

# 倾向评分分配比在流行病学设计中的应用

李智文 张乐 刘建蒙 任爱国

**【导读】** 介绍倾向评分分配比法(PSM)的基本原理、具体方法,并结合实例探讨其在流行病学设计过程中的应用。PSM通过某些观察性研究某些混杂变量与研究因素的关系计算倾向评分,然后从对照组中为处理组每个个体寻找一个或多个倾向评分值相同或非常接近的个体做对照,最终使选取观察对象的混杂变量在处理组和对照组趋于均衡可比。实例分析表明,利用PSM筛选后的研究对象,主要混杂因素在两组中的偏差下降在55%以上。结论:PSM可有效降低观察性研究的混杂偏倚,在流行病学设计阶段使用PSM可使某些观察性研究得到类似随机对照研究的效果。

**【关键词】** 倾向评分分配比;混杂偏倚;流行病学

**Application of propensity score matching in the design of an epidemiological study** Li Zhi-wen, ZHANG Le, LIU Jian-meng, REN Ai-guo. Institute of Reproductive and Child Health, Peking University Health Science Center, Beijing 100191, China

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

**【Introduction】** In this article, we presented the rationale and calculation procedures of the propensity score matching (PSM), and its application in the designing stage of an epidemiological study. Based on existing observational data, PSM can be used to select one or more comparable controls for each subject in 'treatment' group according to the propensity scores estimated by 'treatment' variable and main covariates. The results of an example analysis showed that the bias for main confounders between the treated and control samples declined more than 55% after PMS. Conclusion: PSM can reduce most of the confounding bias of the observational study, and can obtain approximate study effect to the randomized controlled trials when used in the designing of the epidemiological study.

**【Key words】** Propensity score matching; Confounding bias; Epidemiology

在流行病学病因研究中,为了探讨某因素(如暴露或干预,以下统称为处理因素)与结局的关系,需要设立对照进行比较。对照设立的一个重要原则是可比性,也就是说对比组除了所研究的因素之外,其他因素尽可能齐同,这样才能凸显处理因素的效应。随机对照研究通过随机化分配研究对象,使混杂因素在处理组和对照组中的分布趋于平衡,是验证因果联系最理想的设计方法。观察性研究(如队列研究)由于没有随机分配研究对象,常使混杂因素在两组中的分布不均衡,易高或低估暴露效应而引入偏倚。观察性研究对于混杂因素的控制主要在数据的分析阶段采用分层分析或多因素分析,如果对比组之间某些特征差异很大,则可能对研究结果产生较大且无法以分析方法弥补的偏倚。倾向评分分配比(propensity score matching, PSM)可有效降低观察性研究的混杂偏倚,更重要的是该方法可应用于

在流行病学设计阶段,使某些观察性研究得到类似随机对照研究的效果<sup>[1,2]</sup>。在本文中结合实例介绍PSM的基本原理、具体方法及其在流行病学设计中的应用。

## 基本原理

倾向评分(PS)的概念是由Rosenbaum和Rubin于20世纪80年代提出的,是指在一定协变量条件下,一个观察对象可能接受某种处理(或暴露)因素的可能性<sup>[3]</sup>。在观察性研究中,处理组与对照组某些背景特征分布不同,每个个体是否具有“处理”这一特征的概率受其他特征(混杂变量)影响。其概率大小,即PS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,如logistic回归、判别分析等。logistic回归对于自变量没有正态分布的要求,因此更适合于流行病学调查资料中的PS值计算。Rosenbaum和Rubin发现,PS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。PSM 就是从对照组中为处理组每个个体寻找一个或多个 PS 值相同或非常接近的个体做对照,最终两组的混杂变量也趋于均衡可比<sup>[4]</sup>。

PSM 一般分为如下几种类型:

1. 单纯 PSM: 或称 PS 最近法 (nearest available matching on the estimated propensity score), 是 PSM 最基本的方法, 即直接从对照组中寻找出一个或多个与处理组个体 PS 值相同或最相近的个体作为配比对象。具体方法是: 首先根据协变量计算 PS 值, 将处理组观察对象与对照组分开; 然后, 将两组观察对象按照 PS 值大小排序, 从处理组中依次选出一个研究对象, 从对照组中寻找 PS 值与处理组每个对象最相近的一个对象作为配比个体。如果对照组中同时有两个或两个以上 PS 值相同的个体, 则按随机的原则进行选择。配比成功的对象从原人群中移去, 然后进入下一个处理对象的配比过程, 直到处理组中的全部对象完成匹配。配比时一般应规定配比精度, 如 PS 值相差  $<0.01$  或  $<0.001$ 。这一方法类似随机对照研究中的单纯随机分配, 容易实现。

2. 分层 PSM: 单纯 PSM 尽管可以使协变量总体趋于平衡, 但不能保证每个协变量分布完全一致。如果研究者希望某个重要变量精确匹配, 则可以使用分层 PSM。具体方法是: 首先根据该变量 (如性别) 分层后, 分别对每层人群进行单纯 PSM; 然后将配比人群合并, 这样最终的两组研究人群中该变量分布完全相同。这一方法类似于随机化对照研究中的区组随机化, 能够更好地保证最主要变量的平衡性, 但要求精确匹配因素不能太多。

3. 与马氏配比结合的 PSM: 马氏配比是通过矩阵计算两个观察对象的马氏距离的一种匹配办法<sup>[5]</sup>。马氏距离是由印度统计学家 Mahalanobis 提出, 表示  $m$  维空间中两个点之间的协方差距离, 不受量纲的影响, 还可以排除变量之间相关性的干扰。PSM 与马氏配比结合后可以增加个别重点变量平衡的能力。具体结合方法有两种: 一种是把 PS 值作为一个变量同其他重点平衡的变量一起估计马氏距离, 然后进行马氏配比; 另外一种是首先在一定精度的 PM 差值范围内选择对照组中全部可以匹配的对象, 然后根据少数重点变量计算马氏距离, 选择马氏距离最小的对象作为最终的对照。这一方法要求计算马氏距离的变量不能太多, 实现过程比较复杂。

### 实例分析

某医院欲利用一项“妇女孕期妊娠高血压综合

征调查研究”资料进一步随访研究妇女孕早期先兆子痫 (PE) 对儿童某些发育状况的影响。该项研究包括 187 例先兆子痫妇女和 4722 名正常妇女。如果全部随访, 随访任务大。因此, 需要从中选取部分可比的个体进行随访。两组人群一些主要特征的分布存在显著差异, 影响效应的估计, 这些因素包括妇女年龄、怀孕前后 BMI、文化程度、产次 (初产/非初产)、胎数 (单胎/多胎)。现采用单纯的 PSM 和按产次分层 PSM 两种方法选取可比的个体作为随访对象。单纯 PSM 配比首先以 PE 变量 ( $1=PE, 0=正常$ ) 为因变量, 以需要调整的上述变量为自变量构建 logistic 回归模型, 求出每个研究对象的 PM 值。将研究对象按 PE 变量分为两组, 然后用 SPSS 软件程序进行匹配, 配比精度为  $PS < 0.01$ 。为了保证较好的配比效果, 采用精度由高到低 ( $PS < 0.0001 \rightarrow < 0.001 \rightarrow < 0.01$ ) 进行配比。分层 PSM 匹配时, 首先按产次变量将人群分为初产和非初产两组人群, 再以单纯 PSM 方法匹配。配比后的结果见表 1。单纯 PSM 和按产次分层 PSM 分别有 181 和 176 对妇女配比成功, 配比后各变量分布明显趋于一致。利用 Baser<sup>[6]</sup> 推荐的方法对配比效果进行评价: 首先求出每个协变量在配比前后的均数, 无序分类变量需要转换为哑变量进行计算; 然后计算各变量在配比前后均数的偏差; 最后根据偏差计算偏差下降的百分率。从表 2 可见, 单纯 PSM 每个变量偏差下降均在 50% 以上, 其中年龄变量均衡效果最好, 偏差下降达 92.4%; 按照配比分层 PSM 每个变量偏差下降均在 68% 以上, 其中产次在两组中分布完全一致。

### 讨 论

随机对照研究尽管是流行病学病因研究中最理想的设计方法, 但其在人群研究中受到诸多条件的限制, 如实施费用昂贵、医学伦理问题以及某些因素 (如疾病) 无法随机化等。而观察性研究由于不受上述限制而在人群研究中得到广泛应用, 并且随着信息技术的发展, 观察性数据无论是数量还是准确性都在不断增加。因此, 将 PSM 应用于观察性研究的设计阶段将会在人群流行病学病因研究中起到重要作用。通过文献检索可以发现, 国际上越来越多的研究者将 PSM 应用到流行病学、健康服务、经济学以及社会科学诸多领域<sup>[1,7-10]</sup>, 而国内类似的应用研究尚罕见。

队列研究是一种重要的流行病学病因研究方法, 研究者在基线调查的基础上利用 PSM 选择随访

表 1 PSM 前后两组人群主要特征的分布

变 量	配比前			单纯 PSM 配比后			按产次分层配比后		
	PE 组	正常组	P 值	PE 组	正常组	P 值	PE 组	正常组	P 值
人数	187	4722		181	181		176	176	
年龄(岁, $M \pm s$ )	26.3 ± 3.4	24.8 ± 3.0	<0.000 01	26.2 ± 3.3	26.3 ± 3.6	0.75	26.1 ± 3.2	26.0 ± 3.3	0.90
BMI(kg/m <sup>2</sup> , $M \pm s$ )	21.1 ± 2.8	19.9 ± 2.4	<0.000 01	20.9 ± 2.7	21.3 ± 2.8	0.17	20.9 ± 2.6	20.9 ± 2.5	0.95
文化程度*			0.003 6			0.44			0.28
小学及以下	8.0	14.3		7.2	6.6		7.4	9.1	
初中	53.5	56.9		54.7	48.6		54.5	46.0	
高中及以上	38.5	28.8		38.1	44.8		38.1	44.9	
产次*			0.016			0.42			1.00
初产	83.4	75.8		82.9	79.6		83.0	83.0	
非初产	16.6	24.2		17.1	20.4		17.0	17.0	
胎数*			<0.000 01			0.36			0.74
多胎	4.8	0.5		3.9	2.2		2.8	2.3	
单胎	95.2	99.5		96.1	97.8		97.2	97.7	
PS( $M \pm s$ )	0.10 ± 0.04	0.04 ± 0.04	<0.000 01	0.09 ± 0.10	0.09 ± 0.10	0.98	0.08 ± 0.09	0.08 ± 0.09	0.99

注: \* 数据为构成比(%)

表 2 PSM 前后两组人群各变量偏差的下降水平

变 量	配比前			单纯 PSM 配比后			偏差下降 (%)	按产次分层配比后			偏差下降 (%)
	PE 组( $M \pm s$ )	正常组( $M \pm s$ )	偏差	PE 组( $M \pm s$ )	正常组( $M \pm s$ )	偏差		PE 组( $M \pm s$ )	正常组( $M \pm s$ )	偏差	
年龄(岁)	26.29 ± 3.36	24.77 ± 2.96	1.514	26.19 ± 3.33	26.31 ± 3.61	0.115	92.4	26.06 ± 3.25	26.01 ± 3.27	0.044	97.1
BMI(kg/m <sup>2</sup> )	21.07 ± 2.81	19.91 ± 2.36	1.158	20.92 ± 2.68	21.31 ± 2.80	0.394	66.0	20.91 ± 2.65	20.93 ± 2.52	0.017	98.5
文化程度(1/2/3)	2.30 ± 0.61	2.15 ± 0.64	0.160	2.31 ± 0.60	2.38 ± 0.61	0.072	55.1	2.31 ± 0.60	2.36 ± 0.64	0.051	68.1
初产妇(1/2)	1.17 ± 0.37	1.24 ± 0.43	0.077	1.17 ± 0.38	1.20 ± 0.40	0.033	56.8	1.17 ± 0.38	1.17 ± 0.38	0.000	100.0
单胎(1/2)	1.95 ± 0.21	1.99 ± 0.07	0.043	1.96 ± 0.19	1.98 ± 0.15	0.017	61.5	1.97 ± 0.17	1.98 ± 0.15	0.006	86.0

注: 偏差 = 正常组与 PE 组均数差值的绝对值; 偏差下降(%) = (配比前偏差 - 配比后偏差) / 配比前偏差 × 100%

对象,即可以节省随访费用,又可有效降低混杂倚倚。从本文实例中可见,经过 PSM 配比后的研究对象,主要混杂变量基本趋于平衡,各混杂变量偏差下降均在 50% 以上。配比前分布越不平衡的变量,偏差下降的百分比越大。如果采用分层 PSM 或与马氏匹配相结合的方法,则个别重点变量的匹配效果将更好。PSM 还可以用于病例对照研究,如在巢式病例对照研究中用 PSM 从随访人群中选取可比的对照,然后测量病例和对照在某些暴露(如基因型、生化指标等)上的差别。这种情况下,分组变量为疾病有无,协变量是需要平衡的混杂变量。通常情况下,logistic 回归要求因变量只有两种取值(二分类),当因变量的取值有三组及以上时(如血压分正常血压、高血压前期、高血压三组),可以用多分类 logistic 回归分析计算 PS 值。

要想在设计中合理应用 PSM,研究者必须首先明确所获取的资料是否适合进行 PSM 配比。一般来说,PSM 适合于下列几种情况:①处理因素(或病例)在人群中的比例远低于非处理因素(或对照),这

样保证有足够的对照人群可供选择和配比,对照人群越大,配比效果越好;②需要平衡的因素较多,如 March of Dimes 一项对过期产和正常足月产对象的配比研究中对多达 13 个变量进行配比<sup>[7]</sup>;③研究的结局变量的调查难度较大或费用较高,选择部分可比的观察对象无疑会保证研究的可行性和结果的准确性。此外,合理选择配比因素也是值得注意的问题。不能仅仅依据统计学显著性差异来确定配比因素,更重要的是要结合专业知识进行判断。既要避免重要混杂因素的落选,又要注意配比过度。

不管是何种配比方法,均可能涉及个别极端病例无法找到对照的问题,比如本文实例分析中原始数据病例为 187 例,采用单纯和产次分层倾向评分配比后为 181 例。如果处理组中例数减少太多,则会降低检验效能。PSM 配比过程中配比不成功的主要原因包括:①用于计算 PS 的主要协变量有缺失。一个观察单位只要任何一个协变量数据缺失,则无法估计其 PS 值,因而也就无法进行配比。②处理组中部分研究对象 PS 值太大或太小(极端值),较

远地落在了两组倾向评分重叠区范围之外,因此没有可以与之匹配的对照选择。③配比精度过高。在样本量固定的情况下,如果想提高匹配的精确性,必然会减少匹配成功的例数,反之,如果想尽量多保留能够匹配的对象,则匹配精度自然会降低。因此,在配比过程中研究者要结合资料的具体情况、专业知识来选择配比变量及决定匹配的精度和样本量。如果病例或对照配比成功的例数较少,可以适当放宽配比条件,排除配比变量中缺失值较多的变量,以增加配比成功的对象。此外,在同样精度的情况下,采用多阶段配比的办法会增加配比效果。如果希望最终精度为 PM 差值  $< 0.01$ , 那么在进行匹配时先从精度较高水平开始匹配,如  $< 0.0001 \rightarrow < 0.001 \rightarrow < 0.01$ 。

与传统的配比方法相比,PSM 的优势是同时匹配许多混杂因素并不增加匹配的难度和效果。传统分层匹配的方法要根据每个变量取值分层后进行匹配,如果需要平衡的变量个数或水平较多,则分层数成倍增加,往往难以实现。马氏配比是通过计算两个观察对象的马氏距离进行配比,随着配比维数的增加,不但运算量大大增加,而且马氏距离均值也增加,使配比效果下降<sup>[5]</sup>。而倾向评分分配比将所有的协变量综合为一个尺度变量,因此协变量个数增加并不增加配比的难度<sup>[7]</sup>。尽管倾向评分分配比能够同时平衡较多的变量,但其只局限于已知的混杂变量,而许多未知的混杂变量可能仍然会对最终的结果产生影响。因此,其组间均衡性不可能完全达到随机对照研究的均衡性<sup>[7]</sup>。

#### 参 考 文 献

- [1] Austin PC. A critical appraisal of propensity score matching in the medical literature between 1996 and 2003. *Stat Med*, 2008, 27

(12):2037-2049.

- [2] Rubin DB, Thomas N. Matching using estimated propensity scores: relating theory to practice. *Biometrics*, 1996, 52: 249-264.
- [3] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, 70:41-55.
- [4] Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*, 1985, 39:33-38.
- [5] Rubin DB. Bias reduction using Mahalanobis metric matching. *Biometrics*, 1980, 36:293-298.
- [6] Baser O. Too much ado about propensity score models? Comparing methods of propensity score matching. *Value Health*, 2006, 9(6): 377-385.
- [7] D'Agostino RB. Propensity score methods for bias reduction in the comparison of a treatment to a nonrandomized control group. *Stat Med*, 1998, 17(19):2265-2281.
- [8] Jones AS. Maternal alcohol abuse/dependence, children's behavior problems, and home environment: estimates from the National Longitudinal Survey of Youth using propensity score matching. *J Stud Alcohol Drugs*, 2007, 68(2):266-275.
- [9] Trujillo AJ, Portillo JE, Vernon JA. The impact of subsidized health insurance for the poor: evaluating the Colombian experience using propensity score matching. *Int J Health Care Finance Econ*, 2005, 5(3):211-239.
- [10] Do MP, Kincaid DL. Impact of an entertainment-education television drama on health knowledge and behavior in Bangladesh: an application of propensity score matching. *J Health Commun*, 2006, 11(3):301-325.

(收稿日期:2008-11-03)

(本文编辑:张林东)

#### · 书 讯 ·

## 《传染病流行病学》简介

《传染病流行病学》是国内出版的第一部集理论、方法及应用于一体的传染病流行病学专著。全书共三篇二十五章,涉及的内容涵盖了传染病流行病学的基本原理、研究方法和防制技术。此书专著的“原理篇”全面、系统地论述了传染病流行病学的生物学基础、流行过程、传染病监测、新传染病的出现与应对、传染病流行的数学模型等。“方法篇”对传染病流行病学涉及的各种调查研究方法作了重点而翔实的介绍,包括现场调查、传播特征调查、自然疫源地调查、血清流行病学、分子流行病学、地理信息系统的应用及现场干预研究等。“预防控制篇”涉及消毒、杀虫、灭鼠、预防接种和医院感染控制等内容。此外,该书结合当前传染病防制领域出现的新问题,增设了卫生流行病学侦察、生物袭击的流行病学调查和突发公共卫生事件应对策略和措施等章节。全书紧密围绕传染病和突发公共卫生事件的相关流行病学问题,主题鲜明,信息丰富,内容新颖,逻辑严谨,是相关专业研究生的教学用书,并可供科研及各级疾病预防控制机构专业人员培训和卫生行政部门、医疗机构、医药院校有关人员工作时参考使用。该书现已由高等教育出版社出版。定价:52.00元。联系人:安琪,电话:010-58581118。

本刊编辑部