P, et al. What do case-control studies estimate? Survey of methods and assumptions in published case-control research[J]. *Am J Epidemiol*, 2008, 168(9):1073–1081. DOI:10.1093/aje/kwn217.
- [74] Chinn S. A simple method for converting an odds ratio to effect size for use in Meta-analysis[J]. *Stat Med*, 2000, 19(22):3127–3131.
- [75] Cleophas TJ, Zwinderman AH. Transforming Odds Ratios into Correlation Coefficients [M]. *Modern Meta-Analysis* Springer: Cham, 2017.
- [76] da Costa BR, Rutjes AW, Johnston BC, et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: Meta-epidemiological study[J]. *Int J Epidemiol*, 2012, 41(5):1445–1459. DOI:10.1093/ije/dys124.
- [77] Gielen C, Dekkers O, Stijnen T, et al. The effects of pre- and postoperative fibrinogen levels on blood loss after cardiac surgery: a systematic review and Meta-analysis[J]. *Interact Cardiovasc Thorac Surg*, 2014, 18(3):292–298. DOI:10.1093/icvts/ivt506.
- [78] Hartemink N, Boshuizen HC, Nagelkerke NJ, et al. Combining risk estimates from observational studies with different exposure cutpoints: a Meta-analysis on body mass index and diabetes type 2 [J]. *Am J Epidemiol*, 2006, 163(11):1042–1052. DOI:10.1093/aje/kwj141.
- [79] Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to Meta-analysis[J]. *Am J Epidemiol*, 1992, 135(11):1301–1309. DOI:10.1093/oxfordjournals.aje.a116237.
- [80] Balshem H, Helfand M, Schunemann 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.
- [81] Morgan RL, Thayer KA, Bero L, et al. GRADE: Assessing the quality of evidence in environmental and occupational health[J]. *Environ Int*, 2016, 92–93611–616. DOI:10.1016/j.envint.2016.01.004.
- [82] Wensley F, Gao P, Burgess S, et al. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data[J]. *BMJ*, 2011, 342d548. DOI:10.1136/bmj.d548.
- [83] Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations [J]. *Int J Epidemiol*, 2004, 33(1):30–42. DOI:10.1093/ije/dyh132.
- [84] Boef AG, Le Cessie S, Dekkers OM. Mendelian randomization studies in the elderly [J]. *Epidemiology*, 2015, 26(2):e15–16. DOI:10.1097/ede.0000000000000243.
- [85] Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians[J]. *BMJ*, 2018, 362k601. DOI:10.1136/bmj.k601.
- [86] Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology [J]. *Int J Epidemiol*, 2016, 45(6):1866–1886. DOI:10.1093/ije/dyw314.
- [87] Vandenbroucke JP. Commentary: ‘Smoking and lung cancer’—the embryogenesis of modern epidemiology [J]. *Int J Epidemiol*, 2009, 38(5):1193–1196. DOI:10.1093/ije/dyp292.
- [88] Maclure M. Demonstration of deductive Meta-analysis: ethanol intake and risk of myocardial infarction [J]. *Epidemiol Rev*, 1993, 15(2):328–351. DOI:10.1093/oxfordjournals.epirev.a036124.
- [89] Renehan AG, Zwahlen M, Egger M. Adiposity and cancer risk: new mechanistic insights from epidemiology [J]. *Nat Rev Cancer*, 2015, 15(8):484–498. DOI:10.1038/nrc3967.
- [90] Greenland S. Epidemiologic measures and policy formulation: lessons from potential outcomes[J]. *Emerg Themes Epidemiol*, 2005, 25. DOI:10.1186/1742–7622–2–5.
- [91] Renehan AG, Tyson M, Egger M, et al. Body-mass index and incidence of cancer: a systematic review and Meta-analysis of prospective observational studies[J]. *Lancet*, 2008, 371(9612):569–578. DOI:10.1016/s0140–6736(08)60269–x.
- [92] Hernan MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions[J]. *Int J Obes (Lond)*, 2008, 32 Suppl 3:S8–14. DOI:10.1038/ijo.2008.82.
- [93] de Souza RJ, Mente A, Maroleanu A, et al. Intake of saturated and trans unsaturated fatty acids and risk of all cause mortality, cardiovascular disease, and type 2 diabetes: systematic review and Meta-analysis of observational studies [J]. *BMJ*, 2015, 351h3978. DOI:10.1136/bmj.h3978.

(收稿日期:2019–10–24)

(本文编辑:李银鸽)