

On the Restricted Mean Survival Time Curve Survival Analysis

Lihui Zhao* Brian Claggett† Lu Tian‡
Hajime Uno** Marc A. Pfeffer†† Scott D. Solomon‡‡
Lorenzo Trippa§ L. J. Wei¶

*Northwestern University, lihui.zhao@northwestern.edu

†Brigham & Women's Hospital, Harvard Medical School

‡Stanford University

**Dana-Farber Cancer Institute

††Brigham & Women's Hospital, Harvard Medical School

‡‡Brigham & Women's Hospital, Harvard Medical School

§Harvard University, ltrippa@jimmy.harvard.edu

¶Harvard University, wei@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/harvardbiostat/paper187>

Copyright ©2014 by the authors.

ON THE RESTRICTED MEAN SURVIVAL TIME CURVE IN SURVIVAL ANALYSIS

Lihui Zhao¹, Brian Claggett², Lu Tian³, Hajime Uno⁴, Marc A. Pfeffer², Scott D. Solomon², Lorenzo Trippa⁵, and L. J. Wei^{5*}

SUMMARY

For a study with an event time as the endpoint, its survival function theoretically contains all the information regarding the temporal profile of this outcome variable. The survival probability at a specific time point, say t , however, does not transparently capture the temporal profile of this endpoint up to t . An alternative is to use the restricted mean survival time (RMST) at time t to summarize the profile. The RMST is simply the area under the survival curve up to t . The advantages of using such a quantification over the survival rate have been extensively discussed for the setting of a fixed-time analysis. In this article, we generalize this concept by considering a curve based on the RMST over time as an alternative summary to the survival function. Inference, for instance, based on simultaneous confidence bands for a single RMST curve and also the difference between two RMST curves are proposed. The latter is quite informative for evaluating two groups under an equivalence or non-inferiority setting, and provides a clinically meaningful interpretation for the difference of two groups in a time scale. The proposal is illustrated with the data from two clinical trials, one from oncology and the other from cardiology.

Keywords: Equivalence/noninferiority study; Gaussian process; Martingale; Simultaneous confidence band; Survival function.

¹Department of Preventive Medicine, Northwestern University

²Division of Cardiovascular Medicine, Brigham & Women's Hospital, Harvard Medical School

³Department of Health Research and Policy, Stanford University

⁴Department of Medical Oncology, Dana-Farber Cancer Institute

⁵Department of Biostatistics, Harvard University

1 Introduction

In a longitudinal study, often the primary endpoint is the time to a specific event. The corresponding survival function contains all the information about the stochastic features of this endpoint. Empirically this function can be estimated consistently with the Kaplan-Meier (KM) curve with censored event time data (Kaplan and Meier, 1958). When the event rate is low, the empirical cumulative incidence curve, which is one minus the KM estimate of the survival function, is generally utilized to display observed failure probabilities over time. The survival/event probability at a specific time point t is a summary measure, though it does not contain information regarding the event time distribution profile during the time interval $(0, t)$. One may instead utilize the entire survival curve up to time t to have a visual interpretation of the pattern of events during this time period. However, it may be more informative to quantify such “survivorship” history with a summary measure beyond only reporting the event rate at time t .

A summary measure, which provides more clinically interpretable information than the event rate at time t , is the so-called restricted mean survival time (RMST) (Irwin, 1949; Zhao and Tsiatis, 1997, 1999; Murray and Tsiatis, 1999; Chen and Tsiatis, 2001; Andersen *and others*, 2004; Royston and Parmar, 2011; Zhao *and others*, 2012; Tian *and others*, 2014; Uno *and others*, 2014). That is, if all the subjects in the population of interest are followed up to time t , the RMST is the mean of the survival time. The RMST can be estimated well with the area under the corresponding KM curve up to time t . Note that the area above the survival curve is the restricted mean time lost, RMTL, which is $t - \text{RMST}$. The RMTL is a counterpart for the cumulative incidence rate function. With the RMST or RMTL summary measure, one can then construct the corresponding mean survival time curve $\text{RMST}(\cdot)$ or $\text{RMTL}(\cdot)$ over time, which provides a transparent, easily understood temporal profile for evaluating, for example, the benefit or safety of a new therapy.

For illustration, we use the data from a recent study conducted by the Eastern Cooperative Oncology Group (ECOG) for comparing two groups of patients treated by low-

and high-dose dexamethasone for newly diagnosed multiple myeloma (Rajkumar *and others*, 2010). For this trial, there were 445 enrolled patients: 222 were assigned to the low-dose and 223 to the high-dose group. Figure 1(a) shows the Kaplan-Meier (KM) curves of overall survival based on the data collected by November 2008. The survival curve for the low-dose group (dashed line) is always above the one for the high-dose group (solid line), except at the end of the follow-up. In Figure 1(b), for the KM curve of the low dose group, the area under the curve (darker shaded region) up to $t = 40$ months is 35.4. That is, on average, the future patients are expected to be alive for 35.4 out of the 40 months of followup. The corresponding estimated RMTL is $40 - 35.4 = 4.6$ (months). Figures 1(c) and 1(d) show the corresponding estimated processes $RMST(\cdot)$ and $RMTL(\cdot)$, say, $\widehat{RMST}(\cdot)$ and $\widehat{RMTL}(\cdot)$, respectively, for both the low and high dose groups. This type of the curve provides an interesting alternative to the KM counterpart using a time scale on the y -axis instead of the event or event-free probability. This time scale may be clinically appealing to practitioners with respect to the cost-risk-benefit perspectives for evaluating different treatments. For example, in Figure 1(d), through month 40, the empirical average loss times are 4.6 and 6.7 (months) for the low and high dose groups, respectively, and the difference is 2.1 months with 0.95 confidence interval of (0.1, 4.2) months. That is, with this followup duration, on average, a patient in the high dose group is alive for 2.1 fewer months. On the other hand, the survival rate difference at this time point is -0.04 with a 0.95 confidence interval of (-0.14, 0.06), indicating that there is no difference between the two groups. This conclusion with event rates may not reflect the treatment difference appropriately, especially when the study followup time is long and the disease is lethal so that the two survival curves meet at the end of study.

In Section 2, we present the inference procedures for the $RMST(\cdot)$ and $RMTL(\cdot)$. Specifically, the pointwise and simultaneous confidence interval estimation procedures for the difference of the two RMSTs over a specific time interval are proposed and illustrated with the data from the above oncology study. This type of analytic tool is quite useful for evaluating

the between-group differences under the equivalence/non-inferiority setting. In Section 3, we use the data from a large cardiovascular study to illustrate such an application. The conventional approach for summarizing the between-group difference in survival analysis under the equivalence/non-inferiority setting is to use a confidence interval estimation procedure for the hazard ratio. If the upper bound of the interval is less than a pre-specified value, one may claim that the therapy is non-inferior compared with its control counterpart. Using the data from the above cancer study, we show that this approach can be quite misleading. In general, using a single summary measure for claiming that the two groups are similar may be problematic. A more rigorous way is to examine the confidence band of the difference of two entire survival curves or their RMST curves. More discussion is provided in the Remarks section.

2 Pointwise and Simultaneous Confidence Interval Estimates

First we consider the case using event time data from a single group, followed by the case for the comparison of two groups. Let T be the time to an event of interest, which may be censored by an independent variable C . For each individual, the observable quantities are (X, Δ) , where $X = \min(T, C)$ and $\Delta = I(T \leq C)$. The data, $\{(X_i, \Delta_i); i = 1, \dots, n\}$, consist of n independent copies of (X, Δ) .

Now, let $S(t)$ and $\hat{S}(t)$ be the survival function for T and the corresponding KM estimate at time t , respectively. The corresponding $RMST(t)$ and $\widehat{RMST}(t)$ are the areas under $S(\cdot)$ and $\hat{S}(\cdot)$ up to t , respectively. It follows from the uniform consistency property of the KM estimator (Gill, 1983) that $\widehat{RMST}(t)$ is uniformly consistent for $RMST(t)$. To derive an approximation to the distribution of $\widehat{RMST}(t)$, we first use the following approximation

(Fleming and Harrington, 1991, Page 98):

$$\sqrt{n} \left\{ \hat{S}(t) - S(t) \right\} = -\sqrt{n} S(t) \sum_{i=1}^n \int_0^t \frac{dM_i(u)}{\bar{Y}(u)} + o_p(1),$$

where $M_i(t) = N_i(t) - \int_0^t Y_i(u) \lambda(u) du$, $N_i(t) = I\{X_i \leq t, \Delta_i = 1\}$, $Y_i(t) = I\{X_i \geq t\}$, $\bar{Y}(t) = \sum_{i=1}^n Y_i(t)$ and $\lambda(t)$ is the hazard function for T . Note that due to right censoring, this approximation is only valid for the interval $t \in [0, \tau]$, where $\text{pr}(X > \tau) > 0$. It follows from the martingale central limit theorem (Fleming and Harrington, 1991, Chapter 5) that $\sqrt{n} \left\{ \hat{S}(t) - S(t) \right\}$ converges weakly to a Gaussian process over the interval $[0, \tau]$.

With the functional δ -method, the distribution of $\sqrt{n} \left\{ \widehat{RMST}(\cdot) - RMST(\cdot) \right\}$ can be approximated asymptotically by that of a mean-zero Gaussian process $G(\cdot)$. When the sample size is large, this limiting distribution can be approximated well via a perturbation-resampling method (Lin *and others*, 1993; Parzen *and others*, 1997; Zhao *and others*, 2012). Specifically, let $\{Z_i, i = 1, \dots, n\}$ be n random samples from $N(0, 1)$, which are independent of the data. Then the distribution of $\sqrt{n} \left\{ \hat{S}(t) - S(t) \right\}$ can be approximated by the distribution (conditional on the data) of

$$L^*(t) = \sqrt{n} \hat{S}(t) \sum_{i=1}^n Z_i \int_0^t \frac{dN_i(u)}{\bar{Y}(u)},$$

where $\hat{S}(\cdot)$, $\bar{Y}(\cdot)$ and $N_i(\cdot)$ denote the observed quantities from the data for $\hat{S}(\cdot)$, $\bar{Y}(\cdot)$ and $N_i(\cdot)$, respectively. That is, the distribution of $\sqrt{n} \left\{ \hat{S}(t) - S(t) \right\}$ can be approximated using a large number of sets for the perturbation weights $\{Z_i, i = 1, \dots, n\}$ given the observed data. This technique has been successfully utilized in many applications in survival analysis (Park and Wei, 2003; Tian *and others*, 2005; Cai *and others*, 2010).

Now, to approximate $G(\cdot)$, the limiting Gaussian process, one can simply consider the random process over $t \in [0, \tau]$:

$$\int_0^t L^*(s) ds. \tag{1}$$

Note that the conditional asymptotic distribution of (1) is $G(\cdot)$. Such a conditional distribution can be approximated by its empirical counterpart based on realizations from M different sets of random perturbation weights $\{Z_i : i = 1, \dots, n\}$. Let the corresponding standard deviation estimate for the distribution of $G(\cdot)$ be denoted by $\hat{\sigma}_R(\cdot)$. The pointwise confidence interval estimate for the $RMST(\cdot)$ can be constructed based on the standard normal approximation to $\sqrt{n} \left(\widehat{RMST}(t) - RMST(t) \right) / \hat{\sigma}_R(t)$. Specifically, for any $\alpha \in (0, 1)$, a two-sided $1 - \alpha$ confidence interval for $RMST(t)$ is

$$\left(\widehat{RMST}(t) - z_{(1-\alpha/2)} n^{-1/2} \hat{\sigma}_R(t), \widehat{RMST}(t) + z_{(1-\alpha/2)} n^{-1/2} \hat{\sigma}_R(t) \right),$$

where $z_{(1-\alpha/2)}$ is the $100(1 - \alpha/2)$ -th percentile of the standard normal distribution. Note that in theory, the above confidence interval estimation procedure is valid for $t \in [\eta, \tau]$, where $\text{pr}(T < \eta) > 0$ and $\text{pr}(X > \tau) > 0$. The corresponding simultaneous, equal precision confidence interval (Nair, 1984) estimate for $RMST(t)$ over $[\eta, \tau]$ would be

$$\left(\widehat{RMST}(t) - c_\alpha n^{-1/2} \hat{\sigma}_R(t), \widehat{RMST}(t) + c_\alpha n^{-1/2} \hat{\sigma}_R(t) \right),$$

where the cutoff value c_α is chosen such that

$$\text{pr} \left(\sup_{t \in [\eta, \tau]} \left| \frac{\int_0^t L^*(s) ds}{\hat{\sigma}_R(t)} \right| \leq c_\alpha \right) \geq 1 - \alpha.$$

In practice, the time interval $[\eta, \tau]$ can be chosen by requiring that the estimated probabilities for both $\text{pr}(T < \eta)$ and $\text{pr}(X > \tau)$, which can be obtained from the KM estimates for the distributions of T and the empirical cumulative distribution function of X , respectively, are greater than a small positive number d . This truncation is to ensure the positivity of $\hat{\sigma}_R(t)$ and the statistical validity of the interval estimation procedure. Note that this requirement is satisfied as long as η is above the smallest observed event time and τ is below the maximum follow up time. The pointwise and simultaneous confidence intervals for $RMSTL(t)$

can be similarly constructed using the fact that $RMTL(t) = t - RMST(t)$.

As an example, with the data from the low dose group from the cancer study and with $M = 1000$ realizations of $\{Z_i : i = 1, \dots, n\}$ from the standard normal distribution, the 0.95 pointwise (dashed lines) and simultaneous (shaded area) confidence intervals for the $RMTL(\cdot)$ are given in Figure 2(a). Here, the time interval for the simultaneous confidence intervals is $[\eta, \tau] = [2.6, 40]$ months, where the choice of $\eta = 2.6$ and $\eta = 40$ satisfies the condition that the estimated probabilities of both $\text{pr}(T < \eta)$ and $\text{pr}(X > \tau)$ are positive and has been used previously (Uno *and others*, 2014). These confidence bands are quite informative, for instance, at Month 40, the possible values for RMTL are between 3.1 months and 6.1 months based on the simultaneous confidence band.

Now, suppose we are interested in estimating the difference of two RMST curves. Let all the aforementioned random variables and observed quantities be indexed by $k = 1, 2$ for the two treatment groups. The data are $\{(X_{1i}, \Delta_{1i}); i = 1, \dots, n_1\}$ for group 1 and $\{(X_{2j}, \Delta_{2j}); j = 1, \dots, n_2\}$ for group 2. Let $D(t) = RMST_2(t) - RMST_1(t)$. Then, it follows that $\hat{D}(t) = \widehat{RMST}_2(t) - \widehat{RMST}_1(t)$ is uniformly consistent for $D(t)$, for $t \in [0, \tau]$, where $\text{pr}(X_{11} > \tau)\text{pr}(X_{21} > \tau) > 0$. Moreover, $n^{-1/2}(\hat{D}(\cdot) - D(\cdot))$ converges weakly to the Gaussian process $G_2(\cdot) - G_1(\cdot)$, whose distribution can be approximated by the realizations of the difference of the two corresponding processes (1) for the two treatment groups.

Let the standard deviation estimate for $G_2(\cdot) - G_1(\cdot)$ be denoted by $\hat{\sigma}_D(t)$ via the perturbation method with M independent sets of random weights $\{Z_{1i}, i = 1, \dots, n_1; Z_{2j}, j = 1, \dots, n_2\}$. Then the pointwise confidence interval estimate for $D(t)$ is

$$\left(\hat{D}(t) - z_{(1-\alpha/2)} n^{-1/2} \hat{\sigma}_D(t), \hat{D}(t) + z_{(1-\alpha/2)} n^{-1/2} \hat{\sigma}_D(t) \right),$$

and the corresponding simultaneous, equal precision confidence interval estimate over the range $[\eta, \tau]$ is

$$\left(\hat{D}(t) - \tilde{c}_\alpha n^{-1/2} \hat{\sigma}_D(t), \hat{D}(t) + \tilde{c}_\alpha n^{-1/2} \hat{\sigma}_D(t) \right), t \in [\eta, \tau],$$

where $\text{pr}(T_{11} < \eta)\text{pr}(T_{21} < \eta) > 0$, and the cutoff point \tilde{c}_α satisfies that

$$\Pr \left(\sup_{t \in [\eta, \tau]} \left| \frac{\int_0^t [L_2^*(s) - L_1^*(s)] ds}{\hat{\sigma}_D(t)} \right| \leq \tilde{c}_\alpha \right) \geq 1 - \alpha.$$

Empirically, $[\eta, \tau]$ can be chosen such that the estimated probabilities of both $\text{pr}(T_{11} < \eta)\text{pr}(T_{21} < \eta)$ and $\text{pr}(X_{11} > \tau)\text{pr}(X_{21} > \tau)$ are above a small positive number d .

As an example, with the data from the cancer study, Figure 2(b) shows the 0.95 pointwise and simultaneous confidence interval estimates for $D(t)$. Here, the time interval for simultaneous confidence intervals is $[\eta, \tau] = [2.6, 40]$ months. It is interesting to note that for the cancer study, the hazard ratio estimate is 0.87 in favor of the low dose, but with a 0.95 confidence interval of (0.60, 1.27) that does not detect a significant difference between groups, possibly due to the fact that the two hazard functions are crossed. On the other hand, the confidence interval estimate in Figure 2(b) shows that the low dose group is uniformly better than the high dose group over the entire time interval $[2.6, 40]$ months. For instance, at month 20, the difference in RMST is 1.1 months with a 0.95 pointwise confidence interval of (0.4, 2.0) months and a 0.95 simultaneous confidence interval of (0.1, 2.2) months. The concerns and issues have been discussed extensively for using the hazard ratio as a between-group difference measure and model-free alternatives including inference based on the RMST difference at a specific time point have been proposed (Uno *and others*, 2014).

3 An application of the simultaneous confidence band estimation procedure for comparative studies under an equivalence/non-inferiority setting

In comparing a new treatment with a control with an event time as the endpoint, the standard inference procedure is based on the hazard ratio estimate (that is, assuming that the ratio of two hazard functions would be constant over the entire study followup time). To evaluate if

the new treatment is equivalent or non-inferior to the control, for example, with respect to an event time outcome, a single summary measure such as the hazard ratio as a comparison tool can be rather misleading. The above cancer study provides a clear example in which the two dose groups may be claimed to be “equivalent” on the basis of the confidence interval for the hazard ratio, but in fact, these two groups appear to be quite different based on the simultaneous confidence band for the difference of two survival functions or RMST curves.

In this section, we use the data from a cardiovascular trial: “Valsartan In Acute Myocardial Infarction (VALIANT) Study” (Pfeffer *and others*, 2003) to illustrate our inference proposal for evaluating equivalence of two treatment groups with respect to the patient’s survival. There were three arms for the study, the patients in the first group were treated by valsartan, the second group was with captopril and the third one was with a combination of these two drugs. One of the study goals was to investigate if the two mono-therapies were equivalent with respect to the overall survival. The study was conducted from 1999 to 2003 with a total of 9818 patients equally assigned to the above two mono-therapy groups. The median follow-up time is 24.7 months after randomization. Figure 3 shows the KM curves for two arms, which visually are overlapped with each other over 46 months of follow-up. The hazard ratio estimate is 1.02 with a 0.95 confidence interval of (0.93, 1.11). Using the technique discussed in Section 2, the 0.95 pointwise and simultaneous confidence intervals for two RMST functions are given in Figure 4. The largest upper bound and smallest lower bound are 0.9 and -0.9 (months), respectively, indicating that the expected survival times differ by no more than +/- 0.9 months over the duration of the study, less than 2% of the follow-up time. This tight band suggests that there may be no clinically meaningful difference between two mono-therapies with respect to the patient’s survival. Note that the confidence bands are constructed using $M = 1000$ realizations of the standard normal random sample for the resampling method and $[\eta, \tau] = [0.1, 45.5]$ months, which is the maximum-length time interval such that the estimated probabilities for both $\text{pr}(T_{11} < \eta)\text{pr}(T_{21} < \eta)$ and $\text{pr}(X_{11} > \tau)\text{pr}(X_{21} > \tau)$ are positive empirically.

4 Remarks

Although in theory, the survival curve estimate in Figure 1(a) contains all the information empirically regarding the temporal profile of the survivorship of patients in each dose group for the ECOG study, its visual interpretation can be different from that based on the corresponding RMST curve estimates presented in Figure 1(c). For instance, the observed survival probability for the low dose group at month 40 is numerically smaller than that for the high dose group (0.70 vs. 0.74). However, this could also be equally true under a scenario where 30% of low-dose patients died on day 1 of the study as well as the scenario where all low-dose patients survive until month 39. On the other hand, the RMST evaluated at month 40 can distinguish between these two scenarios. Note that the RMST curve for the low dose group is above that of the high dose group over the entire 40 month follow-up. It is not apparent how to quantify such a superiority via the KM curves at this time point. Visually the RMST curve estimates quantify the relative merit between two interventions based on the totality of information up to each time point. Both types of curve estimates can be useful, depending on the clinical question of interest, and may complement each other in helping to describe and interpret the survivorship patterns associated with the intervention(s) of interest .

For the purpose of evaluating whether two regimens are equivalent with respect to an event time outcome, the conventional approach is to utilize an event-driven study and summarize the comparison via, for example, an estimated confidence interval for the hazard ratio. This approach is not ideal and may be misleading. For example, for the cancer study example, a 0.95 confidence interval for the hazard ratio between two dose groups is (0.60, 1.27), which includes the null value one. One may interpret this result to indicate that there is not enough information to make a decision or that there is no difference statistically. A single summary such as the hazard ratio estimate may not adequately capture the differential temporal patterns between two groups. Neither the traditional Kaplan-Meier curves nor the estimated hazard ratio are able translate an estimated treatment effect to the time scale, while the median survival time is presented only in settings with suitably high event

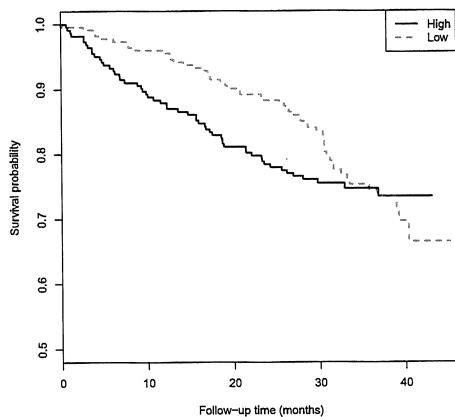
rates. The simultaneous band for the difference of two RMST (RMTL) curves discussed in this article appears to be a useful alternative to address the equivalence or non-inferiority question via a clinically meaningful metric.

References

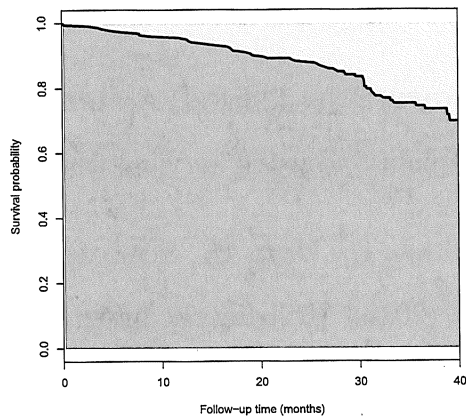
- ANDERSEN, P.K., HANSEN, M.G. AND KLEIN, J.P. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis* **10**(4), 335–350.
- CAI, T., TIAN, L., UNO, H., SOLOMON, S.D. AND WEI, L.J. (2010). Calibrating parametric subject-specific risk estimation. *Biometrika* **97**(2), 389–404.
- CHEN, P.Y. AND TSIATIS, A.A. (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* **57**(4), 1030–1038.
- IRWIN, JO. (1949). The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *Journal of Hygiene* **47**(02), 188–189.
- KAPLAN, EDWARD L AND MEIER, PAUL. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association* **53**(282), 457–481.
- LIN, D.Y., WEI, L.J. AND YING, Z. (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**(3), 557–572.
- MURRAY, S. AND TSIATIS, A.A. (1999). Sequential methods for comparing years of life saved in the two-sample censored data. *Biometrics* **55**(4), 1085–1092.
- NAIR, VIJAYAN N. (1984). Confidence bands for survival functions with censored data: a comparative study. *Technometrics* **26**(3), 265–275.

- PARK, Y. AND WEI, L.J. (2003). Estimating subject-specific survival functions under the accelerated failure time model. *Biometrika* **90**(3), 717–723.
- PARZEN, MI, WEI, LJ AND YING, Z. (1997). Simultaneous confidence intervals for the difference of two survival functions. *Scandinavian Journal of Statistics* **24**(3), 309–314.
- PFEFFER, MARC A, MCMURRAY, JOHN JV, VELAZQUEZ, ERIC J, ROULEAU, JEAN-LUCIEN, KØBER, LARS, MAGGIONI, ALDO P, SOLOMON, SCOTT D, SWEDBERG, KARL, VAN DE WERF, FRANS, WHITE, HARVEY *and others*. (2003). Valsartan, captopril, or both in myocardial infarction complicated by heart failure, left ventricular dysfunction, or both. *New England Journal of Medicine* **349**(20), 1893–1906.
- RAJKUMAR, S VINCENT, JACOBUS, SUSANNA, CALLANDER, NATALIE S, FONSECA, RAFAEL, VESOLE, DAVID H, WILLIAMS, MICHAEL E, ABONOUR, RAFAT, SIEGEL, DAVID S, KATZ, MICHAEL AND GREIPP, PHILIP R. (2010). Lenalidomide plus high-dose dexamethasone versus lenalidomide plus low-dose dexamethasone as initial therapy for newly diagnosed multiple myeloma: an open-label randomised controlled trial. *The lancet oncology* **11**(1), 29–37.
- ROYSTON, P. AND PARMAR, M.K.B. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in medicine* **30**(19), 2409–2421.
- TIAN, LU, ZHAO, LIHUI AND WEI, L.J. (2014). Predicting the restricted mean event time with the subject’s baseline covariates in survival analysis. *Biostatistics* **15**(2), 222–233.
- TIAN, L., ZUCKER, D. AND WEI, L.J. (2005). On the cox model with time-varying regression coefficients. *Journal of the American statistical Association* **100**(469), 172–183.
- UNO, HAJIME, CLAGGETT, BRIAN, TIAN, LU, INOUE, EISUKE, GALLO, PAUL, MIYATA, TOSHIO, SCHRAG, DEBORAH, TAKEUCHI, MASAHIRO, UYAMA, YOSHIAKI, ZHAO, LI-

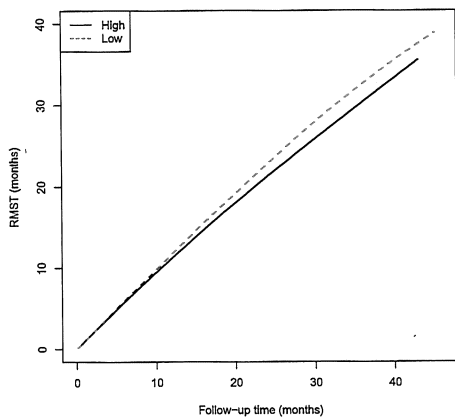
- HUI *and others.* (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of Clinical Oncology* **32**(22), 2380–2385.
- ZHAO, HONGWEI AND TSIATIS, ANASTASIOS A. (1997). A consistent estimator for the distribution of quality adjusted survival time. *Biometrika* **84**(2), 339–348.
- ZHAO, HONGWEI AND TSIATIS, ANASTASIOS A. (1999). Efficient estimation of the distribution of quality-adjusted survival time. *Biometrics* **55**(4), 1101–1107.
- ZHAO, L., TIAN, L., UNO, H., SOLOMON, S.D., PFEFFER, M.A., SCHINDLER, J.S. AND WEI, L.J. (2012). Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical study. *Clinical Trials* **9**(5), 570–577.



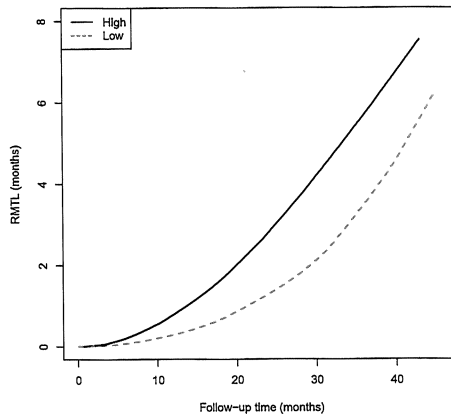
(a) Kaplan-Meier curves



(b) Estimated RMST and RMTL for low-dose group

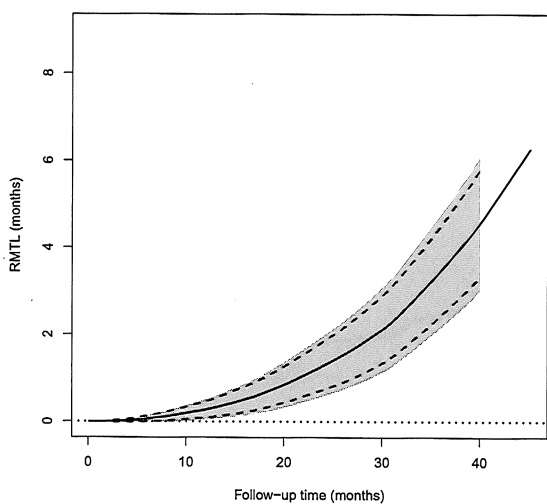


(c) Estimated RMST curve

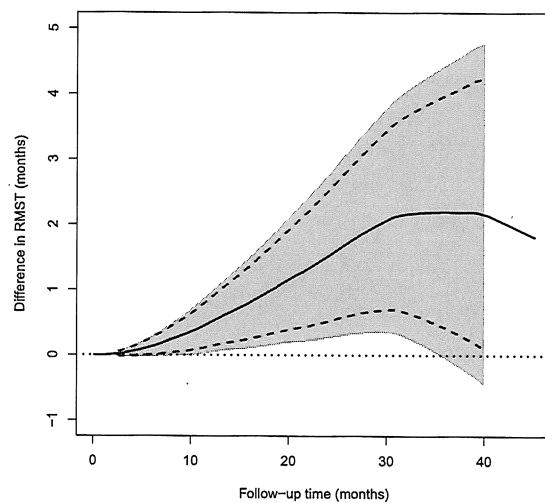


(d) Estimated RMTL curve

Figure 1: Results based on the data from ECOG (E4A03): (a) Kaplan-Meier curves for high-dose (solid line) and low-dose (dashed line) groups; (b) The estimated RMST (darker shaded region below the solid line) and RMTL (lighter shaded region above the solid line) up to 40 months for the low-dose group; (c) The estimated RMST curve for high-dose (solid line) and low-dose (dashed line) groups; (d) The estimated RMTL curve for high-dose (solid line) and low-dose (dashed line) groups.



(a) RMTL curve for low-dose group



(b) Difference in RMST curves between low-dose and high-dose groups

Figure 2: Results based on the data from ECOG (E4A03): (a) RMTL curve for low-dose group; (b) Difference in RMST curves between low-dose and high-dose groups. Solid lines: point estimates, dashed lines: pointwise 0.95 confidence intervals, shaded region: simultaneous 0.95 confidence intervals, dotted line: zero reference line.

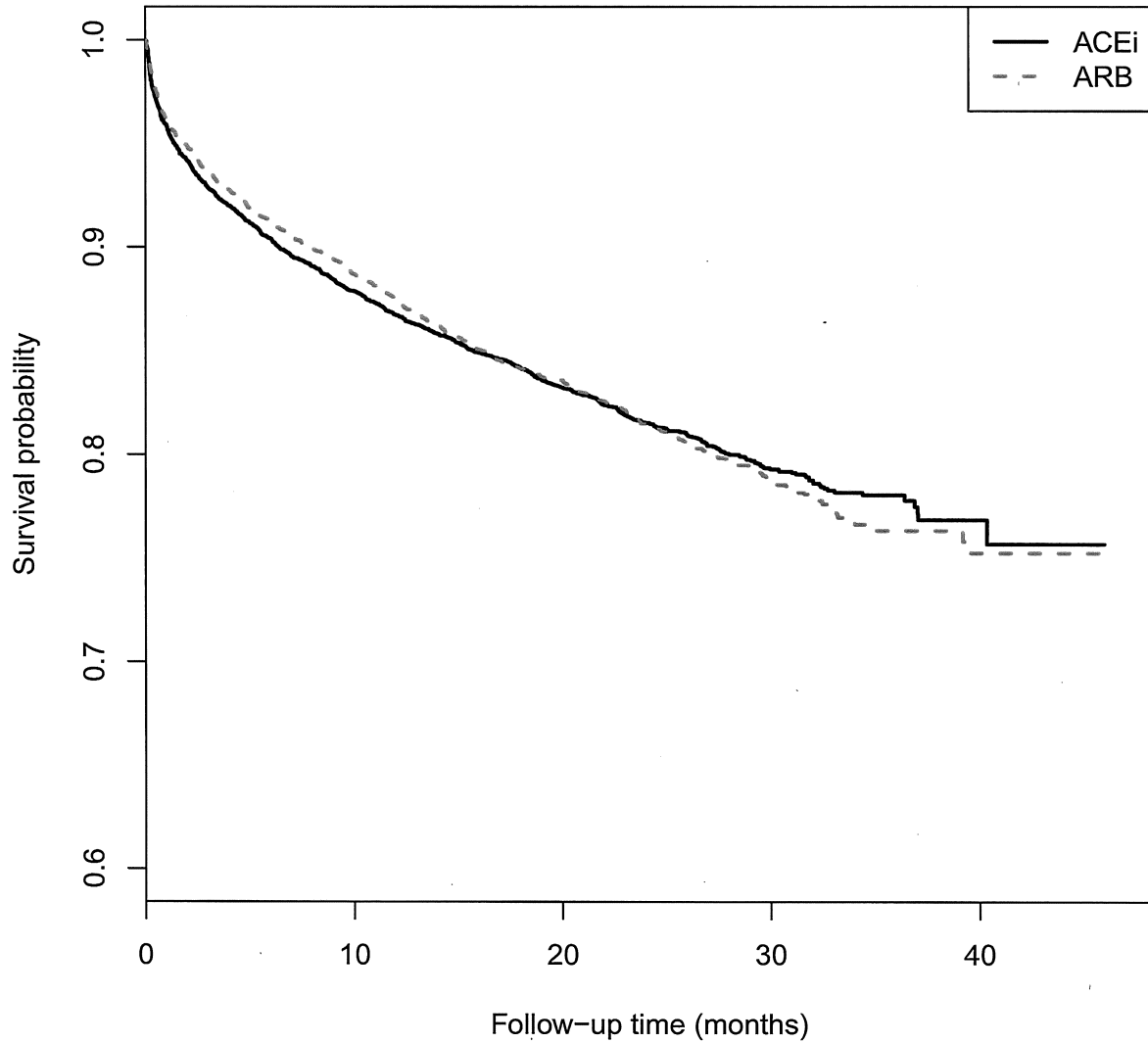


Figure 3: The Kaplan-Meier estimates for the survival functions of patients taking ARB or ACEi in the VALIANT study.

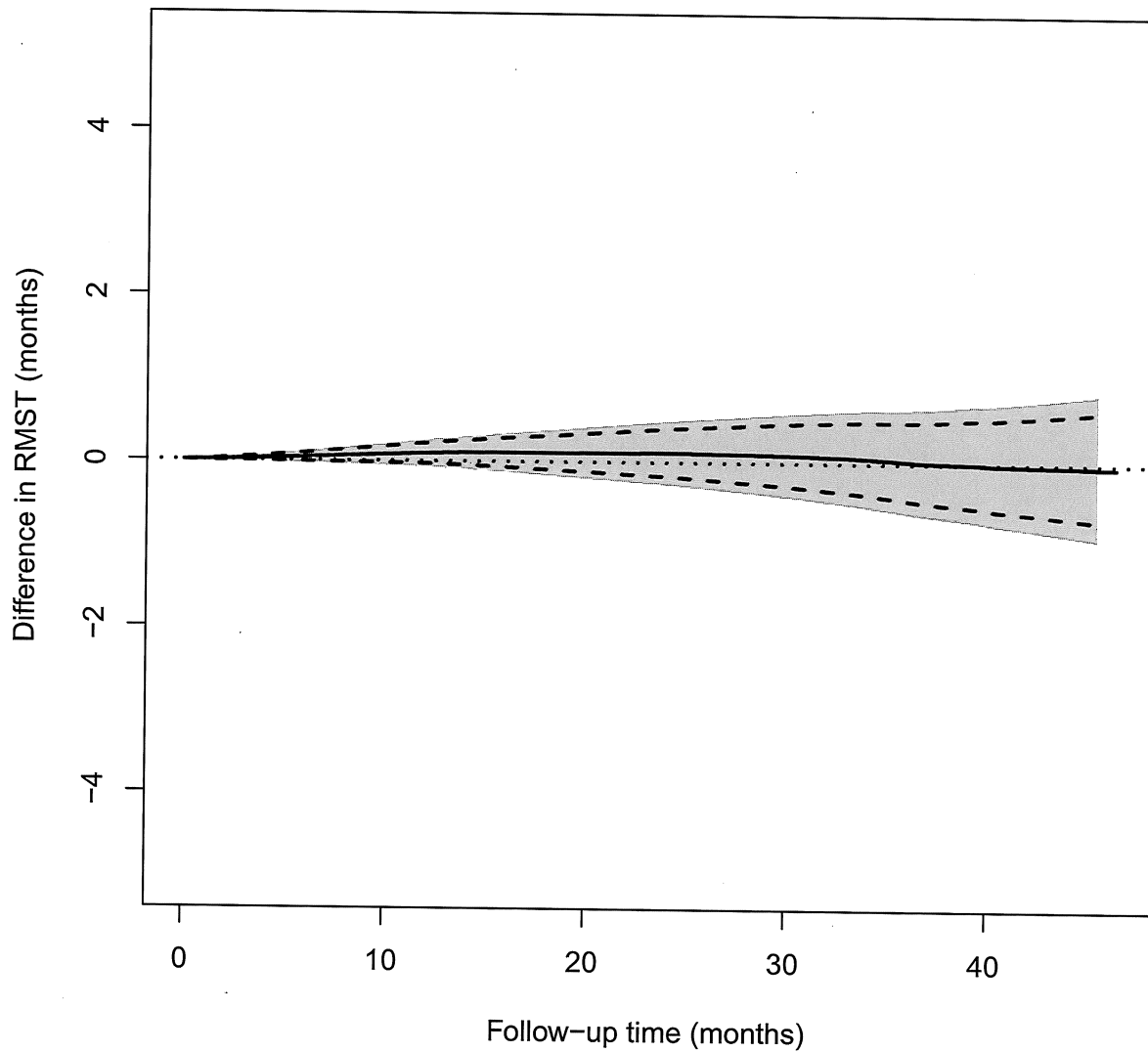


Figure 4: The difference in RMST curves between ARB and ACEi in RMST up to time t of patients in the VALIANT study (Solid: point estimate; Dashed: 0.95 pointwise confidence interval; Shaded: 0.95 simultaneous confidence interval; dotted: zero reference line).

