

DOI: 10.13288/j.11-2166/r.2019.11.006

思路与方法

干预性队列研究的偏倚控制及其在中医领域的特殊性

张彤¹, 何文婷², 费宇彤³, 张颖³, 杨宇飞[✉]

(1. 中国中医科学院西苑医院, 北京市海淀区西苑操场 1 号, 100091; 2. 新疆维吾尔自治区中医医院; 3. 北京中医药大学循证医学中心)

摘要 作为真实世界研究方法之一, 干预性队列研究具有较好的外部真实性, 其证据级别仅次于随机对照试验, 但最大的争议是偏倚。从研究设计、实施和统计分析的角度探讨如何控制干预性队列研究的常见偏倚。对于选择性偏倚, 强调研究者的角色依然是观察者, 不应干涉受试者入组意愿; 清晰地界定暴露因素及强度有助于研究者判断受试者的最终暴露归属; 部分特殊情况下适当采用随机化原则可减少选择性偏倚发生; 可采用匹配法增加组间均衡性。对于信息性偏倚, 建议采用客观、量化的指标收集资料, 若难以实现, 则采用“盲法”收集; 在广泛收集资料的同时, 需注意内容清晰易懂且不可过于繁琐, 以免造成不应答偏倚。对于混杂偏倚, 需要设置适当、清晰的纳入排除标准, 排除已明确的混杂因素; 在分析阶段, 校正分析、亚组分析、多元回归分析、倾向性评分等都可用于控制偏倚。此外, 中医药的干预性队列研究有其自身的特殊性, 选择分中心时需注意不同地区的搭配及中医院、西医院及综合医院的搭配等。期待通过控制偏倚、优化设计以增加干预性队列研究的内部真实性及证据级别, 使研究者在面对中医治疗这样个体化强、干预体系复杂、无法随机化、存在伦理学问题等情况下, 转换研究思路, 让更多患者因临床试验获益。

关键词 干预性队列研究; 偏倚; 真实世界研究; 方法学

临床医师阅读医学文献时需要思考两个重要问题: 这篇报道是否可信? 如果可信, 该报道是否和我的临床实践有关? 这两个问题即研究的内部真实性和外部真实性。安慰剂对照的随机对照试验(randomized controlled trial, RCT) 因为采用了随机、盲法、安慰剂对照等方法, 能克服选择性偏倚、信息偏倚和混杂偏倚, 最大限度地保障内部真实性; 然而 RCT 的合格性标准往往过于严格, 导致高龄、未成年、伴有并发症或共患病、妊娠期、哺乳期、多重用药等患者往往被排除。对于如何在保障内部真实性的同时提高外部真实性, 干预性队列研究不失为一个好的选择。

1 干预性队列研究的发展历程

自 20 世纪 80 年代起, 国外研究者开始将队列研究用于评价干预措施的疗效, 即把干预措施作为

暴露因素, 通过收集资料及随访评价该暴露因素与结局之间的关系, 称为干预性队列研究, 国内研究者近几年也逐渐开始重视这种研究方法^[1-3]。作为真实世界研究, 干预性队列研究更加贴近临床实际, 具有较强的外部真实性, 尤其适用于干预措施较为复杂、无法实施随机和盲法, 以及涉及伦理学问题的情况。如中药预防重症急性呼吸综合征的前瞻性队列研究^[4], 即在控制了偏倚和失访的前提下, 获得了具有较高信度的结果。

在循证医学证据评价等级体系中, 队列研究对某种干预措施的疗效评价证据级别弱于 RCT, 公认的原因是 RCT 随机化原则消除了组间差异, 从而最大程度地控制了偏倚。然而随机化并不是真的无可替代。有研究者对 1985 年至 1998 年发表的 136 篇观察性研究和 RCT 研究结果进行比较, 发现两者的结果相似^[5]。1990 年国外研究者设计了限制性队列研究和扩展性队列研究^[6], 采用限制性队列研究的方法观察了 β 受体阻滞剂降低急性心肌梗死后死亡率, 并将取得的试验结果与多中心、

基金项目: 国家重点研发计划(2017YFC1700604); 首都卫生发展科研专项(2016-I-4171)

* 通讯作者: yyf93@vip.sina.com

随机、双盲、安慰剂对照临床试验 (简称 BHAT 研究)^[8] 的数据进行比较。结果发现, 该限制性队列研究得到了和 BHAT 研究相似的结果。因此, 通过优化队列研究的设计、提高质量, 使得队列研究的证据级别更加接近 RCT 并不是没有可能。

干预性队列研究也有其局限性, 如研究周期长、样本量大、容易失访等, 其最大的争议就是偏倚问题。偏倚是系统误差, 贯穿于研究的整个过程, 从研究的设计、实施到资料收集, 甚至统计分析阶段都可能产生, 但并不是无法控制和纠正的。现就干预性队列研究常见的几种偏倚及控制方法进行探讨, 以期优化干预性队列研究设计。

2 干预性队列研究的偏倚及控制

2.1 选择性偏倚

选择性偏倚是由于科研设计不严谨、选择受试者方法不当等所导致的被选入与未选入受试者间在某些与研究相关的重要特征存在的系统差别。选择性偏倚会使研究结果偏离真实情况。

选择性偏倚发生于研究设计和实施阶段。首先, 必须要改变“临床医师就是研究者”这种国内较为普遍的观点和现象。干预性队列研究仍然属于观察性研究, 研究者的身份仍然是观察者, 仅能以观察者的角度客观评价暴露因素与结局事件之间的关系, 不能干预受试者入组情况。受试者入组需要强调的两个要点: 1) 有无暴露或暴露的等级是队列研究中分组的唯一依据, 干预性队列研究中即为有无目标干预措施或干预措施的强度。在研究设计之初应该清晰明确地界定“暴露”, 便于研究者判断。2) 有学者认为对照组的设定应该具备以下特征: 尽量不设安慰剂对照, 而是疗效、适应症、禁忌症、不良事件与治疗组较相似的阳性药物^[9]。由于队列研究不设盲, 探讨治疗作用时此法可以在一定程度上避免对照组受试者发生队列迁移及退出试验的情况; 但是探讨预防作用时可以采用安慰剂对照, 如同 Horwitz 等^[7] 研究。

关于受试者的入组方式, 有学者认为依次推荐患者自主选择、医患协商、随机选择^[10]。首先应由临床医师、药师充分告知各组的利弊后由患者自主选择干预措施; 当患者没有自主意见时可由医患双方协商决定, 这些是目前队列研究常用的做法, 但均无法避免选择性偏倚的发生。当医患双方对于干预措施均没有选择性倾向时, 可以根据随机数字来决定接受何种干预措施以减少选择性偏倚, 这是值得

推荐的方法。除此以外, 我们推荐采用对结局影响最大的因素进行匹配的方法来减少选择性偏倚、增加组间均衡性。如瑞典进行了一项“切除子宫是否会增加患者压力性尿失禁风险”的队列研究, 研究纳入了 165 260 例子宫切除术后的女性作为暴露组, 按照出生年份和居住地匹配了未行子宫切除术的女性 479 506 例作为非暴露组, 经过 30 年的随访观察两组压力性尿失禁的手术比例, 最终认为暴露组受试者压力性尿失禁的手术风险是非暴露组的 2.4 倍 (95% 置信区间 2.3 ~ 2.5)^[11]。匹配的因素不宜过多, 一般为 2 ~ 3 个。选择匹配因素也很重要, 如探讨某种干预措施对肠癌患者生存期的影响, 需要选择患者年龄、肿瘤分期、病变部位左右半结肠这样明确的预后影响因素来匹配, 这样可以很大程度上增加组间均衡性。还需强调的是, 招募受试者时应按照纳入排除标准筛选连续性病例, 直至完成预定筛选的受试者数量, 这是在非随机状态下减少选择性偏倚的方法。

另外, 队列研究中受试者的暴露状态可能会改变, 即出现“队列迁移”。这类人群不可简单剔除, 需继续观察至终点事件发生。待研究结束以后, 再按照设计之初对“暴露”的界定来判定受试者应该属于哪一个组。当然, 在签署知情同意书时应告知受试者尽量维持初始暴露, 不要轻易改变, 更不要频繁改变, 否则将会对数据分析带来很大的困难。在实施过程中也可以给予受试者关于出院管理、饮食调护、康复等方面实质性的专业指导来增加受试者依从性, 以减少队列迁移发生。

2.2 信息性偏倚

信息性偏倚产生于资料收集过程中, 可以源于调查者、受试者, 也可以源于测量方法及工具。由于两组间收集资料的方式不一致, 从而导致了研究结果与真实结果间的系统误差, 一旦出现, 很难估计其对真实情形曲解程度的大小和方向^[12]。

回顾性队列研究的信息性偏倚几乎不可纠正, 而前瞻性队列研究却可以在研究设计之初采用策略来控制。首先, 在研究设计时尽量采用客观、量化的指标收集资料, 如肿瘤的研究多采用生存期、疾病进展时间、理化检查等作为主要指标, 尽量避免受主观影响的指标。其次, 若无法采用客观、量化指标收集资料, 则建议采用盲法收集资料。王梅等^[13]采用非随机对照研究评价工具 Newcastle-Ottawa 量表对截至 2015 年 12 月 31 日发表的中医药疗效的队列研究文献质量进行评价时发现, 85.84% 的

学者赞同采用盲法收集资料以评价结局事件的方法,这也再次强调了团队协作的必要性。再次,广泛地收取资料,即研究者不仅要收集必需的资料(如患者疗效),同时也要收集一些“虚变量”(如人口学特征资料、不良反应、患者生活质量等),以分散被调查者对必需的“真变量”的注意力,防止患者对此疗法有偏爱或偏见,主观夸大或降低此疗法的效果,避免偏倚^[2]。无论采用何种方法收集资料,事先对调查者进行筛选并统一培训是必要的,尤其是对调查项目的解释及敏感问题同义转化的技巧。最后,不应答偏倚通常是因为受试者不理解调查目的及内容,认为调查存在不恰当之处,内容过于繁琐等造成的^[4]。在设计及实施过程中需注意避免此类问题以提高受试者应答率。

2.3 混杂偏倚

混杂偏倚是研究某因素与疾病之间的关系时,由于在设计和分析阶段未加控制,使得一个或多个潜在混杂因素的影响掩盖或夸大了研究因素与疾病之间的联系^[3]。混杂偏倚出现在研究的设计、分析阶段,是唯一可以通过统计学方法控制的偏倚。

首先,应设置适当、清晰的纳入排除标准。对于明确影响结局的混杂因素可在研究设计时不予纳入,以减少混杂偏倚的发生。清晰的排除标准则可确保临床医师按照同样的原则从两队列中排除受试者,以避免排除偏倚。

其次,在分析阶段,除校正分析之外,对于可能影响结局的混杂因素常采用亚组分析、多元回归分析、倾向性评分等统计分析方法。倾向性评分是由 Rosenbaum 等^[6-17]在 20 世纪 80 年代提出的,是一个在缺乏随机化的情况下提高组间均衡性的创新统计方法,近年来越来越被研究者们所重视。倾向性评分尽可能地将所有可能的混杂因素综合为一个变量,通过平衡两组的倾向性评分有效地均衡混杂变量的分布,从而使组间具有可比性,使研究结果接近随机对照的结果,其具体方法包括倾向评分配比、倾向评分分层、倾向评分回归调整 and 倾向评分加权^[8]。一项观察参附注射液对肝功能影响的队列研究^[9]中,采用经典 Logistic 回归方法时,在不考虑混杂因素的情况下无论是否使用参附注射液,两队列患者的门冬氨酸氨基转氨酶(AST)差异均无统计学意义;而采用了广义增强模型(generalized boosted models, GBM)倾向性评分加权 Logistic 回归平衡了大部分混杂因素后,结果提示参附注射液可造成患者 AST 异常升高。与其他几种

方法相比,倾向性评分最大的优势在于可以同时调整多个变量,更加适用于类似中医治疗这种复杂干预措施的疗效评价,而其局限性在于所有与干预措施相关的特征变量需要精确测量,且两组间的特征变量需要有足够的重叠^[2]。

再次,由于队列研究样本量大、观察时间长,脱落则是较常见的问题,尤其是在前瞻性队列研究中,脱落数据的处理十分重要。研究者常常容易混淆“脱落”“退出”“失访”这 3 个反映受试者依从性的术语。“脱落”是指受试者由于各种原因不能完成研究规定的全部流程提前退出^[2],而“退出”和“失访”是其最常见的两种表现形式。队列研究中非暴露组的脱落率往往较高,对于缺失数据的不同处理方法也会对偏倚产生不同的影响。忽视法和填补法多用于完全随机缺失(missing completely at random, MCAR)和随机缺失(missing at random, MAR)。忽视法即简单删除,是处理缺失数据最原始的方法。填补缺失数据的方法很多,目前尚没有一个适合所有情况的金标准,应用较为广泛的是沿用末次观测值(last observation carried forward, LOCF)的方法。对非随机缺失(missing not at random, MNAR),缺失数据依赖于不完全变量本身,数据的处理则需要增加重视。意向性分析主要用于脱落病例,包括发生队列迁移的病例,研究者较为熟悉故不再赘述。敏感性分析主要用于失访病例,旨在探讨失访病例对主要结局的影响,其结果与主要结果不一致或处理效应估计值显著降低时,则需要分析脱落对研究结果的影响^[2],下结论也应该更加慎重。肿瘤研究中常用的生存分析是结合生存时间和结局事件进行分析的统计方法,可同时分析完整数据和截尾数据,从而减少偏倚的发生^[2]。

3 中医药干预性队列研究的特殊性

中医药干预性队列研究有其自身的特殊性。首先,临床医师和受试者对中医药的认识及偏好容易左右受试者的暴露归属,如中医院受试者可能存在疾病分期偏晚、一般情况及预后普遍偏差的情况^[2]等,除了文中提到的控制选择性偏倚及混杂偏倚的方法以外,还需要注意选择临床中心时不同地区之间的搭配及中医院、西医院及综合医院的搭配。其次,西医队列的受试者也常常会联合使用一些中成药,因而研究设计时应清晰地界定暴露因素及暴露强度,便于判断最终暴露归属。

我们曾对中国知网(CNKI)中 1996 年至 2016

年发表的中医药治疗结直肠癌的 RCT 论文以改良 Jadad 评分量表进行质量评价, 文献评分均为 1~3 分, 为低质量文献^[23]。国内中医药 RCT 的质量参差不齐, 这已经引起了国内外研究者的广泛关注^[24]。而究其原因, 除了研究者对 RCT 论文报告规范不够重视之外, 研究方法是否适宜也是值得商榷的^[25]。有学者检索了 2016 年 PubMed 和中文期刊数据库发表的 RCT 和观察性研究的文献数量, 发现国内 RCT 和观察性研究的文献数量比例为 14.7 : 1, 远远超过了国外文献的 1.3 : 1。这可能与国内研究者对观察性研究了解较少, 而对 RCT 盲目依赖有关^[26]。尽管 RCT 是验证干预措施疗效因果关系推断的“金标准”, 然而并不是所有真实世界的临床问题都能因此解决。设计优良、实施严谨、数据可靠的干预性队列研究, 作为真实世界研究为广大临床医师提供真实世界证据 (real-world evidence, RWE), 越来越受到研究机构重视, 英国制药工业协会 2011 年明确提出 RWE 与 RCT 证据是相互补充的关系^[27]。前者在科研场景为制定指南提供证据, 后者在真实场景获得更符合临床实际的证据, 使研究结果更容易转化到临床实践, 并可用于检验指南的可实践性。美国更是在 2016 年底颁布法案, 要求食品药品监督管理局将 RWE 纳入到现有的医疗产品审批和监管程序中^[28]。RCT 作为评价干预措施的效力的“金标准”目前地位仍无法撼动, 但我们期待通过控制偏倚、优化设计增加干预性队列研究的内部真实性, 使得研究者在面对类似中医治疗这样个体化并且复杂干预体系、无法随机化、存在伦理学问题等情况下, 转换研究思路, 让更多患者因临床试验获益。

致谢: 新疆医科大学附属中医医院医学研究设计与数据处理中心孙奇老师对此稿进行了修改并提出了有益的建议, 在此致谢!

参考文献

- [1] WING YK, LAM SP, LI SX, et al. Sleep-related eating disorder and zolpidem: an open interventional cohort study [J]. J Clin Psychiatry, 2010, 71(5): 653-656.
- [2] PEARSON ER, FLECHTNER I, NJØLSTAD PR, et al. Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations [J]. N Engl J Med, 2006, 355(5): 467-477.
- [3] MOUNTAIN D, ERCLEVE T, ALLELY P, et al. Reated-reducing acute chest pain time in the ED: a prospective pre-/post-interventional cohort study, stratifying risk using early cardiac multi-markers, probably increases discharges safely [J]. Emerg Med Australas, 2016, 28(4): 383-390.
- [4] LAU JT, LEUNG PC, WONG EL, et al. The use of an herbal formula by hospital care workers during the severe acute respiratory syndrome epidemic in Hong Kong to prevent severe acute respiratory syndrome transmission, relieve influenza-related symptoms, and improve quality of life: a prospective cohort study [J]. J Altern Complement Med, 2005, 11(1): 49-55.
- [5] 杨保华, 李国胜, 张义峰, 等. 经皮椎体成形和椎体后凸成形治疗骨质疏松性胸腰椎骨折疗效的回顾性队列研究 [J]. 中国循证医学杂志, 2015, 15(4): 377-380.
- [6] BENSON K, HARTZ AJ. A comparison of observational studies and randomized, controlled trials [J]. N Engl J Med, 2000, 342(25): 1878-1886.
- [7] HORWITZ RI, VISCOLI CM, CLEMENS JD, et al. Developing improved observational methods for evaluating therapeutic effectiveness [J]. Am J Med, 1990, 89(5): 630-638.
- [8] BYINGTON RP. Beta-blocker heart attack trial: design, methods, and baseline results. Beta-blocker heart attack trial research group [J]. Control Clin Trials, 1984, 5(4): 382-437.
- [9] 韩梅, 陈薇, 刘建平, 等. 比较效果研究常用方法之二: 队列研究设计 [J]. 现代中医临床, 2015, 22(3): 20-23.
- [10] 费宇彤, 张颖, 刘建平. 再论“队列研究”在中医药临床疗效评价中的应用 [J]. 世界中医药, 2014, 9(10): 1261-1263.
- [11] ALTMAN D, GRANATH F, CNATTINGIUS S, et al. Hysterectomy and risk of stress-urinary-incontinence surgery: nationwide cohort study [J]. Lancet, 2007, 370(9597): 1494-1499.
- [12] 梁万年, 刘民. 肿瘤病因研究中的信息偏倚及其控制 [J]. 中国肿瘤, 1995, 4(7): 11-12.
- [13] 王梅, 王建华, 张抗, 等. 中医药疗效评价队列研究的方法学质量评价 [J]. 中医杂志, 2016, 57(16): 1379-1383.
- [14] 宋永霞, 洪静芳, 谢伦芳, 等. 临床癌症患者研究中存在的偏倚及控制 [J]. 中华护理杂志, 2015, 50(6): 750-754.
- [15] 董卫. 临床研究中常见的偏倚及其控制 [J]. 中国实用内科杂志, 2007, 27(24): 1985-1986.
- [16] ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects [J]. Biometrika, 1983, 70(1): 41-55.
- [17] ROSENBAUM PR, RUBIN DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score [J]. Am Stat, 1985, 39(1): 33-38.
- [18] 叶晓勤. 基于倾向性评分的中医复杂干预临床疗效评价 [J]. 中国中医基础医学杂志, 2012, 18(2): 218-

- 220.
- [9] 张英,杨薇,姜俊杰,等. 运用倾向性评分方法分析参附注射液对肝功能影响的队列研究 [J]. 辽宁中医药大学学报, 2016, 18(9): 66-70.
- [20] 张颖,李迅,费宇彤,等. 倾向指数用于中医药非随机设计临床疗效评价的思考 [J]. 世界中医药, 2014, 9(10): 1272-1274.
- [21] WAHLBECK K, TUUNAINEN A, AHOKAS A, et al. Dropout rates in randomised antipsychotic drug trials [J]. Psychopharmacology (Berl), 2001, 155(3): 230-233.
- [22] GAMBLE C, HOLLIS S. Uncertainty method improved on best-worst case analysis in a binary meta-analysis [J]. J Clin Epidemiol, 2005, 58(6): 579-588.
- [23] 蔡泳. 病例随访资料的统计分析方法: 生存分析 [J]. 上海口腔医学, 2004, 13(4): 315-317.
- [24] ZHANG T, HE WT, ZI MJ, et al. Cohort study on prognosis of patients with metastatic colorectal cancer treated with integrated chinese and western medicine [J]. Chin J Integr Med, 2018, 24(8): 1-6.
- [25] 何文婷,张彤,杨宇飞,等. 中医药治疗结直肠癌临床疗效 Meta 分析及证型分析 [J]. 中医杂志, 2018, 59(22): 1929-1936.
- [26] 刘建平. 中医药临床试验的方法学问题与挑战: 循证医学的观点 [J]. 中西医结合学报, 2006, 4(1): 1-6.
- [27] 张颖,李迅,费宇彤,等. 倾向指数用于中医药非随机设计临床疗效评价的思考 [J]. 世界中医药, 2014, 9(10): 1271-1274.
- [28] 田磊,管欣,马爱霞. 随机对照试验研究与观察性研究的系统比较 [J]. 中国药房, 2018, 29(4): 493-496.
- [29] LEVIN M. 21st century cures act requires FDA to expand the role of real world evidence [EB/OL]. (2016-12-19) [2017-08-30]. <http://www.jdsupra.com/legal-news/21st-century-cures-act-requires-fda-to-56141/>.
- [30] U. S. House Energy and Commerce Committee. Text of house amendment to the senate amendment to H. R. 34, tsunami warning, education and research act of 2015 [EB/OL]. (2016-12-25) [2017-03-21]. <http://www.congress.gov/114/bills/hr34/BELLS-114hr34enr.pdf>.

Bias Control of Interventional Cohort Study and the Particularity in the Field of Traditional Chinese Medicine

ZHANG Tong¹, HE Wenting², FEI Yutong³, ZHANG Ying³, YANG Yufei¹

(1. Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing 100091; 2. Xinjiang Uygur Autonomous Region Hospital of Traditional Chinese Medicine; 3. Evidence-Based Medicine Center, Beijing University of Chinese Medicine)

ABSTRACT As a real-world study, interventional cohort study has good external authenticity, whose evidence level is lower only than randomized controlled trials (RCTs), but its greatest controversy is bias. This paper discussed how to control common biases from the perspective of research design, implementation and statistical analysis. In controlling selective bias, the author emphasizes that the researcher is an observer and should not interfere with the intervention assignment. Exposure factors and intensity should be clearly defined to help researchers to determine the final exposure attribution of subjects. In some special cases, the randomization principle can reduce the occurrence of selective bias. Matching method can increase the balance between groups. In controlling information bias, it is suggested to use objective and quantitative indicators to collect data. If this is difficult to achieve, then "blinded" data collection method shall be used. For wide collection of information, the content should be clear and easy to understand, in order to avoid too cumbersome to create non-response bias. In controlling compounding bias, it is necessary to properly and clearly set the inclusion and exclusion criteria. The clear confounding factors can be excluded to reduce the occurrence of confounding bias. In the analysis phase, the correction analysis, subgroup analysis, multivariate regression analysis, and propensity scores are commonly used for controlling bias. In addition, the interventional cohort study of traditional Chinese medicine has its own particularity. When choosing sub-centers, researchers should focus on the match of different regions, as well as traditional Chinese medicine hospitals, western medicine hospitals and general hospitals. The author expects to improve the internal authenticity and evidence level of interventional cohort studies by controlling bias and optimizing design. Researchers can change research ideas in case of such individualized and complicated intervention systems like traditional Chinese medicine treatment, which does not allow randomization, and often with ethical issues, and let more patients benefit from clinical trials.

Keywords intervention cohort study; bias; real-world research; methodology

(收稿日期: 2018-10-17; 修回日期: 2018-12-26)

编辑: 柴倩云