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 useful when the treatment sample size is small and the control sample size is large. It is particularly useful when additional information is needed to confirm the assignment of a patient to the treated group. Of course, it suffers in that 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,

their uses are restricted by their parametric assumptions therefore they are not included in the bibliography. Since our main focus is 1-1 paired data analysis, many of the methods cited below were derived specifically for paired data. However, methods derived for clustered (1-many) time-to-event data that can be applied to paired data are also included.

The effect of treatment on survival is typically quantified by the difference between two survival curves. References for various approaches to compare survival curves for paired or clustered data are given in Section 2. In studies where treatment hazards are non-proportional or where survival curves are expected to cross, the clinicians may be interested in the effect of treatments at a pre-specified time point. Section 3 gives references to two papers describing methods to compare survival probabilities at a fixed point in time for clustered survival data. Section 4 provides references to current approach to analyze clustered competing risks data. For each reference cited, a brief summary and key words describing the method and its associated assumptions are given.

2. COMPARING SURVIVAL CURVES

The most common approach proposed to compare survival curves are sign and rank-based tests (references [1]-[13]). These sign and rank-based tests extend nonparametric tests for independent survival data to paired data. Some of these methods use sign test like inference where the ranks are computed by ignoring treatment assignment, i.e., pooled rank, then the scores are computed from the rank differences within a pair. Other methods use a form of weighted or modified log-rank test where the survival differences between treatments are estimated ignoring the pairing, then robust variance estimators are used to adjust for the within pair dependence. Performances of various sign and rank-based tests are compared by Lachenbruch and Woolson (1985, [14]) and Woolson and O'Gorman (1992, [15]). Another common approach consists of tests based on a marginal model ([16]-[19]). The marginal approach is mainly based on the Cox proportional hazards model. Inferences for this approach are based on robust variance estimators. Other existing approaches include weighted Kaplan-Meier estimators (Murray, 2001, [20]), within-pair comparisons (Dabrowska, 1989, 1990; [21] and [22]), frailty models (Hougaard, 2000; Wienke, 2011; [23] and [24]), and classical stratified tests (Klein and Moeschberger, 2003, [25]).

In many of these methods, doubly censored pairs do not contribute to the test statistics. Therefore, the inference is based on a reduced sample. Some of the methods listed below can be extended beyond pairs to k-sample data where each member of the group is assigned to one of k treatments or to clusters of observations with different sizes. Many methods for paired data require that the observations within pairs have common censoring times whereas methods for clustered data generally allow the observations within a cluster to have different censoring times.

Sign and Rank-based Tests

1. Akritas, M. Rank transform statistics with censored data. *Statistics and Probability Letters* 1992; **13**: 209-221.

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, unpaired data included, equal censoring, Prentice-Wilcoxon score, Akritas score, pooled rank statistic, permutation test

5. Gangnon, R. E. and Kosorok, M. R. Sample-size formula for clustered survival data using weighted log-rank statistics. *Biometrika* 2004; **91**: 263-275.

    A class of weighted log-rank tests for clustered survival data with variable cluster size is presented. A consistent variance estimator accounting for the within-cluster correlation and its limiting distribution are given. A sample-size formula based on simplified assumptions of the weighted log-rank tests is also given.

    Keywords: clustered survival data, variable cluster size, unequal censoring, weighted log-rank test, sample-size formula

6. Jeong, J. H. and Jung, S. H. Rank tests for clustered survival data when dependent subunits are randomized. *Statistics in Medicine* 2006; **25**: 361-373.

    This paper derives the adjusted variance for censored data weighted log-rank tests when data are paired.

    Key words: clustered survival data, variable cluster size, unequal censoring, weighted log rank test

7. Jones, M. P. and Woo, D. Linear sign-rank tests for paired-survival data subject to a common censoring time. *Lifetime Data Analysis* 2005; **11**: 351–365.

    A version of the signed-rank test based on generalized ranks is presented for paired data. The test is based on the differences in the logarithms of the survival times of the treatment and control patients. Assuming a common censoring time this leads to four types of data where these differences are completely known, right-censored, left-censored or completely unknown. The generalized ranks of the absolute values of the differences are computed using techniques similar to those of Prentice (1978, [42]). A sign-rank like test is constructed using the generalized sign rank likelihood.

    Key words: paired survival data, equal censoring, sign-rank test, reduced sample inference

8. Jung, S. H. Rank tests for matched survival data. *Lifetime Data Analysis* 1999; **5**: 67-79.

    The paper presents a class of rank statistics for paired survival data. A consistent variance estimate is given to account for the within pair dependency. The test statistics include a predictable process as a weight function. The log-rank test, the Gehan-Wilcoxon test, and the Prentice-Wilcoxon test are special cases of this particular class of rank tests. The test is generalized to k matched samples when *k* treatments are considered.

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 paired control patient and subtracting this from the count of the number of pairs where the treatment score is smaller than the control score. Under the null hypothesis of no treatment effect, this difference should be zero. In this paper, the scores are computed using the Prentice-Wilcoxon scores (Prentice, 1978, [42]). The method requires that observations within a pair are all censored at the same time. The method is compared to a similar statistic based on the the Gehan-Wilcoxon scores of Wei (1980, [12]) and the log rank scores discussed by Mantel and Ciminera (1979, [9]).

    Key words: paired survival data, equal censoring, pooled rank, sign test, Prentice-Wilcoxon test

11. Schoenfeld, D. A. and Tsiatis, A. A. A modified log-rank test for highly stratified data. *Biometrika* 1987; **74**: 167-175.

    A modified log-rank test is proposed for highly stratified data. The log-rank test statistic is modified to accommodate imbalance between treatment groups within stratum and to allow for censoring distribution that depends on treatment. Under the assumption that the censoring distribution depends on either treatment or stratum but not both, the test statistic has an asymptotic normal distribution with mean zero under the null hypothesis. Simulation studies show that this test is more efficient than the usual stratified log-rank test (Klein and Moeschberger, 2003, [25]) when the number of patients in each stratum is small and when the strata effect is not large. When the strata effect is very large, the stratified log-rank test maintains its power better than the modified log-rank test.

    Key words: clustered survival data, variable cluster size, unequal censoring, modified log-rank test

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* 1985; **14**: 2109-2127.

    This article focuses on examining small sample properties of the generalized signed rank (GSR) and generalized sign (GS) tests proposed for matched pair studies with censored observations by Woolson and Lachenbruch (1980, [13]). Demonstrated simulation study suggests that the GSR is more powerful than the GS, and that censoring does not affect power.

    Key words: paired survival data, equal censoring, generalized sign test, sign-rank test, reduced sample size, small sample properties, simulation study

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 potentially equal cluster sizes, regression models that allow for either observation-specific or cluster-specific time varying covariates are developed using a modified generalized estimating equation approach. A modified sandwich estimator for the variance of the estimators is proposed. Point and interval estimation is also proposed for the predicted survival function.

    Key words: clustered survival data, equal cluster size, Cox model, proportional odds model, linear transformation models, modified sandwich estimator, regression, marginal model

17. Holt, J. D. and Prentice, R. L. Survival analyses in twin studies and matched pair experiments. *Biometrika* 1974; **61**: 17-30.

    Proportional hazards models for paired survival data are studied. The models studied include stratified Cox model assuming pair-specific baseline hazards and more restrictive exponential and Weibull models.

    Key words: paired survival data, equal censoring, stratified Cox model, exponential model, Weibull model, marginal likelihood, reduced sample inference, regression

18. Lee, E. W., Wei, J. L., and Ying, Z. Linear regression analysis for highly stratified failure time data. *Journal of the American Statistical Association* 1993; **88**: 557-565.

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 approach is superior compared to the weighted log-rank test. However, when the error distribution is not normal, the weighted log-rank methods outperform the Buckley-James method.

Key words: clustered survival data, linear regression, marginal model, log-rank statistic, generalized Wilcoxon statistic

19. Lee, E. W., Wei, L. J., and Amato, D. A. Cox-type regression analysis for large numbers of small groups of correlated failure time observations. *Survival Analysis: State of the Art* (Klein and Goel Ed). Kluwer Academic Publishers, 1992; 237-247.

This paper presents the marginal approach to clustered survival analysis. In this approach, a stratified Cox model which ignores dependencies between observations within strata is fit. A robust variance estimator is constructed to account for the correlation between individuals within a cluster. The resulting inference scheme should be more powerful than the independent or stratified Cox model when there are a large number of strata with few observations in each stratum.

Key words: clustered survival data, unequal censoring, marginal model, Cox proportional hazards model, independence working model, robust variance estimator, regression

Weighted Kaplan-Meier

20. Murray, S. Using weighted Kaplan-Meier statistics in nonparametric comparisons of paired censored survival outcomes. *Biometrics* 2001; **57**: 361-368.

A test to compare weighted integrated survival curves for paired data is proposed. It is an extension of Pepe-Fleming's test (Pepe and Fleming, 1987, [40]) with variance adjusted to reflect dependence between paired survival times. This is the censored data paired t-test. Since this method compares the area under the survival curves, it performs better than rank-based tests under crossing hazards and performs comparatively well under proportional hazards. The test also allows for the inclusion of singleton members to contribute to the test statistic. Simulation studies show that size and power increase when the paired test is used for positively correlated data and the inclusion of singletons increases the power when correlation between survival times within pair is low to moderate.

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

Tests Based on Within-pair Comparisons

21. Dabrowska, D. M. Rank tests for matched pair experiments with censored data. *Journal of Multivariate Analysis* 1989; **28**: 88-114.

This method first ranks the uncensored observations in the pooled sample among themselves. Next, each censored observation is assigned the same rank as the nearest uncensored observation on the left. This produces a pair of ranks for the observations within a pair. Using a rank based score, the test statistic is computed as the sum of the differences in ranks of the treated and control observations within a pair. The asymptotic properties of this statistic are derived. These include an estimator of the variance which accounts for the within pair covariance.

Key words: paired survival data, unequal censoring, bivariate symmetry, within-pair comparison, linear rank tests, log-rank test, pooled rank

22. Dabrowska, D. M. Signed-rank tests for censored matched pairs. *Journal of the American Statistical Association* 1990; **85**: 476-485.

A censored data version of the (weighted) signed-rank test for paired data is presented. The test is based on the differences in treated and control survival times within a pair. When the smaller of the two observations within a pair is censored, the pair contributes no information to the test. Assuming a common censoring time within each pair, counts are made for pairs with both observations uncensored and the treated group smaller ($N_1(t)$) or larger ($N_2(t)$) than the control and for those that are singly censored if the treatment or control observation is censored and hence larger than the treated observation ($N_3(t)$) or vice versa ($N_4(t)$). The censored weighted log rank test is the weighted sum of $N_1(t) - N_2(t)$ plus $N_3(t) - N_4(t)$. Weights give a censored data version of the sign test, the sign-rank test and the signed-normal scores test. The asymptotic variance is derived and the test is shown to be asymptotically normal under the null hypothesis.

Key words: paired survival data, equal censoring, bivariate symmetry, within-pair comparisons, conditional model, sign-rank tests, reduced sample inference

Shared Frailty Models

Another common approach to analyzing paired or clustered survival data uses a shared frailty models. These approaches are discussed in detail for example in books by Hougaard (2000, [23]) or Wienke (2011, [24]). In such model, a common random frailty multiplies each hazard rate within a pair. Given the frailty the survival times within a pair are independent. The most common shared frailty models assume the frailty follows either a gamma, a normal, or a positive stable distribution. An advantage of the positive stable frailty model is that if the conditional hazards are proportional then the marginal hazards are also proportional.

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 log-rank statistic is computed in each pair and these are summed over the strata. Only pairs where the shorter of the two observations is uncensored contribute to the statistic. This statistics reduces to the difference in the number of deaths in the two samples which occur while both patients in the pair are at risk given the appropriate weight. Other stratified tests are the score, Wald, or likelihood ratio tests from the Cox model.

25. Klein, J. P. and Moeschberger, M. L. *Survival Analysis: Statistical Methods for Censored and Truncated Data 2nd Edition*. Springer-Verlag, 2003.

    Key words: paired survival data, clustered survival data, stratified log-rank test, stratified Cox model, stratified regression

3. COMPARING SURVIVAL CURVES AT A FIXED POINT IN TIME

Comparisons of survival probabilities at a prespecified time-point are done using naïve, transformed or weighted Kaplan-Meier estimators. Fixed time survival probabilities can also be compared using the pseudo-values approach proposed by Andersen et al. (2003, [34]).

26. Galimberti, S., Sasieni, P., and Valsecchi, M. G. A weighted Kaplan-Meier estimator for matched data with application to the comparison of chemotherapy and bone marrow transplantation in leukemia. *Statistics in Medicine* 2002; **21**: 3847-3864.

    The problem of analyzing data where there is retrospective matching between one treated patient and one or more control patient is considered. The authors propose a weighted Kaplan-Meier estimator of the survival function of the treatment group constructed using the average number of deaths and the average number at risk in each stratum. This estimator can then be compared to the Kaplan-Meier estimator of the survival function of the control group at a fixed point in time. A bootstrap variance estimator is considered for the weighted Kaplan-Meier estimator based on a sample of strata. A permutation test or a bootstrap

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 competing causes of failure and right censoring is presented. The model for the sub distributional hazard rate within a group contains a lognormal random frailty to account for correlated observations from clustered data. The results focus on clustering as a center effects.

    Key words: clustered competing risks data, sub-distributional hazard, unequal censoring, Fine and Gray model, frailty model

30. Logan, B., Klein, J. P. and Zhang, M. J. Marginal models for clustered time to event data with competing risks using pseudo-values. *Biometrics* 2011; **67**: 1-7.

    The paper considers regression models for the cumulative incidence function for clustered competing risks data. In this approach, pseudo-observations of Klein and Andersen (2005, [39]) are computed at a grid of time points using the weighted difference between the complete sample cumulative incidence function and the leave-one-out estimate of the cumulative incidence function. These pseudo-observations are computed ignoring the possible association between individuals within a cluster. A generalized estimating equation model is used to compare treatments. A robust variance model is used to account for association within groups. The technique is particularly useful for comparing cumulative incidence functions with clustered data at a single point in time.

    Key words: clustered competing risks data, cumulative incidence function, variable cluster size, fixed time, unequal censoring, marginal, robust variance estimator, regression, pseudo-values

31. Scheike, T. H., Sun, Y., Zhang, M. J., Jensen, T. K. A semiparametric random effects model for multivariate competing risks data. *Biometrika* 2010; **97**: 133-145.

    A two stage procedure is used to develop a marginal model for the cumulative incidence function for clustered competing risks data. The first stage is to estimate parameters in the additive model of Scheike et al. (2008, [43]) using an estimating equation approach. A robust adjusted variance to account for association between individuals within groups is used to make inference about model parameters. The second stage estimates the dependence parameters.

    Key words: clustered competing risks data, cumulative incidence function, variable cluster size, unequal censoring, random effects, marginal model, semiparametric model, estimating equations, inverse censoring probability weighting

32. Zhou, B., Latouche, A., Rocha, V., and Fine, J. Competing risks regression for stratified data. *Biometrics* 2011; **67**: 661-670.

Stratified regression models using the Fine and Grey (1999, [36]) sub-distributional hazard function are discussed. Inference is based on a proportional sub-distributional hazards model with a distinct baseline rate for each stratum. The inverse probability of censoring weighting (IPCW) technique of Robins and Rotnitzky (1992, [44]) is used to obtain an estimating equation for right censored data. Two types of stratification are studied. The first is the usual stratification where the strata sizes are large and can grow asymptotically. Here the IPCW weights are based on the Kaplan-Meier estimator in each stratum. On the other hand, highly stratified data where there are many small strata of a fixed size (such as matched pairs) uses a weight based on the Kaplan-Meier estimator in the complete sample. Both inference for the risk factors and adjusted estimation of the cumulative incidence for the two types of data are studied.

Key words: clustered competing risks data, sub-distributional hazard, variable cluster size, unequal censoring, marginal, Fine and Gray model, stratified regression, inverse censoring probability weighting

5. SUMMARY

The references cited in this bibliography indicate that paired survival data problem has been well explored. Numerous sign and rank-based tests have been proposed. Marginal models, within-pair comparisons, and frailty models are alternative approaches to paired survival analysis. Surprisingly, there are few options for the analyses of studies where each case is matched to m controls. Existing methods to analyze 1-m matched data are limited to marginal and frailty models, while there is a lack of rank-based methods.

ACKNOWLEDGEMENTS

ADDITIONAL REFERENCES

33. Albers, W. and Akritas, M. C. Combined rank tests for the two-sample problem with randomly censored data. *Journal of the American Statistical Association* 1987; **82**: 648-655.

34. Andersen, P. K., Klein J. P., and Rosthøj S. Generalized linear models for correlated pseudo-observations with applications to multi-state models. *Biometrika* 2003; **90**:15-27.

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. Weighted Kaplan-Meier statistics: A class of distance tests for censored survival data. *Biometrics* 1987; **45**: 497-507.

41. Pepe, M. S. and Mori, M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics and Medicine* 1993; **12**: 737-751.

42. Prentice, R. L. Linear rank tests with right censored data (Corr: V70 p304). *Biometrika* 1978; **65**: 167-180.

43. Scheike, T. H., Zhang, M-J., and Gerds, T. A. Predicting cumulative incidence probability by direct binomial regression. *Biometrika* 2008; **95**: 205–220.

44. Robins, J. M. and Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers. *Aids Epidemiology-Methodological Issues* (Eds. Jewell, Dietz, and Farewell). Birkhauser, Boston, 1992; 297-331.