

# 基于个体的标准化法——倾向评分加权

李智文 刘建蒙 任爱国

**【导读】** 倾向评分加权是利用倾向评分值对每个观察单位进行加权调整。由于倾向评分将许多协变量综合为一个变量,因此通过倾向评分加权可以使各混杂变量在两组人群中的分布趋于一致。根据调整后标准人群的不同分为两种加权方法:逆处理概率加权法(IPTW)和标准化死亡比加权法(SMRW)。本文实例分析表明,用IPTW和SMRW加权调整后处理组和对照组妇女各混杂变量的分布均趋于一致,两种方法调整后的效应估计基本相同。本文介绍倾向评分加权法的基本原理、具体方法,并结合实例探讨了其在流行病学中的应用。

**【关键词】** 倾向评分加权; 标准化法; 混杂偏倚

**Introduction to an individual-based standardization method — propensity score weighting** LI Zhi-wen, LIU Jian-meng, REN Ai-guo. Institute of Reproductive and Child Health, Peking University, Beijing 100191, China

Corresponding author: REN Ai-guo, Email: renaiguo@gmail.com

**【Introduction】** In this article, we presented the rationale and calculation procedures of a propensity score weighting method, with its application in epidemiological studies. The rationale for propensity score weighting method is similar to those for traditional standardization methods. Propensity score is used to estimate the weight for each individual. As the propensity score serves the function of observed covariates, the propensity score weighting can balance the distribution of the observed covariates between the comparison groups. There are two weighting methods according to the target standard populations: the Inverse probability of treatment weighting (IPTW) and the Standardized mortality ratio weighting (SMRW). Results of the example show that the distribution of the covariates tended to be consistent after weighting, and the IPTW and SMRW methods showed similar effect estimates. Propensity score weighting method can effectively balance the distribution of the confounding factors between the compared groups in non-randomized controlled trials.

**【Key words】** Propensity score weighting; Standardization method; Confounding bias

倾向评分(propensity score, PS)方法是新出现的一种处理非随机对照研究中混杂偏倚的有力工具<sup>[1-5]</sup>。该方法能够有效地均衡各对比组间协变量的分布,在组间协变量均衡的基础上进一步评价处理因素的效应,从而得到接近随机对照研究的效果。倾向评分法主要包括倾向评分分配比、倾向评分分层、倾向评分回归调整和倾向评分加权<sup>[4,5]</sup>。本文主要介绍倾向评分加权及其应用。

## 基本原理

倾向评分是指在一定协变量(混杂变量)条件下,一个观察对象可能接受某种处理(或暴露)因素的可能性<sup>[1-5]</sup>。在非随机对照研究中,处理组与对照组某些背景特征分布不同,个体是否具有“处理”这

一特征的概率受其他混杂变量的影响。倾向评分是反映所有观察到的混杂变量在两组间均衡性的一个近似函数。它最大限度地概括了混杂变量的作用,因而可以有效地保持处理组和对照组间混杂变量的均衡性,使两组间各个特征变量均衡一致。倾向评分值相同的两个个体,其协变量的分布也趋于一致,此时可以认为这两个个体背景条件相同,接受“处理”因素与否接近随机<sup>[1]</sup>。在混杂变量没有缺失值的情形下,倾向评分可以通过将处理因素作为应变量,其他混杂因素作为自变量建立模型来估计。logistic回归和判别分析是常用的模型。此外Probit回归、非参数回归、半参数回归等可计算概率的模型也可以用来计算倾向评分值<sup>[6]</sup>。

倾向评分加权(propensity score weighting)是边缘结构模型(marginal structural models)的一种简单应用,其基本原理实际上与传统的标准化法(standardization method)类似<sup>[7-9]</sup>。标准化法常用于

内部构成不同的两个或多个率的比较。标准化法的基本思想就是指定一个统一的“标准人口”，按指定“标准人口”计算调整率，使之具备可比性之后再比较，以消除比较组之间由于内部混杂因素构成不同对总率比较的影响。如在比较两组人群的死亡率时，年龄往往是重要的混杂因素，老年人口的死亡率高于青年人口的死亡率。如果两组人群年龄构成不同，就不宜直接比较两组人群的总死亡率，而应统一各组的内部构成，使资料间具有可比性后再进行比较。但当比较组中需要调整的混杂因素较多时或混杂因素为连续性变量时，传统的标准化法由于需要分层太多，可能导致某些层内没有观察对象而无法调整。倾向评分将许多变量综合为一个变量，通过调整倾向评分这一个变量来使各混杂变量在两组人群中分布趋于一致，因此简化了调整过程<sup>[7-9]</sup>。该方法将每一观察单位看作一层，不同倾向评分值预示这一观察单位在两组中的概率的不同。假定不存在未识别的混杂因素，加权调整是基于两种在一定条件下反事实事件的概率对数据进行调整：即假设每个观察对象均接受处理因素和均不接受处理因素两种情况<sup>[10]</sup>。利用倾向评分估计的权重对各观察单位加权产生一个虚拟的标准人群，在该人群中，两组混杂因素趋于一致，均近似于某一预先选定的标准人口分布。

调整方法因选择的目标标准人群不同而分为两种加权方法：逆处理概率加权法 (inverse probability of treatment weighting, IPTW) 和标准化死亡比加权法 (standardized mortality ratio weighting, SMRW)。IPTW 是以所有观察对象 (处理组与对组合并的人群) 为“标准人群”进行调整。Robins 等<sup>[7]</sup> 给出的加权系数 (W) 计算方法是：处理组观察单位的权数  $W_t = 1/PS$ ，对照组观察单位的权数  $W_c = 1/(1 - PS)$ 。PS 为观察单位的倾向评分值。此方法得到的人群往往与原来人群的数量不同，因此虚拟人群各变量的方差大小可有变化。为了得到与原人群的样本量相同的标准人群，Hernan 等<sup>[9]</sup> 对计算方法进行调整，将整个研究人群的处理率和非处理率加入公式调整后得到稳定权数 (stabilized weights)。具体方法是：处理组观察单位的权数  $W_t = Pt/PS$ ，对照组观察单位的权数  $W_c = (1 - Pt)/(1 - PS)$ 。Pt 为整个人群中接受处理因素的比例。SMRW 是将处理组观察对象作为“标准人群”进行调整。Sato 和 Matsuyama<sup>[8]</sup> 给出的加权系数计算方法是：处理组观察单位的权数  $W_t = 1$ ，对照组观察单位的权数  $W_c =$

$PS/(1 - PS)$ 。同样，由于对照组得到的人群数与处理组样本量一致，而与原对照组样本量不同，因此需要进一步作如下调整，计算其稳定权数： $W_c = [PS(1 - Pt)] / [(1 - PS) Pt]$ 。当每一个观察单位的权数计算出来后，就可以对每个观察单位加权后用传统的方法 (如直接效应比较或 logistic 回归) 进行效应估计。

### 实例分析

为了研究孕期增补某种营养素是否对妊娠高血压有预防作用，研究者采用非随机干预方法，在我国部分地区募集孕早期妇女。按照知情选择的原则，妇女决定服用或不服用这种营养素。研究期间共募集并随访到 57 333 名妇女的妊娠结局。根据专业知识确定需要平衡的混杂因素有母亲年龄、孕早期 BMI、文化程度、职业、是否初产、多胎妊娠和民族。表 1 为服用和未服用营养素的妇女各种混杂因素的分布，可见大部分变量在两组中的分布极不平衡。

表 1 加权调整前服用和未服用营养素的妇女各种协变量的分布

变量	服用组 (n=36 271)		未服用组 (n=21 062)		F 或 χ <sup>2</sup> 值	P 值
	n	$\bar{x} \pm s$ 或 %	n	$\bar{x} \pm s$ 或 %		
母亲年龄(岁)	36 271	23.02±2.04	21 062	23.43±2.65	437.9	<0.001
孕早期 BMI	36 271	20.53±2.28	21 062	21.11±2.50	801.9	<0.001
文化程度					623.1	<0.001
小学及以下	5 560	15.3	3 125	14.8		
初中	26 660	73.5	14 018	66.6		
高中及以上	4 051	11.2	3 919	18.6		
职业					940.4	<0.001
农民	13 443	37.1	10 307	48.9		
工人	16 154	44.5	6 832	32.4		
其他	6 674	18.4	3 923	18.6		
非初产	1 670	4.6	2 335	11.1	861.7	<0.001
多胎妊娠	215	0.590	108	0.51	1.52	0.22
非汉族	248	0.68	133	0.63	0.55	0.46

分别用 IPTW 和 SMRW 两种倾向评分加权方法对上述因素进行调整，然后分析增补该营养素与妊娠高血压发生危险性之间的关系。

首先以是否服用营养素为因变量、上述混杂变量为自变量进行 logistic 回归分析，求出每个研究对象的倾向评分值。根据对照组和处理组的人数求出人群中暴露者的比例 (Pt)， $Pt = 36 271 / (36 271 + 21 062) = 0.633$ 。计算 IPTW 调整稳定权数，例如，本研究中第一个观察对象属于对照组，其 PS 值为 0.771 08，则  $W = (1 - Pt) / (1 - PS) = (1 - 0.633) /$

$(1-0.771\ 08)=1.6032$ 。本例中第九个观察对象属于处理组,其PS值为0.708 08,则 $W=Pt/PS=0.633/0.708\ 08=0.8940$ 。

SMRW调整过程基本相同,只是在计算权重数时候不同。表2和表3分别为用IPTW和用SMRW加权调整后处理组和对照组妇女各混杂变量的分布。可以看到调整后两组人群各变量分布趋于一致。不同的是IPTW调整后,混杂变量分布与合并后的总人口一致,而SMRW调整后,对照组协变量分布与处理组一致。最后以加权变量对每一个观察单位加权后,分析数据的均衡性。

表2 IPTW加权调整后服用和未服用营养素的妇女各种协变量的分布

变量	服用组*		未服用组*		F或χ <sup>2</sup> 值	P值
	n	$\bar{x}\pm s$ 或%	n	$\bar{x}\pm s$ 或%		
母亲年龄(岁)	36 325	23.18±2.26	21 056	23.16±2.40	0.39	0.53
孕早期BMI	36 325	20.78±2.70	21 056	20.74±2.33	3.20	0.07
文化程度					0.11	0.95
小学及以下	5 520	15.2	3 198	15.2		
初中	25 779	71.0	14 965	71.1		
高中及以上	5 025	13.8	2 893	13.7		
职业					0.17	0.92
农民	15 008	41.3	8 669	41.2		
工人	14 582	40.1	8 489	40.3		
其他	6 734	18.5	3 898	18.5		
非初产	2 565	7.1	1 476	7.0	0.05	0.82
多胎妊娠	203	0.56	115	0.55	0.04	0.84
非汉族	240	0.66	139	0.66	0	0.99

注: \*由于小数取整,小组合计可能不等于总数

表3 SMRW加权调整后服用和未服用营养素的妇女各种协变量的分布

变量	服用组*		未服用组*		F或χ <sup>2</sup> 值	P值
	n	$\bar{x}\pm s$ 或%	n	$\bar{x}\pm s$ 或%		
母亲年龄(岁)	36 271	23.02±2.04	21 052	23.01±2.23	0.41	0.52
孕早期BMI	36 271	20.53±2.28	21 052	20.53±2.20	0.00	0.99
文化程度					0.87	0.65
小学及以下	5 560	15.3	3 240	15.4		
初中	26 660	73.5	15 514	73.7		
高中及以上	4 051	11.2	2 298	10.9		
职业					0.94	0.62
农民	13 443	37.1	7 719	36.7		
工人	16 154	44.5	9 449	44.9		
其他	6 674	18.4	3 884	18.4		
非初产	1 670	4.6	978	4.6	0.05	0.82
多胎妊娠	215	0.590	118	0.56	0.24	0.63
非汉族	248	0.68	142	0.67	0.02	0.90

注: \*同表2

最后,可以用单因素分析方法直接比较,也可以用多因素分析方法同时调整重点变量。本文采用logistic回归方法。结果表明,采用倾向评分加权调

整后,尚不能认为孕早期增补这种营养素可以降低妊娠高血压综合征的风险。本文IPTW和SMRW调整结果基本相同(表4)。

表4 不同调整方法计算的OR值

调整方法	OR值 95%CI
未调整	1.04(0.98 ~ 1.10)
直接logistic回归	1.05(1.00 ~ 1.12)
IPTW	1.05(1.00 ~ 1.11)
SMRW	1.05(1.00 ~ 1.11)

## 讨 论

传统标准化法以混杂变量将人群分层,然后调整使两组各层人群的比例趋于一致以达到可比。当对比组中需要调整的混杂因素较多时或为连续性变量时,该法由于需要分层太多而可能导致某些层面没有观察对象而无法调整。倾向评分由于将许多变量综合为一个变量,通过调整倾向评分这一个变量来使各混杂变量在两组人群中的分布趋于一致。因此可以同时调整较多的混杂因素或连续变量而并不增加分析的难度。此外,倾向评分加权不但能够对计数资料(如率)进行标化,且可应用于计量资料(如均数)的标化。

如果倾向评分估计和多变量效应估计所用的协变量和模型相同,则直接用各协变量进行调整后的效应点值应该与用倾向评分调整后的效应点值相同。但倾向评分的优势是研究者可以首先构建复杂的模型(如纳入较多的变量或增加复杂多级交互项)来计算倾向评分,然后在最后的效应模型中使用少量的重点变量与倾向评分共同调整。由于倾向评分综合了全部混杂因素的共同作用,将众多的因素综合为一个变量,简化了最终估计因果联系的模型,对模型的诊断比同时纳入较多变量要容易和可靠;同时避免了效应估计时参数过多及共线性的问题所导致的偏差<sup>[3]</sup>。

同其他倾向评分方法一样,倾向评分加权也有局限性,应用过程中值得注意。首先该方法是有前提假设的,即假设在现有观察变量条件下,处理因素的分配是随机的。这种假设又称为强烈可忽略假设(strongly ignorable)<sup>[3]</sup>。因此该方法只能调整观察到的变量,而不能像随机化那样同时平衡所有变量的分布,包括观察到的和未观察到的变量。此外,该方法需要较大的样本量,且要求比较组间倾向评分有足够的重叠范围,否则无法做出有效的平衡。在某些情况下,对于样本含量较小的研究或混杂变量组间差异过大的研究(倾向评分重叠范围

小),即使使用倾向评分法,也可能仍然无法消除组间的不均衡性。

一般情况下,选择 IPTW 和 SMRW 两种方法调整混杂因素的结果基本一致。但当混杂因素对处理效应具有较强的效应修饰作用时, IPTW 和 SMRW 两种方法的调整结果之间将存在较大的差异<sup>[10,11]</sup>。如 Kurth 等<sup>[11]</sup>在研究使用组织纤维蛋白溶解酶原活化剂(tissue plasminogen activator)与缺血性脑卒中患者死亡危险性的关系时发现:如果不调整混杂因素,其  $OR=3.35$  (95%CI: 2.28 ~ 4.91),用 IPTW 调整  $OR=10.77$  (95%CI: 2.47 ~ 47.04),而用 SMRW 调整  $OR=1.11$  (95%CI: 0.67 ~ 1.84),两者相差约 10 倍。究其原因,是由于部分混杂因素存在较强的效应修饰作用,通过倾向评分分层可以发现各层  $OR$  值存在较大差别。在这种情况下, SMRW 调整的  $OR$  值与倾向评分配比及随机对照研究的结果相似。因为,倾向评分配比和 SMRW 均以处理组作为参照,而随机对照研究由于规定了部分入选条件,其研究对象也趋于与处理组一致。而 IPTW 是以整个人群为参照,更全面地考虑了一般人群的特征,因此在效应估计上可能不及前面几种方法稳定,但在识别效应修饰因子或没有纳入的重要变量或交互项方面则具有较大优势<sup>[11]</sup>。虽然 SMRW 与倾向评分配比的平衡结果基本一致,但 SMRW 在数据分析阶段更具优势,这是因为:① 倾向评分配比只是选择了部分对照个体,而 SMRW 利用了全部对照个体的信息,其方差与原人群相近;② SMRW 过程要比倾向评分配比过程容易实现<sup>[10]</sup>。

近年来,国际上越来越多的研究者将倾向评分法应用到流行病学、卫生服务以及经济学、社会科学等领域。倾向评分加权法作为倾向评分应用的一种重要手段,在平衡混杂作用方面将具有广阔的应用前景。

## 参 考 文 献

- [1] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, 70:41-55.
- [2] Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*, 1985, 39:33-38.
- [3] D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*, 1998, 17(19):2265-2281.
- [4] Zhao SJ, Zhang Y, Wang XY, et al. An effective method to reduce bias between two compared groups: propensity score. *Chin J Epidemiol*, 2003, 24(6):516-519. (in Chinese)  
赵守军, 张勇, 汪贻怡, 等. 均衡组间差异的有效方法:倾向评分. *中华流行病学杂志*, 2003, 24(6):516-519.
- [5] Li ZW, Liu JM, Zhang L, et al. Application of propensity score matching in designing of the epidemiological study. *Chin J Epidemiol*, 2009, 30(5):514-517. (in Chinese)  
李智文, 刘建蒙, 张乐, 等. 倾向评分配比在流行病学设计中的应用. *中华流行病学杂志*, 2009, 30(5):514-517.
- [6] Baser O. Too much ado about propensity score models? Comparing methods of propensity score matching. *Value Health*, 2006, 9(6):377-385.
- [7] Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 2000, 11(5):550-560.
- [8] Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology*, 2003, 14(6):680-686.
- [9] Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 2000, 11(5):561-570.
- [10] Sturmer T, Rothman KJ, Glynn RJ. Insights into different results from different causal contrasts in the presence of effect-measure modification. *Pharmacoepidemiol Drug Saf*, 2006, 15:698-709.
- [11] Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol*, 2006, 163(3):262-270.

(收稿日期:2009-08-14)

(本文编辑:张林东)

## · 会 讯 ·

### 第二届北京热带医学与寄生虫学论坛通知

为促进我国热带医学与寄生虫学研究的学术交流,首都医科大学附属北京友谊医院、北京热带医学研究所定于 2010 年 5 月 21-23 日在北京举办国家级继续医学教育项目“第二届北京热带医学与寄生虫学论坛”,期满将授予国家级继续教育 I 类学分 6 分。

本届论坛将邀请国内相关领域知名专家共聚一堂,从不同角度切磋热带医学领域的热点课题,以加强交流,增进友谊。欢迎参会人员积极投稿,论文将编入本届论坛汇编。

报名联系人:谷俊朝;电话:010-63025849,63138570;传真:010-63139265; Email:reyansuo2008@sohu.com

首都医科大学附属北京友谊医院  
北京热带医学研究所