

倾向指数

第一讲 倾向指数的基本概念和研究步骤

王永吉 蔡宏伟 夏结来 蒋志伟

【关键词】 倾向指数; 观察性研究; 统计学方法

Scores on propensity (I) Conception and measures WANG Yong - ji¹, CAI Hong - wei², XIA Jie - lai¹, JIANG Zhi - wei¹.

1 Department of Health Statistics, Faculty of Preventative Medicine, Fourth Military Medical University, Xi'an 710032, China; 2 Information Center, School of Stomatology, Fourth Military Medical University

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 (No. 30800952)

【Key words】 Propensity score; Observational studies; Statistical method

随机对照试验(RCT)被认为是临床试验的金标准,但受样本、时间以及费用等因素的影响,RCT研究的疾病种类、应用范围受到很大限制;而且由于伦理学因素的影响,参加试验的研究对象需要经过纳入、排除标准的筛选,使其结论外推受到限制。大样本的观察性研究是RCT的有益补充,但是选择偏倚以及重要预后因素在组间分布的不均衡性是困扰观察性研究的主要问题^[1]。传统多元模型、工具变数(instrumental variable)和倾向指数(propensity score)等方法是针对该问题的常用研究方法。由于倾向指数方法易于理解、研究步骤标准化程度高,近些年来在欧美被广泛应用于大样本、非随机的观察性研究中^[2,3]。

1. 基本概念:倾向指数的概念由Rosenbaum和Rubin在1983年首次提出^[4],它是多个协变量的一个函数,通过倾向指数的变化可以表示多个协变量共同作用的结果,来均衡化处理组和对照组间的协变量分布。该方法主要用于观察性研究组间混杂因素的事后均衡,对观察性研究的非研究混杂因素进行类似随机化的均衡处理,减少选择偏倚。

倾向指数的定义:在给定一组变量(x_i)条件下,将任意一个研究对象($i=1,2,\dots,N$)划分到处理组($Z_i=1$)或者对照组($Z_i=0$)的条件概率^[4]。

第*i*个研究对象被分配到处理组的概率可以表达为:

$$e(x_i) = P(Z_i = 1 | x_i)$$

假定分组变量 Z_i 和特征变量 x_i 相互独立,则对任意一个观察变量 x_i :

$$P(z_1, \dots, z_n | x_1, \dots, x_n) = \prod_{i=1}^n e(x_i)^{z_i} \{1 - e(x_i)\}^{1-z_i}$$

其中, $e(x_i)$,即 P 被称为倾向指数^[4]。

在完全随机化情况下,研究对象的分组与自身特征变量无关(临床干预因素是随机分组后施加的),因此在分为两组时,每名病例的倾向指数均为0.5。但是对于观察性研究,由于某些因素的影响,一些受试者更倾向于进入处理组或对照组。倾向指数的研究方法认为,假如某个受试对象进入处理组的倾向指数 $e(x_i) = 0.2$,此时恰好有另外一名受试者,虽然拥有不同的特征变量,但如果其进入处理组的倾向指数也是 $e(x_i) = 0.2$,就认为该倾向指数代表的多个协变量在它们之间是相同的。如果将倾向指数相同或相近的研究对象在不同的组间进行匹配,则在总体上处理组和对照组各个特征变量的分布是均衡的。也就是说,不同处理组之间混杂因素的不均衡性对研究结果的干扰被抵消了^[4]。倾向指数是所有协变量的一个函数,在大样本情况下,经过倾向指数调整的组间个体,除了处理因素和结果变量分布不同外,其他协变量应当均衡可比,相当于“事后随机化”,使观察性数据达到“接近随机分配数据”的效果。应用较多的倾向指数方法包括匹配(matching)、分层(stratification)和回归校正(regression adjustment)等^[5,6]。

2. 倾向指数研究现状:在流行病学研究和日常的临床实践中,大多数数据都是观察性数据,这些数据虽然没有RCT获得的数据整齐、规范,但数据最能代表在医疗实践中实际存在的病例谱和疾病谱,且获取成本低。国外已经将倾向指数方法应用在医学和经济学等领域研究中,而且近年来使用倾向指数方法进行研究的文献呈不断增长的趋势。通过Pub Med检索“propensity score”,2000年之前共有72篇,而2009年1—10月就达到344篇。另外截至目前,在New England Journal of Medicine上共发表使用倾向指数方法的相关文章69篇。国内关于倾向指数方法的研究较少,从“中国学术期刊全文数据库”中只检索到相关文献4篇,主要是倾向指数方法介绍和模拟测试,未检索到在国内运用倾向指数方法研究观察性临床数据的实际应用。

3. 研究步骤:①根据临床经验和实际要求,以处理因素为应变量,混杂因素为自变量来构建logistic或者probit模型;②由大量观察性数据拟合模型的参数;③根据拟合的模

DOI: 10.3760/cma.j.issn.0254-6450.2010.03.026

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),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

型计算每个个体的倾向指数,指数范围为 0~1 之间,反映个体被分到某组的概率;④以倾向指数为依据,通过倾向指数匹配或分层等方法来均衡组间协变量的分布;⑤选择合适的方法来评价倾向指数方法应用前后组间协变量的均衡性,均衡性评价是衡量倾向指数方法应用效果的重要指标;⑥根据数据类型选择相应的传统统计方法进行分析,估计处理效应,但是要考虑到应用倾向指数方法后样本的配对特征。

4. 实例应用:以一篇近期发表的文献为例^[7],阐述倾向指数方法的应用步骤。研究背景:低钾血症在心力衰竭患者中很常见,临床上常以口服钾来维持心力衰竭患者的钾水平,但是口服钾对患者是否有不良影响还不明确。研究目的是评价口服钾和心力衰竭患者死亡率以及再次入院率之间的关系。研究样本为 7788 例心力衰竭患者,其中 2199 例(28%)接受口服钾治疗,5589 例(72%)未接受口服钾治疗。

试验前并没有针对是否接受口服钾治疗进行随机分组,所以组间协变量分布可能不均衡,为了减少这种不均衡对评价口服钾和死亡率以及再次入院率之间可能存在的偏倚,该文献采用倾向指数匹配法来均衡组间协变量。研究步骤:①以是否接受口服钾治疗为应变量,种族、年龄等 34 个混杂因素为自变量来构建 logistic 回归模型。②根据研究样本 7788 例心力衰竭患者来拟合 logistic 模型参数。③根据 logistic 回归模型计算每例患者的倾向指数,该研究中倾向指数的含义是在现有协变量的条件下,患者接受口服钾治疗的概率。④采用倾向指数最近匹配法进行组间个体配对,以口服钾组为基准组进行 1:1 匹配,匹配成功比例为 97%(2199 例中的 2131 例匹配成功),匹配后的样本量为 4262 例。⑤采用国外比较推荐的方法(标准化差异)比较匹配前后组间协变量的均衡性,评价匹配效果。该研究的 34 个混杂因素中,在使用倾向指数方法匹配前,有 15 个是不均衡的。在匹配后 34 个混杂因素在组间均达到均衡,其中大部分变量均衡效果良好,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。⑥匹配完成后,根据匹配数据集,采用配对 Cox 回归来分析口服钾和心力衰竭死亡率以及再次入院率之间的关系,结论是不能说明口服钾和心力衰竭患者死亡率有关联,但提示口服钾治疗可能会增加患者的再次入院率^[7]。

5. 适用范围和局限性:该方法的优点,一是能减少非随机观察性研究中的选择性偏倚。通过倾向指数方法调整来均衡处理组和对照组间的协变量分布,减少估计处理效应时的选择性偏倚。二是通过倾向指数调整组间的混杂因素,使海量的临床观察性数据可以成为循证医学的诊疗证据,而这些数据获取成本低,更能够反映医疗实践中实际存在的疾病谱。三是该方法适用于混杂因素很多,而结局变量发生率很低的情况,而传统多元模型并不适合。四是在无法实现随机化的药物临床试验以及医疗器械临床试验中,可以通过倾向指数方法均衡组间混杂因素;另外,在意向性治疗

(intention-to-treat, ITT)分析中,综合考虑脱落病例的基线水平与结局发生情况,采用倾向指数的方法对其完成临床试验的条件概率进行估计并纳入 ITT 分析,与传统分析中对于脱落病例只采用末次观察推进法(last observation carried forward)进行数据接转完成 ITT 分析相比,对外部人群具有更强的外推性^[8]。

倾向指数方法也有局限。一是该方法只能均衡可观测变量,对潜在的未知混杂因素引起的偏倚无能为力^[9];二是样本量较小时,即使通过倾向指数方法调整,组间协变量的分布也不能达到满意的均衡效果;三是如果匹配后样本占匹配前样本的比例过小,会改变样本构成,从而影响对处理效应的估计;四是当处理组和对照组倾向指数没有重叠或者重叠范围较少时,组间缺乏可比性,无法进行合适的匹配^[10]。总之,倾向指数方法在大样本观察性临床研究中的应用日益广泛,但在运用时,要考虑其适用范围。

参 考 文 献

- [1] Stürmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*, 2006, 59:437-447.
- [2] Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med*, 2008, 27:2037-2049.
- [3] 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.
- [4] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, 70:41-55.
- [5] Katherine HH, Thomas AL. Propensity score modeling strategies for the causal analysis of observational data. *Biostatistics*, 2002, 2:179-193.
- [6] Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. *Stat Med*, 2006, 25:2084-2106.
- [7] 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*, 2009.
- [8] 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.
- [9] Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Prac Epidemiol*, 2006, 163:1149-1156.
- [10] Perkins SM, Tu W, Underhill MG, et al. The use of propensity scores in pharmacoepidemiologic research. *Pharmacoepidemiol Drug Safety*, 2000, 9:93-101.

(收稿日期:2009-09-11)
(本文编辑:张林东)