TECHNICAL REPORT #56
MARCH 2012

Analysis of Paired and Clustered Time-to-Event Data: An Annotated Bibliography

Jennifer Le-Rademacher, John P. Klein, Ruta Brazauskas, and Aaron Katch
Division of Biostatistics, Medical College of Wisconsin

1. INTRODUCTION

In comparative studies, paired data arise when treatments are prospectively assigned to pairs of experimental units which are biologically linked such as pairs of eyes from the same patients, skin grafts on the same patients, sets of twins, or litter mates in animal studies. In these studies each treated patient has its own control which hopefully is similar in their survival rate save possibly for the treatment. In many of these experiments a common censoring time may preclude observation of one or the other (or both) of the event times of interest for members of the pair.

Paired data techniques are often suggested as an approach to comparing two treatments in large retrospective studies. Here, a patient given the treatment is artificially matched with a control patients based on a set of key characteristics. While the event times for the treated and control patients within a pair are independent, the baseline hazard rates for the pair may differ from pair to pair.

This retrospective matched pairs design assumes that when patients are matched on one set of covariates they will also be matched on a larger set of covariates. It again allows simple comparisons of like (except for the treatment effect) patients as in the prospective matching design and requires similar methods for analysis. It is 1:1 matching so some patients will be discarded since they are either treated patients for which a control cannot be found or they are extra control cases.

Methods to analyze paired data are well studied for categorical and numerical data. However, when the outcome of interest is survival where censoring is a common occurrence, paired data analysis is more complicated.

This annotated bibliography focuses on nonparametric methods for right censored paired survival data. Although many parametric methods for this type of data exist in the literature,

These tests are constructed by first ranking the data ignoring treatment assignment and pair. The ranking is performed using a redistribute to the right procedure where censored observations are assigned the average rank computed as if they were failures at some time beyond their on-study time. These 'ranks' then replace the original data and the usual paired t-test is computed on the ranks. While derivations assume equal censoring in the two treatments, the author claims that the resulting test is valid in more general censoring schemes.

Key words: paired survival data, k-sample, equal censoring, rank transformation, paired t-test, pooled rank, average rank, redistribute-to-the-right procedure

2. Albers, W. Combined rank tests for randomly censored paired data. *Journal of the American Statistical Association* 1988; **83**: 1159-1162.

   The test proposed in this paper is an extension of the two-sample rank test of Albers and Akritas (1987, [33]). The test computes ranks separately for censored and uncensored observations using the pooled sample and a rank based score is then computed for each observation. The test statistic is calculated from the differences in scores within a pair using a variance adjusted for dependence within a pair. The test assumes a common censoring distribution for all observations. The paper gives optimal score functions for survival times with logistic location alternative and for exponential scale alternatives. An example shows that the result from this test is similar to those of O'Brien and Fleming's test (O'Brien and Fleming, 1987, [10]).

   Key words: paired survival data, equal censoring, rank test, pooled rank

3. Cheng, K. F. Asymptotically nonparametric tests with censored paired data. *Communication in Statistics: Theory and Methods* 1984; **13**: 1453-1470.

   This paper extends the sign rank test based on scores of Wei (1980, [12]) to a more general class of score functions.

   Key words: paired survival data, unpaired data included, unequal censoring, sign rank test

4. Dallas, M. J. and Rao, P. V. Testing equality of survival functions based on both paired and unpaired censored data. *Biometrics* 2000; **56**: 124-159.

   The problem of comparisons of two treatments for data consisting of both matched pairs and independent samples is considered. For the matched pairs, a common censoring time is assumed for members within a pair. A class of permutation tests is constructed using the O'Brien and Fleming (1987, [10]) or the Akritas (1992, [1]) scores from the pooled sample. Permutation tests are performed by looking at all possible permutations of the data between the two samples.

Key words: paired survival data, k-sample survival data, unequal censoring, Gehan-Wilcoxon test, Prentice-Wilcoxon test, log-rank test, consistent variance estimator

9. Mantel, N. and Ciminera, J. L. Use of log-rank scores in the analysis of litter-matched data on time to tumor appearance. *Cancer Research* 1979; **39**: 4308-4315.

The method assigns a censored data log-rank scores to the pooled sample ignoring pairs. Scores for uncensored observations are the expected order statistics of a unit exponential random variable. The scores for censored observations are the score of the closest uncensored observation less than the censored observation inflated by one. Once the scores are assigned, a sign test is constructed based on a comparison of the magnitude of the scores in the two groups within a pair.

Key words: paired survival data, equal censoring, pooled rank, log-rank scores, sign test

10. O'Brien, P. C. and Fleming, T. R. A paired Prentice-Wilcoxon test for censored paired data. *Biometrics* 1987; **43**: 169-180.

Tests are constructed by defining a score for each observation using all observations ignoring pairings. A sign test like statistic is obtained by counting the number of pairs where the score from treatment patients is larger than the score of the (nf)3(l)-3(g(e)4(o)gs10(n t)-6e)4(of)3(t)-2(he)4(s)-

12. Wei, L. J. A generalized Gehan and Gilbert test for paired observations that are subject to arbitrary right censorship. *Journal of the American Statistical Association* 1980; **75**: 634-637.

The test is based on the usual two sample Gehan's Wilcoxon (Gehan, 1965, [37]) test for right censored data. The test uses the numerator of that statistic with a variance corrected for the correlation between pairs.

Key words: paired survival data, unequal censoring, pooled rank, sign test, Gehan-Wilcoxon test

13. Woolson, R. F. and Lachenbruch, P. A. Rank tests for censored matched pairs. *Biometrika* 1980; **67**: 597-606.

Under an assumption of equal censoring for the treated and control subjects within a pair, a generalized rank test for the difference in survival times is computed. Pairs where both observations are censored are removed. For the remaining data the absolute value of the difference between the observed treatment and control on study time is computed. The generalized rank of these right censored observations is computed as is the distribution of these generalized ranks given the signs of the observations. For this data the assumption of common censoring for treatment and control allows for ascertainment of the sign of the differences with singly censored data. Using the joint distribution of the signs and the ranks of the differences, a score test is constructed for the hypothesis of no treatment effect.

Key words: paired survival data, equal censoring, generalized sign test, sign-rank test, reduced sample size, Weibull distribution, double exponential distribution, logistic distribution, score test

Rank-based Tests Performance

14. Lachenbruch, P. A. and Woolson, R. F. On small sample properties of the generalized signed rank and generalized sign tests. *Communications in Statistics - Theory and Methods*f1985; (p

15. Woolson, R. F. and O'Gorman, T. W. A comparison of several tests for censored paired data. *Statistics in Medicine* 1992; **11**: 193-208.

The size and power of several tests for paired survival data are compared in various simulation scenarios. These methods include the paired Prentice Wilcoxon test (O'Brien and Fleming, 1987, [10]), the paired Gehan-Wilcoxon test, generalized signed rank test on the logs of the times and generalized signed rank test on observed times (Woolson and Lachenbruch, 1980, [13]) and Akritas' paired t-test on the ranks (Akritas, 1992, [1]). All tests had the targeted Type I error. The paired t-test on the ranks and the Prentice-Wilcoxon test were found to be slightly more powerful than the other tests.

Key words: paired survival data, equal censoring, Prentice-Wilcoxon test, Gehan-Wilcoxon test, Akritas test, generalized sign-rank test

Tests Based on a Marginal Model

16. Cai, T., Wei, L. J. and Wilcox, M. Semi-parametric regression analysis of clustered failure time data. *Biometrika* 2000; **87**: 867-878.

Inference in a class of linear transformation models is studied for data that consists of many small clusters of observations. This class of models includes the Cox and the proportional odds model as special cases. Data are marginally associated within pairs. Assuming e(e)6w 13-

The paper presents inference procedures for population-averaged regression models of highly stratified failure time data. The models assume linear covariate effects on the log failure times. Inference procedures were developed based on weighted log-rank test statistics with special cases including log-rank statistic and generalized Wilcoxon statistic. The paper also introduces an additional approach using the Buckley-James (Buckley and James, 1979, [35]) estimating equation. Simulation studies show the weighted log-rank and the Buckley-James tests are more efficient than the stratified log-rank test (Klein and Moeschberger, 2003, [25]). When the error distribution is normal, the Buckley-James appro4 Tc 0.00 lati906(r)3( )-2(on)3(dl)-6oeapes t

Key words: paired survival data, unpaired data included, unequal censoring, weighted Kaplan-

23. Hougaard, P. *Analysis of Multivariate Survival Data*. Springer: New York, 2000.

24. Wienke, A. *Frailty Models in Survival Analysis.* Chapman&Hall: Boca Raton, 2011.

Key words: paired survival data, clustered survival data, shared frailty model, gamma frailty, positive stable frailty

Classical Stratified Tests

Classical stratified tests have often been used for paired survival data. These can be found in most standard survival analysis text book such as Klein and Moeschberger (2003, [25]). Included in this category is the weighted stratified log-rank test and the stratified Cox model. For the weighted stratified log-rank test a weighted log4.)

variance method is used to provide critical values for the comparisons between the treatment and control survival functions.

Key words: paired survival data, clustered survival data, variable cluster size, fixed time, weighted Kaplan-Meier, bootstrap variance estimator

27. Su, P. F., Chi, Y., Li, C. I., Shyr, Y., and Liao,Y. D. Analyzing survival curves at a fixed point in time for paired and clustered right-censored data. *Computational Statistics and Data Analysis* 2011; **55**: 1617-1628.

    The problem of comparing two survival curves at a single point in time is considered for paired and clustered survival data. Tests are based on the difference between two Kaplan-Meier estimators. The variance of this difference is computed as the sum of the two Kaplan-Meier variances minus twice the covariance of the two estimators. The needed covariance was originally computed by Murray (2001, [20]). Tests based on comparisons of the transformed (as log, cloglog, logit, and arcsin functions) Kaplan-Meier estimators and the pseudo-values are also computed.

    Key words: paired survival data, clustered survival data, variable cluster size, unequal censoring, fixed time, transformed Kaplan-Meier estimator, pseudo-values

4. ANALYZING CLUSTERED COMPETING RISK DATA

While numerous methods have been proposed for paired survival analysis, methods for paired competing risks analysis remain limited. Existing methods in this area include marginal models or stratified models comparing the cumulative incidence functions or the sub-distributional hazards. These methods were derived for clustered competing risks data with variable cluster sizes. The within cluster dependence is accounted for either by robust variance estimators or by frailty parameters.

28. Chen, B. E., Kramer, J. L., Greene, M. H., and Rosenberg, P. S. Competing risks analysis of correlated failure time data. *Biometrics* 2008; **64**: 172-179.

    The problem of estimation and testing for clustered competing risks data is considered in a marginal model. In this approach the test statistics for the hypothesis of no difference in cumulative incidence between two treatment groups is constructed ignoring the cluster effect. Here either Gray's test (Gray, 1988, [38]) or Pepe and Mori's test (Pepe and Mori, 1993, [41]) is used with a robust variance estimator which adjusts for possible association within clusters.

    Key words: clustered competing risks data, cumulative incidence function, variable cluster size, fixed time, unequal censoring, marginal model, robust variance estimator, Gray's test, Pepe-Mori's test

29. Katsahian, S., Resche-Rigon, M., Chevret, S., Porcher, R. Analysing Multicentre competing risks data with a mixed proportional hazards model for the subdistribution. *Statistics in Medicine* 2006; **25**: 4267-4278.

A frailty model for the sub-distribution hazard of the cause of interest in the presence of

35. Buckley, J. and James, I. Linear regression with censored data. *Biometrika* 1979; **66**: 429-436.

36. Fine, J. P. and Gray, R. J. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* 1999; **94**: 496-509.

37. Gehan, E. A. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 1965; **52**: 203-223.

38. Gray, R. J. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *Annals of Statistics* 1988; **16**: 1141-1151.

39. Klein, J. P. and Andersen, P. K. Regression Modeling of Competing Risks Data Based on Pseudovalues of the Cumulative Incidence Function. *Biometrics* 2005; **61**: 223-229.

40. Pepe, M. S. and Fleming, T. R.