

倾向指数

第三讲 应用中的关键问题

王永吉 蔡宏伟 夏结来 蒋志伟 王陵

【关键词】 倾向指数; 变量选择; 均衡; 处理效应; 灵敏度分析

Propensity score (III) Key issues in practice of propensity score WANG Yong-ji¹, CAI Hong-wei², XIA Jie-lai¹, JIANG Zhi-wei¹, WANG Ling¹. 1 Department of Health Statistics, Faculty of Preventative Medicine, 2 Information Center, School of Stomatology, Fourth Military Medical University, Xi' an 710032, China

Corresponding author: CAI Hong-wei, Email: hwcai@fmmu.edu.cn; XIA Jie-lai, Email: xiajielai@fmmu.edu.cn

This work was supported by a grant from the National Natural Science Foundation of China (No. 30800952).

【Key words】 Propensity score; Variable selection; Balance; Treatment effect; Sensitivity analysis

前两讲介绍了倾向指数(编者注:国内有学者译为倾向评分)的基本概念、研究步骤及其常用研究方法。本讲介绍应用倾向指数方法中的几个关键问题:一是倾向指数模型变量的选择;二是暴露组 and 对照组间可观测变量均衡性的评价;三是应用倾向指数方法后处理效应的估计;四是倾向指数模型的灵敏度分析。

1. 变量的选择:倾向指数方法应用中,首要步骤就是构建倾向指数模型(logistic 回归模型),而模型的构建涉及变量的选择,但该问题一直存在争议,随着倾向指数方法近年来应用日益广泛,国外学者对变量的选择问题进行了一些研究,如Rubin和Thomas^[1]建议倾向指数模型应该纳入所有与结局有关的变量,而不考虑变量与暴露因素的关系;也有学者提出只纳入混杂因素^[2],即与结局和暴露因素都有关的变量。

根据变量与结局变量和暴露因素的关系,可以把变量分为四类:一是混杂因素,即与暴露因素和结局都有关的变量;二是只与暴露因素有关的变量;三是只与结局有关的变量;四是与暴露因素和结局都无关的变量。倾向指数模型变量的选择标准是“怎样能更好地控制混杂因素,减少其对处理

效应估计的干扰”。Brookhart等^[3]在倾向指数模型中分别纳入以上四种变量的不同组合,通过Monte Carlo模拟比较不同模型估计处理效应的偏性和精度,从而确定模型变量的选择对处理效应的影响。模拟研究表明,最优模型是纳入所有与结局有关的变量,而不管其是否与暴露因素有关。纳入与处理有关而与结果无关的变量,会降低估计处理效应的精度,同时不会降低偏倚;纳入与处理无关而与结局有关的变量,可以增加估计处理效应的精度,同时不会增加偏倚。在构建倾向指数模型时,会因为某个重要混杂因素的遗漏而导致该变量在暴露组和对照组间的不均衡,以及处理效应的有偏估计,因而在重要混杂因素未知或者无法获取时倾向指数方法的应用受到限制。为了避免遗漏重要混杂因素,有学者在构建倾向指数模型时,纳入尽可能多的变量,然后通过logistic逐步回归进行筛选,这种方法并不恰当,因为倾向指数模型的应变量是暴露因素,logistic逐步回归会筛选掉与暴露因素弱相关,但与结果强相关的变量,而这类变量属于研究中的混杂因素,应当被纳入模型中。根据国外的研究,倾向指数模型变量选择的标准应该是纳入所有与结局变量有关的变量,具体应结合相关学科专业知识进行选择。

Austin等^[4]也对倾向指数模型变量的选择问题进行了Monte Carlo模拟研究,比较变量的选择对匹配集大小的影响,同时还比较了选择相同变量时倾向指数匹配法和分层法的区别,对变量选择的标准与Rubin和Brookhart等的研究结果一致,认为合理地建模是纳入所有与结果有关的变量,而不考虑变量与暴露因素的关系。

Austin等^[4]的模拟研究还得到了一些对于倾向指数实际应用有意义的结论:一是即使建模合理,倾向指数分层法也会得到处理效应的有偏估计,这也是本系列讲座第二讲中提到的匹配法优于分层法的原因之一;二是如果建模合理可能会多产生20%以上的可匹配样本,而建模不合理会减小匹配集,从而降低处理效应估计的精度;三是ROC曲线面积(等价于logistic回归中的统计量c)与倾向指数模型平衡协变量的能力之间无直接关系,如果纳入一个与暴露因素有关而与结果无关的变量,会增加ROC曲线面积,而这个变量本身对倾向指数模型并没有贡献,所以用ROC曲线面积来评价倾向指数模型是不合理的;四是多结局变量研究的变量选择。影响不同结局变量的混杂因素存在差异,所以应该针对不同的结局变量,构建不同的倾向指数模型,多结果研究的倾向指数模型变量选择问题还有待进一步研究。

2. 可观测变量均衡性的评价:观察性研究数据未经过随

DOI:10.3760/cma.j.issn.0254-6450.2010.07.025

基金项目:国家自然科学基金(30800952)

作者单位:710032 西安,第四军医大学预防医学系卫生统计学教研室(王永吉、夏结来、蒋志伟、王陵),口腔医学院信息中心(蔡宏伟)

通信作者:蔡宏伟, Email: hwcai@fmmu.edu.cn; 夏结来, Email: xiajielai@fmmu.edu.cn

机化,暴露组和对照组间变量的分布可能存在不平衡性。倾向指数方法成功应用的关键是对组间变量均衡的效果,如果暴露组和对照组间协变量的分布均衡,组间混杂因素对结果效应的干扰就被抵消。

以往的研究都是用假设检验方法评价组间可观测变量的均衡性^[5-8],在使用倾向指数匹配法后,随着样本量的减少检验效能降低,有研究者提出组间协变量均衡性的评价是针对样本而言,而不是总体,而假设检验的对象是样本所在的总体,因而采用假设检验来评价组间协变量的均衡性并不合理,为此提出了评价协变量均衡性的统计方法应该遵守的两个原则:一是评价方法应该是针对样本而不是总体;二是检验统计量应该与样本量大小无关。如使用假设检验统计量作为评价均衡性的指标,可能仅仅因为匹配后的样本量变小而导致P值增大^[4]。基于以上考虑,国外学者提出一些评价组间协变量均衡性的方法,包括标准差异(standardized difference)^[9]、方差比法和图示法等^[10]。也有研究提供了假设检验和标准差异两种方法的评价结果^[11,12]。标准差异在近年的研究中应用较多^[13,14],对于连续性变量,其定义是:

$$d = \frac{100 \times |\bar{x}_{treatment} - \bar{x}_{control}|}{\sqrt{\frac{s^2_{treatment} + s^2_{control}}{2}}}$$

式中, $\bar{x}_{treatment}$ 和 $\bar{x}_{control}$ 分别表示暴露组和对照组某变量的均值, $s^2_{treatment}$ 和 $s^2_{control}$ 分别表示暴露组和对照组某变量的方差。对于分类变量,其定义是:

$$d = \frac{100 \times |p_{treatment} - p_{control}|}{\sqrt{\frac{p_{treatment}(1-p_{treatment}) + p_{control}(1-p_{control})}{2}}}$$

式中, $p_{treatment}$ 和 $p_{control}$ 分别表示暴露组和对照组某变量的率。一般认为,当标准差异小于10%时,认为组间变量的均衡性较好^[11]。

方差比法主要是针对连续性变量而言,通过组间协变量方差的比较来衡量均衡性。图示法是通过Q-Q图直观地表示,缺点是不能量化。但是不管选用哪种方法,变量基本的描述性统计都是不可或缺的。多分组资料组间协变量的均衡性评价方法还未提出,有待进一步研究。

本系列讲座第一讲中介绍的倾向指数方法应用实例——探讨口服钾和心力衰竭患者死亡率、住院率的关系,就是采用标准差异法来评价组间协变量的均衡性^[13],文献中以图1的方式直观地表示组间协变量应用倾向指数方法前后的均衡性(匹配前34个协变量中有15个变量的标准差异大于10%,匹配后34个协变量的标准差异均小于10%,其中大部分协变量的标准差异小于5%),也就是说,通过倾向指数匹配,达到了均衡组间协变量的目的。

3. 处理效应的估计:倾向指数方法的主要目的就是减少处理效应估计的偏倚,选用何种方法来估计处理效应,显得至关重要。如本讲第一部分所述,倾向指数模拟研究常通过比较处理效应的偏性和精度来评价建模的合理性。观察性研究中,暴露组和对照组间协变量分布一般不均衡,应用倾

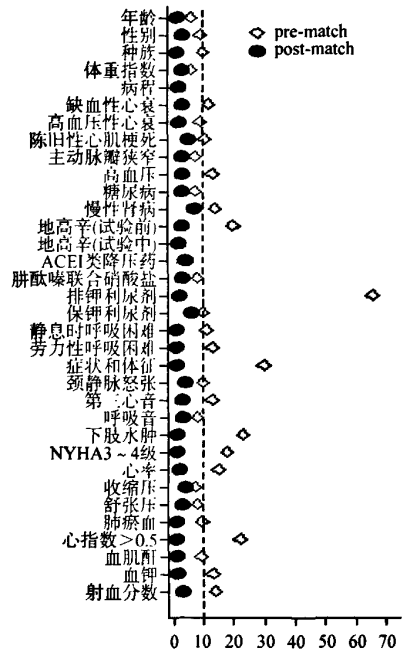


图1 匹配前后标准差异比较(%) (引自:Ekundayo等^[13])

向指数方法后,根据 Rosenbaum 和 Rubin^[15]的假设,“经过倾向指数方法调整后,暴露因素和倾向指数代表的一组协变量之间的条件是独立的”,这组协变量在暴露组和对照组匹配个体之间的分布应当均衡,个体之间有配对特征,暴露组和对照组也不再是两个独立的样本,所以在选择估计处理效应的方法时,必须考虑到样本的配对特征。如对于连续性变量资料可以根据数据分布类型的不同,选用配对t检验或者Wilcoxon符号秩检验;对二分类变量,比较绝对危险度,可以采用 McNemar 检验;比较相对危险度,可以采用 Agresti 和 Min^[16]提出的分析方法;对于生存数据可以采用 Cox 比例风险模型进行分析^[17]。

4. 灵敏度分析:灵敏度分析用来评价潜在的未知混杂因素对研究结论的影响^[18]。经过倾向指数调整,均衡了组间协变量分布,排除了可观测变量引起的偏倚,但是不能排除可能有潜在的混杂因素引起的偏倚,因此,需要用灵敏度分析来量化隐藏偏倚。灵敏度分析虽然不能确定是否存在潜在的混杂因素,但是可以估计隐藏偏倚对研究结论的影响。不同的研究对偏倚的敏感性存在差异,对偏倚不敏感并不表示偏倚不存在,反之亦然^[19]。倾向指数灵敏度分析常用的方法是 Wilcoxon 符号秩检验。

倾向指数方法与随机化方法的区别在于,随机化方法均衡了组间已知以及未知的所有因素,而倾向指数方法仅仅均衡了组间的已知因素。在理论上,如果倾向指数模型纳入了所有影响结果的因素,则通过匹配后得到的数据与经过随机化得到的数据性质应相同。但是实际上,这并不可能,也无法证明在倾向指数模型中纳入了所有应当纳入的因素,所以倾向指数方法只是提供了一种在现有条件下能够做到的最好途径。

参 考 文 献

[1] Rubin DB, Thomas N. Matching using estimated propensity scores: relating theory to practice. *Biometrics*, 1996, 52: 249-264.

[2] Perkins SM, Tu W, Underhill MG, et al. The use of propensity scores in pharmacoepidemiologic research. *Pharmacoepidemiol Drug Safety*, 2000, 9: 93-101.

[3] Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Pract Epidemiol*, 2006, 163: 1149-1156.

[4] Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med*, 2007, 26: 734-753.

[5] Wahle C, Adamopoulos C, Ekundayo OJ, et al. A propensity-matched study of outcomes of chronic heart failure (HF) in younger and older adults. *Arch Gerontol Ger*, 2009, 49: 165-171.

[6] Akazawa M, Biddle AK, Stearns SC. Economic assessment of early initiation of inhaled corticosteroids in chronic obstructive pulmonary disease using propensity score matching. *Clin Therap*, 2008, 30: 1003-1015.

[7] Adamopoulos C, Pitt B, Sui XM, et al. Low serum magnesium and cardiovascular mortality in chronic heart failure: a propensity-matched study. *Int J Cardiol*, 2009, 136: 270-277.

[8] Ekundayo OJ, Markland A, Lefante C, et al. Association of diuretic use and overactive bladder syndrome in older adults: a propensity score analysis. *Arch Gerontol Ger*, 2009, 49: 64-68.

[9] Flury BK, Reidwyl H. Standard distance in univariate and multivariate analysis. *Am Statistician*, 1986, 40: 249-251.

[10] Ho DE, Imai K, King G, et al. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit Anal*, 2007, 15: 199-236.

[11] D'Agostino RB. Propensity score methods for bias reduction in

the comparison of a treatment to a non-randomized control group. *Stat Med*, 1998, 17: 2265-2281.

[12] Yea Y, Kaskutas LA. Using propensity scores to adjust for selection bias when assessing the effectiveness of alcoholics anonymous in observational studies. *Drug Alcoh Depn*, 2009, 104: 56-64.

[13] Ekundayo OJ, Adamopoulos C, Ahmed MI, et al. Oral potassium supplement use and outcomes in chronic heart failure: a propensity-matched study. *Int J Cardiol*, 2010, 141(2): 167-174.

[14] Ekundayo OJ, Dell'italia LJ, Sanders PW, et al. Association between hyperuricemia and incident heart failure among older adults: a propensity-matched study. *Int J Cardiol*, 2009 (Epub ahead of print).

[15] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, 70: 41-55.

[16] Agresti A, Min Y. Effects and non-effects of paired identical observations in comparing proportions with binary matched-pairs data. *Stat Med*, 2004, 23: 65-75.

[17] Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. *J Thorac Cardiovasc Surg*, 2007, 134: 1128-1135.

[18] Rosenbaum PR. Sensitivity analysis for matching with multiple controls. *Biometrika*, 1988, 75: 577-581.

[19] Rosenbaum PR. Discussing hidden bias in observational studies. *Ann Intern Med*, 1991, 115: 901-905.

(收稿日期: 2009-12-28)

(本文编辑: 张林东)

[编者结语: 至此本系列讲座已全部刊出。近年来越来越多的学者将倾向指数(或译为倾向评分)方法应用到流行病学、卫生服务以及经济学、社会科学领域, 有兴趣的读者可与本系列讲座的通信作者联系]

· 消 息 ·

中华医学会系列杂志标注数字对象惟一标识符

数字对象惟一标识符(digital object identifier, DOI)是对包括互联网信息在内的数字信息进行标识的一种工具。

为了实现中华医学会系列杂志内容资源的有效数字化传播, 同时保护这些数字资源在网络链接中的知识产权和网络传播权, 为标识对象的版权状态提供基础, 实现对数字对象版权状态的持续追踪, 自 2009 年第 1 期开始, 中华医学会系列杂志纸版期刊和数字化期刊的论文将全部标注 DOI。即中华医学会系列杂志除科普和消息类稿件外, 其他文章均需标注 DOI, DOI 标注于每篇文章首页脚注的第 1 项。由中华医学会杂志社各期刊编辑部为决定刊载的论文标注 DOI。

参照 IDF 编码方案(美国标准 ANSI/NISO Z39.84-2000)规定, 中华医学会系列杂志标注规则如下: “DOI: 统一前缀/学会标识. 信息资源类型. 杂志 ISSN. ****-****. 年期. 论文流水号”。即: “DOI: 10.3760/cma.j.issn. ****-****. yyyy.nn.zzz”。

中华医学会系列杂志标注 DOI 各字段释义: “10.3760”为中文 DOI 管理机构分配给中华医学会系列杂志的统一前缀; “cma”为中华医学会(Chinese Medical Association)缩写; “j”为 journal 缩写, 代表信息资源类别为期刊; “issn. ****-****”为国际标准连续出版物号(ISSN); “yyyy”为 4 位出版年份; “nn”为 2 位期号; “zzz”为 3 位本期论文流水号。