Communications for Statistical Applications and Methods
2015, Vol. 22, No. 4, 313–320

DOI: http://dx.doi.org/10.5351/CSAM.2015.22.4.313
Print ISSN 2287-7843 / Online ISSN 2383-4757

A Comparison Study of the Test for Right Censored and Grouped Data

Hyo-Il Park[1,a]

[a]Department of Statistics, Chong-Ju University, Korea

Abstract

In this research, we compare the efficiency of two test procedures proposed by Prentice and Gloeckler (1978) and Park and Hong (2009) for grouped data with possible right censored observations. Both test statistics were derived using the likelihood ratio principle, but under different semi-parametric models. We review the two statistics with asymptotic normality and consider obtaining empirical powers through a simulation study. The simulation study considers two types of models the location translation model and the scale model. We discuss some interesting features related to the grouped data and obtain null distribution functions with a re-sampling method. Finally we indicate topics for future research.

Keywords: additive hazards model, grouped data, log-rank test, proportional hazards function, score function

1. Introduction

In the survival analysis, the proportional hazards model (PHM) has been used frequently and applied successfully since Cox (1972) proposed a PHM that has been developed extensively and modified successfully for various statistical situations. However when the proportionality among hazard functions may be suspicious, one may consider an additive hazards model (AHM) as an alternative to analysis convenience and a possibly easy interpretation of the inferential result. To review the PHM and AHM in some detailed fashion, let λ_0 be the baseline hazard function and z, the $p \times 1$ regression vector, which is independent of the time t. Then the hazard function $\lambda(t, z)$ for the PHM and AHM can be represented with the $p \times 1$ regression coefficient vector β as;

$$\lambda(t, z) = \lambda_0 + \exp[\beta' z], \tag{1.1}$$

$$\lambda(t, z) = \lambda_0 + \beta' z, \tag{1.2}$$

where the prime represents the transpose of a vector or matrix. We note that the AHM (1.2) can be reminiscent of the shock model in reliability theory assuming that all arriving shocks to the system are independent.

The PHM (1.1) is well-known to every statistician in the survival field; consequently, it would be redundant to discuss the procedure any further. As an alternative model to the PHM (1.1), the AHM (1.2) has not been widely used and popular because the conditional likelihood method proposed by Cox (1972) cannot be applied to the AHM due to the structure of the hazard function (1.2). The AHM (1.2) was initiated by Aalen (1980, 1989), who considered an inference procedure for λ_0 and β by

[1] Department of Statistics, Cheongju University, Cheongju 360-764, Korea. E-mail: hipark@cju.ac.kr

Published 31 July 2015 / journal homepage: http://csam.or.kr

applying the least squares method instead of using the likelihood principle. Huffer and McKeague (1991) and McKeague (1988) considered weighted least squares estimates for some optimality consideration. Lin and Ying (1994) also proposed an estimate procedure for β using a counting process used for the PHM as an ad hoc approach. McKeague and Sasieni (1994) developed partial parametric AHM. Scheike (2002) investigated the AHM in this direction and proposed an inferential procedure. For the multivariate data case, Yin and Cai (2004) considered inferences based on a marginal AHM approach. Martinussen and Scheike (2006) reviewed extensively the AHM and suggested directions for the research and application; however, Zeng and Cai (2010) considered a recurrent event for AHM. Martinussen *et al.* (2011) provided the estimation of the treatment effect for the AHM while Gerds *et al.* (2013) considered an estimation for a time-dependent concordance index for survival prediction models with covariate dependent censoring.

It is easy to observe objects whether they fail or not periodically or under a time-schedule. For example, after being exposed to the human immunodeficiency virus (HIV), the observation must be carried out periodically since it usually takes several months for blood test results to indicate a HIV negative or HIV positive status. In this case, the corresponding data set contains significant tied value observations even though the underlying life-time distribution is continuous. This type of data set is called grouped data and can be analyzed by a data-specific method. Heitjan (1989) reviewed extensively the methodology and suggested several research directions. For right censored data, Prentice and Gloeckler (1978) considered inferences about β under the PHM. Park (1993) proposed a class of nonparametric tests for the linear model versus Neuhaus (1993) who modified the log-rank tests for the grouped data. Park and Hong (2009) obtained test statistics for the grouped data with AHM under the two sample scheme. Then it would be worthwhile to investigate and compare the efficiencies of two test procedures under the PHM and AHM.

In this study, we consider to compare the efficiency between two nonparametric tests for the AHM and PHM by obtaining the empirical powers through a simulation study. The rest of this paper is organized in the following order. In the next section, we review the two test statistics for the AHM and PHM with a discussion of the limiting distributions under the null hypothesis. Then we compare the efficiency between the two tests under the location translation and scale models in Section 3. In Section 4, we discuss some interesting features concerning the two models and suggest possible future research topics.

2. Score Statistics for Testing $H_0 : \beta = 0$

For this study, we consider $p = 1$. Suppose that we observe life time T_i for the i^{th} individual with some specific constant covariate, z_i, $i = 1, \ldots, n$. We assume that each subject is prone to be censored; consequently, the data set can be represented as $\{(T_i, \delta_i, z_i), i = 1, \ldots, n\}$, where δ_i stands for the censoring status with values 0 or 1 if censored or not. We are concerned with the grouped data; therefore, we assume the positive half real line, $[0, \infty)$ is partitioned into k number of sub-intervals such as $[0, \infty) = \bigcup_{l=1}^{k} [a_{l-1}, a_l)$, with $a_0 = 0$ and $a_k = \infty$. Then one can only have the information that T_i is contained in one of the k sub-intervals for all i. We denote D_l and C_l as the indicate sets for the uncensored and censored observations in the l^{th} sub-interval $[a_{l-1}, a_l)$, respectively. We also denote R_l as the risk set of the l^{th} sub-interval. Finally we denote d_l and r_l as the sizes of D_l and R_l, respectively, $l = 1, \ldots, k$. In this grouped continuous data, we assume that all the censorings occur at the end of a sub-interval and that all deaths proceed any censoring in the same sub-interval. We assume that all observations in the last sub-interval $[a_{k-1}, \infty)$ are censored at a_{k-1} for technical reason with any assumptions disccussed in detail later in the section. Finally we assume that the survival function and

censoring distribution function are independent to avoid identifiability problem. The discrete model in Kalbfleisch and Prentice (1980) that led Prentice and Gloeckler (1978) to propose a score statistic to test $H_0 : \beta = 0$ under the PHM, C_n as follows.

$$C_n = \sum_{l=1}^{k-1} \log\left(\frac{r_l}{r_l - d_l}\right)\left\{\frac{r_l - d_l}{r_l}\sum_{i\in D_l} z_i - \sum_{i\in R_l/D_l} z_i\right\},$$

where log means the natural logarithm and R_l/D_l is the difference set between R_l and D_l for each l, $l = 1, \ldots, k-1$. Then the variance σ_C^2 of C_n can be consistently estimated as

$$\hat{\sigma}_C^2 = \sum_{l=1}^{k-1} \log^2\left(\frac{r_l}{r_l - d_l}\right)\frac{r_l(r_l - d_l)}{r_l}\left\{\frac{1}{r_l}\sum_{i\in R_l} z_i^2 - \bar{z}_l^2\right\},$$

where $\bar{z}_l = r_l^{-1}\sum_{i\in R_l} z_i$. Park and Hong (2009) also proposed a score statistic W_n in the spirit of Kalbfleisch and Prentice (1980) to test $H_0 : \beta = 0$ under the AHM as follows.

$$\begin{aligned} W_n &= \sum_{l=1}^{k-1}\left\{(a_l - a_{l-1})\frac{r_l}{d_l}\sum_{i\in D_l} z_i - (a_l - a_{l-1})\sum_{i\in R_l} z_i\right\} \\ &= \sum_{l=1}^{k-1}(a_l - a_{l-1})\frac{r_l}{d_l}\left\{\sum_{i\in D_l} z_i - \frac{d_l}{r_l}\sum_{i\in R_l} z_i\right\} \\ &= \sum_{l=1}^{k-1}(a_l - a_{l-1})\left\{\frac{r_l - d_l}{d_l}\sum_{i\in D_l} z_i - \sum_{i\in R_l/D_l} z_i\right\}. \end{aligned} \tag{2.1}$$

Then a consistent estimate $\hat{\sigma}_W^2$ of the limiting variance would be of the form

$$\hat{\sigma}_W^2 = \sum_{l=1}^{k-1}(a_l - a_{l-1})^2\frac{r_l(r_l - d_l)}{d_l}\left\{\frac{1}{r_l}\sum_{i\in R_l} z_i^2 - \bar{z}_l^2\right\}.$$

We note that the difference between C_n and W_n is the weight or score. C_n uses $\log[r_l/(r_l - d_l)]$ while W_n does $a_l - a_{l-1}$, the length of the l^{th} sub-interval. Then the two standardized forms

$$\frac{C_n}{\sqrt{\hat{\sigma}_C^2}} \text{ and } \frac{W_n}{\sqrt{\hat{\sigma}_W^2}}$$

converge in distribution into standard normal random variables. We can now test $H_0 : \beta = 0$ with the choice of an appropriate statistic. In the next section, we illustrate the procedures with a dataset and compare the efficiency between C_n and W_n under various scenarios for the model by obtaining empirical powers through a simulation study.

3. An Example and a Simulation Study

For the illustration of two procedures, we consider data reported by Embury *et al.* (1977) for the length of remission (in weeks) for the two groups (maintenance chemotherapy and control) with

Table 1: Some related quantities for W_n and C_n

Test	Value	Variance	p-value
W_n	27.50	253.51	0.083
C_n	4.03	4.81	0.051

Table 2: Exponential distribution

Test	(n_1, n_2)	β					
		0.0	0.1	0.2	0.3	0.4	0.5
W_n	(20, 20)	0.045	0.071	0.189	0.279	0.533	0.607
	(20, 30)	0.055	0.097	0.198	0.297	0.494	0.610
C_n	(20, 20)	0.046	0.071	0.116	0.198	0.319	0.432
	(20, 30)	0.070	0.094	0.149	0.240	0.351	0.459

Table 3: Weibull ($\alpha = 2$) distribution

Test	(n_1, n_2)	β					
		0.0	0.1	0.2	0.3	0.4	0.5
W_n	(20, 20)	0.040	0.083	0.208	0.403	0.608	0.802
	(20, 30)	0.041	0.075	0.195	0.408	0.626	0.795
C_n	(20, 20)	0.062	0.106	0.209	0.373	0.565	0.727
	(20, 30)	0.060	0.115	0.228	0.398	0.594	0.769

acute myelogenous leukemia patients. The length of remission for each patient was measured by week; consequently, the data set contains several tied observations and it would be suitable to use the test procedures based on W_n or C_n. The objective of the experiment was to see if the maintenance chemotherapy prolongs the length of remission. The data has been summarized as:

$$\text{Control group : } 5, 5, 8, 8, 12, 16+, 23, 27, 30, 33, 43, 45,$$
$$\text{Maintenance group : } 9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+,$$

where + indicates censored observation. We note that this is a two-sample problem. Therefore by allocating 0 or 1 to covariate z_i for the i^{th} individual according as from the control or maintenance chemotherapy group in (2.1). Table 1 summarized all results.

In the following, we conduct a simulation study to compare the efficiency between C_n and W_n under the two-sample problem setting. Therefore one can choose 0 or 1 for the value of a covariate z_i according as the observation T_i comes from the first or second sample. For this study, we considered two cases of models for any two random variables X and Y that have some real number β,

$$Y = \beta + X \tag{3.1}$$

and

$$Y = (1 + \beta)X. \tag{3.2}$$

Tables 2–5 summarize the results under the model (3.1) and Tables 6–9, those under the model (3.2). We note that β in (3.1) is the location translation parameter while β in (3.2) acts as a scale parameter. Thus we compare the efficiency between C_n and W_n by varying the values of β. For the underlying distributions, we considered Weibull and gamma distributions. The Weibull distribution has its probability density function defined as for any $x > 0$ and $\alpha > 0$,

$$f(x) = \alpha x^{\alpha-1} \exp\left[-x^{\alpha}\right].$$

Table 4: Weibull ($\alpha = 4/5$) distribution

Test	(n_1, n_2)	β					
		0.0	0.1	0.2	0.3	0.4	0.5
W_n	(20, 20)	0.059	0.074	0.183	0.243	0.498	0.546
	(20, 30)	0.045	0.091	0.188	0.251	0.509	0.592
C_n	(20, 20)	0.056	0.085	0.145	0.201	0.306	0.392
	(20, 30)	0.060	0.094	0.140	0.222	0.327	0.422

Table 5: Gamma distribution

Test	(n_1, n_2)	β					
		0.0	0.1	0.2	0.3	0.4	0.5
W_n	(20, 20)	0.052	0.087	0.170	0.332	0.487	0.663
	(20, 30)	0.059	0.078	0.407	0.468	0.699	0.898
C_n	(20, 20)	0.074	0.085	0.150	0.249	0.389	0.521
	(20, 30)	0.067	0.092	0.335	0.468	0.697	0.783

Table 6: Exponential distribution

Test	(n_1, n_2)	β					
		0.0	0.1	0.2	0.3	0.4	0.5
W_n	(20, 20)	0.045	0.051	0.069	0.080	0.106	0.140
	(20, 30)	0.055	0.074	0.098	0.125	0.150	0.176
C_n	(20, 20)	0.046	0.060	0.073	0.101	0.131	0.163
	(20, 30)	0.070	0.084	0.103	0.135	0.166	0.214

We note that $\alpha = 1$ implies exponential distribution. We considered three different values of α, $\alpha = 1$ (Tables 2 and 6), $\alpha = 2$ (Tables 3 and 7) and $\alpha = 4/5$ (Tables 4 and 8) in this simulation study. For the gamma distribution (Tables 5 and 9), we chose the following one. For $x > 0$, we have

$$f(x) = \frac{x^{-\frac{1}{2}}}{\Gamma(1/2)2^{\frac{1}{2}}} \exp\left[-\frac{x}{2}\right].$$

For the censored distribution, we considered the exponential distribution with a mean 2 for all cases in order to avoid excessive censoring. Sample sizes were chosen as (20, 20) and (20, 30) and we varied the value of β from 0 to 0.5 by increment with 0.1 for the first sample while fixed as 0 for the second. Consequently, the two distributions F and G coincides when $\beta = 0$ in the tables for the first sample and also when the null hypothesis holds when $\beta = 0$. We also chose a partition of $[0, \infty)$ for grouping as $[0, 0.2), \ldots, [1.8, 2.0), [2.0, \infty)$, i.e., 11 sub-intervals. For each case, we obtained empirical power based on 1,000 simulations. The simulations were conducted with SAS/IML on PC version and the nominal significance level is 0.05.

First, we should note that we cannot compare empirical powers among distributions since the random numbers for each case have not been generated under a unified standard version because the mean and variance of the Weibull distribution cannot be obtained explicitly except for $\alpha = 1$. The two different sample sizes show similar trends in varying with empirical powers. However, we often see that W_n achieves higher performance under the location translation model (3.1) whereas C_n shows better performance for (3.2) as expected. Therefore a test based on W_n may be a reliable alternative when the proportional hazards assumption fails (especially when the location shift holds) and will be examined further in the next section.

Table 7: Weibull ($\alpha = 2$) distribution

Test	(n_1, n_2)	β					
		0.0	0.1	0.2	0.3	0.4	0.5
W_n	(20, 20)	0.040	0.057	0.098	0.149	0.218	0.287
	(20, 30)	0.041	0.053	0.097	0.157	0.224	0.303
C_n	(20, 20)	0.062	0.093	0.154	0.261	0.391	0.490
	(20, 30)	0.060	0.108	0.187	0.298	0.432	0.540

Table 8: Weibull ($\alpha = 4/5$) distribution

Test	(n_1, n_2)	β					
		0.0	0.1	0.2	0.3	0.4	0.5
W_n	(20, 20)	0.059	0.063	0.066	0.087	0.106	0.122
	(20, 30)	0.045	0.074	0.091	0.106	0.130	0.145
C_n	(20, 20)	0.056	0.064	0.073	0.092	0.117	0.144
	(20, 30)	0.060	0.078	0.097	0.114	0.131	0.155

Table 9: Gamma distribution

Test	(n_1, n_2)	β					
		0.0	0.1	0.2	0.3	0.4	0.5
W_n	(20, 20)	0.052	0.061	0.085	0.087	0.131	0.233
	(20, 30)	0.059	0.065	0.085	0.089	0.146	0.241
C_n	(20, 20)	0.074	0.080	0.103	0.170	0.245	0.324
	(20, 30)	0.067	0.077	0.114	0.189	0.263	0.341

4. Concluding Remarks

In this section, we discuss some interesting aspects for the tests under the models (1.1) and (1.2). For this, we consider the case of equal length of sub-intervals. Then under the two-sample problem setting, we note that W_n in (2.1) can be re-written as

$$W_n = \sum_{l=1}^{k-1} (r_{2l} d_{1l} - r_{1l} d_{2l}), \tag{4.1}$$

where r_{jl} and d_{jl} denote the size of risk set and the number of deaths in the l^{th} sub-interval of the j^{th} sample, respectively, $j = 1, 2$. We note that W_n in (4.1) is just the Gehan statistic for the grouped data. Therefore one may consider that (2.1) is a modification of the Gehan statistic for the grouped case. The Gehan test is an extension of the Wilcoxon test for censored data; consequently, the Gehan test must be locally the most powerful against the location translation alternatives (Gill, 1980). Therefore the AHM has more empirical power than PHM under the model (3.1) and PHM does more power under (3.2) in the previous section. When d_l is small relative to r_l, we note that we may have

$$\log \frac{r_l}{r_l - d_l} \approx \frac{d_l}{r_l - d_l}. \tag{4.2}$$

By substituting (4.2) in C_n for $\log(r_l/(r_l - d_l))$ for each l, $l = 1, \ldots, k-1$, we can see that C_n is exactly the same statistic as proposed by Cox (1972). Therefore, when the length of sub-intervals are very fine, then a consideration of C_n instead of Cox's form would be meaningless. Finally we note that there is one uncensored observation at most in each sub-interval that corresponds to the no tied-value case; consequently, the assumption for the allowance of discontinuity of hazard function disappears. When we construct the test statistic W_n for the model (1.2), we assumed that all observations in the

last sub-interval $[a_{k-1}, \infty)$ are censored at a_{k-1}, which is the beginning point of the last sub-interval. This means that the last sub-interval $[a_{k-1}, \infty)$ should not contain any observations. The reason for this is as follows. First, we note that the length of the last sub-interval is infinity. If there is any censored or uncensored observation in the last sub-interval, then the length of the last sub-interval should be included in W_n, which is an absurd expression; in addition, the derivation of W_n becomes impossible for the censored observations in the last sub-interval if we maintain the assumption that the censoring occurs at the end of each sub-interval. However such an assumption becomes insignificant and cannot be applied for the real world in the real experiment because a researcher always observes the objects during a finite time period. For the null distribution, we considered the asymptotic normality based on large sample approximation, which is the standard way of consideration for the null distribution of any given test statistic when dealing with the data included in censored observations. One may consider a re-sampling approach such as the permutation principle (Good, 2000) to obtain a null distribution. Park (1993) and Neuhaus (1993) considered the application of the permutation principle to obtain the null distribution of test statistics for right censored and grouped data. However, one must include the assumption of the equality of unknown censoring distributions (which are of nuisance in the statistical inferences) in the null hypothesis if one applies the permutation principle for the censored data. The resulting permutation test is known as exact but conditional. For the AHM (1.2), we note that if the two hazard functions $\lambda_0(t)$ and $\lambda_1(t)$ have the following relation,

$$\lambda(t) = \lambda_0(t) + \lambda_1(t),$$

then one may consider the survival function $S(t)$ corresponding to $\lambda(t)$ as

$$S(t) = S_0(t)S_1(t),$$

where $S_0(t)$ and $S_1(t)$ are the independent survival functions that correspond to $\lambda_0(t)$ and $\lambda_1(t)$ respectively. One may conclude that the AHM is a sum of several hazard functions whose distributions are independent and it is worthwhile to investigate this relationship more deeply in the near future.

Acknowledgment

The author wishes to express his sincere appreciation to the two referees for pointing out errors and providing valuable suggestions to improve the quality of this paper.

References

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In W. Klonecki, A. Kozek and J. Rosinski (Eds.), *Mathematical Statistics and Probability Theory*, Springer, New York, 1–25.

Aalen, O. O. (1989). A linear regression model for the analysis of life times, *Statistics in Medicine*, **8**, 907–925.

Cox, D. R. (1972). Regression models and life-tables, *Journal of Royal Statistical Society Series B (Methodological)*, **34**, 187–220.

Embury, S. H., Elias, L., Heller, P. H., Hood, C. E., Greenberg, P. L. and Schrier, S. L. (1977). Remission maintenance therapy in acute myelogenous leukemia, *Western Journal of Medicine*, **126**, 267–272.

Gerds, T. A., Kattan, M. W., Schumacher, M. and Yu, C. (2013). Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring, *Statistics in Medicine*, **32**, 2173–2184.

Gill, R. D. (1980). Censoring and stochastic integrals, *Statistica Neerlandica*, **34**, 124.

Good, P. (2000). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses (2nd ed.)*, Springer, New York.

Heitjan, D. F. (1989). Inference from grouped continuous data: A review, *Statistical Science*, **4**, 164–179.

Huffer, F. W. and McKeague, I. W. (1991). Weighted test squares estimation for Aalen's additive risk model, *Journal of the American*, **86**, 114–129.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model, *Biometrika*, **81**, 61–71.

Martinussen, T. and Scheike, T. H. (2006). Additive hazards models, In *Dynamic Regression Models for Survival Data*, Springer, New York, 103–173

Martinussen, T., Vansteelandt, S., Gerster, M. and von Hjelmborg, J. (2011). Estimation of direct effects for survival data by using the Aalen additive hazards model, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**, 773–788.

McKeague, I. W. (1988). A counting process approach to the regression analysis of grouped survival data, *Stochastic Processes and Their Applications*, **28**, 221–239.

McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model, *Biometrika*, **81**, 501–514.

Neuhaus, G. (1993). Conditional rank tests for the two-sample problem under random censorship, *Annals of Statistics*, **21**, 1760–1779.

Park, H. I. (1993). Nonparametric rank-order tests for the right censored and grouped data in linear model, *Communications in Statistics-Theory and Methods*, **22**, 3143–3158.

Park, H. I. and Hong, S. M. (2009). A test procedure for right censored data under the additive model, *Communications of the Korean Statistics Society*, **16**, 325–334.

Prentice, R. L. and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data, *Biometrics*, **34**, 57–67.

Scheike, T. H. (2002). The additive nonparametric and semiparametric Aalen model as the rate function for a counting process, *Lifetime Data Analysis*, **8**, 247–262.

Yin, G. and Cai, J. (2004). Additive hazards model with multivariate failure time data, *Biometrika*, **91**, 801–818.

Zeng, D. and Cai, J. (2010). A semiparametric additive rate model for recurrent events with an informative terminal event, *Biometrika*, **97**, 699–712.

Received December 7, 2014; Revised June 8, 2015; Accepted June 15, 2015