

如何利用临床资料进行回顾性队列研究

王小钦

复旦大学循证医学中心，上海 200040
复旦大学附属华山医院血液科，上海 200040

电话：021-52888252，E-mail：wangxiaoqin@shmu.edu.cn

【摘要】 临床病例资源丰富，但往往存在基线资料不全、非随机分组导致混杂因素多、失访率高等问题，限制了临床医生利用这些数据开展研究。采用限制、配对、分层、多因素分析、倾向性评分等多种方法可以克服已有临床病例资料的固有缺陷，从而得出可靠结论。本文从队列研究的设计、分析、论文撰写等方面阐述如何进行高质量的临床研究。

【关键词】 临床研究；回顾性队列研究；混杂因素；倾向性评分

【中图分类号】 R-1 **【文献标志码】** A **【文章编号】** 1674-9081(2019)01-0073-04

DOI： 10.3969/j.issn.1674-9081.2019.01.011

How to Use Clinical Data to Do Retrospective Cohort Study?

WANG Xiao-qin

Evidence-based Medicine Centre of Fudan University, Shanghai 200040, China
Department of Hematology, Huashan Hospital, Fudan University, Shanghai 200040, China

Tel: 86-21-52888252, E-mail: wangxiaoqin@shmu.edu.cn

【Abstract】 Clinical data are abundant, but there are some problems that limit clinicians' use of these data for clinical research, such as incomplete baseline data, confounding bias due to nonrandom grouping, and higher loss of follow-up. The inherent defects of retrospective design can be overcome by limitations, matching, stratification, multivariate analysis, propensity score, and other methods. This article elaborated how to do a high-quality cohort study from the perspectives of design, analysis and paper writing.

【Key words】 clinical research; retrospective cohort study; confounding factors; propensity score analysis

Med J PUMCH, 2019, 10(1):73-76

临床病例资料数据庞大、资源丰富，如果临床医生合理利用其进行临床研究，则可节约大量时间和经费，很快得出结论和成果。但已有临床病例资料往往存在基线数据不全、不详细，治疗方案不统一，随访间期不等，失访率高等问题，故需要良好的研究设计和统计分析来克服这些缺陷，才能得出可靠的结论。

临床研究根据有无人为设计的干预因素分为试验性研究和观察性研究，观察性研究又根据有无对照组，分为描述性和分析性研究，队列研究、病例-对照研究和横断面研究是常用的分析性研究设计方法^[1]。除了不能进行试验性研究和前瞻性队列研究以外，现有临床资料可进行病例报道和病例分析，也可进行病例-对照研究、回顾性队列研究和横断面研究，故可设

计的研究类型非常多。例如<10例的罕见病可进行病例报道，≥10例可进行病例分析。病因学研究多采用病例-对照方法进行设计，疗效评价、预后研究和病因学研究可采用回顾性队列研究设计，诊断试验评价可采用横断面研究方法。由于回顾性队列研究应用范围最广，故本文重点介绍如何利用临床病例资料进行回顾性队列研究的设计和统计分析。

1 设计阶段需注意的几个问题

1.1 提出 PICO 问题

回顾性队列研究的基本步骤包括：(1) 明确研究目的，确立结局指标和研究因素，提出 PICO (P: 研究对象, I: 干预或暴露因素, C: 对照组, O: 结局指标) 问题；(2) 确定研究对象的纳入和排除标准；(3) 计算样本量；(4) 收集病例的基线资料；(5) 随访病例结局；(6) 进行统计分析，校正各种混杂因素；(7) 总结、撰写论文。

写出 PICO 问题非常重要，其可帮助研究者理清思路和明确研究目的。例如，研究 A 基因突变是否是急性髓系白血病复发的高危因素，P (研究对象) 是某种类型的急性髓系白血病，I (暴露因素) 是有 A 基因突变，C (对照组) 是无 A 基因突变，O (结局指标) 是白血病复发，在统计分析阶段需要校正的因素有年龄、白细胞计数、原始细胞计数、染色体核型、化疗方案等影响结局的因素。

1.2 基线数据缺失问题

针对基线资料不全的缺陷，在设计课题阶段可采用限制方法，即设立纳入和排除标准，把关键资料缺失的病例排除。为了样本有代表性，一般设立一个病例纳入的年限，如纳入 2015 年 1 月 1 日至 2018 年 12 月 31 日符合纳入条件的所有病例（连续样本），且排除标准不能太严格。如果排除的病例较多，最好进行一个缺失数据模式分析，如为完全随机缺失或随机缺失，删去缺失病例对结果的影响较小。也可比较纳入对象和剔除对象的基线临床特征，如无临床特征显著差异，则说明纳入病例的代表性较好，基本可以代表总体病例。

1.3 计算样本量

回顾性分析中只要有对照组，需要进行统计学分析、计算 P 值的研究都需要计算样本量，因为样本量太小，有可能得到假阴性的结果。样本量的计算取决于主要结局指标^[2]，例如研究某一疾病的缓解率、无进展生存率、总生存率的样本量是完全不同的，样本

量依次递增。故研究中设计主要结局指标和次要结局指标很重要。

1.4 配对方法

队列研究中，可以把无暴露因素（或治疗方法）的所有病例作为对照组，也可在设计阶段采用配对方法进行对照组病例选择，例如进行年龄、重要的预后因素（疾病分期、亚型、并发症等）配对，根据不同疾病而定，使两组可比性较好。如果应用手工配对方法，一般配对的因素为 2~4 个，否则很难配到合适的对照组。如果应用一些统计方法进行配对，如下面案例应用的迭代扩展半径法 (iterative expanding radius matching)，可以适当增加配对因素。配对的比例一般为 1:1 至 1:4，配对比例达到 1:4 以后，再增加配对例数，统计效能增加不多。

案例分析：发表在 *N Engl J Med* 的一项多中心、回顾性队列研究题为“应用人类白细胞抗原 (human leukocyte antigen, HLA) 不相合活体供体进行肾移植有生存获益”^[3]，该研究目的是探讨与单纯等待患者或等待后接受死者供体肾移植患者相比，HLA 不相合的活体供体肾移植是否有生存获益。数据来源于美国移植登记数据库，其中有 1025 例 HLA 不相合的活体肾移植病例，对照组采用 1:5 配对，配对因素为年龄、性别、种族、既往进行肾移植的次数、HLA 抗体水平、有无糖尿病、肾脏替代治疗时间 7 个因素。配对方法采用迭代扩展半径法。从单纯等待未进行肾移植的患者中选择 5125 例作为对照组 1，从等待后进行了 HLA 相合死亡供体肾移植中选择 5125 例作为对照组 2 (表 1)。然后进行统计分析，分析总生存期差异，发现与等待患者相比，及早应用 HLA 不相合的活体肾供体可提高总生存期。

2 分析阶段需要注意的几个问题

在统计分析阶段可采用分层分析、多因素分析、倾向性评分等多种方法校正各种混杂因素，以明确研究因素是否为独立危险因素或预后因素。

2.1 分层分析

在分析阶段，对一些最重要的影响结局、预后因素进行分层分析，目的是观察消除该影响因素以后，两组结局比较是否有变化。例如上文中，对 HLA 抗体阳性的不同水平进行分层分析，发现不同抗体水平下，HLA 不相合活体供体肾移植仍有生存获益。

表 1 HLA 不相合活体肾移植患者与配对对照组的临床特征比较^[3]

配对因素	HLA 不相合活体肾移植患者 (n=1025)	单纯等待未进行肾移植的对照组 1 (n=5125)	等待后进行 HLA 相合死亡供体肾移植的对照组 2 (n=5125)
移植时年龄 ($\bar{x} \pm s$, 岁)	45.0±12.8	45.9±12.1	46.6±12.0
女性患者 [n (%)]	687 (67.0)	3444 (67.2)	3452 (67.4)
黑人患者 [n (%)]	168 (16.4)	844 (16.5)	840 (16.4)
既往进行过肾移植的次数 [n (%)]			
0	612 (59.7)	3323 (64.8)	3321 (64.8)
1	350 (34.1)	1584 (30.9)	1552 (30.4)
2	57 (5.6)	202 (3.9)	231 (4.5)
≥3	6 (0.6)	16 (0.3)	21 (0.4)
中位抗体水平 [$M (Q), \%$]	66 (22, 94)	68 (20, 95)	68 (21, 95)
糖尿病 [n (%)]	210 (20.5)	1047 (20.4)	1047 (20.4)
肾脏替代治疗时间 ($\bar{x} \pm s$, 年)	7.3±7.6	7.3±7.4	7.4±7.4

HLA：人类白细胞抗原

案例分析：在一项比较输血依赖的骨髓增生异常综合征应用去铁治疗和不去铁治疗是否生存期不同的研究中^[4]，结论是去铁治疗可以延长生存期。因输血量是疾病严重程度的一个标志，疾病越严重每月输血量越多，生存期越短。故对每月输血量（≤3 U 红细胞和>3 U 红细胞）进行分层分析，发现不同输血量情况下，去铁治疗均可延长生存期（图 1），更证明了结论的可靠性。

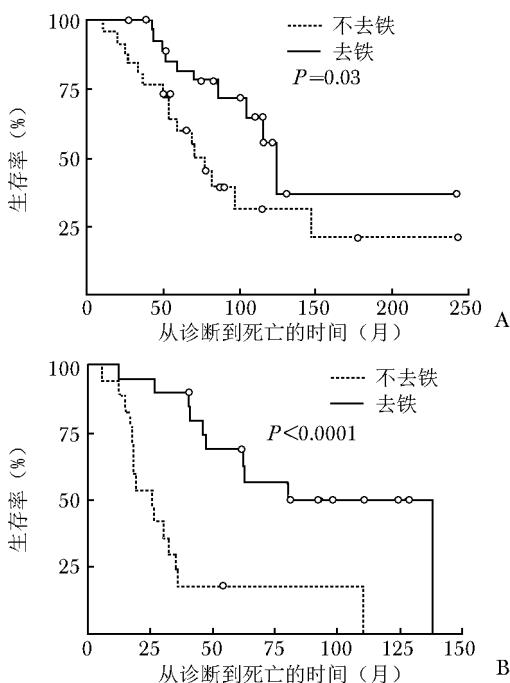


图 1 去铁治疗和不去铁治疗对骨髓增生异常综合征患者生存期的影响^[4]

A. 红细胞输注量≤3 U/月；B. 红细胞输注量>3 U/月

2.2 多因素分析

2.2.1 多因素分析方法

因影响结局的混杂因素较多，多因素分析可同时校正多个混杂因素，故其是队列研究中常用的统计方法。常用的多因素分析方法包括多元回归分析、Logistic 回归分析、COX 回归模型。当结局变量为连续变量时，选择多元回归分析。当结局变量为分类变量，同时又要考虑时间对该结局变量的影响（有时间变量），可选择 COX 回归模型。当结局变量为分类变量，不考虑时间变量时，可以选择 Logistic 回归分析。

案例分析：在一项“老年贫血患者应用促红细胞生成素的治疗反应和心血管结局”的回顾性队列研究中^[5]，应用 Logistic 回归分析应用促红细胞生成素后治疗反应的影响因素，应用 COX 回归分析应用促红细胞生成素后心血管事件的发生结局。治疗反应统一在治疗 2 个月后进行评价，故无需考虑时间变量。而应用促红细胞生成素后心血管事件的发生需要考虑时间变量，有的患者使用后很短时间内就发生了心血管事件，有的患者很久以后才发生心血管事件。研究结论是：不论是不进行校正的单因素分析 [$HR = 1.21$, 95% CI: 0.88~1.66, $P = 0.234$]，还是用 5 个因素（存在其他血细胞减少、血红蛋白水平、年龄、性别、并发症指数）进行校正的多因素分析 [$HR = 0.94$, 95% CI: 0.67~1.31, $P = 0.693$]，均认为应用促红细胞生成素后并不增加心血管事件的发生风险。该文根据不同结局变量，选择了 2 种多因素分析方法。

2.2.2 多因素分析变量的选择

选择多因素分析变量时主要考虑临床意义，从文献或临床经验中确定需要进行校正的因素，即可

能会影响结局的变量。如果样本量有限，要控制进入多因素分析的变量数量，可先通过单因素分析进行初步筛选，然后再纳入多因素分析。文献[4]的研究中，即采用先进行单因素分析，把单因素分析有统计学意义的变量纳入多因素COX回归分析的方法。而文献[5]则根据临床意义选择5个校正变量，未经单因素分析筛选，直接进行多因素分析。在临床研究中，可根据实际情况选择其中一种方法，或2种方法联用。

2.3 倾向性评分法

倾向性评分法可在分析阶段有效平衡非随机对照研究中的混杂偏倚，使研究结果接近随机对照研究的效果，故有学者称其为事后随机化法。其通过配对、分层、回归等分析方法来校正混杂因素，最常用的是配对法。与普通配对法不同的是，倾向性评分法可同时进行多个变量配对，而普通配对法的配对变量是有限的。

案例分析：为比较费城染色体阳性的急性淋巴细胞白血病（acute lymphoblastic leukemia, ALL）应用Hyper-CVAD+帕纳替尼化疗方案与Hyper-CVAD+达沙替尼化疗方案的疗效，但两组基线临床特征有一定差异，故研究者采用倾向性评分配对的方法进行了此项研究^[6]。首先选择配对的8个因素包括年龄、体能状态、白细胞计数、染色体危险度分组、BCR-ABL1的转录类型、是否应用利妥昔单克隆抗体、CD20阳性的原始细胞比例、是否有中枢神经系统累及。然后应用Logistic回归方法计算这8个因素的倾向分值。再应用最近邻居方法对评分进行1:1配对，原有110例ALL病例，经过倾向性评分配对后，41对（82例）病例配对成功。最后，对这82例病例进行统计分析，发现应用Hyper-CVAD+帕纳替尼化疗方案组无事件生存率和总体生存率均优于达沙替尼组。

倾向性评分法能否很好控制混杂因素取决于纳入哪些协变量计算倾向分值，目前尚无公认的标准，但一般认为应纳入已知的、可能影响结局的所有混杂因素。

2.4 失访率问题

回顾性队列研究存在失访是不可避免的，一般要求失访率控制在10%之内，最多不能超过20%，否则会存在选择性偏倚，影响结果，故应尽量控制失访。如果有较多的病例失访，可进行敏感性分析，把失访病例的结局分别按照最差结局和最好结局计算，如结论无显著差异，认为失访对结论的影响较小，结论比

较可靠。否则，结论的可靠性较差。

3 论文撰写规范

队列研究属于观察性研究，撰写论文时要按照观察性研究报告规范（strengthening the reporting of observational studies in epidemiology, STROBE）来进行论文写作。详细的STROBE声明清单见www.strobe-statement.org网站。在进行研究设计之前应先阅读该声明，按照清单上的内容进行设计，才能保证研究质量和论文质量。

4 小结

临床医生可以利用现有病例资料进行各种研究设计，根据不同研究目的选择不同研究方法，回顾性队列研究是常用的设计方法。两组不均衡可比，存在很多混杂因素是回顾性队列研究中比较突出的问题，可应用上述介绍的各种方法来校正混杂因素，同一研究中可同时应用多种方法进行校正。

参 考 文 献

- [1] Schulz KF, Grimes DA.《柳叶刀》临床研究基本概念 [M]. 王吉耀(译). 北京: 人民卫生出版社, 2010: 4-9.
- [2] 曾于珍, 陈世耀. 临床研究结局指标选择与样本量估计 [J]. 协和医学杂志, 2018, 9: 87-92.
- [3] Orandi BJ, Luo X, Massie AB, et al. Survival Benefit with Kidney Transplants from HLA-Incompatible Live Donors [J]. N Engl J Med, 2016, 374: 940-950.
- [4] Rose C, Brechignac S, Vassilieff D, et al. Does iron chelation therapy improve survival in regularly transfused lower risk MDS patients? A multicenter study by the GFM (Groupe Francophone des Myélodysplasies) [J]. Leuk Res, 2010, 34: 864-870.
- [5] Gowanlock Z, Sriram S, Martin A, et al. Erythropoiesis-stimulating agents in elderly patients with anemia: response and cardiovascular outcomes [J]. Blood Adv, 2017, 1: 1538-1545.
- [6] Sasaki K, Jabbour EJ, Ravandi F, et al. Hyper-CVAD plus ponatinib versus hyper-CVAD plus dasatinib as frontline therapy for patients with Philadelphia chromosome-positive acute lymphoblastic leukemia: A propensity score analysis [J]. Cancer, 2016, 122: 3650-3656.

(收稿日期: 2018-05-20)