

如何撰写高质量的流行病学研究论文

第一讲 观察性流行病学研究报告规范——STROBE 介绍

王波 詹思延

【导读】 观察性研究是流行病学研究的重要组成部分,主要用于探索和检验疾病与暴露之间的因果关联。观察性研究报告应当提供评价研究潜在偏倚和研究结论适用性的重要信息。不完整、不规范的报告会限制对研究有效性的评价。本文分析了影响观察性流行病学研究有效性的主要因素,在此基础上特别介绍了国际上为改善观察性研究论文的质量而建立的报告规范——加强观察性流行病学研究报告的质量(STROBE)的制定过程及清单内容。

【关键词】 观察性研究;质量;偏倚

加强观察性流行病学研究报告的质量(strengthening the reporting of observational studies in epidemiology, STROBE)是由流行病学家、方法学家、统计学家、研究者和编辑组成的一个国际性合作小组共同起草,主要目的是为观察性流行病学研究论文提供报告规范,从而改进这类研究报告的质量。

1. 影响观察性流行病学研究有效性的主要因素:在评价任何一个观察性流行病学研究结果时,主要的问题是判断观察到的暴露与疾病之间的关联是否有因果关系,即这种关联是否由偏倚、随机误差或混杂所造成。选择研究对象的途径(选择偏倚)或信息获得的方式(信息偏倚)都有可能造成偏倚。当研究样本比较小的时候,随机误差就会产生重要的作用。同时,一个观察到的关联还可能是由于混杂因素的分布在比较组之间不一致且未进行统计学调整造成的。此外,还需要判断研究结果的适用性。

影响观察性流行病学研究有效性的主要因素及当前报告中存在的问题如下^[1]:

(1) 研究设计的选择:研究设计的选择直接影响着研究结果能够在多大程度上有效地回答研究问题。观察性研究的主要设计类型是病例对照研究、队列研究和横断面研究。研究设计的选择受诸多因素影响,如疾病/暴露是否罕见,时间顺序是否已明确,是探索还是检验病因假设,是研究一种疾病的多种可能危险因素还是一种暴露的多种可能后果,可以使用的资源有多少(时间、经费和信息)等。因此,高质量的研究报告应当阐明选择研究设计的理由,以便读者能够判断这种选择是否合理。

(2) 研究人群的定义和研究对象的选择:研究人群和研究对象的选择直接影响研究结果的代表性、适用性以及选择偏倚和随机误差的作用。一个研究的代表性取决于源人群

和研究对象对源人群的代表性^[2]。在研究报告中区别研究人群(目标人群)、样本人群(即目标人群中被选择纳入研究的人群)和分析人群(样本人群中可以获得数据从而纳入分析的人群)有助于读者评价结果的适用性,即研究结果对哪个人群有效及可被推广的程度。

观察性研究常常基于暴露组与非暴露组或疾病组与非疾病组之间的比较。为了获得有效的结果,正确选择比较组是很重要的^[3]。以病例对照研究为例,病例组的暴露状况应与除不患有目标疾病以外其他方面均与病例组相似的人群的暴露状况进行比较。理论上,通过对源人群进行随机抽样可以获得一个最好的对照组,但在实际的研究中很难实现。这样一来,判断样本人群的选择是否引起偏倚就成为一个重要问题。读者需要知道研究人群和样本人群在年龄分布、居住地、居住时间、种族和民族等方面的详细描述,以及样本人群和分析人群的数量,从而能够确定样本人群是否为研究人群的一个随机样本,即研究人群里的所有个体具有同样的概率被选择进入样本人群。

研究对象较少参与或不应答,可能威胁到研究结果的有效性,即无应答偏倚(non-respondent bias)。一个高质量的研究性研究报告应当描述研究对象的应答情况,但 Slattery 等^[4]注意到在已发表的研究报告中判断应答情况很难,因为各个研究在如何报告应答水平方面一致性很差,在应答的报告方面还没有一个统一的标准^[5]。Olson^[6]对病例对照研究的应答状况进行的综述也证实了这一点。

观察性研究需要谨慎地确定样本量。样本量应该足够大以使研究能够发现重要的关联,但又不能过于大以至超过所需要的。读者可以根据把握度和实际理由(如无应答率、亚组分析等)判断样本大小是否合适。

(3) 暴露或结局的测量:数据收集过程中,暴露或结局测量工具的有效性可能会造成不同程度的信息偏倚,这种偏倚是流行病学研究结果发生变异的一个重要来源。在病例对照研究中,如果病例组和对照组使用同一个测量工具,则称

作者单位:100083 北京大学医学部公共卫生学院流行病与卫生统计学系

通讯作者:詹思延, Email:siyan-zhan@bjmu.edu.cn

为非差异性错分偏倚。这种情况下,估计的关联值向无效应的方向偏倚。如果病例组和对照组的错分是不同的,则偏倚的效应可能是夸大或缩小真实效应,甚至估计的关联与真实效应方向相反。如在病例组使用问卷进行暴露状况评价,而在对照组通过电话访问的方式。在队列研究中,如果一组使用的是死亡证明而另一组使用健康保险资料,同样可能出现差异性错分。通常,如果数据收集方法已经在预试验里进行了验证,或者已经有研究对测量方法的有效性进行了评价,那么数据收集的质量会得到改善。因此,为了评价错分是否发生以及由此导致的偏倚及其类型,读者需要知道数据收集方法在两个比较组间是否一致,以及这种方法的有效性研究结果等信息。

在处理计量暴露时,研究者通常采用选择切点(cutoff)对暴露进行分组。读者常常需要知道这种情况下是如何将计量资料转化为分类变量资料(通常是等级资料变量)的,以及这种分类的依据。这对研究结果的解释以及研究之间的比较具有重要的意义。然而,Pocock等^[7]在一些主流期刊中进行的一项调查表明,42篇研究报告中有20篇没有对计量暴露的分类情况以及分类依据进行报告和解释。

(4)统计学分析:统计学分析提供研究的主要结果,并有助于判断混杂对研究结果的影响。首先需要数据的详细描述,包括应答率、入选研究人群数量和纳入分析人群数量;之后需要反映疾病与暴露之间关联的分析和根据可能的混杂变量分层分析的结果,其他统计模型的分析结果以及结果中由机会可能造成的差异。

混杂可能是影响观察性流行病学研究结果真实性最重要的原因^[8]。在混杂变量的控制方面,需要向读者交代潜在的混杂变量是如何选择的(如事先确定、事后选择或统计分析发现的)。在效应修正的报告方面,应使读者知道这种效应的“异质性”是否经过了统计学的检验。Pocock等^[7]的调查发现,67篇对潜在混杂因素进行控制的研究报告中,大多未提及潜在混杂因素是如何选择的。在34篇声称发现了可能的效应修正的报告里,只有8篇进行了交互作用的统计学检验。

此外,观察性研究常常探讨几个暴露、结局或亚组存在的关联,可能导致对假设的多重检验和假阳性结果概率增多。在许多研究中^[9],有20%或者更多的结果可能有错误,远远超过了预期5%的假阳性概率。应向读者表明研究者对多重检验可能导致的问题。

(5)讨论部分:该部分集中体现作者的研究结果与现有观点、已有的研究结果是否一致,研究可能的偏倚和设计、数据收集及统计分析方面的缺陷等,这有助于读者对该研究的偏倚、随机误差、混杂及结果的适用性做出最终评价。另一方面,伦理学问题在流行病学研究中日益受到关注^[10]。可见,观察性流行病学研究资料不完整妨碍了对研究结果的严格评价与合理解释,并且使得循证医疗卫生和预防实践更为复杂。对研究质量的恰当评价,是进行干预研究和观察性研究系统综述及Meta分析的重要步骤,如果“原始材料”就有

缺陷,那么系统综述结论的真实性也会受到影响。然而,当前的调查表明^[7],即使在主流的医学期刊上,一些观察性流行病学研究报告仍然存在着诸多的缺陷和不完整,从而导致可能错误的结论。因此,观察性流行病学研究必须完整、详细的报告与研究结果、偏倚、适用性等密切相关的重要内容,使读者能够对研究的内部和外部有效性进行评价。这就需要规范观察性研究报告。

2. STROBE 的制定过程:

2003年,一些流行病学家、方法学家和医学统计学家提出STROBE项目^[11]。他们通过检索数据库、查找参考文献等进行相关文献的汇总,在此基础上形成一个可能纳入观察性研究报告的条目表(草案清单)。这些条目在队列研究、病例对照研究或横断面研究中是重要的成分。2004年9月在英国Bristol大学召开第一次国际会议。会议的主要内容就是对草案清单中的每个条目进行讨论以达成共识。同年12月在伦敦,根据对第1版的意见和建议对草案清单进行修改,并在此基础上形成第2版清单。同样起草第3版时又考虑了进一步的评论和意见。由于观察性流行病学研究常常包含数种研究设计和诸多的主题,因此建议将STROBE限定在三种研究设计(队列研究、病例对照研究和横断面研究),并制定出一种通用的格式。以后可以进一步扩展到其他的研究设计,如巢式病例对照研究(nested case-control study)和专门的主题领域,如遗传分子流行病学。同时,STROBE的制定还是一个不断更新的过程,工作小组会根据评论、评价和新证据定期推出更新的建议版本。

3. STROBE 清单内容:

STROBE清单所涉及的条目分为题目和摘要、前言、方法、结果和讨论5个部分,共计22个条目。其中题目和摘要1个条目,前言部分包括背景/原理和目标2个条目;方法部分包括研究设计、研究现场、研究对象、研究变量、测量、偏倚、样本大小、统计方法、计量变量和资助情况等10个条目;结果部分包括研究对象、描述性资料、结局资料、主要结果和其他分析等5个条目;讨论部分包括重要结果、局限性、可推广性和解释等4个条目。所有条目针对三种主要的观察性研究类型(病例对照研究、队列研究和横断面研究)提出了各自的报告要求,其中通用的要求进行了合并,无法合并的则各自表述(表1)。

目前已有两家流行病学期刊(Epidemiology and the Bulletin of the World Health Organization)在给作者的建议里推荐使用STROBE清单。国际医学期刊编辑委员会已把STROBE列为“生物医学期刊投稿的统一要求”。关于STROBE的更多背景和详细内容可以从<http://www.strobe-statement.org/>获得。

毫无疑问,随着STROBE的不断改进与广泛应用,特别是STROBE清单解释与说明文件的推出,观察性流行病学研究的报告将进一步全面、完善,从而改善观察性研究实施的质量和结果评价,最终导致更科学的医疗卫生和预防实践。

表1 STROBE 声明:必需项目清单 第3版 (2005年9月)

条目	队列研究	病例对照研究	横断面研究
题目和摘要	1 ①在题目或摘要中有“队列研究” ②摘要应当是全文的一个内容丰富、结构化的摘要,包括了清单里的重要项目	①在题目或摘要中有“病例对照研究”	①在题目或摘要中有“横断面研究”
前言			
背景/原理	2 对所报告的研究背景和原理进行解释		
目标	3 阐明研究目标,包括任何预先确定的假设		
方法			
研究设计	4 陈述研究设计中的重要内容,如果文章是来自正在进行研究的系列文章之一,应陈述原始研究的目的		
研究现场	5 描述研究现场,数据收集的具体场所和时间范围		
研究对象	6 ①描述纳入和排除标准,研究对象的来源和选择方法 ②描述随访的时间范围和方法	①分别给出病例和对照的纳入和排除标准,来源和选择方法 ②给出精确的病例诊断标准和对照选择的原理 ③对匹配研究,应描述匹配标准和每个病例匹配的对照数	描述纳入和排除标准,研究对象的来源和选择方法
研究变量	7 对所有感兴趣的研究变量列出明确定义,并区分结局、暴露、潜在预测因子、潜在的混杂因子或效应修正因子		
测量	8* 对每个研究变量,描述详细的测量方法,还应描述各组之间测量方法的可比性		
偏倚	9 对可能的潜在偏倚进行描述		
样本大小	10 描述决定样本大小的原理,包括统计学计算和实际考虑		
统计学方法	11 ①描述统计方法,包括控制混杂的方法 ②描述对失访和缺失值的处理 ③如果可能,应描述亚组分析和敏感性分析的方法	②描述匹配和缺失值的处理	②描述设计效应和缺失值的处理
计量变量	12 ①解释计量变量如何分析,如怎样选择分组 ②如果可能,给出连续分析和分组分析的结果		
资助	13 给出当前研究的资助来源和资助者(如果可能,给出原始研究的资助情况)		
结果			
研究对象	14* ①报告研究的各个阶段研究对象的数量,如可能合格的数量、被检验是否合格的数量、证实合格的数量、纳入研究的数量、完成随访的数量和分析的数量 ②描述各个阶段未能参与者的原因 ③推荐使用流程图 ④报告研究对象征集的时间范围 ⑤匹配研究应给出每个病例对应对照数量的分布		
描述性资料	15* ①描述研究对象的特征(如人口学、临床和社会特征)以及关于暴露和潜在混杂因子的信息 ②指出每个研究变量数据的完整程度 ③总结平均的和总的随访数量以及随访天数		
结局资料	16* 报告发生结局事件的数量或综合指标	报告各个暴露类别的数量	报告结局事件的数量或综合指标
主要结果	17 ①陈述未调整的和按照混杂因子调整的关联强度、精确度(如 95% CI)。阐明按照哪些混杂因素进行调整以及选择这些因素,未选择其他因素的原因 ②对计量变量分组进行的比较要报告每组观察值的范围或中位数 ③对有意义的危险因素,可以把相对危险度转化成绝对危险度 ④报告按照实际目标人群的混杂因子和效应修正因子的分布进行标化的结果		
其他分析	18 报告进行的其他分析,如亚组分析和敏感性分析		
讨论			
重要结果	19 概括与研究假设有关的重要结果		
局限性	20 ①结合潜在偏倚和不精确的来源,讨论研究的局限性,以及分析、暴露和结局存在多样性时出现的问题;讨论所有可能偏倚的方向和大小 ②关于研究局限性的讨论不应取代定量的敏感性分析		
可推广性	21 讨论研究结果的可推广性(外推有效性)		
解释	22 结合当前证据和研究局限,谨慎给出一个总体的结果解释,并注意其他可替代的解释		

* 在病例对照研究中分别给出病例和对照的信息,如果可能,在队列研究和横断面研究里给出暴露组和未暴露组的信息

参 考 文 献

1 Maria B, Carsten H, Oliver R. Critical reading of epidemiological papers. A guide. Eur J Public Health, 2001, 11: 97-101.

2 Olson SH, Voigt LF, Begg CB, et al. Reporting participation in case-control studies. Epidemiology, 2002, 13: 123-126.

3 Wacholder S. Design issues in case control studies. Stat Methods Med Res, 1995, 4: 293-309.

4 Slattery ML, Edwards SL, Caan BJ, et al. Response rates among control subjects in case-control studies. Ann Epidemiol, 1995, 5: 245-249.

5 Olson SH, Voigt LF, Begg CB, et al. Reporting participation in case-control studies. Epidemiology, 2002, 13: 123-126.

6 Olson SH. Reported participation in case-control studies: changes over time. Am J Epidemiol, 2001, 154: 574-581.

7 Pocock SJ, Collier TJ, Dandreo KJ, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. BMJ, 2004, 329: 883-889.

8 Erik von E, Matthias E. The scandal of poor epidemiological research. Reporting guidelines are needed for observational epidemiology. BMJ, 2004, 329: 868-869.

9 Ottenbacher KJ. Quantitative evaluation of multiplicity in epidemiology and public health research. Am J Epidemiol, 1998, 147: 615-619.

10 Good Epidemiological Practice (GEP); proper conduct in epidemiological research. IEA European Federation. <http://www.dundee.ac.uk/iea/GoodPract.htm>

11 History of STROBE. <http://www.strobe-statement.org/History.html>

(收稿日期:2006-04-07)
(本文编辑:张林东)