

如何控制观察性疗效比较研究中的混杂因素:(二)未知或未测量混杂因素的统计学分析方法

黄丽红¹ 魏永越² 陈峰²

¹复旦大学附属中山医院生物统计室, 上海 200032; ²南京医科大学公共卫生学院生物统计学系 211166

通信作者:黄丽红, Email:huang.lihong@zs-hospital.sh.cn

【摘要】 观察性疗效比较研究作为随机对照研究的证据补充,其应用价值越来越受到关注。未测量混杂因素的统计学分析方法是观察性疗效比较研究中的重大挑战,本文对观察性疗效比较研究中未知或未测量的混杂因素控制的统计分析方法进行述评。未测量混杂因素的统计学方法包括工具变量法、本底事件率比较正法和双重差分模型及其衍生方法。工具变量法模型构造巧妙,但满足条件的工具变量在实际研究中并不易得;本底事件率比较正法和双重差分模型均要求研究数据有干预前信息,有些实际研究中往往无法满足。未测量混杂因素对统计学方法提出了新的要求、新的挑战,有待国内外统计学者的进一步完善和研究。

【关键词】 观察性疗效比较研究; 现实世界研究; 未测量混杂; 控制; 统计方法

基金项目:国家自然科学基金青年科学基金(81903407)

DOI:10.3760/cma.j.issn.0254-6450.2019.11.020

Confounder adjustment in observational comparative effectiveness researches: (2) statistical adjustment approaches for unmeasured confounders

Huang Lihong¹, Wei Yongyue², Chen Feng²

¹Department of Biostatistics, Zhongshan Hospital, Fudan University, Shanghai 200032, China; ²Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, China

Corresponding author: Huang Lihong, Email: huang.lihong@zs-hospital.sh.cn

【Abstract】 Observational study of therapy efficacy comparison has been widely conducted to provide the additional efficacy evidence to support randomized control study. Statistical adjustment for unmeasured confounders is a major challenge in observational study of therapy efficacy comparison. This paper summarizes and evaluates the relative statistical methods. Currently, the most commonly used methods include instrumental variable, difference-in-differences (DiD) model and prior event rate ratio (PERR) adjustment. The instrumental variable method skillfully escapes unmeasured confounders through model structure, but it is not easy to obtain satisfied instrumental variables. Both PERR and DiD require the data prior to exposure which are not always collected in observational studies. Unmeasured confounders could result in new requirements and pose new challenges for statistical methods, which needs further study and improvement.

【Key words】 Observational study of therapy efficacy comparison; Real world study; Unmeasured confounder; Adjustment; Statistical method

Fund program: National Natural Science Foundation of China Youth Science Foundation (81903407)

DOI:10.3760/cma.j.issn.0254-6450.2019.11.020

人们对事物的认识是一个逐步修正、不断完善的过程。目前,对复杂疾病尚未找到所有的混杂因素,未知的混杂因素,或虽已知但实际测量有困难的混杂因素是观察性疗效比较研究(CER)中的重大挑战,也是混杂因素控制方法学研究的焦点。针对未

知或未测量的混杂,本文从统计学角度就设计良好的观察性CER中如何进行混杂因素的统计分析进行述评,并对其正确应用进行总结。本文中未知混杂和已知但未测量混杂统称为未测量混杂(unmeasured confounder)。

1. 未测量混杂因素的常用控制方法: 根据 Streeter 等^[1]的文献综述, 近年来观察性研究中最常用的分析方法有 3 类, 包括工具变量法 (instrumental variable), 其次双重差分模型 (difference-in-differences, DiD), 本底事件率比较校正法 (prior event rate ratio adjustment, PERR), 及其衍生方法。

(1) 工具变量法: 是由 PG. Wright 于 1928 年首次提出, 最早应用于经济和社会学, 最近开始用于流行病学领域对未测量混杂因素的控制^[2-4]。工具变量需满足: 与所研究的干预相关; 与结局无关, 仅通过干预因素影响结局; 独立于混杂因素。见图 1。



图 1 工具变量

工具变量法的统计原理: 假设 Y 为结局, G 为协变量, 在传统最小二乘法统计模型的左右两边增加工具变量 Z , 同时取协方差, 可得:

$$Cov(Y, Z) = Cov(\beta_0 + \beta_1 G + \varepsilon_i, Z) = \beta_1 Cov(G, Z) + Cov(\varepsilon_i, Z)$$

即 $\sigma_{YZ} = \beta_1 \sigma_{GZ} + \sigma_{\varepsilon Z}$, 两边除以 G 与 Z 的协方差 σ_{GZ} , 则 $\frac{\sigma_{YZ}}{\sigma_{GZ}} = \beta_1 + \frac{\sigma_{\varepsilon Z}}{\sigma_{GZ}}$, 如 $\sigma_{\varepsilon Z} = 0$, 则 $\frac{\sigma_{YZ}}{\sigma_{GZ}} = \beta_1$ 。换言之, 只要设法找到满足条件的一个变量 Z , 就可以得到一个无偏估计的效应量 β_1 。

工具变量法的作用和随机分组异曲同工, 随机分组与所有混杂因素无关, 与实际接受的治疗相关, 只有通过治疗影响结局, 因而事实上, 随机分组变量就是一个工具变量。Brookhart 等^[2]使用此法分析了非选择性 Cox-2 抑制剂、非载体抗炎药与胃肠道并发症的关联, 选择医生的处方习惯作为工具变量, 分析结果与另外两个临床试验结果类似, 但如采用传统的多元回归进行协变量调整, 结果却相反。

工具变量法将变异限定于工具变量, 需要满足单调性假设, 即 $G^{Z=1} \geq G^{Z=0}$, 因而所估计的并非总体平均治疗效应 (average treatment effect) 而是局部平均治疗效应 (local average treatment effect)^[5]。工具变量法已经推广到广义线性和非线性模型中, 其局限在于: 局部平均效应有可能会影响工具变量法估计系数的外推性, 弱工具变量在有限样本情况下性能较差, 可能使工具变量法估计的系数波动性较大, 精准性较差, 任何可能的偏倚都会被过度放大。

近年来迅猛发展的孟德尔随机化 (Mendelian randomization, MR) 方法, 利用遗传变异 (如单核苷

酸多态性, SNP) 作为工具变量进行因果推断^[6], 在病因分析和疾病预后研究中推广应用^[7]。Zhao 等^[8]将工具变量法巧妙应用于不可忽略缺失值 (nonignorable missing data) 的参数识别问题, 完善了加权复合分位数回归 (weighted composite quantile regression analysis) 在不可忽略缺失值分析中的应用, 被称为无应答工具 (nonresponse instrument)。可见, 工具变量法在众多研究领域得到了广泛应用, 备受学者关注。

(2) DiD: 最早由 Ashenfelter^[9] (1976 年) 提出, 并在经济学领域广泛用于政策效果评价, 近年来在观察性研究领域得到了应用^[10-11]。该方法将观察组与对照组的干预前后进行比较, 从而扣除混杂因素。结局变量存在两类差异: 观察组前后的差异 ($A_1 - A_0$) 和对照组前后的差异 ($B_1 - B_0$)。DiD 通过构造双重差分估计量 (DiD estimator), 得到扣除时混混杂后的效应: 观察组前后差异 - 对照组前后差异 $\{(A_1 - A_0) - (B_1 - B_0)\}$ 。见图 2。

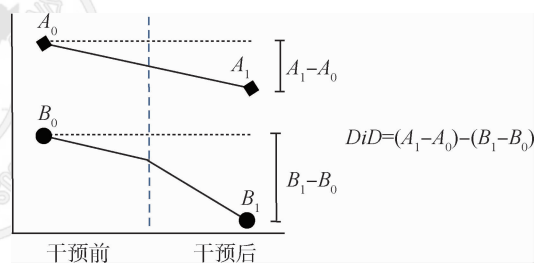


图 2 双重差分模型

DiD 通过将“前后差异”和“有无差异”有效结合, 一定程度上控制了除干预外的混杂因素的影响, 同时可在模型中加入其他可能影响结局变量的协变量, 进一步控制了观察组和对照组中存在的其他影响因素。假设分组变量 $G_i \in \{0, 1\}$, 时间变量 $T_i \in \{0, 1\}$, DiD 的基本模型可表示为:

$$Y_i = \beta_0 + \beta_1 G_i + \beta_2 T_i + \beta_3 (G_i T_i) + \varepsilon_i$$

$T=0$ 表示干预前, $T=1$ 表示干预后; ($G=1, T=1$) 表示干预组; ($G=0, T=1$) 表示对照组。双重差分估计量是横向和纵向比较的结合, 实际上是时间和分组交互项的偏回归系数。当应变量不服从正态分布时, 可对其进行数学变换或采用广义线性模型, 但由于联结函数 (link function) 的存在, 此时交互项的偏回归系数并不直接等同于干预效果, 需进行相应的换算。另外, 应用 DiD 模型评估干预效果时, 还需满足 3 个假设, 干预对对照组不产生影响; 干预之外的因素对观察组和对照组影响相同; 观察组和对照组中观察单位的某些特征分布稳定, 不随时间变化^[12]。

(3)PERR:近年得到推广和应用^[13-15]。该方法基于各组内自身前后对照设计,获得干预前两组发生率比(RR_{prior})和干预后的率比(RR_{post}),进而估计PERR。此时干预前RR仅是观察对象特征的组间差异,而干预后RR除观察对象特征差异外,主要是不同的干预方式。干预的作用通过比值的方式将混杂因素的影响去除,见图3。

PERR可通过事件发生率之比(incidence rate ratios)或风险比(hazard ratio,HR)来估计。方差的可信区间可采用bootstrap法进行估计。PERR的局限性与DiD类似,但当前PERR并不适用于发生率较高的不可逆事件(如死亡或疾病加重),只有当事件发生率很低,观察样本量很大时,影响可忽略。

2012年,Yu等^[13]基于非线性模型、Cox模型、logistic模型,通过模拟试验研究指出Tannen等^[14-15]提出的PERR常低估效应,提出采用配对或分层Cox模型估计 HR_E (干预后HR值)和 HR_{UE} (干预前HR值),通过 $HR_{PERR-ALT} = HR_E / HR_{UE}$ 估计HR值的PERR-ALT法。2016年,Lin和Henley^[16]提出pairwise过程改进PERR和PERR-ALT,并推导了在pairwise过程下干预效应的标准误及可信区间的方法。2017年,Tannen和Yu^[17]针对生存分析中死亡结局事件提出基于Cox模型的PERR法应用思路(post-treated event rate ratio,PTERR),但该思路要求观察组在干预结束后继续收集信息(post-treated period)。

PERR与DiD的区别包括,DiD主要用于连续性结果变量,而PERR适用于分类结果变量;DiD的前后比较是自身配对比较,而PERR不是。

(4)其他:除上述3类及其衍生方法较常用外,还有虽然不常见,但思路值得借鉴的分析方法。

断点回归设计(regression discontinuity design)^[18],是将干预开始时点作为分组界限,干预前作为对照组,干预后作为观察组,通过建立对照组的回归方程预测其到研究结束时的效应,从而通过观察组与对照组的差异估计干预效应。Lee^[19]提出将数据挖掘

中的扰动分析(perturbation analysis)用于未测量混杂的控制,通过模拟试验证明随着干扰变量数量的增加,扰动分析的效能增加,而混杂偏倚明显降低。经济学中常用的Dynamic panel model也被推荐用于滞后观察值(lagged observation)的混杂控制^[20]。

2. 案例分析:传统工具变量在国内医学领域的应用已经非常广泛,而PERR和DiD在观察性CER中报道并不多见,本文将对基于MR的多个SNPs的工具变量法及PERR和DiD进行案例分析。

(1)多个SNPs的工具变量法案例分析:血小板是凝血和炎症机制中的关键因素,活化血小板通过多种机制与癌症风险有关。然而,血小板与肺癌风险之间的因果关系尚不清楚。Zhu等^[21]利用大规模的GWAS(genome-wide association)研究中的统计结果,寻找与血小板计数相关的SNPs作为工具变量集,分析血小板与肺癌风险之间的关联。

该研究寻找到6个SNPs(rs17030845、rs6141、rs3792366、rs210134、rs708382、rs6065)满足工具变量的条件,将其作为工具变量集进行混杂因素控制,肺癌类型考虑非小细胞肺癌(non-small cell lung cancer, NSCLC)、腺癌(adenocarcinoma, AC)、鳞状细胞癌(squamous cell carcinoma, SCC)、小细胞肺癌(small cell lung cancer, SCLC),主要分析结果见图4,其中多重工具变量法为6个SNPs组成的工具变量集的结果,其余为单工具变量分析结果。从结果可知,血小板计数每升高 $100 \times 10^9/L$,NSCLC风险提高62%(95%CI:1.15~2.27, $P=0.005$),SCLC风险提高2倍($OR=3.00$; 95%CI:1.27~7.06, $P=0.01$)。

研究者为了进一步验证结论的稳健性,采用5种敏感性分析方法:惩罚的逆方差加权(penalized inverse-variance weighted)、稳健的逆方差加权(robust inverse-variance weighted)、MR-Egger法、惩罚的MR-Egger法及稳健的MR-Egger法,结果表明NSCLC风险与血小板计数的关联无统计学意义,而SCLC风险与血小板计数存在因果关联。

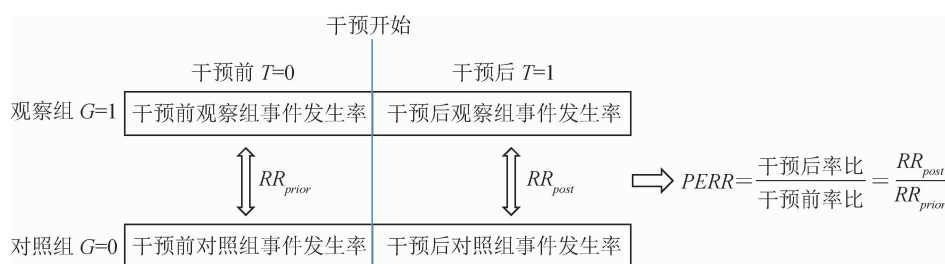


图3 本底事件率比较校正法示意图

(2) 双重差分模型案例分析: Bryson 等^[22]利用2007年1月1日至2008年12月31日 Aetna 保险公司行为健康会员的回顾性队列数据比较药物或心理治疗对酒精滥用患者 (alcohol use disorder, AUD) 的疗效, 其中药物治疗: 普通口服 naltrexone, 缓释 naltrexone、disulfiram 和 acamprosate。研究人群入选标准: 有 AUD 就诊信息; 处方中包括上述药物或心理治疗。排除标准: 6 个月无持续就诊信息; 单次就诊费用超过 25 000 美元; 在就诊前 3 个月接受过 AUD 药物治疗; 就诊后的 6 个月内接受过多种 AUD 药物治疗。最终分析数据集: 缓释 naltrexone 组 211 例, 普通口服 naltrexone 组 1 408 例, disulfiram 组 1 043 例, acamprosate 组 2 479 例, 以及心理治疗组 6 374 例。该研究的主要研究指标有治疗持续时间、就诊部门 (包括心理门诊、行为健康门诊、住院治疗、急诊治疗) 和医疗费用。其中治疗持续时间采用生存分析, 在此不赘述。

基线分析 5 组人群特征不均衡, 包括性别、种

族、保险类型、年龄分组、Charlson 评分、是否有精神分裂症、是否有双相型障碍、是否有重性抑郁症、是否有焦虑性障碍等, 均 $P < 0.001$ 。该研究采用 DiD 分析缓释 naltrexone 与其他各组在就诊部门和医疗费用上差异, 将接受 AUD 治疗的前 6 个月定义为治疗前。DiD 包括 2 个部分, 就诊部门采用 logistic DiD, 对医疗费用则采用线性 DiD 估计, 模型纳入人群基本信息 (性别、年龄、种族、保险类型等), 治疗分组, 时期 (治疗前、治疗后), 以及治疗分组与时期的交互作用。基于回归模型采用循环预测 (recycled predictions) 的方式估计各组不同就诊方式和医疗费用, 用 bootstrap 法估计其可信区间。见表 1。

研究结论: 缓释 naltrexone 组的患者相较于其他治疗组持续治疗时间长 (基于生存分析), 各组门诊就诊量均有不同程度的增加, 缓释 naltrexone 组相较于其他治疗组住院和急诊量, 非药物治疗费用有所下降。

上述案例中对比组人群基线并不均衡, 研究者巧妙运用 DiD, 不仅可控制基线中的已测量混杂, 还

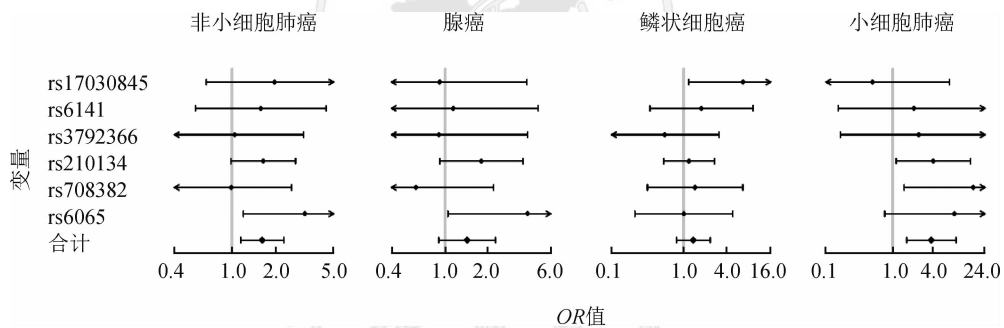


图4 孟德尔随机化工具变量法森林图^[21]

表1 双重差分模型案例部分结果^[22]

分组	治疗组	心理门诊	行为健康门诊	住院	急诊	非药物治疗费用合计
各组前后差异 (95%CI)	缓释 naltrexone	0.97 ^a (0.21 ~ 1.77)	1.41 (-0.14 ~ 3.13)	-0.50 ^a (-0.87 ~ -0.04)	-0.23 ^a (-0.47 ~ -0.01)	-2 152 ^a (-3 924 ~ -443)
	口服 naltrexone	0.81 ^b (0.45 ~ 1.15)	1.24 ^b (0.83 ~ 1.72)	-0.29 ^b (-0.42 ~ -0.15)	-0.16 ^b (-0.22 ~ -0.10)	-843 ^a (-1 580 ~ -68)
	Acamprosate	1.05 ^b (0.85 ~ 1.30)	2.43 ^b (2.08 ~ 2.76)	-0.28 ^b (-0.38 ~ -0.16)	-0.16 ^b (-0.21 ~ -0.11)	-354 (-806 ~ 203)
	Disulfiram	0.68 ^b (0.29 ~ 1.03)	1.43 ^b (0.89 ~ 1.97)	-0.17 ^b (-0.31 ~ -0.01)	-0.11 ^b (-0.18 ~ -0.05)	-639 (-1 441 ~ 14)
	心理治疗	0.03 ^b (0.01 ~ 0.05)	4.28 ^b (3.91 ~ 4.61)	-0.08 ^a (-0.14 ~ -0.01)	-0.16 ^b (-0.19 ~ -0.14)	544 ^b (234 ~ 820)
	缓释 naltrexone 相较于其他组前后差异之差 (95%CI)	口服 naltrexone	0.16 (-0.73 ~ 1.09)	0.17 (-1.52 ~ 1.85)	-0.21 (-0.62 ~ 0.27)	-0.07 (-0.30 ~ 0.17)
	Acamprosate	-0.08 (-0.90 ~ 0.79)	-1.02 (-2.85 ~ 0.56)	-0.22 (-0.61 ~ 0.21)	-0.07 (-0.30 ~ 0.15)	-1 798 (-3 535 ~ 141)
	Disulfiram	0.29 (-0.46 ~ 1.11)	-0.02 (-1.65 ~ 1.54)	-0.33 (-0.73 ~ 0.16)	-0.12 (-0.37 ~ 0.12)	-1 513 (-3 203 ~ 372)
	心理治疗	0.94 ^a (0.18 ~ 1.75)	-2.87 ^a (-4.34 ~ -1.10)	-0.42 (-0.79 ~ 0.08)	-0.07 (-0.30 ~ 0.13)	-2 696 ^b (-4 420 ~ -870)

注: ^a0.01 < P < 0.05, ^bP < 0.01

可控制未测量混杂。但笔者认为,在基线不均衡的情况下可考虑首先进行对比组间匹配(如PS匹配),随后进行DiD分析,如匹配后分析结果与现有结果一致,可提高研究结论的可靠性。

(3)本底事件率比较正法案例分析:Tannen等^[14]利用英国GPRD(general practice research database)数据库,通过比较该电子病历数据库的分析结果与已完成随机对照临床试验(randomized controlled trials,RCTs)的结果,探讨观察性CER的结果是否适合作为RCT结果的证据补充。

基于激素治疗降低绝经期妇女心血管疾病风险的研究目的,研究者收集了6个RCT(Syst-Eur,WHI-Intact uterus,WHI-Hysterectomy,4S,HOPE,EUROPA)研究,并采用类似RCT的设计,利用GPRD数据库分别构建相似的观察性CER。观察组从GPRD数据库中选择所有满足相应入、排标准并服用相同试验药物的患者;对照组则在满足入、排标准但未服用试验药物的患者中,通过匹配年龄、性别进行随机抽取。研究者还将服药过程完全符合方案要求的人群定义进行“模拟的意向性”(simulated intention to treat,sITT)分析,将由于违背方案而提前结束试验的人群定义进行“模拟的实际治疗”(simulated as treated)分析。研究共考察了心肌梗死风险比、卒中风险比和冠状动脉重建术风险比3个终点指标,进行了17项对比(Syst-Eur无冠状动脉重建术相关数据)。采用Cox模型进行风险比估计,并采用PERR对所估计的风险比进行校正。见图5。

17项比较中有9项RCT与GPRD观察性CER的研究结果无差异,另外8项有差异提示混杂因素的影响存在。8项中有7项PERR校正结果与Cox比例风险模型估计的原始HR值不同,PERR校正后的HR值相较于原始HR值更接近RCT的结果。研究结论:PERR在观察性CER中能有效控制未测量混

差,基于电子病历数据库的观察性CER作为RCT的证据补充具有一定应用价值。

该案例借助电子病历数据库GPRD巧妙构建了类似于RCT的观察性CER,既展示了在观察性CER中普遍存在的混杂,又说明了PERR的校正效果,同时也说明了观察性CER的应用价值,值得学习和借鉴。

3. 总结与展望:混杂因素的识别和控制,是医学研究成败的关键。研究开始前,需结合多方信息,尽可能地识别各种潜在的混杂因素,对于观察性CER中已知的、重要的混杂因素,匹配、分层、倾向性评分等方法能有效控制混杂因素^[23],但 these 方法对于未测量混杂因素无能为力。如存在未测量混杂因素,则需同时考虑科学的统计学方法加以控制,显而易见,未测量混杂因素的控制方法通常也能同时控制已测量混杂因素。

当前未测量混杂因素的控制方法并不丰富,选择较少,且尚不完善。PERR和DiD都要求研究数据有干预前的信息,在实际研究中不一定都能满足;工具变量法从方法学角度通过巧妙构造模型来规避所有混杂。MR方法为遗传流行病学研究打开了工具变量的应用之窗。但实际研究中,完全满足应用条件的工具变量并不易得。现有方法在实际应用过程中应在理解的前提下,严格把握适用条件。观察性CER中未测量混杂因素控制主要统计学分析方法总结见表2。

未测量混杂因素在观察性CER中普遍存在,如何控制和减少未测量混杂因素是观察性CER中的重大挑战。因而,未测量混杂因素的统计学方法研究一直以来备受统计学者关注。随着全基因组关联研究(genome wide association analysis,GWAS)的迅速发展,给工具变量的选择提供了广阔的空间,单个遗传变异的解释力度十分有限,在全基因组研究背景下,往往有多个遗传变异和目标表型相关,多个遗

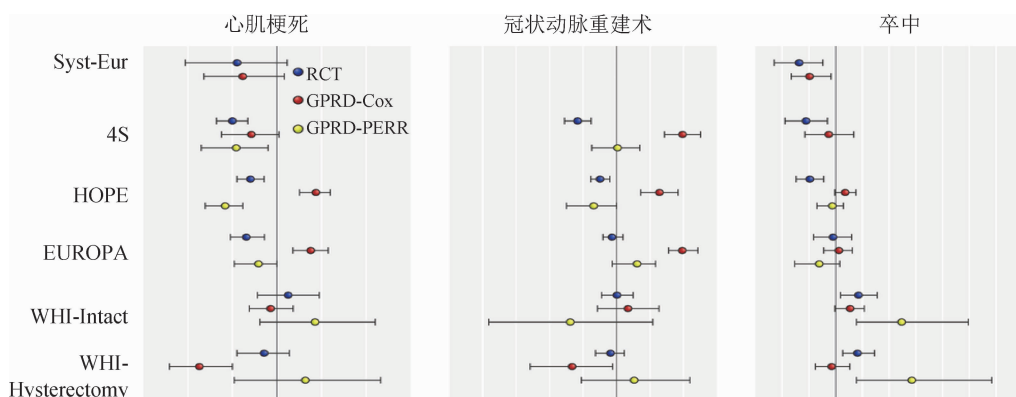


图5 PERR法案例主要分析结果^[14]

表2 观察性疗效比较研究中未测量混杂控制主要统计学分析方法总结

方法	模型要求与应用条件	局限性
工具变量法	<ul style="list-style-type: none"> · 工具变量和所研究的干预因素相关; · 工具变量与结局不相关,仅通过干预因素影响结局; · 工具变量独立于已测量或未测量的混杂因素 	<ul style="list-style-type: none"> · 在实际研究中,寻找满足条件的工具变量非常困难; · 弱工具变量在有限样本情况下性能较差,可能使得估计系数波动大,精准性差; · “局部平均处理效应”有可能会影响估计系数的外推性
双重差分法	<ul style="list-style-type: none"> · 适用于连续性结果变量; · 有干预前的基线信息; · 干预之外的因素对观察组和对照组影响相同; · 观察组和对照组中观察单位的某些特征分布稳定,不随时间变化; · 干预后至少一次测量 	<ul style="list-style-type: none"> · 仅适用于连续性结果变量; · 干预之外的因素如果在观察组和对照组不同,则难以评价结果
本底事件率校正法	<ul style="list-style-type: none"> · 用于率比结局; · 有干预前的本底事件发生率; · 干预之外的因素对观察组和对照组影响相同; · 观察组和对照组中观察单位的某些特征分布稳定,不随时间变化 	<ul style="list-style-type: none"> · 仅适用于事件发生率的比较; · 干预之外的因素如果在观察组和对照组不同,则难以评价结果; · 由于是群体比较,因此不适用于发生率较高的不可逆事件

传变异工具变量联合使用的分析方法正在发展过程中。另外,对于PERR和DiD的统一和改进同样值得深入探索。观察性CER中的未测量混杂因素对统计学方法提出了新的要求、新的挑战,有待国内外统计学者的进一步完善和研究。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

[1] Streeter AJ, Lin NX, Crathorne L, et al. Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review [J]. *J Clin Epidemiol*, 2017, 87: 23-34. DOI: 10.1016/j.jclinepi.2017.04.022.

[2] Brookhart MA, Wang PS, Solomon DH, et al. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable [J]. *Epidemiology*, 2006, 17(3): 268-275. DOI: 10.1097/01.ede.0000193606.58671.c5.

[3] Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review [J]. *J Clin Epidemiol*, 2011, 64 (6) : 687-700. DOI: 10.1016/j.jclinepi.2010.09.006.

[4] Davies NM, Smith GD, Windmeijer F, et al. Issues in the reporting and conduct of instrumental variable studies: a systematic review [J]. *Epidemiology*, 2013, 24 (3) : 363-369. DOI: 10.1097/EDE.0b013e31828abafb.

[5] Jiang W. Have instrumental variables brought us closer to the truth [J]. *Soc Sci Electron Publ*, 2017, 6 (2) : 127-140. DOI: 10.1093/rcfs/cfx015.

[6] Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization [J]. *Stat Methods Med Res*, 2015, 26 (5) : 2333-2355. DOI: 10.1177/0962280215597579.

[7] Wei YY, Tejera P, Wang ZX, et al. A Missense genetic variant in *LRRRC16A/CARMILI* improves acute respiratory distress syndrome survival by attenuating platelet count decline [J]. *Am J Respir Crit Care Med*, 2017, 195 (10) : 1353-1361. DOI: 10.1164/rccm.201605-0946OC.

[8] Zhao PY, Zhao H, Tang NS, et al. Weighted composite quantile regression analysis for nonignorable missing data using nonresponse instrument [J]. *J Nonparametr Stat*, 2017, 29 (2) : 189-212. DOI: 10.1080/10485252.2017.1285030.

[9] Ashenfelter O. Estimating the effect of training programs on earnings [J]. *Rev Econom Stat*, 1976, 60 (1) : 47-57. DOI: 10.2307/1924332.

[10] Patel MS, Volpp KG, Small DS, et al. Association of the 2011 ACGME resident duty hour reforms with mortality and readmissions among hospitalized Medicare patients [J]. *J Am Med Assoc*, 2014, 312 (3) : 2364-2373. DOI: 10.1001/jama.2014.15273.

[11] Rajaram R, Chung JW, Jones AT, et al. Association of the 2011 ACGME resident duty hour reform with general surgery patient outcomes and with resident examination performance [J]. *JAMA*, 2014, 193 (5) : 2374-2384. DOI: 10.1001/jama.2014.15277.

[12] 刘敏学, 胡明, 曾娜, 等. 双重差分模型在医学研究中的应用

[J]. *中国卫生统计*, 2015, 32(3): 528-531.

Shen MX, Hu M, Zeng N, et al. The application of difference-in-differences model in clinical researches [J]. *Chin J Health Stat*, 2015, 32 (3) : 528-531.

[13] Yu MM, Xie DW, Wang XM, et al. Prior event rate ratio adjustment: numerical studies of a statistical method to address unrecognized confounding in observational studies [J]. *Pharmacoepidemiol Drug Safety*, 2012, 21 Suppl 2: 60-68. DOI: 10.1002/pds.3235.

[14] Tannen RL, Weiner MG, Xie DW. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings [J]. *BMJ*, 2009, 338 (7691) : 395-399. DOI: 10.1136/bmj.b81.

[15] Tannen R, Xie DW, Wang XM, et al. A new “comparative Effectiveness” assessment strategy using the THIN database: comparison of the cardiac complications of pioglitazone and rosiglitazone [J]. *Pharmacoepidemiol Drug Safety*, 2013, 22(1) : 86-97. DOI: 10.1002/pds.3360.

[16] Lin NX, Henley WE. Prior event rate ratio adjustment for hidden confounding in observational studies of treatment effectiveness: a pairwise Cox likelihood approach [J]. *Stat Med*, 2016, 35 (28) : 5149-5169. DOI: 10.1002/sim.7051.

[17] Tannen R, Yu MG. A new method to address unmeasured confounding of mortality in observational studies [J]. *Learn Health Syst*, 2017, 1(1) : e10016. DOI: 10.1002/lrh2.10016.

[18] Zuckerman IH, Lee E, Wutoh AK, et al. Application of regression-discontinuity analysis in pharmaceutical health services research [J]. *Health Serv Res*, 2006, 41 (2) : 550-563. DOI: 10.1111/j.1475-6773.2005.00487.x.

[19] Lee WC. Detecting and correcting the bias of unmeasured factors using perturbation analysis: a data-mining approach [J]. *BMC Med Res Methodol*, 2014, 14: 18. DOI: 10.1186/1471-2288-14-18.

[20] Piernas C, Ng SW, Mendez MA, et al. A dynamic panel model of the associations of sweetened beverage purchases with dietary quality and food-purchasing patterns [J]. *Am J Epidemiol*, 2015, 181(9): 661-671. DOI: 10.1093/aje/kwu317.

[21] Zhu Y, Wei YY, Zhang RY, et al. Elevated platelet count appears to be causally associated with increased risk of lung cancer: a Mendelian randomization analysis [J]. *Cancer Epidemiol Biomark Prev*, 2019, 28(5) : 935-942. DOI: 10.1158/1055-9965.EPI-18-0356.

[22] Bryson WC, McConnell J, Korthuis PT, et al. Extended-release naltrexone for alcohol dependence: persistence and healthcare costs and utilization [J]. *Am J Manag Care*, 2011, 17 Suppl 8: S222-234.

[23] 黄丽红, 魏永越, 陈峰. 如何控制观察性疗效比较研究中的混杂因素: (一) 已测量混杂因素的统计学分析方法 [J]. *中华流行病学杂志*, 2019, 40(10) : 1302-1307. DOI: 10.3760/cma.j.issn.0254-6450.2019.10.024.

Huang LH, Wei YY, Chen F. Confounder adjustment in observational comparative effectiveness researches: (1) statistical adjustment approaches for measured confounder [J]. *Chin J Epidemiol*, 2019, 40 (10) : 1302-1307. DOI: 10.3760/cma.j.issn.0254-6450.2019.10.024.

(收稿日期: 2019-03-18)
(本文编辑: 王岚)