

使用 R 和 Stata 软件实现倾向性评分匹配

南通大学公共卫生学院公共卫生与预防医学系(226019) 周洁 张晟 何书 李文超 汪徐林 沈毅[△]

随机对照试验(randomized controlled trial, RCT)目前被认为是临床试验的金标准,但在实际工作中常受到伦理、经济等因素的影响,且因研究对象有严格的纳入排除标准,使其结论外推受到限制^[1]。大样本观察性研究可部分弥补 RCT 的不足,但其研究过程未采用随机对照,基线以及重要预后因素在组间分布的不均衡可能会对结果的真实性造成影响^[2]。倾向性评分匹配(propensity score matching, PSM)是一种均衡混杂因素的方法,近年来广泛用于大样本、非随机观察性研究^[3-5]。目前用于实现 PSM 的软件为数不少^[6-8],R 语言的 Matchit 程序包以及 Stata 软件的 psmatch2 程序包就是专门用于实现 PSM 的易学易用的程序包^[9-10]。本文拟通过实例展示如何在 R 与 Stata 软件上逐步实现 PSM。

倾向性评分的概念与基本原理

倾向性评分(propensity score, PS)的概念是 Rosenbaum 和 Rubin 于 1983 年首次提出,它是指将多个协变量的影响通过 PS 值的变化来表示^[11]。根据 PS 值在组间进行分层、匹配、回归(协变量调整法)或加权分析,即均衡组间协变量的分布,最后在协变量分布均衡的匹配组内或层中进行效应值的分析^[12]。

PS 的基本原理是指在一定可观察协变量(X_i)的条件下,研究对象 i ($i = 1, 2, \dots, n$) 被分配到特定处理组($Z_i = 1$)或对照组($Z_i = 0$)的条件概率。此时,第 i 个研究对象被分配到处理组的概率可以表示为:

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

若给定的特征变量(x_i)与分组变量(Z_i)是相互独立的,则:

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

$e(x_i)$ 即为 PS 值。

Rosenbaum 和 Rubin 提出构建 PS 模型需服从“强烈可忽略”(strongly ignorable)的假设,即计算 PS 值时,协变量应包含所有影响分组的混杂因素,或者说不存在未识别的混杂因素^[11]。它的使用具有两个条件:(1)条件独立性,即观察对象的分组选择只受纳入模型的协变量影响,而不受其他协变量的影响;(2)组

间评分分布具有足够大的重叠区域,即样本量足够大且协变量的 PS 值相近。

在 R 软件上实现倾向性评分匹配

1. 软件安装与程序包加载

R 是一个自由、免费、源代码开放的软件,可用于统计计算、统计编程以及统计绘图^[13]。截止至 2017 年 7 月 25 日,R 软件最新版本为 R3.4.1,用户可从官方网站 <http://www.r-project.org> 上获取最新的版本及相应的安装包。

R 软件安装完毕后,还需要进一步安装和加载所需的 Matchit 程序包,具体安装与加载代码如下:

```
install.packages("Matchit")
library(MatchIt)
```

2. 数据背景与加载

数据选择某医院 2012 年 1 月 1 日至 2015 年 10 月 31 日期间确诊的 603 例妊娠合并乙肝孕妇,探讨 HBeAg 对妊娠合并乙肝孕妇围产期不良结局的影响,其中暴露组(HBeAg 阳性)172 例,对照组(HBeAg 阴性)431 例。由表 1 可见,暴露组与对照组的基线资料无可比性,均衡性较差。

将数据以“HBV”为名、以“.csv”为格式存储在 D 盘里,使用“read.csv()”命令来读取数据,具体读取命令为:

```
HBV <- read.csv("D:/HBV.csv", header = TRUE, sep = ",")
```

上述命令执行后,在命令栏中输入命令“HBV”,即可见到导入数据,表示数据加载成功。

3. 倾向性评分匹配与结果输出

以孕妇是否为 HBeAg 阳性为分组因素,将其余变量作为协变量纳入 PSM 模型之中,并将结果存储在“m.out”这一新变量之中。匹配的算法选用了最常用的最近邻匹配算法(nearest neighbor matching),匹配的比例可通过设定 ratio 的数值来实现,一般设置为 1~5,本例设置为 1:1 匹配。程序命令如下:

```
m.out = matchit(HBeAg ~ age + alt + ast + tbil + glu + wbc + cr + pt, data = HBV, method = "nearest", ratio = 1)
```

值得注意的是 PSM 中匹配算法的选择对结果的

[△]通信作者:沈毅, E-mail: sunny@ntu.edu.cn

表 1 孕妇基线特征表

变量	标签	HBeAg 阳性 (n = 172)	HBeAg 阴性 (n = 431)	P
年龄(岁)	age	26.60 ± 3.56	27.65 ± 4.20	0.002
谷丙转氨酶(U/L)	alt	34.20 ± 54.27	23.57 ± 40.90	0.021
谷草转氨酶(U/L)	ast	35.44 ± 43.65	25.90 ± 28.13	0.008
总胆红素(mg/dL)	tbil	9.52 ± 3.65	10.52 ± 4.79	0.006
血糖(mmol/L)	glu	6.84 ± 2.91	8.07 ± 3.61	<0.001
白细胞计数(/L)	wbc	6.33 ± 5.03	5.15 ± 2.40	0.003
肌酐(mg/dL)	cr	164.64 ± 125.04	208.70 ± 112.62	<0.001
凝血酶原时间(s)	pt	5.78 ± 5.46	3.71 ± 4.76	<0.001

影响十分重要,本文列举了 R 中主要的匹配方法及其适用条件:

精确匹配(exact matching)——该算法会精准地匹配每个协变量值完全相同的处理组与对照组。但当协变量较多或者协变量取值范围较大时,精确匹配可能较难实现。(method = “exact”)

最近邻匹配(nearest neighbor matching)——该算法将处理组与对照组中倾向性评分最接近的研究对象进行匹配。其优点在于可最大化匹配效果,最大程度上保留了研究样本的信息。(method = “nearest”)

优化匹配(optimal matching)——该算法的核心是尽量减少在所有配对组之间的平均绝对距离。需要注意的是使用优化匹配时,需要同时安装 optmatch 程序包。(method = “optimal”)

遗传匹配(genetic matching)——该方法是使用遗传搜索算法找出每个协变量的权重,使得匹配之后的样本达到全局最优平衡。缺点是当变量较多时,需要运行较长的时间。(method = “genetic”)

粗化精确匹配(coarsened exact matching)——这种算法可对特定协变量进行匹配,同时保持其他协变量的平衡。其优点是计算速度快,对多分类变量、极度不平衡资料以及多重填补后的缺失资料有着良好的适应性。使用粗化精确匹配需安装 cem 程序包。(method = “cem”)

```
接着我们可以对结果进行描述与绘图,命令如下:  
summary(m.out)  
plot(m.out,type = "jitter")  
plot(m.out,type = "hist")  
plot(m.out,type = "QQ")
```

图 1 给出了匹配前 PS 模型基本信息(A)、匹配后 PS 模型基本信息(B)、均衡性检验(C)以及样本匹配结果(D)。首先匹配的基本信息包括了暴露组与对照组变量与距离(distance)的均值(means treated 与 means control)、对照组标准差(SD control)、暴露组与对照组的均值差(mean diff)以及 QQ 图中暴露组与对

照组的中位数距离(eQQ med)、均数距离(eQQ mean)和最长距离(eQQ max)。接着均衡性检验提示,本例样本整体距离的均值差由原来的 0.0713 下降为 0.0099,均衡性提升 86.11%,所有变量均衡性提高 50% 以上,匹配效果良好。最后匹配结果包括了暴露组(treated)与对照组(control)匹配前(all)、匹配成功(matched)、匹配失败(unmatched)以及被排除(discarded)的样本量(sample sizes)。

图 2 是 PS 分布抖点图(jitter plot),它表示暴露组与对照组间匹配与未匹配者(unmatched treatment units:未匹配暴露组,matched treatment units:已匹配暴露组,matched control units:已匹配对照组,unmatched control units:未匹配对照组)PS 的分布,从而了解匹配的效果。点的位置表示个体的得分情况;在 1:n 的情况下,点的大小则表示个体的权重。本例结果显示,获得匹配的个体倾向性评分较为接近,匹配效果较好。

图 3 是 PS 分布的直方图,表示暴露组与对照组间匹配前后(raw treated、raw control、matched treated、matched control 分别指匹配前暴露组与对照组以及匹配后暴露组与对照组)PS 值的分布,通过此图可看出暴露组与对照组匹配均衡性及匹配效果。本例中,匹配后对照组的 PS 由匹配前的不均衡变为均衡且与试验组接近,说明匹配效果较好。

图 4 为 PS 模型变量的 QQ 图,它表示暴露组与对照组间各个变量的 PS 分布,可看出单个变量的匹配前后的均衡情况。如年龄变量(age)的 PS 由匹配前的非正态分布变为匹配后的正态分布,说明年龄变量匹配效果较好。

4. 完成匹配数据的导出

匹配完成后,需将成功匹配的数据导出进行效应值的分析。此时可使用 match.data 命令将数据存储在“m.data”这一新变量中,并使用“write.csv”命令将数据以“mHBV.csv”格式导出到 D 盘中,以便进一步的分析,具体命令如下:

A

Summary of balance for all data:

	Means Treated	Means Control	SD	Control	Mean Diff	eQQ Med	eQQ Mean	eQQ Max
distance	0.3362	0.2649	0.1076	0.0713	0.0591	0.0717	0.2701	
age	26.5988	27.6520	4.1952	-1.0531	1.0000	1.0465	4.0000	
alt	34.1977	23.5705	40.8966	10.6271	5.0000	10.6163	98.0000	
ast	35.4419	25.9012	28.1302	9.5407	5.0000	9.5000	131.0000	
tbil	9.5230	10.5230	4.7919	-1.0000	0.8000	1.0839	39.7000	
glu	6.8392	8.0686	3.6123	-1.2293	1.1100	1.2398	8.9400	
wbc	6.3343	5.1464	2.4043	1.1879	0.3550	1.1866	40.6200	
cr	164.6352	208.7027	112.6221	-44.0675	39.2500	48.1589	150.8000	
pt	5.7831	3.7102	4.7620	2.0728	0.0550	2.0992	10.4000	

B

Summary of balance for matched data:

	Means Treated	Means Control	SD	Control	Mean Diff	eQQ Med	eQQ Mean	eQQ Max
distance	0.3362	0.3263	0.1176	0.0099	0.0007	0.0102	0.2197	
age	26.5988	26.3721	3.6564	0.2267	0.0000	0.3198	2.0000	
alt	34.1977	29.2209	49.3561	4.9767	3.0000	6.6279	88.0000	
ast	35.4419	31.5407	36.5910	3.9012	3.0000	4.4128	131.0000	
tbil	9.5230	9.2905	3.8240	0.2325	0.3000	0.4010	5.8000	
glu	6.8392	7.1044	3.0920	-0.2651	0.2500	0.3150	2.9000	
wbc	6.3343	5.8542	2.7225	0.4801	0.0900	0.5332	40.6200	
cr	164.6352	171.5534	117.2241	-6.9181	12.9000	16.1936	149.3000	
pt	5.7831	5.4110	5.4304	0.3721	0.0100	0.4410	9.4200	

C

Percent Balance Improvement:

	Mean Diff.	eQQ Med	eQQ Mean	eQQ Max
distance	86.1113	98.8372	85.8445	18.6767
age	78.4696	100.0000	69.4444	50.0000
alt	53.1695	40.0000	37.5685	10.2041
ast	59.1103	40.0000	53.5496	0.0000
tbil	76.7546	62.5000	63.0068	85.3904
glu	78.4319	77.4775	74.5904	67.5615
wbc	59.5884	74.6479	55.0661	0.0000
cr	84.3010	67.1338	66.3746	0.9947
pt	82.0492	81.8182	78.9896	9.4231

D

Sample sizes:

	Control	Treated
All	431	172
Matched	172	172
Unmatched	259	0
Discarded	0	0

图 1 倾向性评分模型结果、均衡性检验与样本匹配结果

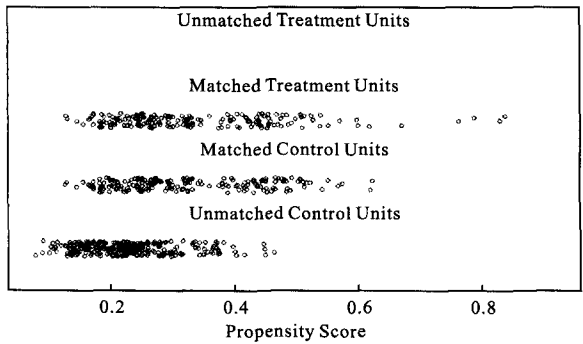


图 2 倾向性评分分布抖点图

```
m. data <- match. data(m. out)
write. csv(m. data, file = "D:/mHBV. csv")
```

数据导出后,软件自动生成了三个新变量: *id* 为每一个观测对象唯一的 ID 号; *distance* 为最近邻匹配的距离值; *weight* 为匹配的权重值。

在 Stata 软件上实现倾向性评分匹配

1. Stata 软件的下载与模块加载

Stata 软件是一套可进行数据分析、数据管理以及绘制专业图表的统计软件。用户既可通过窗口菜单进行操作,也可直接输入编程代码完成数据运算。用户可从 <http://www.stata.com/stata14/> 购买最新版 Sta-

ta14.0, 本文使用的为该版本。

在 Stata 的命令栏输入 `ssc install psmatch2`, 即可自动搜索及安装 `psmatch2` 程序包。

2. 数据录入

分析数据同前,可使用菜单栏中 `import` 过程来导入数据或者直接将 excel 表格中的数据复制粘贴至数据编辑器之中。

3. 倾向性评分匹配与结果输出

同样以孕妇是否为 HBeAg 阳性为分组因素,将其余变量作为协变量纳入 PS 模型之中。使用 `psmatch2` 命令,匹配的算法选用卡钳匹配算法(`caliper`),它是在最近邻匹配的基础上加一个限制条件,即暴露组与对照组 PS 差值需要在一定范围内才能进行匹配,卡钳值根据经验设为 0.03^[14]。匹配的比例可通过设定最紧邻算法后括号中的数值来实现,本例设置为 1:1 匹配。`common` 选项可查看暴露组样本是否支持共同支撑假设,即强制排除暴露组中 PS 值大于对照组最大 PS 值或小于对照组最小 PS 值的样本。`ties` 强制当试验组观测有不只一个最优匹配时同时记录。程序命令如下:

```
psmatch2 hbeag age alt ast tbil glu wbc cr pt, logit
neighbor(1) caliper(.03) common ties
```

pstest, both
psgraph
Stata 中的匹配方法主要为:最近邻匹配、核匹配

(kernel matching)、局部线性匹配法(local linear matching)以及基于马氏距离的匹配方法(Mahalanobis matching)。^[15]

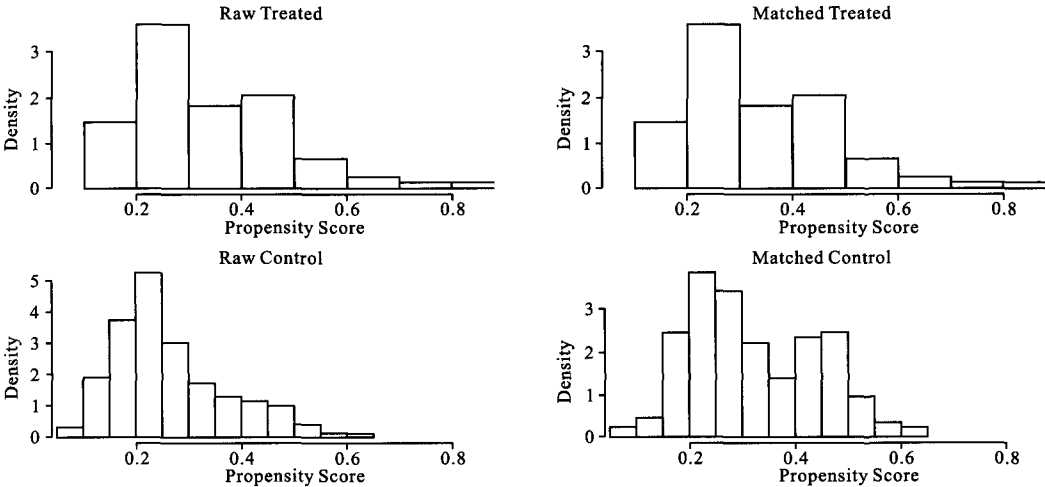


图3 倾向性评分分布直方图

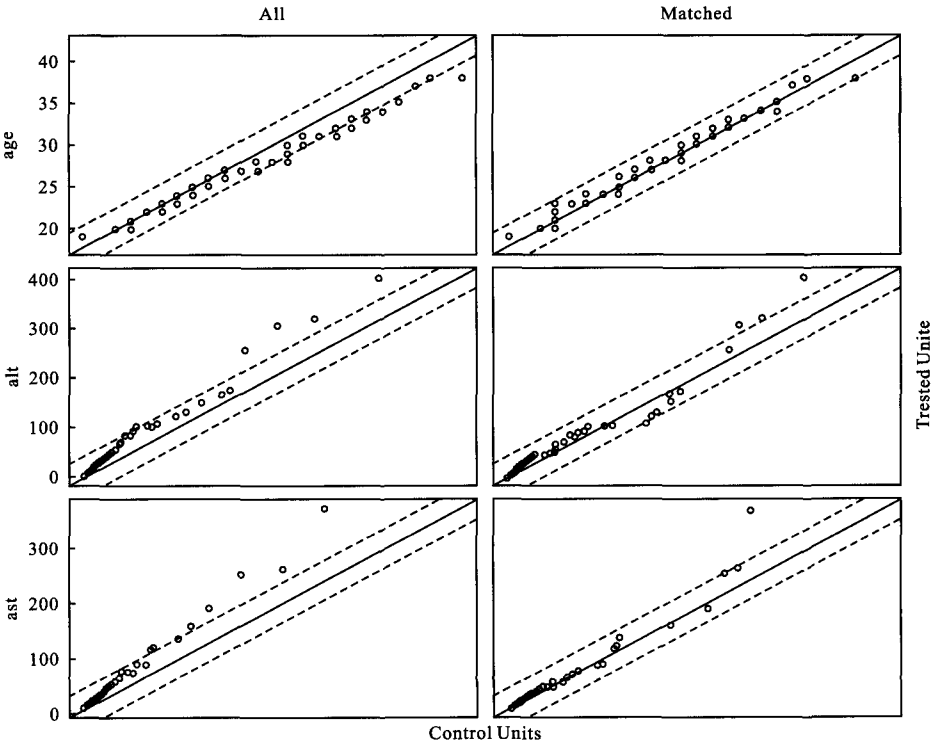


图4 倾向性评分变量 QQ 图

*: 只列举前三个变量(年龄,谷丙转氨酶、谷草转氨酶)

pstest 命令可做出匹配后均衡性检验,理论上 pstest 只能对连续变量做均衡性检验,对分类变量应重新整理数据后运用 χ^2 检验或者秩和检验。psgraph 命令可对匹配结果进行图示。
首先 psmatch2 命令可以看到模型回归结果以及各变量的整体情况。接着打开数据编辑窗口,会发现软件自动生成了一些新变量,其中_pscore 是观察对象的 PS 值;_id 是自动生成的每一个观察对象唯一的 ID 号;_treated 表示观察对象是否为暴露组;_n1 表示的

是观察对象被匹配到的对照的_id;_pdif 表示已经完成匹配的观察对象间概率值之差。
pstest 命令使用 both 选项可得到匹配前后的信息,均衡性检验结果可知,所有变量匹配后均由组间不平衡变为均衡,匹配效果良好(表 2)。
psgraph 命令作图结果可见匹配后样本倾向性评分值大致相同,匹配效果良好。与 R 不同的是,psgraph 命令可绘制出变量是否具有共同支撑区域,本例中有部分暴露组样本被排除(图 5)。

表 2 倾向性评分模型均衡性检验结果

变量	是否匹配	均值	优化程度 (%)		t 检验		暴露对照组 方差比
		暴露组/对照组	相对偏倚 (%)	偏倚	t	P	
年龄(岁)	否	26.599/27.652	-27.1	79.5	-2.9	0.004	0.72 *
	是	26.665/26.449	5.5		0.52	0.601	0.82
谷丙转氨酶 (U/L)	否	34.198/23.571	22.1	84.1	2.61	0.009	1.76 *
	是	28.114/29.808	-3.5		-0.38	0.701	0.57 *
谷草转氨酶 (U/L)	否	35.442/25.901	26	100	3.18	0.002	2.41 *
	是	30.108/30.108	0		0	1	0.63 *
总胆红素 (mg/dL)	否	9.523/10.523	-23.5	91.8	-2.47	0.014	0.58 *
	是	9.4853/9.567	-1.9		-0.2	0.844	0.83
血糖 (mmol/L)	否	6.8393/8.0686	-37.4	95.8	-3.98	<0.001	0.65 *
	是	6.8756/6.9278	-1.6		-0.16	0.872	0.95
白细胞计数 (/L)	否	6.3343/5.1464	30.2	97.3	3.91	<0.001	4.37 *
	是	6.0154/6.0472	-0.8		-0.1	0.924	1.08
肌酐 (mg/dL)	否	164.64/208.7	-37	96.7	-4.2	<0.001	1.23
	是	165.61/167.09	-1.2		-0.11	0.911	1.17
凝血酶原 时间(s)	否	5.7831/3.7102	40.4	85.9	4.62	<0.001	1.32
	是	5.7234/5.4306	5.7		0.5	0.619	1.06

* 未匹配组方差比在[0.74,1.35]之外,匹配组方差比在[0.74,1.36] 之外。

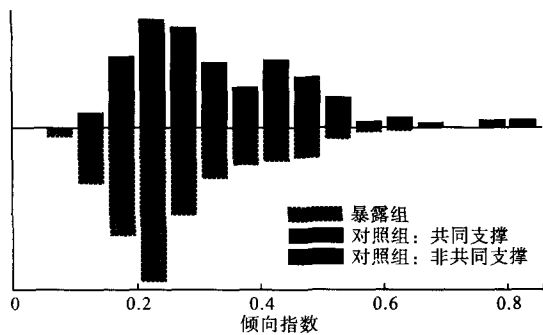


图 5 倾向性评分模型匹配结果直方图

讨 论

PSM 作为一种半参数统计方法,可以更加灵活地处理各种函数与模型,增加了暴露组与对照组合理匹配的可能性,限制也相对较少。尤其在处理较多混杂因素问题时,PSM 与传统方法相比,运算更为便捷,匹配更为合理,可为研究者提供更高效客观的匹配选择。当然,PSM 的应用也有其自身的限制。由于需要组间评分分布具有足够大的重叠区域,需要研究有一定的样本量支持,当样本量较小时,则无法解决协变量失衡的问题。另外,当数据存在缺失值,或有未知混杂因素时,此方法也难以使用。

目前,PSM 的应用已经成为一个热点,但却缺少易学易懂易用的软件。SPSS 软件在 22.0 之后的版本直接纳入了“倾向得分匹配”模块,虽然实现起来比较方便,但存在只能实现 1:1 匹配,匹配方法单一,没有绘图选项等缺点。如果研究者需要解决这些问题,则

需要在 SPSS 中安装 PS Matching 模块,此模块的本质为调用 R 的 Matchit 包实现 PSM,但此模块安装步骤较为复杂^[7]。SAS 软件同样在最新的 9.4 版本中 STAT 14.3 模块下加入了 proc psmatch 命令,使得实现 PSM 更加便捷。而之前的版本只能通过宏程序的方式来实现 PSM^[16]。

R 的 Matchit 程序包与 Stata 的 psmatch2 程序包都有着简明的程序,直观的结果以及理想的绘图。R 与 Stata 相比,有着更加多样的匹配选项和绘图选择;而 Stata 则可在匹配中选择共同支持区域,使结果更加合理。另外,它的均衡性检验可以直接比较匹配前后 P 值的结果,更直观方便。研究者可根据自身的需求选择来应用。

本文通过实例介绍了如何在 R 与 Stata 软件中实现倾向性评分匹配,并提供了详细的程序说明和结果解释,相信可以对运用 PSM 的科研工作者有所帮助。

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. Journal of clinical epidemiology, 2006,59(5):437. e1-437. e24.

[2] Heinze G, Jüni P. An overview of the objectives of and the approaches to propensity score analyses. European heart journal, 2011,32(14):1704-1708.

[3] Melduni RM, Schaff HV, Lee HC, et al. Impact of Left Atrial Appendage Closure During Cardiac Surgery on the Occurrence of Early Postoperative Atrial Fibrillation, Stroke, and Mortality. Circulation, 2017,135(4):366-378.

(下转第 636 页)

- Tests: Results and Implications. *Journal of Educational and Behavioral Statistics*, 1979, 4(3): 207-230.
- [23] Reise SP, Waller NG. Fitting the two-parameter model to personality data: The parameterization of the Multidimensional Personality Questionnaire. *Applied Psychological Measurement*, 1990, 14: 45-58.
- [24] Hambleton RKE. Applications of Item Response Theory. *International Journal of Educational Research*, 1989, 13(6): 818-825.
- [25] Bonsaksen T, Kottorp A, Gay C. Rasch analysis of the General Self-Efficacy Scale in a sample of persons with morbid obesity. *Health and Quality of Life Outcomes*, 2013, 11(1): 202.
- [26] Humphreys LG, Jr RGM. An Investigation of the Parallel Analysis Criterion for Determining the Number of Common Factors. *Multivariate Behavioral Research*, 1975, 10(2): 193-205.
- [27] Sam KL, Wang HY, Li C, et al. Item hierarchy of the Chinese version of cerebral palsy quality of life for children. *European journal of paediatric neurology*, 2013, 17(2): 126-134.
- [28] Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks - Plans for the patient-reported outcomes measurement information system (PRO-MIS). *Medical Care*, 2007, 45(5): S22-S31.
- [29] Chen WH, Thissen D. Local Dependence Indexes for Item Pairs Using Item Response Theory. *Journal of Educational and Behavioral Statistics*, 1997, 22(3): 265-289.
- [30] Yen WM. Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model. *Applied Psychological Measurement*, 1984, 8(2): 125-145.
- [31] Henning G. Meanings and implications of the principle of local independence. *Language Testing*, 1989, 6(1): 95-108.
- [32] Lord FM. Applications of Item Response Theory to Practical Testing Problems. Lawrence Erlbaum Associates, 1980, 1-274.
- [33] Linacre JM. Sample size and item calibration stability. *Rasch measurement transactions*, 1994, 7(4): 328.
- [34] Wright BD, Stone MH. Best Test Design. *Rasch Measurement*, 1979, 222.
- [35] Hulin CL, Lissak RI, Drasgow F. Recovery of Two and Three-Parameter Logistic Item Characteristic Curves: A Monte Carlo Study. *Applied Psychological Measurement*, 1982, 6(3): 249-260.
- [36] Fayers P. Item Response Theory for Psychologists. *Quality of Life Research*, 2004, 13(3): 715-716.
- [37] 闫宇翔, 王洪源. 项目反应理论在医学量表测量误差估计中的应用. *中国卫生统计*, 2004, 21(3): 162-164.
- [38] 杨铮, 戚艳波, 万崇华, 等. 慢性病生命质量量表共性模块项目反应理论分析. *中国公共卫生*, 2012, 28(11): 1477-1480.
- [39] 陈新林, 古模发, 徐谦, 等. 基于项目反应理论的鼻咽癌患者生存质量影响因素分析. *中国肿瘤*, 2012, 21(9): 691-694.
- [40] Shen M, Cui Y, Hu M, et al. Quantifying traditional Chinese medicine patterns using modern test theory: an example of functional constipation. *BMC Complement Altern Med*, 2017, 17(1): 44.
- [41] 雷新勇. 非参数项目反应理论模型及其在教育考试中的应用. 考试研究, 2006, (3): 55-73.

(责任编辑: 刘 壮)

(上接第 632 页)

- [4] Thourani VH, Kodali S, Makkar RR, et al. Transcatheter aortic valve replacement versus surgical valve replacement in intermediate-risk patients: a propensity score analysis. *The Lancet*, 2016, 387(10034): 2218-2225.
- [5] Branch-Elliman W, Ripollone JE, O'Brien WJ, et al. Risk of surgical site infection, acute kidney injury, and Clostridium difficile infection following antibiotic prophylaxis with vancomycin plus a beta-lactam versus either drug alone: A national propensity-score-adjusted retrospective cohort study. *PLoS Medicine*, 2017, 14(7): e1002340.
- [6] Parsons LS. Reducing bias in a propensity score matched-pair sample using greedy matching techniques Proceedings of the Twenty-sixth Annual SAS Users group international conference. SAS Institute Inc, 2001: 214-226.
- [7] 黄福强, 杜春霖, 孙梦辉, 等. 倾向评分分配比在 SPSS 软件上的实现. *南方医科大学学报*, 2015, 35(11): 1597-1601.
- [8] 李智文, 李宏田, 张乐, 等. 用 SPSS 宏程序实现观察对象的倾向评分分配比. *中国卫生统计*, 2011, 28(1): 89-90.
- [9] Ho DE, Imai K, King G, et al. MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 2011, 42(8): 1-28.
- [10] Leuven E, Sianesi B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Statistical Software Components*, 2015.
- [11] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, 70(1): 41-55.
- [12] 王永吉, 蔡宏伟, 夏结来, 等. 倾向指数的基本概念和研究步骤. *中华流行病学杂志*, 2010(3): 347-348.
- [13] Chongsuvivatwong V. Analysis of epidemiological data using R and Epicalc. Thailand: Book Unit, Faculty of Medicine, Prince of Songkla University, 2008.
- [14] 王永吉, 蔡宏伟, 夏结来, 等. 倾向指数常用研究方法. *中华流行病学杂志*, 2010(5): 584-585.
- [15] 焦明旭, 张晓, 刘迪, 等. 倾向性评分匹配在非随机对照研究中的应用. *中国卫生统计*, 2016, 33(2): 350-352.
- [16] 李智文, 任爱国. 倾向评分法在 SAS 软件中的实现. *中国生育健康杂志*, 2010, 21(5): 320-320.

(责任编辑: 邓 妍)