

《因果推断研究中混杂因素控制与结果报告指南》解读

刘颖¹ 刘旭¹ 汪颖¹ 王迪芬¹ 马朋林²

¹ 贵州医科大学附属医院重症医学科, 贵阳 550004; ² 解放军总医院第八医学中心 SICU, 北京 100091

通信作者: 马朋林, Email: mapenglin1@163.com

【摘要】 因果推断研究是一种因果关系检验, 目的是评估暴露因素对结果的影响。试验性研究和观察性研究均可用于检验暴露因素和结局的因果关联。由于伦理或试验条件等因素影响, 试验性研究有时会受到限制。临床研究中观察性研究占有很大比例, 但观察性研究如果设计方案不严谨、混杂因素控制不好, 则会影响观察性研究的有效性和研究价值。来自世界各国的 35 个杂志的 47 位编辑组成了特别小组, 制定了《因果推断研究中混杂因素控制与结果报告指南》, 给予了研究者很好的指导。本文对该指南进行解读, 希望对临床研究者提供帮助。

【关键词】 流行病学; 因果关系; 研究设计; 混杂因素

基金项目: 贵州省高层次人才科研条件特助经费项目(TZJF-2011-25)

DOI: 10.3760/cma.j.cn121430-20201127-00734

Interpretation of Guidelines for controlling confounding factors and reporting results in causal inference studies

Liu Ying¹, Liu Xu¹, Wang Ying¹, Wang Difen¹, Ma Penglin²

¹Department of Intensive Care Unit, the Affiliated Hospital of Guizhou Medical University, Guiyang 550004, Guizhou, China; ²Department of Surgical Intensive Care Unit, the 8th Medical Center of Chinese PLA General Hospital, Beijing 100091, China

Corresponding author: Ma Penglin, Email: mapenglin1@163.com

【Abstract】 Causal inference research is a causal test designed to assess the impact of exposures on outcomes. Both experimental and observational studies can be used to examine causal associations between exposure factors and outcomes. Experimental studies are sometimes limited by factors such as ethics or experimental conditions. Observational studies account for a large proportion in clinical studies, but the effectiveness and research value of observational studies will be affected if the design of observational studies is not rigorous and the confounding factors are not well controlled. The *Guidelines for controlling confounding factors and reporting results in causal inference studies* formulated by a special group of 47 editors from 35 journals from all over the world provide good guidance to researchers. This article interprets the guidelines and hopes to provide help for clinical researchers.

【Key words】 Epidemiology; Causality; Research design; Confounding factor

Fund program: Special Funds for Scientific Research Conditions for High-Level Talents of Guizhou Province of China (TZJF-2011-25)

DOI: 10.3760/cma.j.cn121430-20201127-00734

近年来, 中国医学科学研究所取得的成就令世人瞩目。尽管规模发展迅速, 但我们清醒地认识到, 中国重症医学研究尚缺乏深度和高度。我们拥有全球最大规模的重症医师队伍, 同时担负着最为庞大的重症患者群体, 理应对世界重症医学临床问题的探索作出卓越的贡献。因此, 提升中国重症医学临床研究水平迫在眉睫。

2019年, Lederer等^[1]47位来自世界各国的35个杂志编辑所组成的特别小组制定了《因果推断研究中混杂因素控制与结果报告指南》。该指南代表了当今编辑对因果推断研究方法的最佳阐释和理解, 有助于提醒研究人员针对此类研究在观点上的重大改变。编辑们发现, 既往发表的此类研究存在诸多“缺陷”, 包括研究设计方法不恰当、忽视威胁结果准确性的风险因素(如混杂因素)、报告结果含糊其辞、缺乏重复试验的结果以及不能在观察性研究方法学的框架

下解释所获得的研究结果^[2]。因此, 应该对此类研究方法提出更加严格的要求^[3-4]。只有在设计与实施方法均严谨的情况下, 观察性研究才能够提供有价值的、支持因果关系的证据, 才能够有效提高关于因果关系推断研究的严谨性、有效性以及学术价值。我们相信, 本次解读对渴求进步的中国重症医学研究者将有较大的帮助, 故摘取报告该指南的主要论点介绍如下。

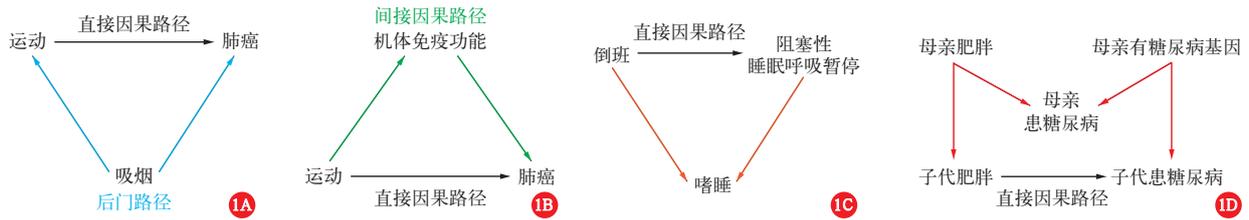
1 指南关于因果推断的定义

因果推断是一种因果关联检验, 目的是评估暴露因素对结果的影响。试验性研究(如随机对照试验)和观察性研究(如队列研究、病例对照研究、横断面研究)均可用于检验暴露因素与结局的因果关联。因果推断研究需要清晰明确的假设、足够的关注度来减少研究中的选择偏倚和信息偏倚, 以及需要一个严谨的设计方案控制混杂因素导致的混杂偏倚。

表 1 《因果推断研究中混杂因素控制与结果报告指南》关于因果推断研究的关键原则

原则 1: 因果推断研究需要详细考虑混杂因素	原则 2: 对于结果的解释不能依赖 P 值的大小	原则 3: 结果应该以清晰明了的方式呈现
<ul style="list-style-type: none"> • 最佳变量选择方法: <ul style="list-style-type: none"> ① 根据既往混杂因素定义有目的地选择变量 ② 因果模型使用直接的有向无环图 • 不能充分控制混杂因素的变量选择方法: <ul style="list-style-type: none"> ③ 基于 P 值或模型的方法 ④ 基于 β 系数改变的方法 ⑤ 变量选择以识别“独立的预测因子” • 用来检验单一因果关系的模型中不要试图做所有的效应评估 	<ul style="list-style-type: none"> • P 值应尽量少地单独出现 • 展示伴有或不伴有 P 值的效应评估和精度测量 • 效应评估的变异性应该与结论一致 • 得出“无关联”结论需要排除有意义的效应范围 • 避免使用“有意义”这一词,而需要更具体地描述 	<ul style="list-style-type: none"> • 运用 STROBE 声明和检查清单 • STROBE “解释与阐述”文件后的模型图 • 定量结果的可视化呈现 <ul style="list-style-type: none"> ◦ 如果可能的话,呈现单独的数据点 ◦ 避免过多的线条、文字、网格和缩略词 ◦ 连续性数据不应用带有标准误差条的柱形图

注: STROBE 为《加强流行病学观察性研究的报告》



注: 图中每个箭头均代表一个因果效应,黑色箭头均代表直接因果路径。A 为开放后门路径中的混杂因素,蓝色箭头代表开放的后门路径,即运动←吸烟→肺癌。吸烟是一个混杂因素,本能地开放后门路径;控制“吸烟”将关闭后门路径,从而消除了这条通路上的混杂因素。

B 为中间变量,绿色箭头代表间接因果路径。“机体免疫功能”部分参与调节运动与肺癌的关联,即运动→机体免疫功能→肺癌;控制“机体免疫功能”是不合理的做法,因为它会部分关闭因果路径,降低试验中观察到的运动与肺癌的因果效应。C 为碰撞变量,橙色箭头代表关闭的后门路径,即倒班→嗜睡←阻塞性睡眠呼吸暂停。“嗜睡”是一个碰撞变量,本能地关闭后门路径。

控制“嗜睡”将开放后门路径,从而产生混杂因素。D 为 M 图形,红色箭头代表一个关闭的后门路径,即子代肥胖←母亲肥胖→母亲患糖尿病←母亲有糖尿病基因→子代患糖尿病。“母亲患糖尿病”是一个碰撞变量,本能地关闭后门路径,控制“母亲患糖尿病”将开放后门路径,从而通过这条后门路径产生混杂因素

图 1 《因果推断研究中混杂因素控制与结果报告指南》推荐在进行因果推断研究时用于识别混杂因素的有向无环图

2 指南关于因果推断研究的 3 个原则

指南提供了 3 个关键原则指导研究者在因果推断研究中的分析和报告(表 1)。

2.1 原则 1——因果推断研究需要详细考虑混杂因素:指南提出,因果推断的观察性研究中需重点关注如何定义和选择混杂因素,有两种方法适用:一是根据以往的经验,二是运用因果模型来选择及识别混杂因素。

2.1.1 既往定义混杂因素的方法:混杂因素会影响目标结局事件,长期以来混杂因素被定义为与所研究的暴露因素相关的外来因素,并且它不是所研究的暴露因素与研究疾病因果链上的中间变量。这种定义有很大局限性。指南建议研究者根据先前的知识及经验,有目的地确定及识别混杂因素,而不是利用现有的数据将与暴露因素和结果相关的变量认定为混杂因素。

2.1.2 运用因果模型识别混杂因素:指南指出,根据经验选择和识别混杂因素的方法仅适用于简单的因果推断研究,对于较为复杂的因果关系研究不合适。因而,指南呼吁研究者在检验因果关联时运用因果模型。由 Greenland 开发、Pearl 等不断完善的因果推断模型可以满足我们目前在进行因果推断研究时混杂因素识别的需要^[5-9],指南推荐应用有向无环图(directed acyclic graph, DAG)。特别是当存在多个混杂因素时,应用 DAG 可将各变量的关系用直观可视的图形表示,以梳理各变量间的关系。DAG 中通过箭头连接节点构

成有方向的路径,但并不形成封闭的循环,其中节点表示变量,一个箭头代表一个变量对另一个变量的直接因果效应。一条“通路”就是连接任意两个变量的一组箭头。“目标因果路径”是所研究暴露因素与目标结局相关性的直接路径。“后门路径”是除目标因果路径外,另一个暴露因素与结局间的路径。“混杂”定义为在所研究暴露因素与目标结局之间至少存在一条“开放”的后门路径,而存在于开放后门路径上的自然变量称为“混杂因素”。任何两个通过开放路径连接的变量之间都存在关联。当研究人员“控制”混杂因素时,后门路径将被“关闭”。

指南展示了 4 种简单的 DAG 结构(图 1)。

例 1: 开放后门路径中混杂因素的识别。假设试验者要论证运动是否能够降低肺癌发生风险。在图 1A 中有一条因果路径:运动→肺癌,一条后门路径:运动←吸烟→肺癌,这条开放的后门路径反映了混杂的存在,因此吸烟就是运动与肺癌发生风险关联的一个混杂因素。值得注意的是,此时我们把混杂因素定义为一个变量,当控制混杂因素时,后门路径将被关闭。

识别是否存在混杂因素可通过以下步骤判断:① 删除所有从暴露因素发出的箭头;② 查看是否有任何从暴露因素到目标结局的开放路径,即暴露因素的直接效应去除后,观察所研究的暴露因素与目标结局是否通过开放的后门路径产生关联。如果步骤②结果为否,那么该研究中就不存在

混杂因素；反之，如果步骤②中从暴露因素到目标结局还存在开放路径，那么就代表存在混杂因素，需要进一步识别混杂因素。

当不只一个变量存在于后门路径时，控制通路上一个混杂因素即足以关闭这条后门路径。在一个有许多条通路的 DAG 上，通常控制少量的变量（混杂因素的最小调整集合）即能够关闭整个后门路径。指南推荐在因果推断研究中使用 DAG 识别后门路径和建立最小混杂因素调整集合。指南推荐的 dagitty.net 为研究者提供了一个简单的界面，通过该界面可以构建 DAG，进而识别后门路径和混杂因素的最小调整集合^[10]。

例 2：在 DAG 上识别中间变量。中间变量存在于所研究暴露因素与目标结局的因果路径上（而非后门路径）。当然，中间变量很有意义，因为它们是疾病的起因和机制。在图 1B 中，假设研究者要论证运动是否能够降低肺癌发生风险，“机体免疫功能”就是中间变量。至少，运动对肺癌发生风险的一些因果效应是由运动引起的机体免疫功能变化介导的，即运动→机体免疫功能→肺癌。一条包括中间变量的因果路径称为间接效应或者间接因果路径。而将运动与肺癌直接联系起来的箭头表示的是运动对肺癌的直接因果效应。

中间变量通常使间接因果路径开放。控制或者调整中间变量将会关闭间接因果路径，影响所研究暴露因素与目标结局间因果关联的真实性。因而需要特别关注中间变量，不应将它们当作混杂因素进行处理。运用 DAG 能够帮助研究者识别中间变量，从而避免在进行因果关联检验时控制这些变量。

例 3：在 DAG 上识别碰撞变量。碰撞变量是 1 个变量有 2 个或 2 个以上的前因，它们位于目标的通路中。当沿着通路的两个箭头都指向一个变量时，即可在 DAG 上识别碰

撞变量（图 1C）。碰撞变量本能地关闭后门路径，而控制一个碰撞变量反而会开放后门路径，从而产生混杂因素。

在图 1C 中，假设研究者要研究倒班与阻塞性睡眠呼吸暂停的因果关联，研究者可能想控制“嗜睡”，因为倒班和阻塞性睡眠呼吸暂停均会导致嗜睡。但是，嗜睡是一个碰撞变量，自然地阻断后门路径：倒班→嗜睡←阻塞性睡眠呼吸暂停，如果控制嗜睡将会开放后门路径，从而产生混杂因素。

例 4：M 图形的识别。“M 图形”是以 DAG 形状命名的。当研究因素或结果都不是碰撞变量的直接原因时，也可能存在碰撞变量。例如：研究者要验证肥胖与糖尿病的关系，可能把“母亲患糖尿病”作为校正因素。因为“母亲肥胖”可引起“母亲患糖尿病”及“子代肥胖”，“母亲有糖尿病基因”会导致“母亲患糖尿病”和“子代患糖尿病”。这些关系可能让研究者认为“母亲患糖尿病”是混杂因素，但实际上它不是。图 1D 显示，“母亲患糖尿病”是后门路径“子代肥胖←母亲肥胖→母亲患糖尿病←母亲有糖尿病基因→子代患糖尿病”的碰撞变量，会自动阻断后门路径。如果将“母亲患糖尿病”当作混杂因素给予控制，会开放后门路径，产生混杂因素；忽略“母亲患糖尿病”则是正确的做法。

指南建议，当研究者想要关闭后门路径时，在控制混杂因素前，要先明确这些附加变量不是中间变量和碰撞变量。

这份简洁的指南未能提供详尽的关于因果推断的讨论，但是这些例子可以鼓励研究者在研究中考虑使用因果模型。指南为研究者推荐了一些相关的优秀资源以供学习参考（表 2）。

2.1.3 不能充分控制混杂因素的变量选择方法：指南提出，基于 P 值或模型的变量选择方法（包括正向、反向和逐步选择），不应该用于因果推断。这些方法忽视了假设下的因果结构，因而不能充分控制混杂因素。那些依赖模型拟合或相关构建的方法（如相关系数 r^2 、赤池信息量准则以及贝叶斯

表 2 《因果推断研究中混杂因素控制与结果报告指南》推荐的因果推断研究相关学习资源

书籍	文献	网站
<ul style="list-style-type: none"> • Pearl J, Mackenzie D. The book of why: the new science of cause and effect [M]. New York: Basic books, 2018. • Pearl J. Causality: models, reasoning and inference [M]. 2nd ed. New York: Cambridge University Press, 2009. • Hernia MA, Robins JM. Causal inference [M/OL]. Boia Baton: Chapman & Hall/CRC, forthcoming, 2018. http://www.humphs.harvard.due/Michelangelo/causal-inference-book. 	<ul style="list-style-type: none"> • Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research [J]. Epidemiology, 1999, 10 (1): 37-48. • Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias [J]. Epidemiology, 2003, 14 (3): 300-306. • Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias [J]. Epidemiology, 2004, 15 (5): 615-625. DOI: 10.1097/01.ede.0000135174.63482.43. • Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies [J]. Epidemiology, 2009, 20 (4): 488-495. DOI: 10.1097/EDE.0b013e3181a819a1. • Morabia A. History of the modern epidemiological concept of confounding [J]. J Epidemiol Community Health, 2011, 65 (4): 297-300. DOI: 10.1136/jech.2010.112565. • Williamson EJ, Aitken Z, Lawrie J, et al. Introduction to causal diagrams for confounder selection [J]. Respirology, 2014, 19 (3): 303-311. DOI: 10.1111/resp.12238. • Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data [J]. Am J Public Health, 2018, 108 (5): 616-619. DOI: 10.2105/AJPH.2018.304337. 	<ul style="list-style-type: none"> • 关于因果推断和直接有向无环图的在线课程： Miguel Hernán. Causal diagrams: draw your assumptions before your conclusions. https://www.edx.org/course/causal-diagrams-draw-assumpti-ns-harvardx-ph559x. • 供建立有向无环图的网页： http://dagitty.net. Textor J, van der Zander B, Gilthorpe MS, et al. Robust causal inference using directed acyclic graphs: the R package 'dagitty' [J]. Int J Epidemiol, 2016, 45 (6): 1887-1894.

信息准则)也不适用于因果推断研究。这些方法都严重依赖现有数据,而这些数据中的因果关系可能不能被发现或者并不明显。

2.1.4 指南对因果关联、因果效应以及声称因果关联的主张:在用观察性研究检验因果关系时,研究者的目的总是寻找证据支撑(或驳斥)暴露因素对结局的因果效应。然而,我们通常无法从任何单一研究中确定这些因果效应。所以指南提出,运用“因果关联”这个词条(而不是“因果效应”)去描述从观察性研究得出的结果是合理的。同时指南提醒研究者,在没有确凿证据证明真正因果效应的情况下,在观察性研究中,当提及因果关联时,使用“效果评估”是合理的。不应该断言暴露一定会对结果产生影响,或者暴露会“保护避免”或“促进”结局的发生。

2.1.5 指南关于控制混杂因素方法的说明:研究者可能在试验设计阶段和数据分析过程中控制混杂因素。采取暴露随机化,运用工具变量,通过倾向评分进行加权回归,使用多因素回归分析进行校正,按混杂因素分层、控制混杂因素纳入(限制)以及匹配混杂因素都是常见的方法^[11]。关于这些方法指南持中立意见。

2.2 原则 2——对于结果的解释不能依赖 P 值的大小:近些年,因果推断中 P 值的优点被质疑^[12-15]。 P 值经常被错误地解释和运用^[16]。指南建议 P 值通常不要单独出现(“组学”研究和交互检验除外)。除了 P 值之外,还应提出效应估计和可信区间。

指南建议在对因果关系作出结论时,应对效应评估的变化幅度进行解释。例如:当危险比为 2.1、95% 可信区间为 0.97~4.20 时,即便相应的 P 值为 0.1 也不应得出没有因果关联的结论,因为最大的危险比 4.2 没有被合理排除,并且至少在这个研究样本中,研究因素与结局的关联确实被观察到。因此,应得出“该暴露因素导致结局事件发生的概率增加 2.1 倍(95% 可信区间为 0.97~4.20)”这个结论才是合适的。在这个例子中,点估计和区间估计提供了丰富的信息。同时,指南反对使用模糊的词语“有意义”和“无意义”,这些词语会导致读者(和作者)含蓄地得出一个关联存在或不存在的结论。指南建议,应该像上述例子中呈现的一样,简单报道定量的结果,这样效应大小的临床、统计学或生物学解释具有更大的价值。

2.3 原则 3——结果应该以清晰明了的方式呈现:2007 年发表的《加强流行病学观察性研究的报告》(*Strengthening the reporting of observational studies in epidemiology*, STROBE)声明,为检验因果关联的临床观察性研究结果的报道提供了清楚和有价值的指导^[17]。指南强烈建议研究者在报告结果时遵循 STROBE 声明,遵循 STROBE“解释与阐述”文件中提供的具体要求^[18],并把下面内容牢记心中:① 在队列研究中,结果的图表展示应该包括事件发生次数、人次、发病率以及每个暴露因素水平对应的未校正和已校正混杂因素后的发病率危险比;② 在横断面调查研究中,结果的图表展示应该包括事件发生次数、发病率以及每个暴露因素水平对

应的未校正和已校正混杂因素后的发病率;③ 在病例对照研究中,结果的图表展示应该包括病例组和对对照组各自的暴露人数、暴露率以及未校正和已校正混杂因素后每组的优势比。指南鼓励研究者采取深思熟虑和谨慎的方法对定量结果进行可视化呈现。有可能的话,单独数据点的表示应加上集中趋势和离散的测量。

综上所述,这份指南给予了研究者很好的建议,通过对指南的学习和解读可帮助研究者提高观察性研究工作的严谨性,提高发表在期刊上的科学研究的价值和有效性。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Lederer DJ, Bell SC, Branson RD, et al. Control of confounding and reporting of results in causal inference studies. Guidance for authors from editors of respiratory, sleep, and critical care journals [J]. *Ann Am Thorac Soc*, 2019, 16 (1): 22–28. DOI: 10.1513/AnnalsATS.201808–564PS.
- [2] Ioannidis JP. Why most clinical research is not useful [J]. *PLoS Med*, 2016, 13 (6): e1002049. DOI: 10.1371/journal.pmed.1002049.
- [3] Kitsios GD, Dahabreh IJ, Callahan S, et al. Can we trust observational studies using propensity scores in the critical care literature? A systematic comparison with randomized clinical trials [J]. *Crit Care Med*, 2015, 43 (9): 1870–1879. DOI: 10.1097/CCM.0000000000001135.
- [4] Prasad VK, Cifu AS. Ending medical reversal: improving outcomes, saving lives [M]. Baltimore: Johns Hopkins University Press, 2015.
- [5] Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias [J]. *Epidemiology*, 2003, 14 (3): 300–306.
- [6] Morabia A. History of the modern epidemiological concept of confounding [J]. *J Epidemiol Community Health*, 2011, 65 (4): 297–300. DOI: 10.1136/jech.2010.112565.
- [7] Williamson EJ, Aitken Z, Lawrie J, et al. Introduction to causal diagrams for confounder selection [J]. *Respirology*, 2014, 19 (3): 303–311. DOI: 10.1111/resp.12238.
- [8] Pearl J. Causality: models, reasoning, and inference [M]. 2nd ed. New York: Cambridge University Press, 2009.
- [9] Pearl J, Mackenzie D. The book of why: the new science of cause and effect [M]. New York: Basic Books, 2018.
- [10] Textor J, van der Zander B, Gilthorpe MS, et al. Robust causal inference using directed acyclic graphs: the R package 'dagitty' [J]. *Int J Epidemiol*, 2016, 45 (6): 1887–1894. DOI: 10.1093/ije/dyw341.
- [11] Gershon AS, Jafarzadeh SR, Wilson KC, et al. Clinical knowledge from observational studies. Everything you wanted to know but were afraid to ask [J]. *Am J Respir Crit Care Med*, 2018, 198 (7): 859–867. DOI: 10.1164/rccm.201801–0118PP.
- [12] Benjamin DJ, Berger JO, Johannesson M, et al. Redefine statistical significance [J]. *Nat Hum Behav*, 2018, 2 (1): 6–10. DOI: 10.1038/s41562-017-0189-z.
- [13] Ioannidis JPA. The proposal to lower P value thresholds to .005 [J]. *JAMA*, 2018, 319 (14): 1429–1430. DOI: 10.1001/jama.2018.1536.
- [14] Lakens D, Adolfs FG, Albers CJ, et al. Justify your alpha [J]. *Nat Hum Behav*, 2018, 2: 168–171.
- [15] Wasserstein RL, Lazar NA. The ASA's statement on p -values: context, process, and purpose [J]. *Am Stat*, 2016, 70: 129–133. DOI: 10.1080/00031305.2016.1154108.
- [16] Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations [J]. *Eur J Epidemiol*, 2016, 31 (4): 337–350. DOI: 10.1007/s10654-016-0149-3.
- [17] von Elm E, Altman DG, Egger M, et al. The strengthening of reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies [J]. *Ann Intern Med*, 2007, 147 (8): 573–577. DOI: 10.7326/0003-4819-147-8-200710160-00010.
- [18] Vandenberg JP, von Elm E, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration [J]. *Ann Intern Med*, 2007, 147 (8): W163–194. DOI: 10.7326/0003-4819-147-8-200710160-00010-w1.

(收稿日期: 2020-11-27)