Biometrics 64, 157–163
March 2008
DOI: 10.1111/j.1541-0420.2007.00826.x

Nonparametric Inference on Median Residual Life Function

Jong-Hyeon Jeong[1,*] Sin-Ho Jung,[2] and Joseph P. Costantino[1]

[1]Department of Biostatistics, University of Pittsburgh,
Pittsburgh, Pennsylvania 15261, U.S.A.
[2]Department of Biostatistics and Bioinformatics, Duke University,
Durham, North Carolina 27710, U.S.A.
[*]*email:* jeong@nsabp.pitt.edu

Summary. A simple approach to the estimation of the median residual lifetime is proposed for a single group by inverting a function of the Kaplan–Meier estimators. A test statistic is proposed to compare two median residual lifetimes at any fixed time point. The test statistic does not involve estimation of the underlying probability density function of failure times under censoring. Extensive simulation studies are performed to validate the proposed test statistic in terms of type I error probabilities and powers at various time points. One of the oldest data sets from the National Surgical Adjuvant Breast and Bowel Project (NSABP), which has more than a quarter century of follow-up, is used to illustrate the method. The analysis results indicate that, without systematic post-operative therapy, a significant difference in median residual lifetimes between node-negative and node-positive breast cancer patients persists for about 10 years after surgery. The new estimates of the median residual lifetime could serve as a baseline for physicians to explain any incremental effects of post-operative treatments in terms of delaying breast cancer recurrence or prolonging remaining lifetimes of breast cancer patients.

Key words: Breast cancer; Censoring; Martingale; Median residual life; Survival data.

1. Introduction

At a diagnosis of cancer, patients may wish to learn from their physicians an estimate of their expected survival and how much their life expectancy may be extended if they undergo a specific cancer treatment regime. For many treatment regimes, a physician can refer to published Kaplan–Meier plots from previous randomized clinical trials to provide estimates of survival probabilities or median failure times at the time of diagnosis. However, when there exists a possibility of benefiting from a new secondary therapy to be given in the middle of the follow-up period, patients may also wish to know their *residual* life expectancy at several years after the initial diagnosis, and how much it can be prolonged by adopting the new therapy. In contrast to the case for the initial diagnosis, the estimates of the residual lifetime after being followed up for a certain number of years cannot be directly obtained from the Kaplan–Meier plots. The need for such estimates is becoming more critical in breast cancer as long-term courses of secondary therapies are now being considered for patients who remain recurrence free after several years of initial treatment. For example, in recent randomized Phase III clinical studies on breast cancer (Goss et al., 2003; Coombes et al., 2004), estrogen receptor positive patients who had been on tamoxifen for up to 5 years without recurrence of the original disease were rerandomized to placebo or an aromatase inhibitor, either letrozole or exemestane. To advise a patient who would be interested in participating in this type of study, a physician would need to explain the potential benefits of the new secondary course of drug in terms of prolonging the patient's remaining lifetime, given the fact that she has survived for, say, 5 years after originally being treated.

Two natural quantitative measures for the remaining lifetimes are the mean residual life function (Chiang, 1960) and the median residual life function (Schmittlein and Morrison, 1981). Many authors studied the mean residual life function (Chen, Hollander, and Langberg, 1983; Berger, Boos, and Guess, 1988; Oakes and Dasu, 1990; Maguluri and Zhang, 1994; Chen et al., 2005), but not much work can be found in the literature on inference for censored survival data via the median residual life function. Wang and Hettmansperger (1990) proposed a confidence interval approach to compare two quantiles from failure time distributions under censoring. Su and Wei (1993) introduced a nonparametric test statistic to improve Wang and Hettmansperger's procedure that requires estimation of the probability density function of failure times under censoring to evaluate the variance of the median failure time. Both methods, however, were intended only for comparing the remaining lifetimes of patients at the *origin* of the follow-up period. Earlier, Berger et al. (1988) proposed a modified test statistic based on Fligner and Rust's (1982) approach to compare two median residual lifetimes under censoring. Unlike Su and Wei's procedure, however, their approach also involves a nonparametric estimation of the probability density function of the failure time distribution under censoring, which is the major drawback in the median-based inference procedures for censored survival data. In this article, we

propose a method for estimating the median residual lifetimes consistently in a single group and generalize the results of Su and Wei (1993) to compare the median residual lifetimes between two groups at *any* fixed time point during the follow-up period.

The new estimation and testing procedures will be applied to one (B-04) of the oldest National Surgical Adjuvant Breast and Bowel Project (NSABP) data sets, which, with more than 25 years of follow-up, is often viewed as a natural history of prognosis in breast cancer because the patients were treated with surgery (± radiotherapy) without any subsequent systematic chemo- or hormonal therapies. Although it is well known that the detection of disease in the axillary lymph nodes is an important prognostic factor in breast cancer, the issue of association of the nodal status with the pattern of a patient's remaining lifetime has not been addressed in the literature. In this article, novel information on effects of the nodal status on the median residual lifetimes of breast cancer patients is evaluated from the NSABP B-04 data set, conditional on the length of follow-up time. This information could serve as a baseline for physicians to explain any additive effects of post-operative therapies in terms of extending the patients' remaining life years in both node-negative and node-positive breast cancer patients.

In Section 2, we propose an approach to the estimation of the median residual lifetime by solving an equation that involves Kaplan–Meier estimators (Kaplan and Meier, 1958), and derive the variance of the estimating function. In Section 3, we propose a two-sample test statistic and associated confidence interval for the ratio of two median residual lifetimes. In Section 4, simulation studies are performed to assess the proposed testing procedure in terms of empirical type I error probabilities and powers. In Section 5, the developed methodology is applied to the NSABP B-04 data.

2. Estimation of Median Residual Lifetime: One-Sample Case

In one sample, the median residual life function at time t is defined as

$$\theta_t = \text{median}(T - t \,|\, T \geq t), \tag{1}$$

which implies the median of remaining lifetimes among survivors beyond time t. The function (1) satisfies the relation $P(T - t \geq \theta_t \mid T \geq t) = 1/2$, implying that $P(T \geq t + \theta_t) = (1/2)\ P(T \geq t)$. The function θ_t has been extensively studied for *noncensored* cases by many authors, especially Schmittlein and Morrison (1981), Arnold and Brockett (1983), Csörgö and Horvath (1983), Gupta and Langford (1984), Joe and Proschan (1984), Csörgö and Csörgö (1987), Alam and Kulasekera (1993), Song and Cho (1995), among others. Gupta and Langford (1984) noted that θ_t does not uniquely determine $S(t) = P(T \geq t)$, but here we propose a method that models $S(t)$ first and then infer θ_t at a fixed time point t_0. Therefore, in the sequel, θ_{t_0} implies the median residual life function evaluated at a specific time t_0.

In this section, a simple approach to consistent estimation of the median residual lifetime under censoring is proposed for a single group by inverting a function of Kaplan–Meier estimators. The asymptotic distribution of the estimating equation that yields the estimator of the median residual life function is derived via martingale representation of the Kaplan–Meier estimator. This variance formula will be used later to construct a test statistic for comparing two median residual lifetimes.

2.1 *Notations*

Suppose that $n_k (k = 1, 2)$ patients are randomized to a group k. Let $n = n_1 + n_2$. In group k, let T_{ki}, $(i = 1, \ldots, n_k)$ be failure times with survivor function $S_k(t)$ and cumulative hazard function $\Lambda_k(t) = -\log S_k(t)$. Because of early termination of study or loss to follow-up, T_{ki}'s may not be observed. In conjunction with the failure time T_{ki}, let C_{ki} be the censoring time. Then, for a patient i from group k, we observe $\{(X_{ki}, \Delta_{ki}),\ i = 1, \ldots, n_k\}$, where $X_{ki} = \min(T_{ki}, C_{ki})$ and $\Delta_{ki} = I(T_{ki} \leq C_{ki})$. We assume that censoring times are independent of failure times. Let $Y_{ki}(t) = I(X_{ki} \geq t)$ and $N_{ki}(t) = \Delta_{ki} I(X_{ki} \leq t)$ be the at-risk and death processes, respectively, for patient i in group k. We also define $Y_k = \sum_{i=1}^{n_k} Y_{ki}$ and $N_k = \sum_{i=1}^{n_k} N_{ki}$.

2.2 *Estimation*

To simplify the notations, we consider one-sample case by dropping the subscript k in this section. The survivor function of the residual lifetime for a patient who has survived beyond time t_0, i.e., $(T - t_0 \mid T > t_0)$, is given as $S(t \mid t_0) = S(t + t_0)/S(t_0)$ for $t_0 \geq 0$. Hence, the median of the residual lifetime distribution at t_0 can be estimated by solving the equation $\hat{u}(\theta_{t_0}) = 0$ for θ_{t_0}, where

$$\hat{u}(\theta_{t_0}) = \hat{S}(t_0 + \theta_{t_0}) - \frac{1}{2}\hat{S}(t_0) \tag{2}$$

and $\hat{S}(t)$ is the Kaplan–Meier estimator of $S(t)$ from survival data $\{(X_i, \Delta_i),\ 1 \leq i \leq n\}$. Let $\hat{\theta}_{t_0}$ denote the solution. Then, by Theorem 3.4.2 of Fleming and Harrington (1991), $\hat{S}(t)$ is uniformly consistent to $S(t)$ over $0 \leq t \leq \xi$, where $\xi = \sup\{t : y(t) > 0\}, y(t) = S(t)G(t)$, and $G(t) = P(C \geq t)$. Hence, for $t_0 + \theta_{t_0} < \xi, \hat{u}(\theta_{t_0})$ uniformly converges to

$$u(\theta_{t_0}) = S(t_0 + \theta_{t_0}) - \frac{1}{2}S(t_0). \tag{3}$$

Now let $\theta_{t_0,0}$ denote the true value of the median residual lifetime at time t_0. Then we have $u(\theta_{t_0,0}) = 0$, and consequently $\hat{\theta}_{t_0}$ is a consistent estimator of $\theta_{t_0,0}$.

By Corollary 3.2.1 of Fleming and Harrington (1991), at the true value $\theta_{t_0,0}$,

$$\hat{u}(\theta_{t_0,0}) = -\sum_{i=1}^{n} S(t_0 + \theta_{t_0,0}) \int_0^{t_0+\theta_{t_0,0}} \frac{dM_i(s)}{Y(s)} + \frac{1}{2}S(t_0)\sum_{i=1}^{n}\int_0^{t_0} \frac{dM_i(s)}{Y(s)} + o_p(n^{-1/2}),$$

where $M_i(t) = N_i(t) - \int_0^t Y_i(s)\, d\Lambda(s)$ is a martingale (Theorem 1.3.2, Fleming and Harrington 1991), so that $E\{dM_i(t) \,|\, \mathcal{F}_{t-}\} = 0$ for a filtration $\{\mathcal{F}_t : t \geq 0\}$. Because $n^{-1}\, Y(t)$ uniformly converges to $y(t)$ over $[0, \xi]$, we have

$$\hat{u}(\theta_{t_0,0}) = \sum_{i=1}^{n} \epsilon_i + o_p(n^{-1/2}),$$

where

$$\epsilon_i = -S(t_0+\theta_{t_0,0})\int_0^{t_0+\theta_{t_0,0}} \frac{dM_i(s)}{ny(s)} + \frac{1}{2}S(t_0)\int_0^{t_0}\frac{dM_i(s)}{ny(s)}. \quad (4)$$

By the definition of the median residual life function at the true value $\theta_{t_0,0}$, we have $S(t_0+\theta_{t_0,0}) = (1/2)S(t_0)$, which simplifies (4) further to

$$\epsilon_i = -\frac{1}{2}S(t_0)\int_{t_0}^{t_0+\theta_{t_0,0}}\frac{dM_i(s)}{ny(s)}.$$

Because $\epsilon_1,\ldots,\epsilon_n$ are independent random variables with mean 0, by the central limit theorem, $\hat{u}(\theta_{t_0,0})$ is approximately normal with mean 0 and variance $\sigma^2_{t_0}$ that can be consistently estimated by $\hat{\sigma}^2_{t_0} = \sum_{i=1}^n \hat{\epsilon}_i^2$. Here $\hat{\epsilon}_i$ is obtained from ϵ_i by replacing the unknown parameters by their consistent estimators, i.e.,

$$\hat{\epsilon}_i = -\hat{S}(t_0+\hat{\theta}_{t_0})\int_0^{t_0+\hat{\theta}_{t_0}}\frac{d\hat{M}_i(s)}{Y(s)} + \frac{1}{2}\hat{S}(t_0)\int_0^{t_0}\frac{d\hat{M}_i(s)}{Y(s)},$$

or

$$\hat{\epsilon}_i = -\frac{1}{2}\hat{S}(t_0)\int_{t_0}^{t_0+\hat{\theta}_{t_0}}\frac{d\hat{M}_i(s)}{Y(s)},$$

where $\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s)\,d\hat{\Lambda}(s)$ and $\hat{\Lambda}(t) = \int_0^t Y^{-1}(s)\,dN(s)$ is the Nelson–Aalen estimator (Nelson, 1972; Aalen, 1978) of the cumulative hazard function. Hence, a $100\times(1-\alpha)\%$ confidence interval for θ_{t_0} can be obtained by

$$\left\{\theta_{t_0} : \hat{\sigma}^{-2}_{t_0}\hat{u}(\theta_{t_0})^2 < \chi^2_{1,1-\alpha}\right\}, \quad (5)$$

where $\chi^2_{1,1-\alpha}$ is the $100\times(1-\alpha)$th percentile of the χ^2-distribution with 1 degree of freedom.

3. Two-sample Test Statistic and Confidence Interval

Now suppose we want to compare median residual lifetimes between two groups at time t_0. For group k, let θ_{k,t_0} be the median residual lifetime at time t_0. For convenience, suppose we are interested in making inference on the ratio of two median residual lifetimes, i.e., $\tau_{t_0} = \theta_{2,t_0}/\theta_{1,t_0}$. A statistical hypothesis can be formulated as $\mathrm{H}_0 : \tau_{t_0} = \tau_{t_0,0}$ versus $\mathrm{H}_1 : \tau_{t_0} \neq \tau_{t_0,0}$, where $\tau_{t_0,0}$ is a specified value of τ_{t_0} under the null hypothesis. When $\tau_{t_0,0} = 1$, it will be tested whether two median residual lifetimes at a given time t_0 are equal or not. For group k, let the estimating function be

$$\hat{u}_k(\theta_{k,t_0}) = \hat{S}_k(t_0+\theta_{k,t_0}) - \frac{1}{2}\hat{S}_k(t_0).$$

Noting that $\theta_{2,t_0} = \tau_{t_0,0}\theta_{1,t_0}$ under $\mathrm{H}_0 : \tau_{t_0} = \tau_{t_0,0}$, we will consider a two-sample test statistic for τ_{t_0}

$$W_{t_0}(\tau_{t_0,0},\theta_{1,t_0}) = \frac{\hat{u}_1^2(\theta_{1,t_0})}{\hat{\sigma}^2_{1,t_0}} + \frac{\hat{u}_2^2(\tau_{t_0,0}\theta_{1,t_0})}{\hat{\sigma}^2_{2,t_0}}, \quad (6)$$

where $\hat{\sigma}^2_{k,t_0}$ is the variance estimate of $\hat{u}_k(\theta_{k,t_0})$ derived in Section 2.2. We note that the statistic (6) reduces to Su and Wei's (1993) when $t_0 = 0$.

Following similar arguments as in Su and Wei (1993), for any given time t_0 it can be generally shown that $Q_{t_0}(\tau_{t_0,0}) = \inf_{\theta_{1,t_0}} W_{t_0}(\tau_{t_0,0},\theta_{1,t_0})$ follows asymptotically χ^2_1-distribution (see Web Appendix). We reject $\mathrm{H}_0 : \tau_{t_0} = \tau_{t_0,0}$ with type I error α if $Q_{t_0}(\tau_{t_0,0}) > \chi^2_{1,1-\alpha}$. As stated in Su and Wei (1993), an important advantage of using this type of statistic is that there is no need for estimating the underlying probability density function of failure times under censoring to make inference about the ratio of the two median residual lifetimes.

From (6), a $100\times(1-\alpha)\%$ confidence interval for τ_{t_0} can be obtained from

$$\left\{\tau_{t_0} : \inf_{\theta_{1,t_0}} W_{t_0}(\tau_{t_0},\theta_{1,t_0}) < \chi^2_{1,1-\alpha}\right\}. \quad (7)$$

Note that, to achieve a confidence interval from (7), the statistic $W_{t_0}(\tau_{t_0},\theta_{1,t_0})$ needs to be minimized over θ_{1,t_0} for each fixed value of τ_{t_0}. Thus, values of τ_{t_0} associated with the minimum values of the statistic that exceeds the absolute values of $\chi^2_{1,1-\alpha}$ will be the lower and upper limits of the confidence interval.

To accommodate heterogeneity in the population such as age at diagnosis of breast cancer, a stratified test statistic can be also constructed. Denoting l to be the number of strata, the stratified test statistic can be formed as

$$R_{t_0}(\tau_{t_0,0}) = \sum_{j=1}^{l} Q^{(j)}_{t_0}(\tau_{t_0,0}), \quad (8)$$

where $Q^{(j)}_{t_0}(\tau_{t_0,0})$ is the statistic $Q_{t_0}(\tau_{t_0,0})$ that corresponds to the jth stratum. The statistic $R_{t_0}(\tau_{t_0,0})$ will asymptotically follow a χ^2-distribution with l degrees of freedom.

4. A Simulation Study

4.1 *Type I Errors*

A simulation study has been performed first to validate the proposed test statistic in terms of type I error probabilities at the significance level of 0.05 and median lengths of 95% confidence intervals for $\tau_{t_0} = \theta_{2,t_0}/\theta_{1,t_0}$. Failure times were generated from a Weibull distribution with the survival function

$$S(t) = \exp\{-(\rho t)^\eta\}, \quad (9)$$

where ρ and η are scale and index parameters, respectively. By setting $u(\theta_{t_0})$ in (3) equal to zero and solving it for θ_{t_0} after replacing $S(\cdot)$ with the survival function of Weibull distribution in (9), we obtain the median residual life function at time t_0

$$\theta_{t_0} = \frac{1}{\rho}\left\{\log(2) + (\rho t_0)^\eta\right\}^{1/\eta} - t_0, \quad t_0 \geq 0. \quad (10)$$

When the true distribution is exponential, i.e., $\eta = 1$, the function (10) reduces to $(1/\rho)\log(2)$, which does not depend on t_0.

Using the probability integral transformation under the Weibull distribution, we have generated failure times for both groups from

$$T_i = (1/\rho)\{-\log(1-U_i)\}^{1/\eta}, \qquad i = 1,\ldots,n, \quad (11)$$

where U is a random variable following the uniform distribution between 0 and 1. The censoring distribution was assumed to follow a uniform distribution between c and 15, where c determines the censoring proportion and the maximum follow-up period was limited to 15 years. The parameter values for

Table 1
Empirical 95% *coverage probabilities and ML of empirical* 95% *confidence intervals of the two-sample test statistic for comparing median residual lifetimes*

		Censoring proportion			
	n	0% (ML)	10% (ML)	20% (ML)	30% (ML)
$t_0 = 0$	50	0.978 (0.66)	0.978 (0.65)	0.981 (0.68)	0.976 (0.74)
	100	0.971 (0.44)	0.970 (0.44)	0.971 (0.48)	0.977 (0.50)
	200	0.968 (0.30)	0.966 (0.30)	0.974 (0.32)	0.975 (0.34)
$t_0 = 1$	50	0.980 (0.83)	0.979 (0.84)	0.981 (0.90)	0.976 (0.96)
	100	0.971 (0.56)	0.973 (0.56)	0.976 (0.61)	0.979 (0.65)
	200	0.961 (0.38)	0.961 (0.38)	0.968 (0.41)	0.973 (0.44)
$t_0 = 2$	50	0.974 (1.04)	0.973 (1.06)	0.977 (1.17)	0.976 (1.24)
	100	0.974 (0.71)	0.976 (0.71)	0.976 (0.78)	0.978 (0.83)
	200	0.969 (0.47)	0.971 (0.48)	0.974 (0.52)	0.976 (0.56)
$t_0 = 3$	50	0.984 (1.31)	0.986 (1.34)	0.977 (1.51)	0.979 (1.62)
	100	0.979 (0.88)	0.981 (0.91)	0.981 (0.99)	0.982 (1.07)
	200	0.964 (0.60)	0.967 (0.62)	0.971 (0.68)	0.976 (0.73)
$t_0 = 4$	50	0.981 (1.66)	0.980 (1.81)	0.979 (1.96)	0.979 (2.16)
	100	0.978 (1.08)	0.975 (1.14)	0.977 (1.29)	0.984 (1.39)
	200	0.966 (0.73)	0.973 (0.78)	0.976 (0.86)	0.973 (0.94)
$t_0 = 5$	50	0.967 (2.11)	0.976 (2.46)	0.962 (2.75)	0.964 (2.94)
	100	0.968 (1.39)	0.973 (1.56)	0.977 (1.77)	0.977 (1.87)
	200	0.970 (0.93)	0.974 (1.04)	0.977 (1.17)	0.977 (1.26)

ρ and η have been chosen as 0.2 and 2, so that the true mean and variance of failure times are 4.43 and 5.37, respectively. One thousand samples were simulated under this scenario.

Table 1 summarizes (i) empirical 95% coverage probabilities of the null hypothesis and (ii) the median lengths (MLs) of empirical 95% confidence intervals, at different time points (t_0) for various sample sizes (n) per group and censoring proportions. The maximum of different time points in Table 1 has been determined as the maximum time point that allows the median failure time to exist among the observed survival times beyond that point. For each sample, the empirical coverage probability was calculated as the percentage of $Q_{t_0}(\tau_{t_0,0})$ being less than or equal to $\chi^2_{1,1-\alpha}$ when $\tau_{t_0,0} = 1$, i.e., under H_0. The length of empirical 95% confidence interval for τ_{t_0} was estimated for each sample by using (7). The median value of the 1000 interval lengths was reported as the ML for each case.

One can observe that the empirical type I error probabilities (=1-empirical coverage probabilities) approach the true value of 0.05 as sample size increases at each fixed time point t_0, but tend to be conservative. The conservativeness slightly increases with the censoring proportion. We note that the results for $t_0 = 0$ are comparable to ones presented in Su and Wei (1993). The MLs of empirical 95% confidence intervals are becoming narrower as sample size increases, but wider as t_0 increases, because the variance of the Kaplan–Meier estimator $\hat{S}_k(t)$ increases in t.

4.2 *Powers*

Power analyses were performed under the parametric proportional hazards model (Cox, 1972, 1975) with a single covariate. In terms of the survival function, the model specifies

$$S(t; z) = \exp\{-(\rho t)^\eta \exp(\beta z)\},$$

where z is a group indicator, i.e., $z = 0$ for the control group and $z = 1$ for a treatment group, and β is a corresponding regression coefficient. Under this model, the median residual life function is

$$\theta_{t_0}(z) = \frac{1}{\rho}\Big\{ \exp(-\beta z)\log(2) + (\rho t_0)^\eta \Big\}^{1/\eta} - t_0, \quad t_0 \geq 0. \quad (12)$$

For the control group, failure times were generated from (11). Failure times for a treatment group were generated from

$$T_i = (1/\rho)\{-\exp(-\beta)\log(1 - U_i)\}^{1/\eta}, \qquad i = 1, \ldots, n.$$

Censoring times were generated from a uniform distribution between c and 15, as before, where c is a constant controlling for the censoring proportion. The parameter values for ρ and η have been similarly fixed as 0.2 and 2. Rejected proportions of the null hypothesis of $H_0 : \tau_{t_0} = 1$ were evaluated at the significance level of 0.05, when the true values of β were 0.6 and 0.9. Note that, at the origin with $t_0 = 0$, $\beta = 0$, 0.6, and 0.9 correspond to the median residual lifetimes of 4.16, 5.62, and 6.53, respectively, from (12). Thus the true values of τ_{t_0} corresponding to $\beta = 0.6$ and 0.9 are 1.35 and 1.57 relative to the control group. Table 2 summarizes the results for the censoring proportions of 10%, 20%, and 30%. Overall, the power tends to increase notably as sample size increases. The censoring proportion has a slightly negative impact on powers. Larger values of the fixed time points are associated with lower powers, as expected, but again the power goes up quickly as sample size increases.

5. Application to NSABP B-04 Data

The selection of an appropriate control group for a clinical study is based on cumulative information for drugs that have been identified as effective at the time when a study is designed. As a result, it is not easy to find recent breast cancer

Table 2
Empirical powers of the two-sample test statistic for comparing median residual lifetimes at the significance level of 0.05

		Censoring Proportion					
		10%		20%		30%	
	n	$\tau_{t_0} = 1.35$	1.57	1.35	1.57	1.35	1.57
$t_0 = 0$	50	0.408	0.796	0.392	0.769	0.351	0.711
	100	0.767	0.983	0.712	0.980	0.639	0.960
	200	0.976	0.999	0.964	0.999	0.951	0.999
$t_0 = 1$	50	0.378	0.774	0.349	0.742	0.323	0.673
	100	0.722	0.979	0.669	0.965	0.603	0.941
	200	0.972	0.999	0.963	0.999	0.936	0.999
$t_0 = 2$	50	0.316	0.676	0.273	0.617	0.245	0.549
	100	0.644	0.946	0.584	0.920	0.508	0.884
	200	0.931	0.999	0.897	0.999	0.849	0.995
$t_0 = 3$	50	0.245	0.494	0.211	0.448	0.190	0.389
	100	0.506	0.830	0.432	0.749	0.379	0.668
	200	0.821	0.996	0.769	0.978	0.692	0.964
$t_0 = 4$	50	0.172	0.341	0.157	0.295	0.144	0.284
	100	0.337	0.603	0.285	0.509	0.231	0.459
	200	0.638	0.896	0.553	0.825	0.490	0.763

studies that include, as one of their study arms, the placebo group without any post-operative therapy. In that regard, with more than a quarter century of follow-up, the information contained in the NSABP B-04 data is often viewed as the natural history of breast cancer because the study population involves breast cancer patients who had been treated with surgery (± radiotherapy) without any subsequent systematic chemo- or hormonal therapies.

The NSABP B-04 study was designed to compare radical mastectomy with a less extensive surgery (total mastectomy) with or without radiation therapy. A total of 1079 women with clinically negative axillary nodes underwent radical mastectomy, total mastectomy without axillary dissection but with post-operative irradiation, or total mastectomy plus axillary dissection if their nodes became positive. A total of 586 women with clinically positive axillary nodes underwent either radical mastectomy or total mastectomy without axillary dissection but with post-operative irradiation. Fisher et al. (2002) reported an analysis of the 25-year follow-up update of the B-04 data. About 90% of all patients were either followed for at least 25 years or were known to have died before that time. The treatment groups are well balanced in terms of the percentage of patients with follow-up of less than 25 years. The proportion of patients still alive is very low after the long-term follow-up, i.e., less than 30% among node-negative patients and less than 20% among node-positive patients. The results showed that there were no significant differences among the three groups of women with negative nodes or between the two groups of women with positive nodes with respect to disease-free survival and overall survival, implying that the total mastectomy worked equivalently as the radical mastectomy in both node-negative and node-positive breast cancer patients.

In this example, the median residual lifetimes of breast cancer patients from the B-04 data are estimated and compared *retrospectively* between node-negative and node-positive patients. Because there was no significant difference in overall survival between radical mastectomy and total mastectomy in each nodal group and the patients were not given any post-operative therapies, this comparison between the two nodal groups may be considered not being confounded by any treatment effect. First, we estimate the median residual lifetimes and their 95% confidence intervals every year in each nodal group by solving equations (2) and (5). Then, at every other year, the median residual lifetimes are compared between two nodal groups by using the confidence interval approach in (7). Even though the median follow-up time was close to 25 years in this data set, the results presented here are only through 12 years, which was the maximum time point that allows the median failure time to exist among survivors in both node-negative and node-positive groups.

Figure 1 shows the estimated median residual lifetimes and their pointwise 95% confidence intervals in node-negative and node-positive patients, respectively. It is interesting to observe that the median residual lifetimes in both groups are increasing, i.e., during the first 5 years in node-negative group and during the first 9 years in node-positive group. This may imply that early deaths of the patients at high risk in both groups possibly increase the median residual lifetimes among survivors. It also suggests that there exist notable differences between the two groups throughout about 10 years, even though they tend to converge at the tail. Table 3 summarizes the ratios of the two median residual lifetimes and their 95% confidence intervals calculated from (7) at every other year. For example, at year 4, the median residual life years among node-negative and node-positive patients were 13.05 and 8.24, respectively, resulting in the ratio of 0.63 with 95% confidence interval of (0.49, 0.81). Overall, the results indicate that the median residual lifetimes are significantly different between node-negative and node-positive breast cancer patients up to year 8 at the 5% significance level. Equivalently, any statistically significant difference in median residual lifetimes between the two nodal groups fades away about 10 years after surgery, if there is no systematic post-operative therapy.

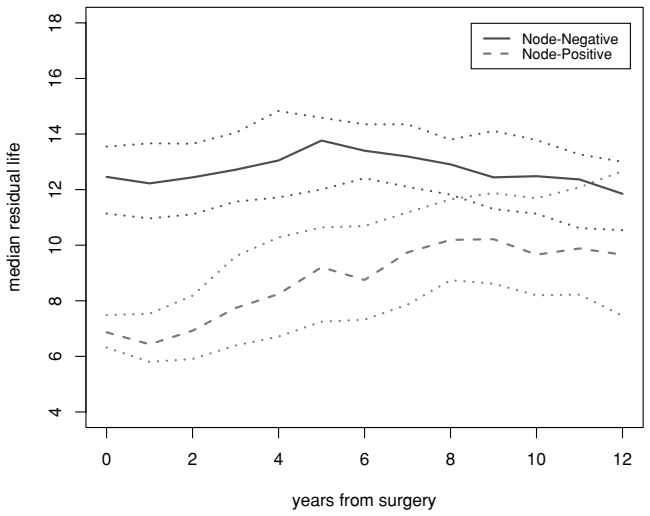

Figure 1. Estimated median residual lifetimes in node-negative and node-positive breast cancer patients from NSABP B-04 data.

Even though it is well known that the nodal status is one of the most important prognostic factors in breast cancer, the information achieved here may have value in that (i) there exists no information on life expectancy among breast cancer patients in terms of their remaining lifetimes as time progresses and (ii) the new information can be used as baseline expectations for evaluating the efficacy of a systematic therapy in terms of prolonging the patients' remaining lifetimes.

As suggested by a referee, similar analysis has been performed stratified by age group by using the test statistic (8). The younger age group included the patients with age less than or equal to 49, and the older group ones with age greater than or equal to 50. The results showed that the difference in median residual lifetimes between node-negative and node-positive was still statistically significant in favor of the node-negative group throughout 6 years, with p-values of <0.0001, <0.0001, 0.00022, and 0.00035 at years 0, 2, 4, and 6, respectively. Here, the maximum of the fixed time points could be extended only to year 6 for the group comparison, due to higher censoring proportion in younger age group among node-negative patients, i.e., about 41% at $t_0 = 0$.

6. Discussion

In this article, we have generalized Su and Wei's test statistic (1993), so that median residual lifetimes can be compared at *any* given time point. The proposed method can be useful when the efficacy of a drug needs to be evaluated in terms of prolonging patients' remaining lifetimes or delaying recurrence of an original disease in clinical trials. Simulation results show that the proposed test statistic performs reasonably well in terms of type I error probabilities and powers. By applying the proposed method to the NSABP B-04 data set, we could estimate the pattern of change in median residual lifetimes from each nodal group of breast cancer patients as time progresses. The results indicate that significant difference in median residual lifetimes between node-negative and node-positive breast cancer patients disappears about 10 years after surgery.

As pointed out by an associate editor, a potential limitation of statistical modeling by using the median residual lifetimes could be the effect of the proportion of censored observations, because the median of failure times in the mixture distribution of failure and censoring times cannot be theoretically defined when the minimum of the estimated survival curve does not reach 0.5. As illustrated in our simulation study and real data example, the median residual lifetimes can be only estimated up to the maximum time point that allows the median failure time to exist among survivors beyond that point. It should also be noted that the comparison in median residual lifetimes between two groups in our example has been performed at each fixed time point, so that the results need to be interpreted accordingly. To compare the median residual life functions over the entire follow-up period based on our results, the issue of multiple comparisons arises, which was not addressed in this article.

Our theoretical results in Section 2.2 are based on the uniform consistency of the Kaplan–Meier estimator $\hat{S}(t)$ over $0 < t < \xi = \sup\{t : S(t)G(t) > 0\}$, which leads to a condition that $t_0 + \theta_{t_0} < \xi$ in our case. Here, ξ may be considered as the maximum follow-up time, which has the lower bound $t_0 + \theta_{t_0}$ assuming that the median residual lifetime θ_t exists at $t = t_0$.

An important extension would be to develop a regression model to take into account continuous prognostic factors for a group comparison. For example, a model can be constructed to compare median residual lifetimes between two groups at a given time, adjusting for age at entry, number of positive

Table 3
Estimated median residual lifetimes in node-negative and node-positive groups, the ratios, and 95% *confidence intervals for the ratios* (*NSABP B-04 data*)

	Median residual lifetime			
t_0	Node-negative	Node-positive	Ratio	95% CI
0	12.46	6.87	0.55	(0.49, 0.63)
2	12.44	6.93	0.56	(0.47, 0.70)
4	13.05	8.24	0.63	(0.49, 0.81)
6	13.40	8.75	0.65	(0.54, 0.81)
8	12.91	10.19	0.79	(0.66, 0.93)
10	12.48	9.66	0.77	(0.62, 1.00)
12	11.85	9.66	0.82	(0.63, 1.08)

lymph nodes, tumor size, and estrogen receptor level as continuous variables. In some cases, however, it may be more appropriate to apply the median residual life regression model not just at fixed time points but over the entire support region as in Gelfand and Kottas (2003) and Chen et al. (2005), which requires semiparametric model assumptions such as accelerated life or proportional median residual life. This issue is under further investigation.

7. Supplementary Materials

The derivation of the distribution of $Q_{t_0}(\tau_{t_0,0})$ in Section 3 is available under the Paper Information link at the *Biometrics* website `http://www.tibs.org/biometrics`.

Acknowledgements

We are grateful to an associate editor and a referee for their valuable comments. This research was supported in part by the Department of Defense (DOD) grant W81XWH-04-1-0605 and National Health Institute (NIH) grants 5-U10-CA69974-09 and 5-U10-CA69651-11.

References

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics* **6,** 701–726.

Alam, K. and Kulasekera, K. B. (1993). Estimation of the quantile function of residual life time distribution. *Journal of Statistical Planning and Inference* **37,** 327–337.

Arnold, B. C. and Brockett, P. L. (1983). When does the βth percentile residual life function determine the distribution? *Operations Research* **31,** 391–396.

Berger, R. L., Boos, D. D., and Guess, F. M. (1988). Tests and confidence sets for comparing two mean residual life functions. *Biometrics* **44,** 103–115.

Chen, Y.-Y., Hollander, M., and Langberg, N. A. (1983). Tests for monotone mean residual life, using randomly censored data. *Biometrics* **39,** 119–127.

Chen, Y. Q., Jewell, N. P., Lei, X., and Cheng, S. C. (2005). Semiparametric estimation of proportional mean residual life model in presence of censoring. *Biometrics* **61,** 170–178.

Chiang, C. L. (1960). A stochastic study of the life table and its applications: I. Probability distributions of the biometric functions. *Biometrics* **16,** 618–635.

Coombes, R. C., Hall, E., Gibson, L. J., Paridaens, R., Jassem, J., Delozier, T., Jones, S. E., Alvarez, I., Bertelli, G., Ortmann, O., Coates, A. S., Bajetta, E., et al. (2004). A randomized trial of exemestane after two to three years of tamoxifen therapy in postmenopausal women with primary breast cancer. *The New England Journal of Medicine* **350,** 1081–1092.

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

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Csörgö, M. and Csörgö, S. (1987). Estimation of percentile residual life. *Operations Research* **35,** 598–606.

Csörgö, S. and Horvath, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **62,** 411–426.

Fisher, B., Jeong, J. H., Anderson, S., Bryant, J., Fisher, E. R. and Wolmark, N. (2002). Twenty-five year findings from a randomized clinical trial comparing radical mastectomy with total mastectomy and with total mastectomy followed by radiation therapy. *The New England Journal of Medicine* **347,** 567–575.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Fligner, M. A. and Rust, S. W. (1982). A modification of Mood's median test for the generalized Behrens-Fisher problem. *Biometrika* **69,** 221–226.

Gelfand, A. E. and Kottas, A. (2003). Bayesian semiparametric regression for median residual life. *The Scandinavian Journal of Statistics* **30,** 651–665.

Goss, P. E., Ingle, J. N., Martino, S., Robert, N. J., Muss, H. B., Piccart, M. J., Castiglione, M., Tu, D., Shepherd, L. E., Pritchard, K. I., Livingston, R. B., Davidson, N. E., Norton, L., et al. (2003). A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. *The New England Journal of Medicine* **349,** 1793–1802.

Gupta, R. C. and Langford, E. S. (1984). On the determination of a distribution by its median residual life function: A functional equation. *Journal of Applied Probability* **21,** 120–128.

Joe, H. and Proschan, F. (1984). Percentile residual life functions. *Operations Research* **32,** 668–678.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimator from incomplete observations. *Journal of the American Statistical Association* **53,** 457–481.

Maguluri, G. and Zhang, C.-H. (1994). Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society Series B* **56,** 477–489.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14,** 945–965.

Oakes, D. and Dasu, T. (1990). A note on residual life. *Biometrika* **77,** 409–410.

Schmittlein, D. C. and Morrison, D. G. (1981). The median residual lifetime: A characterization theorem and an application. *Operations Research* **29,** 392–399.

Song, J. and Cho, G. (1995). A note on percentile residual life. *Sankhya* **57,** 333–335.

Su, J. Q. and Wei, L. J. (1993). Nonparametric estimation for the difference or ratio of median failure times. *Biometrics* **49,** 603–607.

Wang, J. L. and Hettmansperger, T. P. (1990). Two-sample inference for median survival times based on one-sample procedures for censored survival data. *Journal of the American Statistica1 Association* **85,** 529–536.

Received May 2006. *Revised February* 2007.
Accepted February 2007.