

## · 基础理论与方法 ·

# 疾病风险评分在药物流行病学研究中的应用

赵厚宇 詹思延

100191 北京大学公共卫生学院流行病与卫生统计学系

通信作者: 詹思延, Email: siyan-zhan@bjmu.edu.cn

DOI: 10.3760/cma.j.issn.0254-6450.2017.02.025

**【摘要】** 疾病风险评分通过平衡不同组间研究对象的基线疾病风险以控制高维数据结构中的混杂效应,从而减小暴露因素效应估计的偏倚,因此在利用医疗数据库探索药物疗效或不良反应等规律的药物流行病学研究中有重要的应用价值。尽管疾病风险评分方法在很多情况下具有与倾向性评分相似的作用,而且在一些特殊暴露条件的研究中具有倾向性评分与传统混杂控制方法不可比拟的优势,但目前疾病风险评分在药物流行病学研究中应用范围远不及倾向性评分广泛。基于对疾病风险评分方法在药物流行病学研究中应用价值的考量,本文阐述了疾病风险评分的原理、模型构建、评分估计和应用的方法,以期为疾病风险评分方法在药物流行病学研究中的应用提供参考。

**【关键词】** 疾病风险评分; 混杂控制; 药物流行病学; 医疗数据库

**基金项目:** 国家自然科学基金(81473067)

**Application of disease-risk score in pharmacoepidemiologic studies** Zhao Houyu, Zhan Siyan  
*Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China*  
*Corresponding author: Zhan Siyan, Email: siyan-zhan@bjmu.edu.cn*

**【Abstract】** Disease risk score (DRS) can be used to adjust the confounding effects on data with high dimensions and can reduce related bias through balancing the risk or probability, regarding the development of some specific diseases, between the two compared groups. The DRS approach thus can be applied to studies of pharmacoepidemiology when administrative medical database is used for data analysis. Although DRS functions are similarly to the propensity scores (PS) under many situations, even with some advantages over PS or conventional analytical methods in some special exposure settings, the usage of DRS is far limited than the PS method. Considering the important application value of DRS in pharmacoepidemiologic studies, we are introducing the theory, model, estimation and application of DRS, to present reference for the development of DRS method in the pharmacoepidemiologic studies.

**【Key words】** Disease risk score; Confounder control; Pharmacoepidemiologic studies; Administrative medical databases

**Fund program:** National Natural Science Foundation of China (81473067)

## 一、背景

药物流行病学以观察性研究为主,混杂问题在所难免。尤其在大数据时代,利用医保数据、区域医疗数据等管理型数据库研究药物疗效或不良反应逐渐成为药物流行病学研究的重要手段,并越来越受到重视<sup>[1-2]</sup>。这些医疗数据库通常含有大量的协变量信息,数据量大、人群时间和空间异质性高。而药物或其他治疗措施的暴露往往受到患者、医生以及医疗系统等多种因素及其相互作用的影响,这些因

素相互之间可能以复杂的、某种尚未明确的方式相互作用而引起混杂<sup>[3]</sup>。因此,如何处理混杂效应成为研究的关键。同时,在高维数据建模中,暴露因素与其他协变量之间的效应修饰作用变得异常复杂,因而常规的统计模型可能无法胜任<sup>[4]</sup>。倾向性评分(propensity score, PS)方法平衡暴露组和对照组间发生暴露的概率,可以同时降低模型复杂性并控制多种混杂变量,因而在药物流行病学研究中得到广泛应用<sup>[1,5]</sup>。但是,PS方法的基本假设是不存在协变

量的特定组合使得某研究对象以“1”或“0”的确定概率发生暴露,同时样本中结局事件数/协变量数之比需达到一定水平,才可拟合稳健的PS模型<sup>[6]</sup>,因此当暴露不能看作在研究对象间随机发生,或者对于罕见暴露事件,PS方法就可能失效。例如,对于新上市药物,用药者和药物相关结局出现均较少见,而且在上市最初阶段,人群使用新药的倾向可能随时间变化较显著,因而受其他协变量的影响不稳定,这两方面原因使得PS方法在新药上市后安全性和有效性研究中应用价值有限<sup>[7-8]</sup>。又如,当同时比较针对同一疾病的多种药物(或多个暴露水平)疗效或不良反应时,PS模型会变得更为复杂。此时,一种类似于PS的高维混杂控制技术——疾病风险评分(disease risk score,DRS)提供了更为有效的解决方案。

DRS的思想最早在1976年就由Miettinen提出<sup>[9-10]</sup>,不同于PS平衡组间暴露倾向,DRS估计研究对象在特定协变量和假定无暴露的条件下发生某种结局的概率,因而DRS平衡两组间结局的影响因素<sup>[5]</sup>。由于DRS的估计并不直接涉及暴露因素,因此对罕见暴露及多类别、多水平暴露情形同样适用<sup>[11]</sup>。此外,通常混杂因素对疾病风险的影响比对暴露倾向的影响相对于时间变化更加稳定<sup>[12]</sup>,同时,在进行匹配时,组间DRS分布的重叠程度(共同支持域)通常要大于PS分布的重叠程度,因此在采用最邻近匹配等算法进行匹配时,使用DRS匹配可以纳入更多的样本进行分析<sup>[8, 12]</sup>(表1)。目前已有较多理论和实证研究证明了DRS方法的有效性。Hansen<sup>[13]</sup>从理论上证明DRS可以实现混杂平衡,Månnsson等<sup>[14]</sup>利用模拟数据分析认为PS方法在病例对照研究中表现不佳,当暴露效应确实存在时,

PS调整会引入人为的效应修饰作用,而且对混杂的控制不完全。Arbogast和Ray<sup>[4]</sup>建议当PS模型应用条件受限或效果不佳时,选择使用DRS。然而,尽管DRS的思想提出时间早于PS(1983年<sup>[15]</sup>),但是由于DRS模型的构建更为复杂,而最初人们对其认识不充分,致使DRS在药物流行病学及其他领域的应用远不如PS广泛<sup>[2, 5]</sup>。Tadrous等<sup>[5]</sup>对Medline数据库1965—2011年间有关DRS的研究进行了系统综述,结果仅检索到97篇相关文献,实证研究86篇,其中药物疗效研究在2000年前仅占6.4%(3/47),2000年后则上升到46.2%(18/39),是使用该方法最多的流行病学研究领域,由此可见,DRS方法的应用还十分有限,但在药物流行病学研究中有很大的潜在应用空间。

## 二、DRS模型构建与估计

考察药物等暴露因素是否降低或增加结局的风险时需要平衡各组在无暴露因素作用时的基础疾病风险。如果混杂已被充分观测并被DRS平衡,那么暴露组增加或降低的疾病风险就可以认为是暴露因素引起的。因此暴露组的DRS不能直接计算,而应建立预测模型后估计其在非暴露情形下的疾病风险<sup>[12, 16]</sup>,所以DRS方法用于模型构建和DRS估计的数据通常不是同一数据集,甚至不是同一样本。

基于上述思想,DRS的估计可通过两个步骤实现:①模型构建;②DRS估计。根据①中用于模型构建的数据集不同,DRS又可分为四种方法。为了便于表述,现作如下符号规定:假定研究所关注的结局事件为D,暴露或处理因素为T,二者均为二值变量(“1”为发生,“0”为未发生);协变量X为一p维向量 $X=(x_1, x_2, \dots, x_p)$ ,研究样本FC(full-cohort)包含n

表1 DRS和PS方法特点及其影响因素比较

模型要素	疾病风险评分(DRS)	倾向性评分(PS)
估计模型	logistic回归、Poisson回归、Cox比例风险模型、判别函数、分类树模型等	logistic回归、Poisson回归、Cox比例风险模型、判别函数、分类树模型等
估计模型的因变量	结局(通常为二分类)	暴露或处理因素(通常为二分类)
模型构建和评分估计	模型构建和评分估计是相同或不同样本数据集	模型构建和评分估计是同一样本数据集
应用方法	匹配、分层、协变量回归调整	匹配、分层、协变量回归调整、加权
历史数据(暴露前)	非常适合于估计DRS	对变量筛选有提示作用,但通常不用于估计PS
结局罕见	对DRS限制很大,甚至使之失效	PS方法特别有价值,但此时不利于发现PS模型中的工具变量
暴露罕见	影响小	限制PS估计,甚至使之失效
暴露因素随时间变化	影响小	PS可能失效,应考虑时间的交互作用或者依时性PS
多结局或多暴露	对多个暴露因素,DRS可能特别有效;对于多个研究结局,需要多个DRS模型	单个PS可用于多个研究结局,此时PS模型中应纳入对每个结局都有影响的变量
效应修饰作用	DRS可估计自然尺度的效应修饰作用	不能直接估计效应修饰作用
平衡组间疾病风险	用于分层或匹配时,自然尺度	非主要作用,可能是PS分层或匹配的次生作用
平衡组间暴露倾向	非主要作用,可能是DRS分层或匹配的次生作用	分层或匹配时,自然尺度
排除没有对照的个体	排除因为过高或过低疾病风险而没有对照的个体,相对于PS排除的个体更少	排除因为过高或过低暴露倾向而没有对照的个体,相对于DRS排除的个体更多

个暴露对象和 $m$ 个非暴露对象,其中非暴露组的 $m$ 个研究对象组成数据集 $UO$ (unexposed-only);历史数据 $HD$ (historical data)<sup>[8,13]</sup>,以及研究样本外部数据 $AD$ (alternate data)<sup>[17]</sup>,四个数据集的特征和关系见表2。此处以最常用的logistic模型为例来说明DRS模型构建和评分估计的步骤。

### 1. FC方法

(1) 模型构建:研究样本的所有观测参与拟合,协变量及暴露因素为预测因子,构建以下模型(式中, $e$ 为随机误差,下同):

$$\text{logit}\{P[(D|X, T)=1]\} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p + \gamma t + e \quad (1)$$

(2) DRS估计:对任意研究对象 $\in FC$ ,令 $T=0$ ,再将 $FC$ 数据回代(1)式的模型并进行指数转换后得到DRS:

$$\begin{aligned} \text{DRS} = P[(D|X, T=0)=1] = \\ \frac{\exp(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p)}{1 + \exp(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p)} \end{aligned} \quad (2)$$

### 2. UO、HD、AD方法

(1) 模型构建:仅利用 $UO$ 、 $HD$ 、 $AD$ 样本数据,其中没有发生暴露的个体:

$$\text{logit}\{P[(D|X)=1]\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + e \quad (3)$$

(2) DRS估计:将 $FC$ 数据回代(3)式的模型并进行指数转换后分别得到 $UO$ 、 $HD$ 、 $AD$ 三种方法下的DRS:

$$\begin{aligned} \text{DRS} = P[(D|X)=1] = \\ \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)} \end{aligned} \quad (4)$$

在药物流行病学研究中,对照组通常不是空白或安慰剂对照,而是其他药物或疗法,此时,非暴露组应定义为对照药物组<sup>[17]</sup>。Tadrous等<sup>[5]</sup>系统综述结果表明,在报告模型构建方法的77个实证研究中使用logistic模型、判别函数、线性模型、Poisson回归和Cox比例风险模型拟合DRS估计模型的文献分别有46.8%、16.9%、10.4%、9.1%和7.8%,另有9.1%的研究使用其他估计方法<sup>[5]</sup>。

### 三、DRS的应用及实例

与PS类似,DRS常用于分层、匹配或者直接作为连续型协变量与暴露因素一同纳入最终的模型进行调整,但是将DRS作为连续协变量直接纳入模型进行调整对于控制混杂和偏倚表现不佳<sup>[6,18]</sup>。对于分层方法,常先将DRS分段(例如五分法、十分法等),然后作为分类或分层变量纳入暴露因素效应估计模型进行调整<sup>[2,19]</sup>。根据研究设计不同,DRS可进行暴露组间匹配或结局组间匹配,前者在队列研究中常见,而后者则可用于病例对照研究中。根据匹配算法的不同,常用的匹配算法有最邻近距离法、最优匹配等方法,最邻近距离法是按照组间个体DRS距离(如绝对值距离、马氏距离等)值依次递增逐对匹配,从而实现局部最优;最优匹配则实现两组中所有匹配对间DRS距离之和最小,从而达到全局最优<sup>[20-23]</sup>。Deb等<sup>[24]</sup>对PS应用方式的比较研究表明,匹配和加权(逆概率处理加权,IPTW)控制混杂效果优于分层和回归调整。DRS可能有相似的规律,但还缺少相关的研究加以验证。Tadrous等<sup>[5]</sup>的系统综述中73篇文献报道了DRS处理方法,93.2%作为分类变量处理,其中三、四、五和十分类最为常见,仅6.8%作为连续协变量进行调整,7.3%进行了其他处理。83篇报道了分析方法,其中60.2%使用DRS进行分层,34.9%将DRS(连续或分类)作为协变量调整,仅2.4%使用了匹配方法,另有2%使用其他分析方法。

Wyss等<sup>[16]</sup>在一项比较达比加群和华法林预防缺血性脑卒中和全死因死亡联合风险的有效性研究中使用了 $HD$ -DRS的匹配方法。研究者连接了2010年10月19日至2012年12月31日美国Medicare系统的住院、门诊和药房数据并从中随机抽取20%和1%的数据进行分析,入选研究对象全部为65岁以上曾在门诊或住院诊断为房颤或房扑的老年人,在研究开始前一年内(洗脱期)未使用任何口服抗凝药物(新使用者)。利用历史数据(historical data)达比加群上市(2008年1月1日至2010年10月18日)前的

表2 四种DRS建模数据集特征比较

建模数据集	数据集特点或条件及其与 $FC$ 的关系	共同特征
$FC$	研究样本, $t=1$ 及 $t=0$ 的研究对象。	不同数据集间疾病诊断、暴露定义及所有协变量的定义、编码、赋值等特征应相同;不同样本人群间特征应具有可比性,以保证DRS模型可以在不同人群、同一人群不同时期样本间具有可推广性。
$UO$	研究样本 $FC$ 中 $t=0$ 的研究对象。	
$HD$	在暴露因素发生前(如新药上市前), $HD$ 和 $FC$ 来自统一数据库,对任意研究对象 $i \in HD$ , $t=0$ ,可能存在 $i \in (HD \cap FC)$ ,对 $HD$ 方法需规定药物作用洗脱期	
$AD$	没有暴露因素或暴露因素发生率非常低的样本数据,如没有批准研究药物上市地区的人群数据,对 $AD$ , $P(T=1)=0$ ,且一般不存在 $i \in (AD \cap FC)$ 。	

注:  $FC$ (full-cohort);  $UO$ (unexposed-only);  $HD$ (historical data);  $AD$ (alternate data)

系统数据,建立新使用华法林的患者缺血性卒中和全死因死亡联合风险的一年风险模型,然后使用该模型估计研究样本人群的DRS。PS模型则使用样本人群构建。拟合DRS或PS的变量为专家建议的37个变量和根据经验选择的新增200个变量。然后使用最邻近匹配方法,以1:1进行组间匹配,最后用Cox比例风险模型估计两种药物的治疗效果(表3)。使用DRS或PS匹配调整混杂对估计结果有较大程度的改善,PS方法和HD-DRS方法调整混杂有相似的估计结果,但是DRS模型预测能力更好,且组间匹配率高于PS,因此DRS方法估计的结果比PS方法更为精确(95%CI宽度更窄)。

**表3** 达比加群与华法林预防缺血性卒中和全死因死亡联合风险的效果比较

样本量 <sup>a</sup>	协变量 <sup>b</sup>	分析方法	HR值(95%CI) <sup>c</sup>	匹配率 (%) <sup>d</sup>	C统计量 <sup>e</sup>
20%抽样 (n=67 667)		不作调整	0.48(0.46~0.50)		
	37	PS匹配	0.75(0.70~0.80)	99.9	0.68
		DRS匹配	0.73(0.69~0.77)	100.0	0.73
	237	PS匹配	0.88(0.81~0.95)	99.2	0.73
1%抽样 (n=3 383)		DRS匹配	0.87(0.81~0.94)	99.7	0.78
		不作调整	0.47(0.41~0.54)		
	37	PS匹配	0.75(0.57~0.99)	98.5	0.71
		DRS匹配	0.74(0.57~0.98)	99.1	0.73
	237	PS匹配	0.89(0.61~1.29)	92.0	0.79
		DRS匹配	0.85(0.62~1.16)	98.5	0.78

注:<sup>a</sup>20%抽样人群中含11 407例达比加群新使用者,1%抽样人群中含576例达比加群新使用者;<sup>b</sup>参与拟合DRS或PS估计模型的协变量数;<sup>c</sup>Cox比例风险模型,以华法林新使用者为参照;<sup>d</sup>达比加群新使用者与华法林新使用者匹配,DRS和PS方法使用相同的尺度;<sup>e</sup>DRS或PS估计模型的C统计量

#### 四、DRS模型的完善

1. 建模样本的选择:FC方法使用了研究样本的全部观测建立模型,增加了建模的样本量,因此不存在模型外推的困难,但同时也使模型的复杂性增加,而更容易发生模型设定错误。此时,即使DRS预测模型一个小的设定错误就可能产生严重的偏倚<sup>[13]</sup>。相比之下,UO方法仅利用对照人群建立的DRS估计模型对模型误设更稳定,但是该方法又可能导致过度拟合,致使对照组中患病风险高的个体的风险被高估,而患病风险低的个体药物治疗的风险被低估,从而导致偏倚<sup>[13,17]</sup>。但在大样本条件下,过度拟合作用会得到缓和,由此产生的偏倚也会很小<sup>[16]</sup>。在实际应用中,FC和UO方法何者更优并无一致的结论,最早Pike等<sup>[6]</sup>利用模拟数据分析结果表明FC-DRS会高估暴露因素效应的显著性。但Cook和Goldman<sup>[19]</sup>进一步研究表明,只有当结局和暴露因素之间复相关系数超过90%时,FC-DRS方法才会过

高估计暴露因素的效应,但这种情况在实际应用中通常不会出现。因此在绝大多数情况下,FC方法是有效的。Arbogast和Ray<sup>[4]</sup>的模拟研究、Cadarette等<sup>[11]</sup>的多种药物平行比较研究则表明,FC方法比UO方法估计效果更好,偏倚更小。

FC-DRS,UO-DRS与PS相似,建模和评分估计都使用同一研究样本的全部或部分数据集(same-sample estimation),在不存在遗漏混杂的情况下,这种方法是有效的,但实际情况中,通常不可能观测到所有潜在的混杂变量<sup>[16]</sup>,此时可能存在过度拟合。当研究样本对照组与暴露组协变量存在显著差异时,使用FC或UO构建DRS控制混杂的估计结果的精确度甚至不如不进行任何混杂调整的模型<sup>[13]</sup>。HD和AD方法是对UO和FC方法的改进,能够缓和单一样本模型过拟合的问题<sup>[13,17]</sup>。如果协变量对疾病风险的影响在不同人群间是稳定的,那么就可以利用HD和AD来估计DRS。这两种方法的优点在于,一是通过外部人群增加了结局的数量,增大了样本量,这对于利用多个变量预测罕见疾病风险的研究是有利的;其二,由于模型构建所采用的人群数据中暴露率很低或没有暴露发生,因此利用此人群估算DRS可以减少工具变量和遗漏变量引起的偏倚增大效应<sup>[17]</sup>。医疗管理数据通常都是纵向数据,具有很长的记录时间,可以依据药物引进和上市时间划分出历史数据和研究数据集,因此HD方法在采用大型医疗管理数据的药物流行病学研究中特别有利。此时,利用同一研究人群的历史数据增加了DRS建模样本和预测样本之间研究对象和变量定义的可比性,同时减小了模型由建模样本推广到预测样本时过度拟合问题<sup>[8,17]</sup>。但是,我们并不总是能够获得具有长观察时间范围的历史数据,同时如果所采用的医疗数据(如医保数据库等)中变量定义和编码发生了改变或更新(例如疾病编码由ICD-9改为ICD-10)可能引起不同时期数据集变量含义的改变,这通常使得利用历史数据估计DRS变得困难。对此,可利用暴露因素发生率比研究人群低或无该暴露因素的外部人群估计DRS,即用AD方法替代HD方法<sup>[13,17]</sup>。对于FC,UO方法与HD、AD方法的比较研究还较少,模拟数据的研究表明,使用外部样本(HD、AD)构建和估计DRS与使用同一样本(FC、UO)相比,可以控制过拟合从而减小效应估计的I类错误<sup>[13]</sup>。而利用HD方法的实证研究表明HD-DRS可以有效实现组间混杂的平衡<sup>[25~26]</sup>。

2. 建模协变量的筛选:在构建DRS模型时,另

一个关键问题是应该将哪些变量纳入模型参与拟合,即如何选择构建模型的协变量。模型拟合的变量过多或过少可能面临调整过度或调整不足的问题,均会引起结果的偏倚<sup>[3]</sup>。Miettinen<sup>[10]</sup>、Rubin<sup>[27]</sup>等认为如果遗漏重要变量会造成严重混杂偏倚,例如逐步回归等常用变量筛选方法,可能从最终模型中剔除一些重要的协变量,由此导致遗漏关键混杂因子<sup>[4,28]</sup>,因此建议将所有潜在混杂变量都纳入模型中。Hirano 和 Imbeus<sup>[29]</sup>则认为只应纳入那些相关性显著的变量。通常我们会认为模型中调整的变量越多,则混杂控制越好,由混杂引起的偏倚越小。但是许多研究者注意到,如果不对建模变量进行筛选,很可能引入两种偏倚——“Z-偏倚”和“M-偏倚”<sup>[3]</sup>,反而使偏倚增大。无论 DSR 还是 PS 都只能控制参与评分模型构建的变量造成的混杂,对未观测到或未进入模型的变量(即遗漏变量)不能进行控制。当潜在混杂变量中存在工具变量或类工具变量(工具变量指只与暴露因素高度相关,而与结局不相关的变量),调整工具变量使遗漏变量引起的偏倚增大,从而会增加 I 型错误<sup>[13]</sup>。这种偏倚随着未观测变量的增加、工具变量与暴露因素关联性的增强而增大,此即“Z-偏倚”。“M-偏倚”是由于对潜在混杂中的“Collider 变量”(Collider, 当一个变量同时是两个或多个其他变量的结局或效应变量时,该变量就称为 Collider 变量)进行调整时而产生<sup>[17,30]</sup>。Pearl<sup>[31]</sup>证明,调整工具变量引起的偏倚增大随着工具变量的增多而单调增加,而由增加协变量引起的偏倚减小并非单调,因此模型中每增加一个工具变量,引起的偏倚增大的速度比偏倚减小的速度快,总的结果使偏倚增大。为了减小由于工具变量、Collider 变量和未观测变量引起的偏倚,Pearl 等总结前人研究给出了一般性的建议<sup>[31-32]</sup>:①暴露因素的预测因子,即工具变量不应纳入模型进行调整;②对结局相关变量进行调整在减小偏倚方面更安全而且更有效;③在构建模型时应该考虑变量筛选,传统的让所有变量进入模型的方法不可取。对于高维数据,Kumamaru 等<sup>[7]</sup>提出了主成分分析进行降维,lasso 回归和岭回归等特征缩减技术来处理高维协变量估计时的过度拟合问题,对于高维数据,使用主成分、lasso 回归、岭回归等方法估计结果比不使用这些方法更精确,且主成分与特征缩减方法结合分析的效果更优。这些方法的探索丰富和促进了DRS的应用。

由于众多潜在的医疗、社会、个人行为等因素决定病人是否接受某种医疗处理或服用某种药物,加

之传统的医疗数据库存在诸多缺陷,因此有时很难选择一种恰如其分的模型来控制混杂。而且研究者通常无法完全确切地认定某个变量是否为混杂因素、工具变量或者 M-型偏倚的来源,而 FC、UO、AD、HD 等 DRS 方法优劣并无确切的论断,因此 Brookhart 等<sup>[3]</sup>建议研究者应该报告不同模型设定下的统计结果,在处理数据之前制定首要分析方案,然后改变分析方法,并将其他分析方法的结果以附件的形式予以报告。如果研究者不能明确究竟哪种模型的效果更优,那么可以考虑利用 Meta-分析或者贝叶斯方法将不同分析方法的结果进行汇总。总之,DRS 方法在药物流行病学研究中的应用还很欠缺,有待后来的研究者在实践中对 DRS 的理论和应用进一步完善。

利益冲突 无

## 参 考 文 献

- [1] Glynn RJ, Schneeweiss S, Stürmer T. Indications for propensity scores and review of their use in pharmacoepidemiology [J]. Basic Clin Pharmacol Toxicol, 2010, 98 (3) : 253-359. DOI: 10.1111/j.1742-7843.2006.pto\_293.x.
- [2] Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies [J]. Stat Methods Med Res, 2009, 18(1):67-80. DOI: 10.1177/0962280208092347.
- [3] Brookhart MA, Stürmer T, Glynn RJ, et al. Confounding control in healthcare database research: challenges and potential approaches [J]. Med Care, 2010, 48 (S1) : 114-120. DOI: 10.1097/MLR.0b013e3181dbebe3.
- [4] Arbogast PG, Ray WA. Performance of disease risk scores, propensity scores, and traditional multivariable outcome regression in the presence of multiple confounders [J]. Am J Epidemiol, 2011, 174(5):613-620. DOI: 10.1093/aje/kwr143.
- [5] Tadrous M, Gagne JJ, Stürmer T, et al. Disease risk score as a confounder summary method: systematic review and recommendations [J]. Pharmacoepidemiol Drug Saf, 2013, 22 (2):122-129. DOI: 10.1002/pds.3377.
- [6] Pike MC, Anderson J, Day N. Some insights into miettinen's multivariate confounder score approach to case-control study analysis [J]. J Epidemiol Community Health, 1979, 33 (1) : 104-106. DOI: 10.1136/jech.33.1.104.
- [7] Kumamaru H, Schneeweiss S, Glynn RJ, et al. Dimension reduction and shrinkage methods for high dimensional disease risk scores in historical data[J]. Emerg Themes Epidemiol, 2016, 13:5. DOI: 10.1186/s12982-016-0047-x.
- [8] Glynn RJ, Gagne JJ, Schneeweiss S. Role of disease risk scores in comparative effectiveness research with emerging therapies [J]. Pharmacoepidemiol Drug Saf, 2012, 21 (S2) : 138-147. DOI: 10.1002/pds.3231.
- [9] Desai RJ, Glynn RJ, Wang S, et al. Performance of disease risk score matching in nested case-control studies: a simulation study

- [J]. Am J Epidemiol, 2016, 183 (10) : 949–957. DOI: 10.1093/aje/kwv269.
- [10] Miettinen OS. Stratification by a multivariate confounder score [J]. Am J Epidemiol, 1976, 104(6) : 609–620.
- [11] Cadarette SM, Gagne JJ, Solomon DH, et al. Confounder summary scores when comparing the effects of multiple drug exposures [J]. Pharmacoepidemiol Drug Saf, 2010, 19(1) : 2–9. DOI: 10.1002/pds.1845.
- [12] Connolly JG, Gagne JJ. Comparison of calipers for matching on the disease risk score [J]. Am J Epidemiol, 2016, 183 (10) : 937–948. DOI: 10.1093/aje/kwv302.
- [13] Hansen BB. The prognostic analogue of the propensity score [J]. Biometrika, 2008, 95(2) : 481–488. DOI: 10.1093/biomet/asn004.
- [14] Måansson R, Joffe MM, Sun W, et al. On the estimation and use of propensity scores in case-control and case-cohort studies [J]. Am J Epidemiol, 2007, 166 (3) : 332–339. DOI: 10.1093/aje/kwm069.
- [15] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects [J]. Biometrika, 1983, 70(1) : 41–55. DOI: 10.1093/biomet/70.1.41.
- [16] Wyss R, Ellis AR, Brookhart MA, et al. Matching on the disease risk score in comparative effectiveness research of new treatments [J]. Pharmacoepidemiol Drug Saf, 2015, 24(9) : 951–961. DOI: 10.1002/pds.3810.
- [17] Wyss R, Lunt M, Brookhart MA, et al. Reducing bias amplification in the presence of unmeasured confounding through out-of-sample estimation strategies for the disease risk score [J]. J Causal Inference, 2014, 2 (2) : 131–146. DOI: 10.1515/jci-2014-0009.
- [18] Tadrous M, Mamdani MM, Juurlink DN, et al. Performance of the disease risk score in a cohort study with policy-induced selection bias [J]. J Comp Eff Res, 2015, 4(6) : 607–614. DOI: 10.2217/cer.15.40.
- [19] Cook EF, Goldman L. Performance of tests of significance based on stratification by a multivariate confounder score or by a propensity score [J]. J Clin Epidemiol, 1989, 42 (4) : 317–324. DOI: 10.1016/0895-4356(89)90036-X.
- [20] Hansen BB, Klopfer SO. Optimal full matching and related designs via network flows [J]. J Comput Graph Stat, 2012, 15 (3) : 609–627. DOI: 10.1198/106186006X137047.
- [21] Olmos A, Govindasamy P. Propensity scores: a practical introduction using R [J]. J MultiDisciplin Eval, 2015, 11 (25) : 69–88.
- [22] Rosenbaum PR. Optimal matching for observational studies [J]. J Am Stat Assoc, 1989, 84 (408) : 1024–1032. DOI: 10.2307/2290079.
- [23] Stuart EA. Matching methods for causal inference: a review and a look forward [J]. Stat Sci, 2010, 25 (1) : 1–21. DOI: 10.1214/09-STS313.
- [24] Deb S, Austin PC, Tu JV, et al. A review of propensity-score methods and their use in cardiovascular research [J]. Can J Cardiol, 2016, 32(2) : 259–265. DOI: 10.1016/j.cjca.2015.05.015.
- [25] Silber JH, Rosenbaum PR, Trudeau ME. Multivariate matching and bias reduction in the surgical outcomes study. [J]. Med Care, 2001, 39(10) : 1048–1064. DOI: 10.1097/00005650-200110000-00003.
- [26] Toh S, Gagne JJ, Rassen JA, et al. Confounding adjustment in comparative effectiveness research conducted within distributed research networks [J]. Med Care, 2013, 51 (8 Suppl 3) : S4–10. DOI: 10.1097/MLR.0b013e31829b1bb1.
- [27] Rubin DB. Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups [J]. Stat Med, 2009, 28(9) : 1420–1423. DOI: 10.1002/sim.3565.
- [28] Brookhart MA, Schneeweiss S, Rothman KJ. Variable selection for propensity score models [J]. Am J Epidemiol, 2006, 163(12) : 1149–1156. DOI: 10.1093/aje/kwj149.
- [29] Hirano K, Imbens GW. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization [J]. Health Serv Outcom Res Methodol, 2001, 2 (3/4) : 259–278. DOI: 10.1023/A:1020371312283.
- [30] Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider [J]. Int J Epidemiol, 2010, 39 (2) : 417–420. DOI: 10.1093/ije/dyp334.
- [31] Pearl J. Invited commentary: understanding bias amplification [J]. Am J Epidemiol, 2011, 174 (11) : 1223–1227. DOI: 10.1093/aje/kwr352.
- [32] White H, Lu X. Causal diagrams for treatment effect estimation with application to efficient covariate selection [J]. Rev Econom Stat, 2010, 93(4) : 1453–1459. DOI: 10.1162/REST\_a\_00153.

(收稿日期:2016-10-19)

(本文编辑:王岚)