

·基础理论与方法·

如何控制观察性疗效比较研究中的混杂因素:(三)混杂因素控制的敏感性分析方法

黄丽红¹ 赵杨² 魏永越² 陈峰²

¹复旦大学附属中山医院生物统计室,上海200032; ²南京医科大学公共卫生学院生物统计学系 211166

通信作者:黄丽红, Email: huang.lihong@zs-hospital.sh.cn

【摘要】 观察性疗效比较研究中混杂在所难免,在利用一些统计分析方法对已测量或未测量混杂因素加以控制后,是否消除了混杂的影响不得而知,此时需进行敏感性分析。本文介绍混杂因素处理中的敏感性分析方法。基于不同的研究,敏感性分析思路各不相同,对于已测量混杂因素可采用传统的敏感性分析方法,对于未测量混杂因素目前理论相对系统的方法主要有混杂函数法、边界因子法和倾向性评分校正法,另外Monte Carlo 敏感性分析和Bayes 敏感性分析也是近年来备受热议的方法。当敏感性分析结果与主要分析结果一致时,无疑提高了研究结论的可靠性。

【关键词】 观察性疗效比较研究; 混杂; 控制; 敏感性分析方法

基金项目:国家自然科学青年基金(81903407)

DOI:10.3760/cma.j.issn.0254-6450.2019.12.026

How to adjust confounders in studies on observational comparative effectiveness: (3) approaches on sensitivity analysis for confounder adjustment

Huang Lihong¹, Zhao Yang², Wei Yongyue², Chen Feng²

¹Department of Biostatistics, Zhongshan Hospital, Fudan University, Shanghai, 200032, China;

²Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, 211166, China

Corresponding author: Huang Lihong, Email: huang.lihong@zs-hospital.sh.cn

【Abstract】 Confounders are difficult to avoid in studies on observational comparative effectiveness. It is often unclear whether the confounders have been completely eliminated after controlling the measured or unmeasured potential confounding effects or if sensitivity analysis is needed when using the specific statistical methods, under given circumstances. This manuscript summarizes and evaluates the confounding sensitivity analysis methods. Based on different studies, sensitivity analyses need to use different approaches. The traditional sensitivity analysis can be applied for the measured confounders. Currently, the relatively systematic sensitivity analyses for unmeasured confounders would include confounding function, bounding factor and propensity score calibration. Additionally, more investigations are associated with Monte Carlo and Bayesian sensitivity analysis. Reliability of the research conclusion thus may largely be improved when the sensitivity analysis results are consistent with the main analysis.

【Key works】 Observational comparative effectiveness research; Confounder; Adjustment; Sensitivity analysis

Fund program: National Natural Science Foundation for Young Scientists (81903407)

DOI:10.3760/cma.j.issn.0254-6450.2019.12.026

早在1959年,Cornfield等首次介绍了用敏感性分析(sensitivity analysis)评估混杂偏倚^[1]。敏感性分析无法真正控制混杂,但可以通过评估干预与结局的关联从而评价混杂因素的潜在影响^[2]。在经过设计和统计分析的努力之后进行的敏感性分析被誉为控制混杂偏倚的最后一道防线。

混杂因素控制的统计分析方法一直备受学者关注,尤其是未测量混杂因素,新方法也在不断涌现,但有些方法还未经过实践考察,作为敏感性分析方

法提出。敏感性分析可在主要分析的基础上提供补充信息,助力于干预与结局的关联性判定。本文从统计学角度就设计良好的观察性疗效比较研究(CER)中的敏感性分析方法进行述评,并结合案例加以说明。

1. 已测量混杂因素的敏感性分析方法:传统的敏感性分析包括单因素敏感性分析和多因素敏感性分析。单因素敏感性分析即在固定其他因素的条件下,变动其中一个不确定因素;然后,再变动另一个

因素(仍然保持其他因素不变),以此求出某个不确定因素本身对结局的影响程度。多因素敏感性分析法是指在假定其他不确定性因素不变条件下,计算分析两种或两种以上不确定性因素同时发生变动对结局的影响程度,确定敏感性因素及其极限值。

针对已测量混杂因素,可采用传统的敏感性分析方法,通过改变已有的信息来考察研究结果的稳健性,例如纳入不同的协变量组合,比较不同调整情形下研究结果的变化;纳入不同的对照组,比较观察到的潜在偏倚的差别;改变结局的定义(如临床相关结局通过几种不同的方式加以确定),根据这些不同的结局进行再分析,可以揭示最初的结局定义对所研究结局的真实反映程度;采用多种数据来源进行分析,各种数据库中的重复结果有助于理解结论的稳健性和减少结果的偶然性等。如果采用不同的统计分析方法得到了一致的结论,无疑将提高研究结论的可靠性。但笔者建议,敏感性分析在固定其他因素的条件下,每次只考虑变动其中一个不确定因素,以评价该不确定因素对结局的影响;如果改变多个因素,则难以评价是哪个因素导致的不一致结论。

2. 未测量混杂因素的敏感性分析方法:基于不同的研究问题,不同的数据类型,敏感性分析的思路各不相同,本文将介绍几种理论相对成熟系统、应用条件局限性小的敏感性分析新方法。

(1) 混杂函数:2017年,Kasza等^[3]提出了混杂函数(confounding function)的敏感性分析方法,基于反事实(counterfactual)原理,通过构建混杂函数将未观测到的混杂大小度量化,通过调整混杂函数取值评价混杂因素对因果效应值在大小和方向上的影响。

令 Y 表示结局($1=$ 发生, $0=$ 未发生), G 表示二分类的干预变量($0=$ 未干预, $1=$ 干预)。 Y_i^a 表示个体*i*在反事实时的潜在结果(potential outcome)^[4]。所谓反事实是指:如果个体*i*实际接受了干预而假设其未接受干预,或实际没有接受干预而假设其接受干预。显然,个体*i*的潜在结果不受其他个体的处理的影响,即个体是独立的,并且对每个个体和每种处理只有一个潜在结果。由于个体*i*的潜在结果 $Y^1(i)$ 和 $Y^0(i)$ 不能同时观测到,因此进行因果推断时需要考虑到可能存在的未测量混杂导致的偏倚问题。

定义混杂函数:

$$c(a) = \frac{P(Y^a = 1 | G = 1)}{P(Y^a = 1 | G = 0)}, a \in \{0, 1\}$$

当 $c(0)=c(1)=1$ 时,说明研究中不存在混杂;当 $c(x)>1$ 时表示由于混杂效应的存在使得干预组发

生结局的风险高于非干预组;反之,当 $c(x)<1$ 时表示由于混杂效应的存在使得干预组发生结局的风险低于非干预组。因此,在敏感性分析中常借助 $c(0)$ 和 $c(1)$ 的大小考察未测量混杂的效应。应用时应结合专业知识设定 $c(0)$ 和 $c(1)$,考察经混杂函数校正后,所估计的干预效应是否依然存在。

(2) 边界因子:2016年,Ding和Vander weele^[5]提出了边界因子(bounding factor)方法进行敏感性分析,该法无需对未测量混杂因素作任何具体假设,且可评估所有混杂因素的影响。虽然二分类干预比较常见,但该敏感性分析方法可应用于多分类或连续性干预,混杂因素可为多分类或连续性指标,混杂因素可为单个或多个。在此,基于相对危险度(relative risk, RR)结局、二分类的干预因素介绍边界因子的统计原理。

令 U 表示未知混杂因素(可以为任意尺度), RR_{GY}^{obs} 表示干预 G 与结局 Y 的观察效应, RR_{GY}^{true} 表示校正未测量混杂因素后的干预 G 与结局 Y 之间的真正效应, RR_{GU} 为干预 G 与未测量混杂因素 U 之间的效应, RR_{UY} 是未测量混杂因素 U 与结局 Y 之间的效应,则有

$$RR_{GY}^{true} \geq RR_{GY}^{obs} / \frac{RR_{GU} \times RR_{UY}}{RR_{GU} + RR_{UY} - 1}$$

该等式表明,即便在未测量混杂因素存在时,真实 RR 将大于观察 RR ,其中 $\frac{RR_{GU} \times RR_{UY}}{RR_{GU} + RR_{UY} - 1}$ 称为联合边界因子(joint bounding factor)。在应用边界因子进行敏感性分析时,结合实际研究情况,通过假设 RR_{GU}, RR_{UY} 计算联合边界因子,从而得到校正后的真实 RR 。

将上述方法推广到多分类结局、连续性结局和生存结局的情形,从而使得该方法的应用范围更广,具体计算在此不作赘述。

(3) 倾向性评分校正:2005年,Stürmer等^[6]利用回归校正的思想,提出倾向性评分校正法(propensity score calibration, PSC)进行未测量混杂因素的校正,该方法需要借助外部数据库(validation study data)信息。由于未测量混杂因素的作用,倾向性评分(PS)估计值往往存在误差,将基于目标研究数据所估计的 PS 称为误差 PS (error-prone, EP),记为 PS_{EP} ;当外部数据库涵盖了比研究数据更多的信息时,根据外部数据库估计的 PS 称为标准 PS (gold standard, GS),记为 PS_{GS} 。假设 X 为协变量矩阵, G 为干预变量($1=$ 干预, $0=$ 未

干预), $e(X)$ 为根据协变量 X 个体分入观察组的条件概率,即PS。根据研究数据得误差PS估计:

$$PS_{EP}=e(X_{EP})=prob(G=1 \mid X_{EP})$$

根据外部数据库及研究数据库,可得标准PS估计:

$$PS_{GS}=e(X_{GS})=prob(G=1 \mid X_{GS})$$

假设 $e(X_{EP})$ 在给定 X_{GS} 和干预 G 的条件下独立于结局,误差估计模型为

$$E[e(X_{GS}) \mid G, e(X_{EP})] = \lambda_0 + \lambda_E G + \lambda_{EP} e(X_{EP})$$

因而,基于研究数据及 $e(X_{EP})$ 所建模型的干预效应估计 β ,可通过 β/λ_E 加以校正。

关于PS校正的敏感性分析,也有学者提出其他算法^[7],但只能考虑对单个未测量混杂的校正。Stürmer等^[6]的PSC法基于回归校正的思想,成功将PS校正推广到了多个未测量混杂,并且该方法不要求外部数据库中包含结局信息,不需要定义混杂与结局的关联。但该方法的局限性与回归校正相同^[8],假设 $e(X_{EP})$ (根据研究数据协变量 X ,个体分入观察组的条件概率)在给定 X_{GS} (外部数据库协变量 X)和干预 G 的条件下独立于结局,而这个假设在外部数据库缺乏结局信息时无法检验。

(4)其他方法:基于实际研究情况,通过假设一些未知参数,例如未测量混杂所占人群比例、干预和结局与未测量混杂的相关系数等,对未测量混杂进行不同情境下的模拟,随后进行敏感性分析,若发现未测量混杂对结局的影响异常大,将推翻所观察到的干预与结局之间关联的结论,这类敏感性分析思路对未测量混杂的分布没有限制^[2,9-10],也是众多学者热衷的一类方法。

外部数据库信息借用的敏感性分析方法是近年来的研究热点。若外部数据库包含的信息相较研究数据丰富,且包含了结局信息,可将研究数据中的未测量混杂因素视为缺失值加以填补,填补方式可考虑多重填补(multiple imputation)或Bayes填补(Bayesian imputation)等。外部信息借用的敏感性分析还可采用Monte Carlo敏感性分析(Monte Carlo sensitivity analysis, MCSA)和Bayes敏感性分析(Bayesian sensitivity analysis, BSA)。前者通过借助外部信息(例如:文献报道、Meta分析结果等),假定未测量混杂的分布,模拟产生数据从而进行未测量混杂的控制,后者同样需要借助外部信息假定未测量混杂先验分布,结合模拟产生的混杂因素样本数据,导出后验分布从而进行参数估计。这两种模拟数据的敏感性分析思路引发了一系列的探讨^[11-12],如何付诸实际应用有待进一步研究。

3. 案例分析:这里介绍两个利用混杂函数法和PSC法进行敏感性分析的实例。

(1)混杂函数案例分析:Kasza等^[3]基于996位接受冠心病介入治疗患者的数据信息(数据来源于Kereiakes等^[13]的研究),估计血小板聚集抑制剂阿昔单抗(abciximab)使用者6个月内的死亡风险。996名患者中有698名服用阿昔单抗,其中11名在6个月内死亡;另外298名患者未服用阿昔单抗,其中15名在6个月内死亡,未校正死亡风险RR值为0.31(95%CI:0.15~0.67)。主要分析结果基于倾向性评分及逆概率加权法(inverse probability of treatment weighting, IPTW)进行校正,校正因素选取性别、身高、是否患糖尿病,近期是否发生急性心肌梗死等7个协变量,校正后得到RR值为0.19(95%CI:0.08~0.50)。

针对上述主要分析结果,进行敏感性分析。根据临床经验可知,接受阿昔单抗的患者不可能比未接受的死亡风险高,而且接受阿昔单抗的患者通常病情比较重,在反事实的条件下,接受阿昔单抗治疗的患者,在6个月内的死亡风险更高,因而混杂函数 $c(0), c(1) > 1$ 是比较符合实际的假设。图1A中以水平虚线表示 $c(0)=1$;以垂直虚线表示 $c(1)=1$;实线表示 $c(0)=c(1)$, $c^{0.19}(1)$ 标注的线是指当真实效应估计值为0.19时对应 $c(0)$ 和 $c(1)$ 各自不同取值变

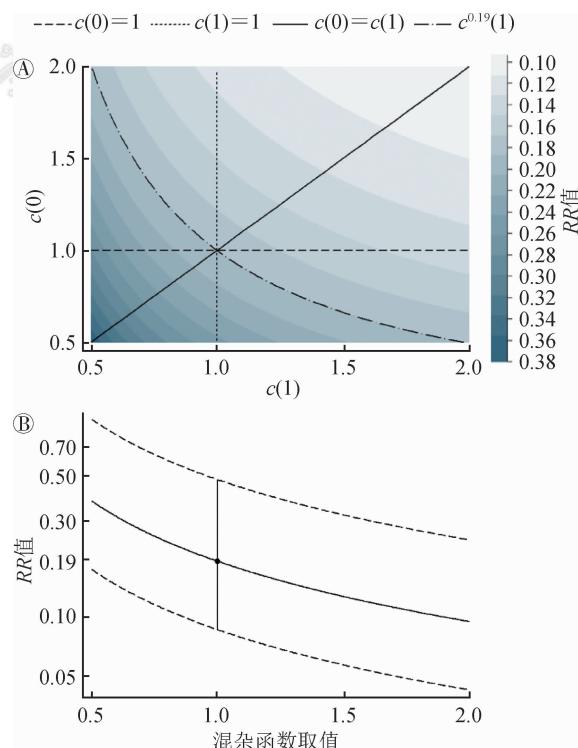


图1 混杂函数案例结果^[3]

化,不同的颜色区域显示不同的 RR 取值范围,可看到当 $c(0)$ 和 $c(1) > 1$ 时,均 RR 值 < 0.2 。图1B中可以看出, RR 值的95%CI均不包含1,当混杂函数 > 1 时,大部分 RR 取值 < 0.19 。在调整已测量和未测量混杂因素后,阿昔单抗可降低冠心病介入治疗患者的死亡风险。

(2)PSC案例分析:Stürmer等^[6]利用回顾性数据库进行老年人非甾体抗炎药(nonsteroidal anti-inflammatory drugs, NSAIDs)与1年内死亡率的关联研究。研究数据库来自纽泽西州老年和残疾项目(Aged and Disabled program),将其中1995年1月1日至1997年12月31日期间,>65岁的关节炎患者纳入研究,研究数据收集的信息包括年龄、性别、种族、药物处方信息、诊断信息、住院次数、门诊次数、自住院开始1年内的随访信息等。外部数据库来自Medicare Current Beneficiary Survey(MCBS),该调查包括老年和残疾项目受益人,收集信息更广泛,除研究数据已有信息外,还包括BMI、受教育情况、收入、吸烟情况、日常行为能力,并且调查数据来自面对面的访谈,应答率很高(85%~95%),数据完整性好,为了让外部数据库与研究数据库具备良好同质性,研究者根据研究数据库人群的年龄和性别特征在MCBS数据库中随机抽取了5 108名患者。

为了评估未测量混杂,该研究利用研究数据计算PS,并定义相同的协变量利用MCBS数据计算误差PS,同时利用MCBS中所有信息(包括研究数据库)计算标准PS,另外还额外将患者自我汇报的患类风湿性关节炎(rheumatoid arthritis, RA)或骨关节炎(osteoarthritis, OA)的发病情况作为医院诊断的补充信息纳入PS估计中作为另一项敏感性分析,部

分结果见表1。随后采用Cox比例风险模型估计NSAIDs使用者的1年内死亡风险,未校正的 HR 值为0.68(95%CI: 0.66~0.71),将协变量纳入Cox模型进行传统校正后的 HR 值为0.80(95%CI: 0.77~0.83),基于研究数据的误差PS校正后的 HR 值为0.81(95%CI: 0.78~0.84),PSC进行未测量混杂校正后 HR 值为0.92(95%CI: 0.88~0.96),结合自我汇报发病情况补充信息的PSC法得 HR 值为1.06(95%CI: 1.00~1.12)。研究结论:基于研究数据的主要分析结果与结合外部数据信息的PSC敏感性分析 HR 值结果一致,尚无足够证据表明服用NSAIDs后会在1年内增加老年人的死亡风险。

4. 总结与展望:在观察性CER中必须考虑并正确处理所有的混杂,才可能得到效应的无偏估计,通过敏感性分析方法来评估未测量混杂因素对结局的潜在影响是值得推荐的统计分析策略。研究者可以评估当违背“不存在未测量混杂因素”这一假设时研究结果的敏感性,当稳健性受到混杂因素的影响时,研究推论可能会改变。由于未测量混杂因素的未知和不确定的性质,基于不同的研究,敏感性分析策略各不相同,敏感性分析方法灵活多变,在实际工作中,仁者见仁智者见智。本文从统计学方法的角度出发,将几种理论体系相对成熟的敏感性分析方法进行总结,见表2,这些敏感性分析方法目前尚无成熟软件包,均需通过编程实现。

当统计分析方法不够成熟时,针对混杂因素的不确定性所开展的敏感性分析就尤显重要。值得一提的是,虽然敏感性分析很有价值,研究者可以选择简明扼要地报告结果,也可以通过大量的敏感性分析得到更多的信息,因此需要在两者之间权衡。对

表1 倾向性评分法估计部分结果^[6]

分组	研究数据		外部数据敏感性分析					
			误差PS		标准PS			
	OR值	95%CI	OR值	95%CI	OR值	95%CI	OR值	95%CI
年龄(年)	0.98	0.98~0.99	0.98	0.97~1.00	0.99	0.98~1.00	0.98	0.97~1.00
女性	1.2	1.2~1.3	1.2	1.0~1.5	1.2	1.0~1.5	1.1	0.9~1.4
诊断								
心肌梗死	0.9	0.8~0.9	1.1	0.8~1.5	1.1	0.8~1.6	1.1	0.7~1.5
充血性心力衰竭	0.9	0.9~1.0	0.9	0.6~1.3	0.8	0.5~1.2	0.8	0.6~1.3
糖尿病	1.0	1.0~1.0	0.9	0.6~1.3	0.7	0.5~1.0	0.7	0.5~1.0
癌症	0.8	0.8~0.8	0.6	0.4~0.9	0.6	0.4~1.0	0.6	0.4~1.0
RA或OA	2.1	2.0~2.2	2.4	1.7~3.4	2.1	1.5~3.0	1.8	1.3~2.5
自我汇报诊断								
RA或OA							4.1	3.1~5.5
.....								
PS评分	0.63		0.60		0.66		0.71	

表2 观察性CER中敏感性分析方法总结

方法	应用条件	局限性
混杂函数(confounding function)	基于反事实原理,混杂因素无需任何假设	干预为二分类,应当结合专业知识设定混杂函数,以评估未测量混杂影响的可能范围,带有主观性
边界因子(bounding factor)	混杂因素无需任何假设。对干预、结局及混杂的数据类型无特殊限制	需设定干预与未测量混杂之间的效应,未测量混杂因素与结局之间的效应,带有一定主观性
倾向性评分校正法(propensity score calibration)	需要有外部数据信息,外部数据中可不包含结局信息	假设 $e(X_{EP})$ (根据研究数据协变量 X ,个体分入观察组的条件概率)在给定 X_{CS} (外部数据库协变量 X)和干预 G 的条件下独立于结局,而这个假设在外部数据库缺乏结局信息时无法检验
Monte-Carlo敏感性分析	需要外部信息提供:未测量混杂因素分布、未测量混杂因素的样本统计量	依赖外部信息;不同的模拟方式结论可能不同
Bayes敏感性分析	需要外部信息提供:未测量混杂因素先验分布、未测量混杂因素的样本统计量	依赖外部信息;不同的 Bayes 模型构建方式结论可能不同

于那些导致结果发生显著变化的敏感性分析,应加以关注。虽然敏感性分析的结果不能作为最终的研究结论,但是不一致的结论,足以影响研究的价值。

需再次强调的是,控制混杂最好的方法是在研究设计时,良好的研究设计可从源头控制各种可能的混杂,选择恰当的研究设计是减少混杂偏倚的必要条件,而因研究设计选择不当所导致的偏倚,或在实施过程中由于质量控制不严引入的偏倚,并不能在统计分析阶段完全消除。

值得一提的是,规范地报告非随机对照研究结果便于读者准确理解研究结果及其局限性,便于资料的汇总和系统综述,越来越多的学者意识到这一工作的重要性,美国CDC牵头发布了《非随机对照设计报告规范》(transparent reporting of evaluations with nonrandomized designs, TREND) 规范或声明^[14-15],为观察性CER的规范化应用奠定了基础。

大数据时代,随着人们对现实世界数据的关注和利用,对混杂因素控制的统计分析方法要求也越来越高,这一方向将是未来统计学研究的一个重要领域。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- Rosenbaum PR. Sensitivity to hidden bias [C]//Observational studies. New York, NY: Springer, 2002: 87-135. DOI: 10.1007/978-1-4757-3692-2_4.
- Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics [J]. Pharmacoepidemiol Drug Saf, 2010, 15 (5) : 291-303. DOI: 10.1002/pds.1200.
- Kasza J, Wolfe R, Schuster T. Assessing the impact of unmeasured confounding for binary outcomes using confounding functions [J]. Int J Epidemiol, 2017, 46 (4) : 1303-1311. DOI: 10.1093/ije/dyx023.
- Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies [J]. J Educ Psychol, 1974, 66 (5) : 688-701. DOI: 10.1037/h0037350.
- Ding P, Vanderweele TJ. Sensitivity analysis without assumptions [J]. Epidemiology, 2016, 27 (3) : 368-377. DOI: 10.1097/EDE.0000000000000457.
- Stürmer T, Schneeweiss S, Avorn J, et al. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration[J]. Am J Epidemiol, 2005, 162 (3) : 279-289. DOI: 10.1093/aje/kwi192.
- Schneeweiss S, Glynn R, Tsai EH, et al. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction [J]. Epidemiology, 2005, 16 (1) : 17-24. DOI: 10.1097/01.ede.0000147164.11879.b5.
- Armstrong B. Measurement error in nonlinear models [J]. Technometrics, 1995, 39 (2) : 231-232. DOI: 10.1080/00401706.1997.10485096.
- Lin NX, Logan S, Henley WE. Bias and sensitivity analysis when estimating treatment effects from the cox model with omitted covariates[J]. Biometrics, 2013, 69 (4) : 850-860. DOI: 10.1111/biom.12096.
- Groenwold RHH, Nelson DB, Nichol KL, et al. Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research[J]. Int J Epidemiol, 2010, 39 (1) : 107-117. DOI: 10.1093/ije/dyp322.
- Greenland S. A commentary on 'A comparison of Bayesian and Monte Carlo sensitivity analysis for unmeasured confounding' [J]. Stat Med, 2017, 36 (20) : 3278-3280. DOI: 10.1002/sim.7370.
- Mccandless LC, Gustafson P. A comparison of Bayesian and Monte Carlo sensitivity analysis for unmeasured confounding [J]. Stat Med, 2017, 36 (18) : 2887-2901. DOI: 10.1002/sim.7298.
- Kereiakes ADJ, Obenchain RL, Barber BL, et al. Abciximab provides cost-effective survival advantage in high-volume interventional practice[J]. Am Heart J, 2000, 140 (4) : 603-610. DOI: 10.1067/mhj.2000.109647.
- Fuller T, Pearson M, Peters JL, et al. Evaluating the impact and use of transparent reporting of evaluations with non-randomised designs (TREND) reporting guidelines [J]. BMJ Open, 2012, 2 (6) : e002073. DOI: 10.1136/bmjopen-2012-002073.
- 罗晓敏,詹思延.如何撰写高质量的流行病学研究论文:第六讲非随机对照试验研究报告规范—TREND介绍[J].中华流行病学杂志,2007,28(4):408-410. DOI: 10.3760/j.issn:0254-6450.2007.04.026.
- Luo XM, Zhan SY. How to write the high quality epidemiological paper: session 6 transparent reporting of evaluations with nonrandomized designs-introduction of TREND [J]. Chin J Epidemiol, 2007, 28 (4) : 408-410. DOI: 10.3760/j.issn: 0254-6450.2007.04.026.

(收稿日期:2019-03-18)

(本文编辑:王岚)