Postprint: How Clinicians Can Utilize Retrospective Clinical Data to Conduct Retrospective Cohort Studies

Authors: Wang Xiaoqin

Date: 2018-12-04T00:00:00+00:00

Abstract

Retrospective clinical case resources are abundant; however, several issues limit clinicians from utilizing these data to conduct clinical research, such as incomplete baseline data, numerous confounding factors due to non-randomized grouping, and high loss-to-follow-up rates. Various methods including restriction, matching, stratification, multivariate analysis, and propensity scoring can overcome the inherent limitations of retrospective data and yield reliable conclusions.

Full Text

How Clinical Physicians Can Utilize Retrospective Clinical Data for Retrospective Cohort Studies

Evidence-based Medicine Centre of Fudan University, Department of Hematology, Huashan Hospital, Fudan University
Telephone: 021-52888252, Email: wangxiaoqin@shmu.edu.cn

Abstract Retrospective clinical case data are abundant in resources; however, several issues limit clinicians from utilizing these data for clinical research, such as incomplete baseline data, numerous confounding factors due to non-random grouping, and high loss-to-follow-up rates. Various methods including restriction, matching, stratification, multivariate analysis, and propensity scoring can overcome the inherent limitations of retrospective data to yield reliable conclusions.

[**Keywords**] Clinical Research; Retrospective Cohort Study; Confounding Factors; Propensity Score Analysis

Abstract Retrospective clinical data are abundant, but there are some problems that limit clinicians' use of these data for clinical research, such as incom-

plete baseline data, confounding bias due to nonrandom grouping, and higher loss of follow-up. The inherent defects of retrospective design can be overcome by the limitations, matching, stratification, multivariate analysis, propensity score and other methods.

[**Key words**] Clinical Research; Retrospective Cohort study; Confounding factors; Propensity score analysis

Retrospective clinical case data are vast and rich in resources. If clinicians can reasonably utilize them for clinical research, they can save substantial time and funding while quickly reaching conclusions and producing results. However, retrospective data suffer from issues such as incomplete and non-detailed baseline data, non-uniform treatment protocols, unequal follow-up intervals, and high loss-to-follow-up rates. Therefore, robust study design and statistical analysis are required to overcome these limitations and obtain reliable conclusions.

Clinical research can be divided into experimental and observational studies based on whether there is a 人为 designed intervention factor. Observational studies can be further categorized into descriptive and analytical studies based on whether a control group exists. Cohort studies, case-control studies, and cross-sectional studies are commonly used analytical study designs[1]. Except for experimental studies and prospective cohort studies, retrospective data can be used for case reports and case series analyses, as well as for case-control studies, retrospective cohort studies, and cross-sectional studies. Therefore, numerous study types can be designed. For instance, rare diseases can be reported as case reports (<10 cases), while cases with 10 patients can be analyzed as case series.

Etiological studies often employ case-control designs, while therapeutic efficacy evaluation, prognosis research, and etiological studies can utilize retrospective cohort study designs. Diagnostic test evaluations can adopt cross-sectional study methods. Since retrospective cohort studies have the broadest application range, the following sections will focus on introducing how to design and statistically analyze retrospective cohort studies to achieve authentic and reliable conclusions.

1.1 Formulating a PICO Question

The fundamental steps of a retrospective cohort study include: clarifying the research purpose, establishing outcome measures and study factors, formulating a PICO question (Population P, Intervention or Exposure I, Comparison C, Outcome O); determining inclusion and exclusion criteria for study subjects; calculating sample size; collecting baseline data of cases; following up case outcomes; conducting statistical analysis to adjust for various confounding factors; and summarizing and writing the manuscript.

Articulating the PICO question is crucial as it helps researchers organize their thoughts and clarify research objectives. For example, when investigating whether mutation A is a high-risk factor for acute myeloid leukemia (AML) relapse, P (population) would be a specific type of AML, I (exposure factor) would be the presence of mutation A, C (comparison group) would be absence of mutation A, and O (outcome) would be leukemia relapse. During the statistical analysis phase, factors that influence outcomes need to be adjusted for, including age, white blood cell count, blast cell count, chromosome karyotype, and chemotherapy regimens.

1.2 Addressing Missing Baseline Data

To address the limitation of incomplete baseline data, restriction methods can be employed during the study design phase by establishing inclusion and exclusion criteria that exclude cases with missing critical information. To ensure sample representativeness, a specific enrollment period is typically defined, such as including all eligible cases (consecutive sample) from January 1, 2015, to December 31, 2018, with exclusion criteria not being overly stringent. If a substantial number of cases are excluded, a missing data pattern analysis should be conducted. If the missing data are either completely random or randomly missing, deleting cases with missing data will have minimal impact on results. Additionally, baseline clinical characteristics can be compared between included and excluded subjects. If no significant differences in clinical characteristics exist, this demonstrates that the included cases have good representativeness and can essentially represent the overall population.

1.3 Sample Size Calculation

Is sample size calculation necessary for retrospective analyses? As long as a control group exists and statistical analysis with P-value calculation is required, sample size calculation is essential, as insufficient sample size may lead to false-negative results.

Sample size calculation depends on the primary outcome measure[2]. For instance, the sample sizes required for studying a disease’ s remission rate, progression-free survival rate, and overall survival rate are entirely different, with sample sizes increasing sequentially. Therefore, designing primary and secondary outcome measures in research is critically important.

1.4 Matching Methods

In cohort studies, all cases without a specific exposure factor (or treatment method) can serve as the control group. Alternatively, matching methods can be employed during the design phase to select the control group, such as matching for age and important prognostic factors (disease stage, subtype, complications, etc.) depending on the specific disease, to improve comparability between groups. If manual matching is applied, generally 2-4 matching factors are used, as it becomes difficult to find appropriate matches with more factors. If statistical methods are used for matching, such as the iterative expanding radius matching method applied in the case study below, the number of matching factors can be appropriately increased. The matching ratio typically ranges from 1:1 to 1:4; beyond a ratio of 1:4, further increasing the number of matched cases yields minimal additional statistical power.

Case Study: A multicenter retrospective cohort study published in the *New England Journal of Medicine* entitled "Survival Benefit with Kidney Transplants from HLA-Incompatible Live Donors" [3] investigated whether HLA-incompatible living donor kidney transplantation provides a survival benefit compared to patients who simply remained on the waiting list (or those who later received deceased donor kidney transplantation). Data were obtained from the United States transplant registry database, which included 1,025 cases of HLA-incompatible living donor kidney transplantation. The control group was matched at a 1:5 ratio using seven matching factors: age, sex, race, number of previous kidney transplants, HLA antibody level, presence of diabetes, and duration of renal replacement therapy. The iterative expanding radius matching method was employed. From patients who remained on the waiting list without receiving kidney transplantation, 5,125 cases were selected as Control Group 1, and from those who later received HLA-matched deceased donor kidney transplantation, 5,125 cases were selected as Control Group 2 (Table 1). Subsequent statistical analysis of overall survival differences revealed that early utilization of HLA-incompatible living donor kidneys could improve survival time compared to waiting patients.

Table 1. Comparison of Clinical Characteristics between HLA-Incompatible Living Donor Kidney Transplant Patients and Matched Control Groups

Characteristic	HLA-Incompatible Living Donor Kidney Transplant (N=1025)	Control Group 1: Waiting List Only (N=5125)	Control Group 2: Deceased Donor Transplant (N=5125)
Age at transplant (years)	45.0±12.8	45.9±12.1	46.6±12.0
Female patients - n (%)	687 (67.0)	3444 (67.2)	3452 (67.4)
Black patients - n (%)	168 (16.4)	844 (16.5)	840 (16.4)

Characteristic	HLA-Incompatible Living Donor Kidney Transplant (N=1025)	Control Group 1: Waiting List Only (N=5125)	Control Group 2: Deceased Donor Transplant (N=5125)
Number of previous kidney transplants - n (%)	612 (59.7)	3323 (64.8)	3321 (64.8)
(subcategory)	350 (34.1)	1584 (30.9)	1552 (30.4)
(subcategory)	57 (5.6)	202 (3.9)	231 (4.5)
(subcategory)	6 (0.6)	16 (0.3)	21 (0.4)
Median antibody level (%) - median (quartiles)	66 (22-94)	68 (20-95)	68 (21-95)
Diabetes - n (%)	210 (20.5)	1047 (20.4)	1047 (20.4)
Duration of renal replacement therapy (years)	7.3±7.6	7.3±7.4	7.4±7.4

2. Key Issues in the Analysis Phase of Retrospective Cohort Studies

During the statistical analysis phase, various methods such as stratified analysis, multivariate analysis, and propensity scoring can be employed to adjust for confounding factors and clarify whether the study factor is an independent risk or prognostic factor.

2.1 Stratified Analysis

During the analysis phase, stratified analysis is performed on some of the most important factors affecting outcomes and prognosis to observe whether comparisons of outcomes between groups change after eliminating the influence of these factors. For example, in the study mentioned above, stratified analysis was conducted on different levels of HLA antibody positivity, revealing that HLA-incompatible living donor kidney transplantation still provided survival benefits across different antibody levels.

Case Study: In a study comparing whether iron chelation therapy versus no iron chelation therapy affects survival time in transfusion-dependent myelodysplastic syndrome[4], the conclusion was that iron chelation therapy prolonged survival. Since transfusion volume is a marker of disease severity—more severe disease requires more monthly transfusions and is associated with shorter survival—stratified analysis was performed based on monthly transfusion volume (3 units of red blood cells vs. >3 units of red blood cells). The analysis revealed that iron chelation therapy prolonged survival across different transfusion volumes (Figure 2 [Figure 2: see original paper]), further strengthening the validity of the conclusion.

2.2 Multivariate Analysis

Since numerous confounding factors influence outcomes, multivariate analysis—which can simultaneously adjust for multiple confounders—is a commonly used statistical method in cohort studies. Commonly employed multivariate analysis methods include multiple regression analysis, logistic regression analysis, and Cox regression models.

How should these three methods be selected? Multiple regression analysis is chosen when the outcome variable is continuous. The Cox regression model is appropriate when the outcome variable is categorical and the effect of time on that outcome needs to be considered (time-dependent variable). Logistic regression analysis is selected when the outcome variable is categorical and time is not a consideration.

Case Study: In a retrospective cohort study on "Treatment Response and Cardiovascular Outcomes of Erythropoiesis-Stimulating Agents in Elderly Anemic Patients" [5], logistic regression analysis was used to examine factors influencing treatment response after erythropoiesis-stimulating agent administration, while COX regression analysis was applied to analyze cardiovascular event outcomes following treatment, investigating whether erythropoiesis-stimulating agents increased cardiovascular events. Treatment response was uniformly evaluated two months after therapy initiation, thus not requiring time variable consideration. However, the occurrence of cardiovascular events after erythropoiesis-stimulating agent administration required time variable consideration, as some patients experienced cardiovascular events shortly after use while others did so much later. The study concluded that erythropoiesis-stimulating agent use did not increase the risk of cardiovascular events, whether in unadjusted univariate analysis [HR=1.21 (95%CI: 0.88-1.66), P=0.234] or in multivariate analysis adjusting for five factors (presence of other cytopenias, hemoglobin level, age, sex, comorbidity index) [HR=0.94 (95%CI: 0.67-1.31), P=0.693]. This article selected two different multivariate analysis methods based on different outcome variables.

How should variables for multivariate analysis be selected? Factors requiring adjustment can be identified from literature or clinical experience—that is, variables that may influence outcomes. These can be preliminarily screened through univariate analysis, or alternatively, all variables can be included in multivariate analysis without univariate screening. Variable selection should primarily consider clinical significance. However, if sample size is limited, the number of variables entering multivariate analysis must be controlled, making univariate screening feasible. In the aforementioned study on iron chelation therapy improving survival in myelodysplastic syndrome[4], the approach of first conducting univariate analysis and then including variables with statistical significance in multivariate COX regression analysis was adopted. In contrast, the article on erythropoiesis-stimulating agent treatment for elderly anemia[5] selected five adjustment variables based on clinical significance without univariate screening. In our own research, we can choose either method or combine both approaches based on actual circumstances, considering both univariate analysis results and including variables that lack statistical significance in univariate analysis but hold substantial clinical significance.

2.3 Propensity Score Methods

Propensity score methods can effectively balance confounding bias in non-randomized controlled studies during the analysis phase, making study results approximate those of randomized controlled trials (RCTs), hence some refer to it as a post-hoc randomization method. Propensity scores can be applied for matching, stratification, regression, and other analytical methods to adjust for confounding factors, with propensity score matching being the most commonly used. The difference between propensity score matching and conventional matching is that propensity score methods can simultaneously match multiple variables, whereas conventional matching is limited in the number of matching variables.

Case Study: In Philadelphia chromosome-positive acute lymphoblastic leukemia (ALL), which treatment yields better response, event-free survival, and overall survival: Hyper-CVAD chemotherapy regimen plus ponatinib versus Hyper-CVAD plus dasatinib? Because baseline clinical characteristics differed between the two groups, the authors employed propensity score matching for their investigation[6]. First, eight matching factors were selected: age, performance status, white blood cell count, chromosome risk group, BCR-ABL1 transcript type, rituximab use, proportion of CD20-positive blasts, and central nervous system involvement. Logistic regression was then applied to calculate propensity scores for these eight factors. The nearest neighbor method was used for 1:1 matching based on these scores. Of the original 110 ALL cases, 41 pairs (82 cases) were successfully matched after propensity score matching. Finally, statistical analysis of these 82 cases revealed that the ponatinib group had superior event-free survival and overall survival rates compared to the dasatinib group. The ability of propensity score methods to adequately control confounding factors depends on which covariates are included in calculating propensity scores. While no universally accepted standard currently exists, it is generally believed that all known confounding factors that may influence outcomes should be included.

2.4 Loss-to-Follow-Up Issues

Loss-to-follow-up is inevitable in retrospective cohort studies. Generally, the loss-to-follow-up rate should be controlled within 10% and should not exceed 20%, as higher rates may introduce selection bias and affect results. Therefore, every effort should be made to control loss-to-follow-up. If substantial case loss occurs, sensitivity analysis can be performed by calculating outcomes for lost cases under both worst-case and best-case scenarios. If conclusions show no significant differences, the impact of loss-to-follow-up on conclusions is considered minimal, and the conclusions are relatively reliable. Otherwise, the reliability of conclusions is poor.

3. Guidelines for Manuscript Writing

Cohort studies are observational studies, and manuscripts should be written according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines. The detailed STROBE statement checklist is available at www.strobe-statement.org. Researchers should review this statement before designing their study and follow the checklist contents to ensure research quality and manuscript quality.

4. Summary

Clinicians can utilize retrospective data for various study designs, selecting different research methods according to study objectives, with retrospective cohort studies being a commonly used design. Imbalanced comparability between groups and numerous confounding factors are prominent issues in retrospective cohort studies. The various methods introduced above can be applied to adjust for confounding factors, and multiple methods can be used simultaneously within the same study for adjustment.

References

1. Schulz KF, Grimes DA. 王吉耀，译. 《柳叶刀》临床研究基本概念 [M]. 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(１):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(10): 940-950. doi:10.1056/NEJMoa1508380.

4. Rose C, Brechignac S, Vassilief 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(7):864-70. doi: 10.1016/j.leukres.2009.12.004.

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(19):1538-1545. doi: 10.1182/bloodadvances.2017007559.

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(23):3650-3656. doi: 10.1002/cncr.30231.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.